Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale — temi e correzione

Mercoledì 20 dicembre 2017

Contenuti

- Testi dei 130 temi d'esame
- Traccia della soluzione degli Esercizi 1 e 2 del Tema 1
- Risposte corrette e commentate alle domande dell'esercizio 3
- Griglie di correzione dei temi

I temi sono basati su uno stesso dataset i cui campioni e attributi vengono riscalati, permutati e leggermante perturbati.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 1

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.10	Grasso	Altruista
2	0.78	Magro	Altruista
3	0.88	Grasso	Egoista

i	x_{i1}	x_{i2}	y_i
4	0.28	Medio	Egoista
5	0.59	Medio	Altruista
6	0.37	Magro	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n-1$.
   (c) $(n-1)^2$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[-1, 1]$.
    (b) $[0, +\infty)$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 2

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	Sedentario	0.84	Altruista
2	Attivo	0.16	Altruista
3	Sedentario	0.43	Egoista

i	x_{i1}	x_{i2}	y_i
4	Sportivo	0.34	Egoista
5	Attivo	0.94	Egoista
6	Sportivo	0.65	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[0, 1]$.
    (b) $[0, +\infty)$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 3

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	5.7	Svogliato
2	Magro	0.8	Svogliato
3	Medio	3.5	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Medio	7.6	Svogliato
5	Magro	8.6	Partecipe
6	Grasso	2.6	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n - 1$.
    (b) $n(n-1)/2$.
    (c) $(n-1)^2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 4

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	8.6	Grasso	Triste
2	2.6	Magro	Triste
3	0.8	Grasso	Felice

i	x_{i1}	x_{i2}	y_i
4	3.5	Medio	Triste
5	7.6	Medio	Felice
6	5.7	Magro	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n-1$.
   (c) $n(n-1)/2$.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
    (b) Che esiste una forte dipendenza lineare fra le due variabili.
    (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 5

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	2.8	Felice
2	Grasso	7.8	Triste
3	Magro	5.9	Triste

i	x_{i1}	x_{i2}	y_i
4	Medio	1.0	Triste
5	Grasso	3.7	Felice
6	Medio	8.8	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[1, +\infty)$.
    (b) $[0, 1]$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 6

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	57	Magro	Svogliato
2	26	Magro	Partecipe
3	35	Medio	Partecipe

i	x_{i1}	x_{i2}	y_i
4	76	Medio	Svogliato
5	8	Grasso	Svogliato
6	86	Grasso	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa della variabile di output nei figli.
    (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (c) ... massimizza l'impurità attesa della variabile di output dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 7

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	0.38	Svogliato
2	Grasso	0.79	Partecipe
3	Medio	0.60	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Magro	0.11	Partecipe
5	Medio	0.29	Svogliato
6	Magro	0.89	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa delle variabili di input dei figli.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n-1)^2$.
   (c) $n(n-1)/2$.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
    (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
    (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 8

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	9.1	Basso	Egoista
2	8.1	Alto	Altruista
3	3.1	Medio	Egoista

i	x_{i1}	x_{i2}	y_i
4	6.2	Medio	Altruista
5	4.0	Alto	Egoista
6	1.3	Basso	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n - 1$.
    (b) $n(n-1)/2$.
    (c) $(n-1)^2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 9

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	8.6	Magro	Egoista
2	0.8	Magro	Altruista
3	2.6	Medio	Egoista

i	x_{i1}	x_{i2}	y_i
4	7.6	Grasso	Altruista
5	3.5	Grasso	Egoista
6	5.7	Medio	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'entropia attesa della variabile di output nei figli.
    (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 10

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	8.5	Alto	Felice
2	0.7	Alto	Triste
3	5.6	Medio	Triste

i	x_{i1}	x_{i2}	y_i
4	3.4	Basso	Felice
5	7.5	Basso	Triste
6	2.5	Medio	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
3. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[0, 1]$.
    (b) $[1, +\infty)$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 11

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	86	Egoista
2	Magro	35	Egoista
3	Medio	26	Egoista

i	x_{i1}	x_{i2}	y_i
4	Medio	57	Altruista
5	Grasso	8	Altruista
6	Magro	76	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n - 1)/2$.
   (c) $(n - 1)^2$.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[-1, 1]$.
    (b) $[0, +\infty)$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 12

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	3.0	Altruista
2	Medio	6.1	Egoista
3	Grasso	3.9	Altruista

i	x_{i1}	x_{i2}	y_i
4	Grasso	8.0	Egoista
5	Magro	1.2	Egoista
6	Magro	9.0	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n-1$.
   (c) $n(n-1)/2$.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[0, +\infty)$.
    (b) $[0, 1]$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 13

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	Attivo	0.60	Partecipe
2	Sportivo	0.38	Svogliato
3	Sedentario	0.11	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Sportivo	0.79	Partecipe
5	Attivo	0.29	Svogliato
6	Sedentario	0.89	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... massimizza l'impurità attesa della variabile di output dei figli.
    (b) ... minimizza l'impurità attesa della variabile di output nei figli.
    (c) ... minimizza l'impurità attesa delle variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 14

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	0.37	Triste
2	Grasso	0.10	Felice
3	Grasso	0.88	Triste

i	x_{i1}	x_{i2}	y_i
4	Medio	0.28	Triste
5	Magro	0.78	Felice
6	Medio	0.59	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
2. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza minore.
    (b) Quelli a distanza maggiore.
    (c) Dipende dal linkage criterion.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 15

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	0.87	Triste
2	Alto	0.09	Felice
3	Medio	0.36	Triste

i	x_{i1}	x_{i2}	y_i
4	Basso	0.27	Triste
5	Medio	0.77	Felice
6	Basso	0.58	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n - 1$.
   (c) $n(n-1)/2$.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza minore.
    (b) Quelli a distanza maggiore.
    (c) Dipende dal linkage criterion.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 16

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Basso	8.3	Felice
2	Medio	3.3	Triste
3	Medio	6.4	Felice

i	x_{i1}	x_{i2}	y_i
4	Alto	1.5	Felice
5	Basso	4.2	Triste
6	Alto	9.3	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n-1$.
   (c) $(n-1)^2$.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ... minimizza l'entropia attesa della variabile di output nei figli.
    (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 17

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	0.55	Svogliato
2	Magro	0.74	Svogliato
3	Grasso	0.06	Svogliato

i	x_{i1}	x_{i2}	y_i
4	Grasso	0.84	Partecipe
5	Magro	0.33	Partecipe
6	Medio	0.24	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
3. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[1, +\infty)$.
    (b) $[-1, 1]$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 18

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	0.09	Partecipe
2	Grasso	0.77	Partecipe
3	Magro	0.27	Svogliato

i	x_{i1}	x_{i2}	y_i
4	Medio	0.87	Svogliato
5	Grasso	0.36	Svogliato
6	Magro	0.58	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n - 1$.
   (c) $n(n-1)/2$.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
    (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 19

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	0.94	Pessimista
2	Basso	0.84	Ottimista
3	Basso	0.43	Pessimista

i	x_{i1}	x_{i2}	y_i
4	Alto	0.16	Ottimista
5	Medio	0.65	Ottimista
6	Medio	0.34	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[0, 1]$.
    (b) $[-1, 1]$.
    (c) $[0, +\infty)$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 20

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	84	Altruista
2	Medio	74	Egoista
3	Medio	33	Altruista

i	x_{i1}	x_{i2}	y_i
4	Grasso	55	Egoista
5	Grasso	24	Altruista
6	Magro	6	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
    (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 21

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	1.3	Egoista
2	Medio	6.2	Egoista
3	Medio	3.1	Altruista

i	x_{i1}	x_{i2}	y_i
4	Grasso	4.0	Altruista
5	Magro	9.1	Altruista
6	Grasso	8.1	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
3. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[1, +\infty)$.
    (b) $[-1, 1]$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 22

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	8.1	Svogliato
2	Grasso	9.1	Partecipe
3	Magro	4.0	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Grasso	1.3	Svogliato
5	Medio	3.1	Partecipe
6	Medio	6.2	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
    (b) Che esiste una forte dipendenza lineare fra le due variabili.
    (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 23

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	4.2	Triste
2	Alto	1.5	Felice
3	Alto	9.3	Triste

i	x_{i1}	x_{i2}	y_i
4	Medio	8.3	Felice
5	Basso	3.3	Triste
6	Basso	6.4	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'entropia attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 24

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	38	Magro	Altruista
2	89	Grasso	Altruista
3	11	Grasso	Egoista

i	x_{i1}	x_{i2}	y_i
4	79	Magro	Egoista
5	60	Medio	Egoista
6	29	Medio	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'entropia attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 25

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	1.2	Basso	Svogliato
2	9.0	Basso	Partecipe
3	8.0	Medio	Svogliato

i	x_{i1}	x_{i2}	y_i
4	6.1	Alto	Svogliato
5	3.0	Alto	Partecipe
6	3.9	Medio	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n - 1$.
   (c) $n(n-1)/2$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[0, 1]$.
    (b) $[0, +\infty)$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 26

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	9.0	Ottimista
2	Medio	6.1	Pessimista
3	Medio	3.0	Ottimista

i	x_{i1}	x_{i2}	y_i
4	Magro	8.0	Pessimista
5	Magro	3.9	Ottimista
6	Grasso	1.2	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (b) Che esiste una forte dipendenza lineare fra le due variabili.
    (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 27

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	Sportivo	29	Egoista
2	Attivo	38	Egoista
3	Attivo	79	Altruista

i	x_{i1}	x_{i2}	y_i
4	Sedentario	11	Altruista
5	Sedentario	89	Egoista
6	Sportivo	60	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[1, +\infty)$.
    (b) $[0, 1]$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 28

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	3.1	Sedentario	Egoista
2	6.2	Sedentario	Altruista
3	4.0	Sportivo	Egoista

i	x_{i1}	x_{i2}	y_i
4	1.3	Attivo	Altruista
5	8.1	Sportivo	Altruista
6	9.1	Attivo	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Dipende dal linkage criterion.
    (b) Quelli a distanza maggiore.
    (c) Quelli a distanza minore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 29

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	14	Ottimista
2	Basso	41	Pessimista
3	Medio	92	Pessimista

i	x_{i1}	x_{i2}	y_i
4	Basso	82	Ottimista
5	Alto	63	Ottimista
6	Alto	32	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa della variabile di output nei figli.
    (b) ... massimizza l'impurità attesa della variabile di output dei figli.
    (c) ... minimizza l'impurità attesa delle variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 30

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	1.0	Sportivo	Svogliato
2	3.7	Attivo	Partecipe
3	8.8	Sportivo	Partecipe

i	x_{i1}	x_{i2}	y_i
4	7.8	Attivo	Svogliato
5	2.8	Sedentario	Partecipe
6	5.9	Sedentario	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[0, +\infty)$.
    (b) $[-1, 1]$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 31

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.14	Grasso	Egoista
2	0.92	Grasso	Altruista
3	0.41	Magro	Altruista

i	x_{i1}	x_{i2}	y_i
4	0.32	Medio	Altruista
5	0.63	Medio	Egoista
6	0.82	Magro	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $(n-1)^2$.
    (b) $n(n-1)/2$.
    (c) $n-1$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 32

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	6.5	Magro	Felice
2	8.4	Grasso	Felice
3	4.3	Grasso	Triste

i	x_{i1}	x_{i2}	y_i
4	3.4	Magro	Triste
5	1.6	Medio	Felice
6	9.4	Medio	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n - 1$.
    (b) $n(n-1)/2$.
    (c) $(n-1)^2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 33

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.28	Grasso	Ottimista
2	0.37	Medio	Ottimista
3	0.78	Medio	Pessimista

i	x_{i1}	x_{i2}	y_i
4	0.88	Magro	Ottimista
5	0.59	Grasso	Pessimista
6	0.10	Magro	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[-1, 1]$.
    (b) $[0, +\infty)$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 34

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	Sedentario	0.25	Egoista
2	Sportivo	0.07	Altruista
3	Sportivo	0.85	Egoista

i	x_{i1}	x_{i2}	y_i
4	Sedentario	0.56	Altruista
5	Attivo	0.34	Egoista
6	Attivo	0.75	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
2. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'entropia attesa della variabile di output nei figli.
    (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 35

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	0.92	Svogliato
2	Magro	0.63	Partecipe
3	Grasso	0.82	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Grasso	0.41	Svogliato
5	Medio	0.14	Partecipe
6	Magro	0.32	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
5. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
    (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 36

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	39	Felice
2	Alto	61	Triste
3	Alto	30	Felice

i	x_{i1}	x_{i2}	y_i
4	Basso	12	Triste
5	Medio	80	Triste
6	Basso	90	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n - 1$.
   (c) $n(n-1)/2$.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[-1, 1]$.
    (b) $[0, 1]$.
    (c) $[1, +\infty)$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 37

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0,1], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.09	Sedentario	Partecipe
2	0.77	Sportivo	Partecipe
3	0.58	Attivo	Partecipe

i	x_{i1}	x_{i2}	y_i
4	0.27	Attivo	Svogliato
5	0.87	Sedentario	Svogliato
6	0.36	Sportivo	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ... minimizza l'entropia attesa della variabile di output nei figli.
    (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 38

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Basso	3.9	Partecipe
2	Alto	1.2	Svogliato
3	Medio	6.1	Svogliato

i	x_{i1}	x_{i2}	y_i
4	Basso	8.0	Svogliato
5	Medio	3.0	Partecipe
6	Alto	9.0	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
3. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n-1$.
   (c) $(n-1)^2$.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...minimizza l'impurità attesa delle variabili di input dei figli.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che esiste una forte dipendenza lineare fra le due variabili.
    (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 39

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.27	Medio	Egoista
2	0.09	Grasso	Altruista
3	0.58	Medio	Altruista

i	x_{i1}	x_{i2}	y_i
4	0.87	Grasso	Egoista
5	0.77	Magro	Altruista
6	0.36	Magro	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
5. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza maggiore.
    (b) Dipende dal linkage criterion.
    (c) Quelli a distanza minore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 40

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	37	Attivo	Felice
2	59	Sedentario	Triste
3	88	Sportivo	Felice

i	x_{i1}	x_{i2}	y_i
4	28	Sedentario	Felice
5	78	Attivo	Triste
6	10	Sportivo	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
    (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (c) Che esiste una forte dipendenza lineare fra le due variabili.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 41

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	63	Sportivo	Altruista
2	41	Sedentario	Egoista
3	14	Attivo	Altruista

i	x_{i1}	x_{i2}	y_i
4	82	Sedentario	Altruista
5	32	Sportivo	Egoista
6	92	Attivo	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ...massimizza l'impurità attesa della variabile di output dei figli.
    (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
    (c) ...minimizza l'impurità attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 42

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.64	Basso	Egoista
2	0.83	Medio	Egoista
3	0.33	Basso	Altruista

i	x_{i1}	x_{i2}	y_i
4	0.93	Alto	Altruista
5	0.15	Alto	Egoista
6	0.42	Medio	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n-1$.
   (c) $n(n-1)/2$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
8. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'entropia attesa della variabile di output nei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 43

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	65	Sedentario	Felice
2	94	Attivo	Triste
3	43	Sportivo	Triste

i	x_{i1}	x_{i2}	y_i
4	16	Attivo	Felice
5	34	Sedentario	Triste
6	84	Sportivo	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
3. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n - 1$.
   (c) $n(n-1)/2$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza minore.
    (b) Quelli a distanza maggiore.
    (c) Dipende dal linkage criterion.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 44

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.37	Medio	Ottimista
2	0.10	Basso	Pessimista
3	0.28	Alto	Ottimista

i	x_{i1}	x_{i2}	y_i
4	0.59	Alto	Pessimista
5	0.88	Basso	Ottimista
6	0.78	Medio	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Dipende dal linkage criterion.
    (b) Quelli a distanza maggiore.
    (c) Quelli a distanza minore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 45

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Sedentario	3.4	Triste
2	Attivo	1.6	Felice
3	Attivo	9.4	Triste

i	x_{i1}	x_{i2}	y_i
4	Sportivo	4.3	Triste
5	Sportivo	8.4	Felice
6	Sedentario	6.5	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[0, +\infty)$.
    (b) $[0, 1]$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 46

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	60	Attivo	Svogliato
2	89	Sedentario	Partecipe
3	38	Sportivo	Partecipe

i	x_{i1}	x_{i2}	y_i
4	79	Sportivo	Svogliato
5	11	Sedentario	Svogliato
6	29	Attivo	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n-1)^2$.
   (c) $n(n-1)/2$.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (b) Che esiste una forte dipendenza lineare fra le due variabili.
    (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 47

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	2.7	Grasso	Partecipe
2	7.7	Medio	Svogliato
3	8.7	Magro	Partecipe

i	x_{i1}	x_{i2}	y_i
4	0.9	Magro	Svogliato
5	5.8	Grasso	Svogliato
6	3.6	Medio	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
3. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n(n-1)/2$.
    (b) $n-1$.
    (c) $(n-1)^2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 48

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	57	Altruista
2	Magro	76	Altruista
3	Grasso	26	Egoista

i	x_{i1}	x_{i2}	y_i
4	Magro	35	Egoista
5	Medio	86	Egoista
6	Medio	8	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'entropia attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 49

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	0.64	Felice
2	Medio	0.93	Triste
3	Magro	0.42	Triste

i	x_{i1}	x_{i2}	y_i
4	Grasso	0.33	Triste
5	Magro	0.83	Felice
6	Medio	0.15	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa delle variabili di input dei figli.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[0, +\infty)$.
    (b) $[-1, 1]$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 50

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Basso	0.06	Ottimista
2	Medio	0.33	Pessimista
3	Medio	0.74	Ottimista

i	x_{i1}	x_{i2}	y_i
4	Basso	0.84	Pessimista
5	Alto	0.55	Ottimista
6	Alto	0.24	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n - 1$.
    (b) $(n-1)^2$.
    (c) $n(n-1)/2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 51

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	94	Basso	Altruista
2	16	Basso	Egoista
3	84	Alto	Egoista

i	x_{i1}	x_{i2}	y_i
4	34	Medio	Altruista
5	65	Medio	Egoista
6	43	Alto	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (b) ... minimizza l'impurità attesa della variabile di output nei figli.
    (c) ... massimizza l'impurità attesa della variabile di output dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 52

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	31	Sportivo	Partecipe
2	91	Attivo	Partecipe
3	81	Sedentario	Svogliato

i	x_{i1}	x_{i2}	y_i
4	62	Sportivo	Svogliato
5	40	Sedentario	Partecipe
6	13	Attivo	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
    (b) Che esiste una forte dipendenza lineare fra le due variabili.
    (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 53

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	0.7	Triste
2	Medio	2.5	Felice
3	Alto	8.5	Felice

i	x_{i1}	x_{i2}	y_i
4	Basso	7.5	Triste
5	Medio	5.6	Triste
6	Basso	3.4	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
3. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[1, +\infty)$.
    (b) $[0, 1]$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 54

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	11	Attivo	Pessimista
2	89	Attivo	Ottimista
3	29	Sportivo	Ottimista

i	x_{i1}	x_{i2}	y_i
4	38	Sedentario	Ottimista
5	60	Sportivo	Pessimista
6	79	Sedentario	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...minimizza l'impurità attesa delle variabili di input dei figli.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ...massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ...minimizza l'entropia attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 55

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	32	Sportivo	Triste
2	14	Sedentario	Felice
3	63	Sportivo	Felice

i	x_{i1}	x_{i2}	y_i
4	41	Attivo	Triste
5	82	Attivo	Felice
6	92	Sedentario	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
8. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $(n-1)^2$.
    (b) $n-1$.
    (c) $n(n-1)/2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 56

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	80	Grasso	Egoista
2	12	Medio	Egoista
3	30	Magro	Altruista

i	x_{i1}	x_{i2}	y_i
4	90	Medio	Altruista
5	61	Magro	Egoista
6	39	Grasso	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
5. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n(n-1)/2$.
    (b) $n-1$.
    (c) $(n-1)^2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 57

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0,1], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.62	Attivo	Partecipe
2	0.40	Sedentario	Svogliato
3	0.81	Sedentario	Partecipe

i	x_{i1}	x_{i2}	y_i
4	0.13	Sportivo	Partecipe
5	0.31	Attivo	Svogliato
6	0.91	Sportivo	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
    (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
    (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 58

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	2.4	Alto	Egoista
2	0.6	Basso	Altruista
3	3.3	Medio	Egoista

i	x_{i1}	x_{i2}	y_i
4	5.5	Alto	Altruista
5	8.4	Basso	Egoista
6	7.4	Medio	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
10. Che significato ha il parametro principale K dell'algoritmo K-means?
    (a) Il numero di iterazioni dell'algoritmo.
    (b) Il numero di cluster in cui suddividere il dataset.
    (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 59

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Attivo	37	Triste
2	Sedentario	88	Triste
3	Sedentario	10	Felice

i	x_{i1}	x_{i2}	y_i
4	Attivo	78	Felice
5	Sportivo	59	Felice
6	Sportivo	28	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...minimizza l'impurità attesa delle variabili di input dei figli.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n(n-1)/2$.
    (b) $(n-1)^2$.
    (c) $n-1$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 60

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	7.7	Grasso	Felice
2	2.7	Medio	Triste
3	5.8	Medio	Felice

i	x_{i1}	x_{i2}	y_i
4	8.7	Magro	Triste
5	3.6	Grasso	Triste
6	0.9	Magro	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n-1$.
   (c) $(n-1)^2$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che esiste una forte dipendenza lineare fra le due variabili.
    (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 61

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	5.8	Pessimista
2	Magro	7.7	Pessimista
3	Medio	8.7	Ottimista

i	x_{i1}	x_{i2}	y_i
4	Grasso	2.7	Ottimista
5	Magro	3.6	Ottimista
6	Medio	0.9	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
2. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n - 1)/2$.
   (c) $(n - 1)^2$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza maggiore.
    (b) Quelli a distanza minore.
    (c) Dipende dal linkage criterion.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 62

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	84	Svogliato
2	Magro	6	Partecipe
3	Medio	55	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Grasso	33	Svogliato
5	Grasso	74	Partecipe
6	Medio	24	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n-1$.
   (c) $(n-1)^2$.
4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[0, 1]$.
    (b) $[1, +\infty)$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 63

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	2.4	Pessimista
2	Alto	7.4	Ottimista
3	Basso	8.4	Pessimista

i	x_{i1}	x_{i2}	y_i
4	Basso	0.6	Ottimista
5	Alto	3.3	Pessimista
6	Medio	5.5	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (b) ... minimizza l'impurità attesa della variabile di output nei figli.
    (c) ... massimizza l'impurità attesa della variabile di output dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 64

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	Sedentario	0.35	Altruista
2	Attivo	0.57	Egoista
3	Sedentario	0.76	Egoista

i	x_{i1}	x_{i2}	y_i
4	Attivo	0.26	Altruista
5	Sportivo	0.86	Altruista
6	Sportivo	0.08	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
3. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'entropia attesa della variabile di output nei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 65

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Basso	61	Svogliato
2	Medio	12	Svogliato
3	Alto	39	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Medio	90	Partecipe
5	Alto	80	Svogliato
6	Basso	30	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[-1, 1]$.
    (b) $[1, +\infty)$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 66

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	8	Grasso	Partecipe
2	86	Grasso	Svogliato
3	57	Medio	Partecipe

i	x_{i1}	x_{i2}	y_i
4	26	Medio	Svogliato
5	35	Magro	Svogliato
6	76	Magro	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n(n-1)/2$.
    (b) $n-1$.
    (c) $(n-1)^2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 67

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	Sportivo	78	Svogliato
2	Sedentario	10	Svogliato
3	Attivo	28	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Attivo	59	Svogliato
5	Sedentario	88	Partecipe
6	Sportivo	37	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
    (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
    (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 68

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	36	Felice
2	Alto	77	Triste
3	Basso	58	Triste

i	x_{i1}	x_{i2}	y_i
4	Medio	9	Triste
5	Medio	87	Felice
6	Basso	27	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
2. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
    (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (c) Che esiste una forte dipendenza lineare fra le due variabili.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 69

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	27	Magro	Ottimista
2	77	Medio	Pessimista
3	87	Grasso	Ottimista

i	x_{i1}	x_{i2}	y_i
4	9	Grasso	Pessimista
5	36	Medio	Ottimista
6	58	Magro	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
5. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ... minimizza l'entropia attesa della variabile di output nei figli.
    (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 70

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	10	Basso	Pessimista
2	28	Medio	Ottimista
3	59	Medio	Pessimista

i	x_{i1}	x_{i2}	y_i
4	78	Alto	Pessimista
5	88	Basso	Ottimista
6	37	Alto	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n - 1$.
    (b) $n(n - 1)/2$.
    (c) $(n - 1)^2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 71

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	8.1	Sedentario	Altruista
2	6.2	Attivo	Altruista
3	1.3	Sportivo	Altruista

i	x_{i1}	x_{i2}	y_i
4	3.1	Attivo	Egoista
5	4.0	Sedentario	Egoista
6	9.1	Sportivo	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
3. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
8. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
    (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 72

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	6.1	Alto	Triste
2	1.2	Medio	Triste
3	3.0	Alto	Felice

i	x_{i1}	x_{i2}	y_i
4	8.0	Basso	Triste
5	9.0	Medio	Felice
6	3.9	Basso	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ... minimizza l'entropia attesa della variabile di output nei figli.
    (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 73

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	0.38	Sportivo	Ottimista
2	0.89	Sedentario	Ottimista
3	0.79	Sportivo	Pessimista

i	x_{i1}	x_{i2}	y_i
4	0.11	Sedentario	Pessimista
5	0.60	Attivo	Pessimista
6	0.29	Attivo	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[-1, 1]$.
    (b) $[1, +\infty)$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 74

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	82	Triste
2	Medio	63	Triste
3	Basso	14	Triste

i	x_{i1}	x_{i2}	y_i
4	Basso	92	Felice
5	Alto	41	Felice
6	Medio	32	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... massimizza l'impurità attesa della variabile di output dei figli.
    (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (c) ... minimizza l'impurità attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 75

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.61	Medio	Pessimista
2	0.39	Grasso	Ottimista
3	0.30	Medio	Ottimista

i	x_{i1}	x_{i2}	y_i
4	0.90	Magro	Ottimista
5	0.12	Magro	Pessimista
6	0.80	Grasso	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (b) ... massimizza l'impurità attesa della variabile di output dei figli.
    (c) ... minimizza l'impurità attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 76

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	57	Magro	Ottimista
2	8	Grasso	Ottimista
3	86	Grasso	Pessimista

i	x_{i1}	x_{i2}	y_i
4	76	Medio	Ottimista
5	35	Medio	Pessimista
6	26	Magro	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
3. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... massimizza l'impurità attesa della variabile di output dei figli.
    (b) ... minimizza l'impurità attesa della variabile di output nei figli.
    (c) ... minimizza l'impurità attesa delle variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 77

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	6.1	Sedentario	Triste
2	9.0	Attivo	Felice
3	1.2	Attivo	Triste

i	x_{i1}	x_{i2}	y_i
4	8.0	Sportivo	Triste
5	3.9	Sportivo	Felice
6	3.0	Sedentario	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'entropia attesa della variabile di output nei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 78

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	41	Partecipe
2	Medio	14	Svogliato
3	Medio	92	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Grasso	63	Svogliato
5	Magro	82	Svogliato
6	Grasso	32	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
2. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che esiste una forte dipendenza lineare fra le due variabili.
    (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 79

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	40	Ottimista
2	Magro	62	Pessimista
3	Grasso	91	Ottimista

i	x_{i1}	x_{i2}	y_i
4	Medio	81	Pessimista
5	Magro	31	Ottimista
6	Grasso	13	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n-1$.
   (c) $n(n-1)/2$.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[-1, 1]$.
    (b) $[1, +\infty)$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 80

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.91	Basso	Ottimista
2	0.40	Medio	Ottimista
3	0.62	Alto	Pessimista

i	x_{i1}	x_{i2}	y_i
4	0.31	Alto	Ottimista
5	0.81	Medio	Pessimista
6	0.13	Basso	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n-1)^2$.
   (c) $n(n-1)/2$.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa della variabile di output nei figli.
    (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (c) ... massimizza l'impurità attesa della variabile di output dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 81

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	1.4	Medio	Ottimista
2	4.1	Magro	Pessimista
3	9.2	Medio	Pessimista

i	x_{i1}	x_{i2}	y_i
4	3.2	Grasso	Pessimista
5	8.2	Magro	Ottimista
6	6.3	Grasso	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
3. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
    (b) ...massimizza l'impurità attesa della variabile di output dei figli.
    (c) ...minimizza l'impurità attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 82

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	3.6	Partecipe
2	Medio	5.8	Svogliato
3	Magro	7.7	Svogliato

i	x_{i1}	x_{i2}	y_i
4	Medio	2.7	Partecipe
5	Grasso	0.9	Svogliato
6	Grasso	8.7	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[-1, 1]$.
    (b) $[0, 1]$.
    (c) $[1, +\infty)$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 83

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	Sedentario	0.90	Pessimista
2	Attivo	0.61	Ottimista
3	Attivo	0.30	Pessimista

i	x_{i1}	x_{i2}	y_i
4	Sedentario	0.12	Ottimista
5	Sportivo	0.39	Pessimista
6	Sportivo	0.80	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n-1$.
   (c) $n(n-1)/2$.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Dipende dal linkage criterion.
    (b) Quelli a distanza minore.
    (c) Quelli a distanza maggiore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 84

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	8.0	Pessimista
2	Medio	3.9	Ottimista
3	Grasso	9.0	Ottimista

i	x_{i1}	x_{i2}	y_i
4	Grasso	1.2	Pessimista
5	Magro	6.1	Pessimista
6	Magro	3.0	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
10. Che significato ha il parametro principale K dell'algoritmo K-means?
    (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
    (b) Il numero di iterazioni dell'algoritmo.
    (c) Il numero di cluster in cui suddividere il dataset.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 85

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	8.2	Medio	Altruista
2	9.2	Grasso	Egoista
3	6.3	Magro	Altruista

i	x_{i1}	x_{i2}	y_i
4	3.2	Magro	Egoista
5	4.1	Medio	Egoista
6	1.4	Grasso	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
2. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n-1)^2$.
   (c) $n(n-1)/2$.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'entropia attesa della variabile di output nei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 86

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	5.7	Pessimista
2	Medio	8.6	Ottimista
3	Grasso	7.6	Pessimista

i	x_{i1}	x_{i2}	y_i
4	Grasso	3.5	Ottimista
5	Medio	0.8	Pessimista
6	Magro	2.6	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[-1, 1]$.
    (b) $[0, 1]$.
    (c) $[1, +\infty)$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 87

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.81	Basso	Pessimista
2	0.91	Alto	Ottimista
3	0.40	Basso	Ottimista

i	x_{i1}	x_{i2}	y_i
4	0.13	Alto	Pessimista
5	0.62	Medio	Pessimista
6	0.31	Medio	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (b) ... massimizza l'impurità attesa della variabile di output dei figli.
    (c) ... minimizza l'impurità attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 88

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	9	Ottimista
2	Medio	36	Pessimista
3	Basso	58	Ottimista

i	x_{i1}	x_{i2}	y_i
4	Medio	77	Ottimista
5	Basso	27	Pessimista
6	Alto	87	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
    (b) Che esiste una forte dipendenza lineare fra le due variabili.
    (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 89

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.84	Medio	Svogliato
2	0.16	Alto	Svogliato
3	0.94	Alto	Partecipe

i	x_{i1}	x_{i2}	y_i
4	0.65	Basso	Svogliato
5	0.43	Medio	Partecipe
6	0.34	Basso	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n - 1)/2$.
   (c) $(n - 1)^2$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[0, 1]$.
    (b) $[-1, 1]$.
    (c) $[1, +\infty)$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 90

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	86	Alto	Pessimista
2	57	Medio	Ottimista
3	8	Alto	Ottimista

i	x_{i1}	x_{i2}	y_i
4	26	Medio	Pessimista
5	35	Basso	Pessimista
6	76	Basso	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza maggiore.
    (b) Quelli a distanza minore.
    (c) Dipende dal linkage criterion.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 91

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	78	Medio	Pessimista
2	88	Alto	Ottimista
3	59	Basso	Pessimista

i	x_{i1}	x_{i2}	y_i
4	37	Medio	Ottimista
5	10	Alto	Pessimista
6	28	Basso	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (b) ... massimizza l'impurità attesa della variabile di output dei figli.
    (c) ... minimizza l'impurità attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 92

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	2.8	Medio	Ottimista
2	3.7	Grasso	Ottimista
3	8.8	Magro	Ottimista

i	x_{i1}	x_{i2}	y_i
4	7.8	Grasso	Pessimista
5	5.9	Medio	Pessimista
6	1.0	Magro	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n - 1)/2$.
   (c) $(n - 1)^2$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'entropia attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 93

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	5.9	Grasso	Pessimista
2	2.8	Grasso	Ottimista
3	3.7	Magro	Ottimista

i	x_{i1}	x_{i2}	y_i
4	8.8	Medio	Ottimista
5	1.0	Medio	Pessimista
6	7.8	Magro	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $(n-1)^2$.
    (b) $n-1$.
    (c) $n(n-1)/2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 94

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	92	Altruista
2	Grasso	63	Egoista
3	Magro	41	Altruista

i	x_{i1}	x_{i2}	y_i
4	Medio	14	Egoista
5	Grasso	32	Altruista
6	Magro	82	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[1, +\infty)$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n - 1$.
    (b) $(n - 1)^2$.
    (c) $n(n - 1)/2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 95

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	8.3	Triste
2	Magro	4.2	Felice
3	Medio	9.3	Felice

i	x_{i1}	x_{i2}	y_i
4	Medio	1.5	Triste
5	Grasso	6.4	Triste
6	Grasso	3.3	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n-1$.
   (c) $n(n-1)/2$.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
3. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
    (c) Che esiste una forte dipendenza lineare fra le due variabili.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 96

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	8.5	Alto	Partecipe
2	3.4	Medio	Partecipe
3	5.6	Basso	Svogliato

i	x_{i1}	x_{i2}	y_i
4	2.5	Basso	Partecipe
5	0.7	Alto	Svogliato
6	7.5	Medio	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
8. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... massimizza l'impurità attesa della variabile di output dei figli.
    (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (c) ... minimizza l'impurità attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 97

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	2.7	Ottimista
2	Grasso	5.8	Pessimista
3	Magro	3.6	Ottimista

i	x_{i1}	x_{i2}	y_i
4	Magro	7.7	Pessimista
5	Medio	8.7	Ottimista
6	Medio	0.9	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ...minimizza l'impurità attesa della variabile di output nei figli.
    (b) ...massimizza l'impurità attesa della variabile di output dei figli.
    (c) ...minimizza l'impurità attesa delle variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 98

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	Sedentario	7.5	Egoista
2	Attivo	2.5	Altruista
3	Sportivo	8.5	Altruista

i	x_{i1}	x_{i2}	y_i
4	Attivo	5.6	Egoista
5	Sedentario	3.4	Altruista
6	Sportivo	0.7	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[0, +\infty)$.
    (b) $[-1, 1]$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 99

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.39	Medio	Pessimista
2	0.30	Basso	Pessimista
3	0.80	Medio	Ottimista

i	x_{i1}	x_{i2}	y_i
4	0.12	Alto	Ottimista
5	0.61	Basso	Ottimista
6	0.90	Alto	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
2. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n - 1$.
   (c) $n(n-1)/2$.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
    (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
    (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 100

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	90	Medio	Triste
2	30	Grasso	Triste
3	61	Grasso	Felice

i	x_{i1}	x_{i2}	y_i
4	12	Medio	Felice
5	80	Magro	Felice
6	39	Magro	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
2. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
    (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
    (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
    (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 101

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	82	Medio	Pessimista
2	92	Basso	Ottimista
3	41	Medio	Ottimista

i	x_{i1}	x_{i2}	y_i
4	14	Basso	Pessimista
5	63	Alto	Pessimista
6	32	Alto	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
2. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
3. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n - 1$.
   (c) $n(n-1)/2$.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
9. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[-1, 1]$.
    (b) $[0, 1]$.
    (c) $[0, +\infty)$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 102

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	0.08	Egoista
2	Basso	0.57	Egoista
3	Medio	0.76	Egoista

i	x_{i1}	x_{i2}	y_i
4	Basso	0.26	Altruista
5	Medio	0.35	Altruista
6	Alto	0.86	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
10. Che significato ha il parametro principale K dell'algoritmo K-means?
    (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
    (b) Il numero di cluster in cui suddividere il dataset.
    (c) Il numero di iterazioni dell'algoritmo.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 103

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	92	Triste
2	Medio	41	Triste
3	Magro	32	Triste

i	x_{i1}	x_{i2}	y_i
4	Medio	82	Felice
5	Grasso	14	Felice
6	Magro	63	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[0, 1]$.
    (b) $[0, +\infty)$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 104

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	0.07	Svogliato
2	Grasso	0.75	Svogliato
3	Magro	0.56	Svogliato

i	x_{i1}	x_{i2}	y_i
4	Medio	0.85	Partecipe
5	Magro	0.25	Partecipe
6	Grasso	0.34	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che esiste una forte dipendenza lineare fra le due variabili.
    (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
    (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 105

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	93	Alto	Altruista
2	64	Basso	Egoista
3	83	Medio	Egoista

i	x_{i1}	x_{i2}	y_i
4	33	Basso	Altruista
5	15	Alto	Egoista
6	42	Medio	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...minimizza l'impurità attesa delle variabili di input dei figli.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
9. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Dipende dal linkage criterion.
    (b) Quelli a distanza minore.
    (c) Quelli a distanza maggiore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 106

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	7.7	Medio	Altruista
2	8.7	Basso	Egoista
3	5.8	Alto	Altruista

i	x_{i1}	x_{i2}	y_i
4	2.7	Alto	Egoista
5	3.6	Medio	Egoista
6	0.9	Basso	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n-1$.
   (c) $(n-1)^2$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[0, +\infty)$.
    (b) $[0, 1]$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 107

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	0.09	Pessimista
2	Basso	0.77	Pessimista
3	Basso	0.36	Ottimista

i	x_{i1}	x_{i2}	y_i
4	Alto	0.87	Ottimista
5	Medio	0.58	Pessimista
6	Medio	0.27	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
6. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $(n-1)^2$.
    (b) $n - 1$.
    (c) $n(n-1)/2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 108

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.33	Alto	Partecipe
2	0.84	Basso	Partecipe
3	0.06	Basso	Svogliato

i	x_{i1}	x_{i2}	y_i
4	0.55	Medio	Svogliato
5	0.74	Alto	Svogliato
6	0.24	Medio	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[0, 1]$.
    (b) $[0, +\infty)$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 109

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	9.4	Egoista
2	Medio	6.5	Altruista
3	Grasso	4.3	Egoista

i	x_{i1}	x_{i2}	y_i
4	Magro	1.6	Altruista
5	Grasso	8.4	Altruista
6	Medio	3.4	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
8. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di cluster in cui suddividere il dataset.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $(n-1)^2$.
    (b) $n(n-1)/2$.
    (c) $n-1$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 110

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Sedentario	33	Triste
2	Sportivo	83	Felice
3	Attivo	93	Triste

i	x_{i1}	x_{i2}	y_i
4	Sedentario	64	Felice
5	Attivo	15	Felice
6	Sportivo	42	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Dipende dal linkage criterion.
    (b) Quelli a distanza minore.
    (c) Quelli a distanza maggiore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 111

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.36	Medio	Pessimista
2	0.58	Alto	Ottimista
3	0.77	Medio	Ottimista

i	x_{i1}	x_{i2}	y_i
4	0.09	Basso	Ottimista
5	0.27	Alto	Pessimista
6	0.87	Basso	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza minore.
   (c) Quelli a distanza maggiore.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa della variabile di output nei figli.
   (b) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $n - 1$.
   (c) $(n-1)^2$.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'entropia attesa della variabile di output nei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 112

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	5.5	Felice
2	Grasso	3.3	Triste
3	Grasso	7.4	Felice

i	x_{i1}	x_{i2}	y_i
4	Medio	2.4	Triste
5	Magro	8.4	Triste
6	Magro	0.6	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'entropia attesa della variabile di output nei figli.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza minore.
    (b) Dipende dal linkage criterion.
    (c) Quelli a distanza maggiore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 113

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 100], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	36	Basso	Pessimista
2	27	Alto	Pessimista
3	87	Medio	Pessimista

i	x_{i1}	x_{i2}	y_i
4	77	Basso	Ottimista
5	58	Alto	Ottimista
6	9	Medio	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 100 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
4. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[0, 1]$.
    (b) $[-1, 1]$.
    (c) $[1, +\infty)$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 114

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	5.5	Basso	Egoista
2	0.6	Alto	Egoista
3	2.4	Basso	Altruista

i	x_{i1}	x_{i2}	y_i
4	3.3	Medio	Altruista
5	7.4	Medio	Egoista
6	8.4	Alto	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
4. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n - 1$.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
    (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
    (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
    (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 115

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	2.4	Magro	Altruista
2	5.5	Magro	Egoista
3	3.3	Medio	Altruista

i	x_{i1}	x_{i2}	y_i
4	0.6	Grasso	Egoista
5	7.4	Medio	Egoista
6	8.4	Grasso	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
4. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
5. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[0, 1]$.
    (b) $[-1, 1]$.
    (c) $[0, +\infty)$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 116

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	2.6	Felice
2	Medio	3.5	Felice
3	Grasso	5.7	Triste

i	x_{i1}	x_{i2}	y_i
4	Medio	7.6	Triste
5	Magro	8.6	Felice
6	Magro	0.8	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa della variabile di output nei figli.
   (c) ...minimizza l'impurità attesa delle variabili di input dei figli.
3. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza minore.
5. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
6. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[1, +\infty)$.
    (b) $[-1, 1]$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 117

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0,1], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.06	Sedentario	Altruista
2	0.24	Sportivo	Egoista
3	0.84	Sedentario	Egoista

i	x_{i1}	x_{i2}	y_i
4	0.74	Attivo	Altruista
5	0.55	Sportivo	Altruista
6	0.33	Attivo	Egoista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
6. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[0, 1]$.
    (b) $[1, +\infty)$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 118

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Grasso	0.84	Altruista
2	Medio	0.34	Egoista
3	Magro	0.94	Egoista

i	x_{i1}	x_{i2}	y_i
4	Magro	0.16	Altruista
5	Grasso	0.43	Egoista
6	Medio	0.65	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
4. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...massimizza l'impurità attesa della variabile di output dei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n-1$.
   (c) $n(n-1)/2$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[0, +\infty)$.
    (b) $[0, 1]$.
    (c) $[-1, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 119

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Basso	0.80	Ottimista
2	Medio	0.61	Ottimista
3	Basso	0.39	Pessimista

i	x_{i1}	x_{i2}	y_i
4	Alto	0.90	Pessimista
5	Alto	0.12	Ottimista
6	Medio	0.30	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
10. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
    (a) $[1, +\infty)$.
    (b) $[-1, 1]$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 120

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad x_{i2} \in [0, 1], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Alto	0.11	Partecipe
2	Basso	0.38	Svogliato
3	Medio	0.29	Svogliato

i	x_{i1}	x_{i2}	y_i
4	Medio	0.60	Partecipe
5	Alto	0.89	Svogliato
6	Basso	0.79	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 1 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.
2. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
7. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
10. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
    (a) ... minimizza l'entropia attesa della variabile di output nei figli.
    (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
    (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 121

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0,1], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.58	Attivo	Felice
2	0.77	Sedentario	Felice
3	0.36	Sedentario	Triste

i	x_{i1}	x_{i2}	y_i
4	0.27	Attivo	Triste
5	0.09	Sportivo	Felice
6	0.87	Sportivo	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
2. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
3. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
4. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
6. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
7. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (b) Il numero di cluster in cui suddividere il dataset.
   (c) Il numero di iterazioni dell'algoritmo.
8. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa della variabile di output nei figli.
   (b) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (c) ...massimizza l'impurità attesa della variabile di output dei figli.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
10. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
    (a) $n(n-1)/2$.
    (b) $n - 1$.
    (c) $(n-1)^2$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 122

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	3.4	Medio	Pessimista
2	0.7	Grasso	Ottimista
3	7.5	Medio	Ottimista

i	x_{i1}	x_{i2}	y_i
4	5.6	Magro	Ottimista
5	8.5	Grasso	Pessimista
6	2.5	Magro	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di iterazioni dell'algoritmo.
   (c) Il numero di vicini da considerare nella costruzione di ciascun cluster.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'entropia attesa della variabile di output nei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
9. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[-1, 1]$.
   (c) $[0, 1]$.
10. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
    (a) $[0, +\infty)$.
    (b) $[-1, 1]$.
    (c) $[0, 1]$.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 123

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Sedentario}, \texttt{Attivo}, \texttt{Sportivo}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$

i	x_{i1}	x_{i2}	y_i
1	0.38	Attivo	Ottimista
2	0.11	Sedentario	Pessimista
3	0.29	Sportivo	Ottimista

i	x_{i1}	x_{i2}	y_i
4	0.60	Sportivo	Pessimista
5	0.89	Sedentario	Ottimista
6	0.79	Attivo	Pessimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
5. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
6. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
7. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
8. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza minore.
   (b) Dipende dal linkage criterion.
   (c) Quelli a distanza maggiore.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
    (a) Che esiste una forte dipendenza lineare fra le due variabili.
    (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
    (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 124

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	6.5	Medio	Svogliato
2	4.3	Alto	Partecipe
3	8.4	Alto	Svogliato

i	x_{i1}	x_{i2}	y_i
4	3.4	Medio	Partecipe
5	9.4	Basso	Partecipe
6	1.6	Basso	Svogliato

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... massimizza l'informazione mutua fra le variabili di input dei figli.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[0, 1]$.
   (b) $[1, +\infty)$.
   (c) $[-1, 1]$.
5. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
6. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n(n-1)/2$.
   (c) $n-1$.
7. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
8. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
9. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Dipende dal linkage criterion.
    (b) Quelli a distanza minore.
    (c) Quelli a distanza maggiore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 125

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 100], \quad y_i \in \{\texttt{Partecipe}, \texttt{Svogliato}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Medio	60	Svogliato
2	Magro	79	Svogliato
3	Medio	29	Partecipe

i	x_{i1}	x_{i2}	y_i
4	Magro	38	Partecipe
5	Grasso	11	Svogliato
6	Grasso	89	Partecipe

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 100 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n-1$.
   (c) $n(n-1)/2$.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y)=0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0,+\infty)$.
   (b) $[0,1]$.
   (c) $[-1,1]$.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y)=1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
7. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ...minimizza l'entropia attesa della variabile di output nei figli.
   (c) ...minimizza l'informazione mutua fra le variabili di input dei figli.
8. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1,1]$.
   (b) $[0,+\infty)$.
   (c) $[0,1]$.
9. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1,+\infty)$.
   (b) $[0,1]$.
   (c) $[-1,1]$.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza maggiore.
    (b) Quelli a distanza minore.
    (c) Dipende dal linkage criterion.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 126

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Ottimista}, \texttt{Pessimista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.40	Magro	Ottimista
2	0.91	Medio	Ottimista
3	0.62	Grasso	Pessimista

i	x_{i1}	x_{i2}	y_i
4	0.13	Medio	Pessimista
5	0.81	Magro	Pessimista
6	0.31	Grasso	Ottimista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $(n - 1)^2$.
   (c) $n(n - 1)/2$.
2. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
3. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
4. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di iterazioni dell'algoritmo.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di cluster in cui suddividere il dataset.
5. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
6. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (b) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
8. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[1, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
9. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
10. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
    (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
    (b) ... massimizza l'impurità attesa della variabile di output dei figli.
    (c) ... minimizza l'impurità attesa della variabile di output nei figli.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 127

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad x_{i2} \in [0, 10], \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	Magro	3.3	Triste
2	Grasso	9.3	Triste
3	Medio	4.2	Triste

i	x_{i1}	x_{i2}	y_i
4	Medio	8.3	Felice
5	Magro	6.4	Felice
6	Grasso	1.5	Felice

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i2} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i1} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}|) & \text{se } \boldsymbol{x}_{i1} = \boldsymbol{x}_{j1} \\ 10 - |\boldsymbol{x}_{i2} - \boldsymbol{x}_{j2}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
3. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... massimizza l'impurità attesa della variabile di output dei figli.
   (c) ... minimizza l'impurità attesa della variabile di output nei figli.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
7. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[0, +\infty)$.
8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $(n-1)^2$.
   (b) $n - 1$.
   (c) $n(n-1)/2$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Quelli a distanza minore.
    (b) Dipende dal linkage criterion.
    (c) Quelli a distanza maggiore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 128

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Alto}, \texttt{Medio}, \texttt{Basso}\}, \quad y_i \in \{\texttt{Felice}, \texttt{Triste}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.39	Alto	Felice
2	0.80	Alto	Triste
3	0.90	Medio	Felice

i	x_{i1}	x_{i2}	y_i
4	0.30	Basso	Felice
5	0.12	Medio	Triste
6	0.61	Basso	Triste

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
2. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, +\infty)$.
   (b) $[0, 1]$.
   (c) $[-1, 1]$.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che esiste una forte dipendenza lineare fra le due variabili.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... massimizza l'impurità attesa della variabile di output dei figli.
7. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n(n-1)/2$.
   (b) $(n-1)^2$.
   (c) $n-1$.
8. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Quelli a distanza maggiore.
   (b) Quelli a distanza minore.
   (c) Dipende dal linkage criterion.
9. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... minimizza l'entropia attesa della variabile di output nei figli.
   (b) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'informazione mutua fra le variabili di input dei figli.
10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
    (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
    (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
    (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 129

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 1], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	0.30	Grasso	Altruista
2	0.61	Grasso	Egoista
3	0.12	Magro	Egoista

i	x_{i1}	x_{i2}	y_i
4	0.80	Medio	Egoista
5	0.90	Magro	Altruista
6	0.39	Medio	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 1 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[-1, 1]$.
   (c) $[0, +\infty)$.
2. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
3. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n - 1)/2$.
   (c) $(n - 1)^2$.
4. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
5. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ... massimizza l'impurità attesa della variabile di output dei figli.
   (b) ... minimizza l'impurità attesa della variabile di output nei figli.
   (c) ... minimizza l'impurità attesa delle variabili di input dei figli.
6. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ... massimizza l'informazione mutua fra le variabili di input dei figli.
   (b) ... minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ... minimizza l'entropia attesa della variabile di output nei figli.
7. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
8. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[-1, 1]$.
   (b) $[0, +\infty)$.
   (c) $[0, 1]$.
9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che esiste una forte dipendenza lineare fra le due variabili.
   (c) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
10. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
    (a) Dipende dal linkage criterion.
    (b) Quelli a distanza minore.
    (c) Quelli a distanza maggiore.

Algoritmi Avanzati / Machine Learning for Computer Science, A.A. 2017–2018

Seconda prova parziale, tema 130

Mercoledì 20 dicembre 2017

- **Nota bene: chi non segue queste indicazioni rischia l'annnullamento della prova.**
- Al termine dello svolgimento della prova, è necessario riconsegnare **tutti** i fogli, comprese le brutte copie e il presente testo.
- Il presente foglio non deve riportare alcuna scritta.
- Riportare il proprio nome, cognome e numero di matricola e il numero di tema in testa a tutti i fogli protocollo, di bella e di brutta copia.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 11 punti ciascuno. Le 10 domande dell'esercizio 3 valgono 1 punto ciascuna ($+1$ se la risposta è corretta, -1 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset di $m = 6$ campioni, $n = 2$ attributi scalari e una variabile dipendente categorica:

$$x_{i1} \in [0, 10], \quad x_{i2} \in \{\texttt{Grasso}, \texttt{Medio}, \texttt{Magro}\}, \quad y_i \in \{\texttt{Altruista}, \texttt{Egoista}\} \qquad i = 1, \ldots, 6.$$

i	x_{i1}	x_{i2}	y_i
1	2.9	Magro	Altruista
2	1.1	Grasso	Egoista
3	6.0	Magro	Egoista

i	x_{i1}	x_{i2}	y_i
4	3.8	Medio	Altruista
5	7.9	Medio	Egoista
6	8.9	Grasso	Altruista

1.1) Stimare il coefficiente di impurità di Gini e l'entropia della variabile di uscita sulla base delle 6 osservazioni.
1.2) Costruire la radice di un albero di decisione basato sull'impurità di Gini. Per la variabile x_{i1} considerare solo una divisione dicotomica basata sulla mediana; per la variabile x_{i2} considerare un figlio per ciascuno dei tre valori (in stile ID3).
1.3) Completare l'albero di decisione basando il secondo livello sulla variabile non utilizzata nel primo. Qual è l'impurità di Gini di ciascuna foglia?

Esercizio 2

Utilizzando lo stesso dataset dell'esercizio 1, consideriamo solamente le variabili di ingresso x_{i1}, x_{i2} e definiamo la seguente funzione di similarità fra gli elementi del dataset:

$$\text{sim}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 2 \cdot (10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}|) & \text{se } \boldsymbol{x}_{i2} = \boldsymbol{x}_{j2} \\ 10 - |\boldsymbol{x}_{i1} - \boldsymbol{x}_{j1}| & \text{altrimenti.} \end{cases}$$

2.1) Costruire la matrice delle distanze ed eseguire l'algoritmo di clustering agglomerativo gerarchico utilizzando il single linkage criterion per la similarità fra cluster. Disegnare il dendrogramma risultante.
2.2) Ripetere l'esercizio utilizzando il complete linkage criterion.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Non segnare in alcun modo le domande e le risposte su questo foglio **pena l'annullamento della prova**.
In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo...
   (a) ...minimizza l'entropia attesa della variabile di output nei figli.
   (b) ...minimizza l'informazione mutua fra le variabili di input dei figli.
   (c) ...massimizza l'informazione mutua fra le variabili di input dei figli.
2. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 0$?
   (a) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (b) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.
   (c) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
3. Che significato ha il parametro principale K dell'algoritmo K-means?
   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.
4. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?
   (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.
   (c) Che esiste una forte dipendenza lineare fra le due variabili.
5. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?
   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.
6. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo...
   (a) ...minimizza l'impurità attesa delle variabili di input dei figli.
   (b) ...massimizza l'impurità attesa della variabile di output dei figli.
   (c) ...minimizza l'impurità attesa della variabile di output nei figli.
7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?
   (a) Dipende dal linkage criterion.
   (b) Quelli a distanza maggiore.
   (c) Quelli a distanza minore.
8. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?
   (a) $[-1, 1]$.
   (b) $[1, +\infty)$.
   (c) $[0, 1]$.
9. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?
   (a) $n - 1$.
   (b) $n(n-1)/2$.
   (c) $(n-1)^2$.
10. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?
    (a) $[0, 1]$.
    (b) $[-1, 1]$.
    (c) $[0, +\infty)$.

Traccia della soluzione del Tema 1

La soluzione è applicabile anche agli altri temi: considerando che gli elementi e le coordinate sono permutati casualmente, leggermente perturbati e riscalati, i risultati sono gli stessi, anche se l'ordine può cambiare. In particolare, nel primo esercizio la variabile da usare alla radice è sempre quella numerica e l'albero di decisione termina al secondo livello con sei foglie pure.

Esercizio 1

1.1) La variabile di uscita, y_i, è equidistribuita fra due valori, quindi il suo coefficiente di impurità vale:

$$GI(Y) = 1 - \Pr(Y = \texttt{Altruista})^2 - \Pr(Y = \texttt{Egoista})^2 = 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2}.$$

Allo stesso modo, l'entropia vale

$$H(Y) = -2\left(\frac{1}{2}\log_2\frac{1}{2}\right) = 1.$$

1.2) Per quanto riguarda la variabile numerica x_{i1}, la mediana θ lascia per definizione metà dei valori in un nodo e l'altra metà nell'altro. In questo caso, i tre valori y_i corrispondenti a $x_{i1} \leq \theta$ sono:

$$y_1 = \texttt{Altruista}, \quad y_4 = \texttt{Egoista}, \quad y_6 = \texttt{Egoista},$$

con un coefficiente di impurità di Gini pari a

$$GI(Y|X_1 \leq \theta) = 1 - \Pr(Y = \texttt{Altruista}|X_1 \leq \theta)^2 - \Pr(Y = \texttt{Egoista}|X_1 \leq \theta)^2 = 1 - \frac{1}{9} - \frac{4}{9} = \frac{4}{9}.$$

Allo stesso modo, i tre valori y_i corrispondenti a $x_{i1} > \theta$ sono:

$$y_2 = \texttt{Altruista}, \quad y_3 = \texttt{Egoista}, \quad y_5 = \texttt{Altruista};$$

Dato che i valori di probabilità sono nuovamente $1/3$ e $2/3$, il coefficiente di Gini è lo stesso del caso precedente:

$$GI(Y|X_1 > \theta) = 1 - \frac{1}{9} - \frac{4}{9} = \frac{4}{9}.$$

Di conseguenza, l'impurità di Gini attesa in seguito all'uso della prima variabile nel nodo radice è

$$GI(Y|X_1) = \frac{4}{9}.$$

Se invece usiamo la seconda colonna come radice, osserviamo che i tre figli risultanti contengono i seguenti campioni:

- Per $x_{i2} = \texttt{Grasso}$: $y_1 = \texttt{Altruista}$, $y_3 = \texttt{Egoista}$;
- Per $x_{i2} = \texttt{Medio}$: $y_4 = \texttt{Egoista}$, $y_5 = \texttt{Altruista}$;
- Per $x_{i2} = \texttt{Magro}$: $y_2 = \texttt{Altruista}$, $y_6 = \texttt{Egoista}$.

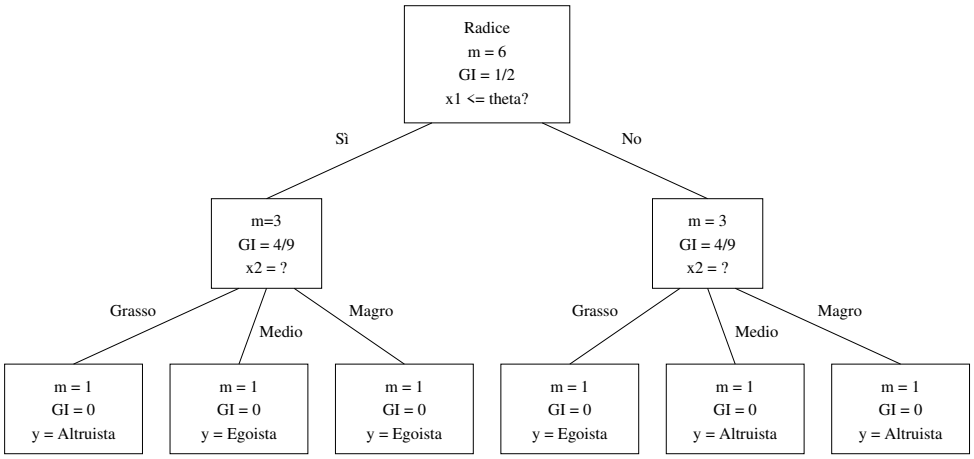
In tutt'e tre i nodi la distribuzione dell'output è uniforme, quindi l'impurità attesa di Gini resta

$$GI(Y|X_2) = \frac{1}{2},$$

senza nessun guadagno rispetto alla situazione iniziale.

Scegliamo dunque la prima colonna (quella numerica) per la radice dell'albero.

1.3) Usando la seconda colonna al livello successivo dell'albero, il dataset risulta spezzato in sei foglie pure:



Esercizio 2

2.1) La funzione di similarità si basa sulla distanza fra le coordinate numeriche (decresce quando la distanza cresce) e raddoppia se le coordinate categoriche dei due elementi sono uguali. Ad esempio:

$$\text{sim}(\boldsymbol{x}_1, \boldsymbol{x}_2) = 1 - |0.1 - 0.78| = 0.32;$$

$$\text{sim}(\boldsymbol{x}_1, \boldsymbol{x}_3) = 2(1 - |0.1 - 0.88|) = 2 \cdot 0.22 = 0.44.$$

La tabella completa (tralasciando per comodità le simmetrie, e indicando ogni elemento con il suo indice) è la seguente:

	2	3	4	5	6
1	0.32	0.44	0.82	0.51	0.73
2		0.90	0.50	0.81	1.18
3			0.40	0.71	0.49
4				1.38	0.91
5					0.78

Il primo passo consiste ovviamente nella scelta della massima similitudine. In questo caso,

$$\text{sim}(\boldsymbol{x}_4, \boldsymbol{x}_5) = 1.38;$$

Una volta raccolti i due elementi in un cluster, ricelcoliamo le distanze del cluster appena formato dagli altri elementi sulla base del single linkage criterion. Ad esempio,

$$\text{sim}\big(\{\boldsymbol{x}_4, \boldsymbol{x}_5\}, \boldsymbol{x}_1\big) = \max\big\{\text{sim}(\boldsymbol{x}_4, \boldsymbol{x}_1), \text{sim}(\boldsymbol{x}_5, \boldsymbol{x}_1)\big\} = \max\{0.82, 0.51\} = 0.82.$$

Dopo la prima unione, la tabella delle similarità è quindi la seguente:

	2	3	45	6
1	0.32	0.44	0.82	0.73
2		0.90	0.81	1.18
3			0.71	0.49
45				0.91

Ora la massima similitudine è

$$\text{sim}(\boldsymbol{x}_2, \boldsymbol{x}_6) = 1.18,$$

e in seguito all'unione di questi due cluster otteniamo:

	3	45	26
1	0.44	0.82	0.73
3		0.71	0.90
45			0.91

Il passo successivo vede l'unione dei due cluster appena formati:

$$\text{sim}\big(\{\boldsymbol{x}_2, \boldsymbol{x}_6\}, \{\boldsymbol{x}_4, \boldsymbol{x}_5\}\big) = 0.91.$$

Una volta uniti i due cluster, le similitudini sono:

	3	4526
1	0.44	0.82
3		0.90

In seguito si unisce l'elemento $\boldsymbol{x}_3$ al cluster appena formato, con similitudine

$$\text{sim}\big(\boldsymbol{x}_3, \{\boldsymbol{x}_2, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6\}\big) = 0.90.$$

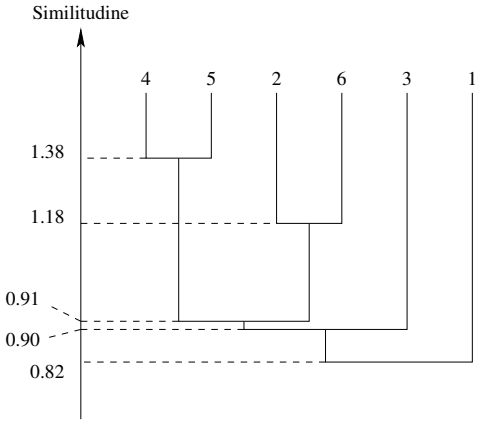
Ecco la nuova tabella:

	34526
1	0.82

Quindi si unisce $\boldsymbol{x}_1$ al resto con similitudine

$$\text{sim}\big(\boldsymbol{x}_1, \{\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6\}\big) = 0.82.$$

Il dendrogramma risultante è dunque:



2.2) Il primo passo consiste comunque nell'unione degli elementi $\boldsymbol{x}_4$ e $\boldsymbol{x}_5$ con similitudine 1.38. Cambia però la rideterminazione delle similitudini fra cluster, questa volta basate sul complete linkage criterion. Ad esempio,

$$\text{sim}\big(\{\boldsymbol{x}_4, \boldsymbol{x}_5\}, \boldsymbol{x}_1\big) = \min\big\{\text{sim}(\boldsymbol{x}_4, \boldsymbol{x}_1), \text{sim}(\boldsymbol{x}_5, \boldsymbol{x}_1)\big\} = \min\{0.82, 0.51\} = 0.51.$$

La tabella risultante dalla prima aggregazione è

	2	3	45	6
1	0.32	0.44	0.51	0.73
2		0.90	0.50	1.18
3			0.40	0.49
45				0.78

$$\text{sim}(\boldsymbol{x}_2, \boldsymbol{x}_6) = 1.18$$

	3	45	26
1	0.44	0.51	0.32
3		0.40	0.49
45			0.50

$$\text{sim}\big(\boldsymbol{x}_1, \{\boldsymbol{x}_4, \boldsymbol{x}_5\}\big) = 0.51$$

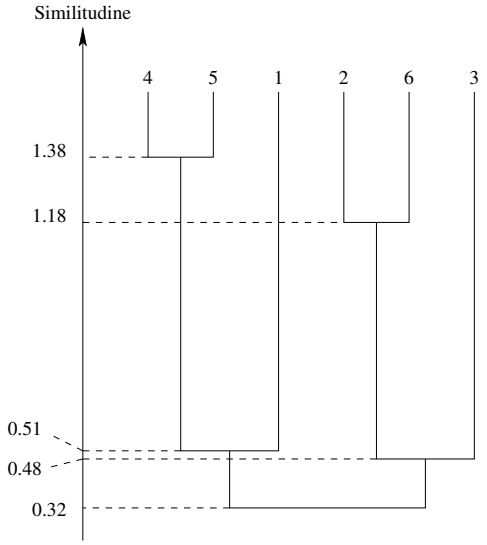
	3	26
145	0.40	0.32
3		0.49

$$\text{sim}\big(\boldsymbol{x}_3, \{\boldsymbol{x}_2, \boldsymbol{x}_6\}\big) = 0.49$$

	326
145	0.32

$$\text{sim}\big(\{\boldsymbol{x}_1, \boldsymbol{x}_4, \boldsymbol{x}_5\}, \{\boldsymbol{x}_3, \boldsymbol{x}_2, \boldsymbol{x}_6\}\big) = 0.32.$$

Il dendrogramma diventa dunque:



Esercizio 3

Nel seguente elenco la risposta corretta è riportata per prima.

1. In un albero di decisione addestrato in base all'information gain, la decisione attribuita a un nodo…

   (a) …minimizza l'entropia attesa della variabile di output nei figli.

   (b) …massimizza l'informazione mutua fra le variabili di input dei figli.

   (c) …minimizza l'informazione mutua fra le variabili di input dei figli.

   Il fattore da valutare è sempre l'entropia della varibile di output, in quanto misura dell'incertezza del valore da prevedere.

2. In un albero di decisione addestrato in base all'impurità di Gini, la decisione attribuita a un nodo…

   (a) …minimizza l'impurità attesa della variabile di output nei figli.

   (b) …massimizza l'impurità attesa della variabile di output dei figli.

   (c) …minimizza l'impurità attesa delle variabili di input dei figli.

   L'obiettivo di un albero di decisione è di avere nodi puri, quindi di minimizzare l'impurità. Come nella domanda precedente, la variabile di cui ci interessa valutare l'incertezza è sempre l'output.

3. Qual è l'intervallo di variabilità dell'entropia di una distribuzione di probabilità discreta?

   (a) $[0, +\infty)$.

   (b) $[0, 1]$.

(c) $[-1, 1]$.

L'entropia di una variabile discreta non è mai negativa, e può assumere qualsiasi valore, a partire da 0 (esito certo). Per rendersi conto che il suo valore non è limitato, basta considerare la sua interpretazione come "numero di bit" necessari a rappresentare l'informazione.

4. Qual è l'intervallo di variabilità dell'impurità di Gini di una distribuzione di probabilità discreta?

   (a) $[0, 1]$.
   (b) $[0, +\infty)$.
   (c) $[-1, 1]$.

   L'impurità di Gini è una probabilità, qindi varia tra 0 e 1. In realtà, il valore 1 non è ottenibile.

5. Qual è l'intervallo di variabilità dell'indice di correlazione di Pearson fra due variabili casuali discrete?

   (a) $[-1, 1]$.
   (b) $[0, 1]$.
   (c) $[1, +\infty)$.

   La correlazione è una covarianza normalizzata, e può assumere valori negativi.

6. Che significato ha il parametro principale K dell'algoritmo K-means?

   (a) Il numero di cluster in cui suddividere il dataset.
   (b) Il numero di vicini da considerare nella costruzione di ciascun cluster.
   (c) Il numero di iterazioni dell'algoritmo.

   K rappresenta il numero di centroidi o prototipi. Da non confondere, ovviamente, con l'omonimo parametro dell'algoritmo KNN. Il numero di iterazioni non è generalmente prefissato.

7. Quali due cluster vengono uniti in un'iterazione dell'algoritmo di clustering agglomerativo gerarchico?

   (a) Quelli a distanza minore.
   (b) Quelli a distanza maggiore.
   (c) Dipende dal linkage criterion.

   I due cluster da unire sono sempre i più simili (o meno distanti), indipendentemente dal linkage criterion, che entra in gioco solo nella determinazione di queste distanze.

8. Quante iterazioni sono necessarie per un'esecuzione completa dell'algoritmo di clustering gerarchico agglomerativo su un insieme di n elementi?

   (a) $n - 1$.
   (b) $n(n - 1)/2$.
   (c) $(n - 1)^2$.

   Si parte da n cluster e ad ogni iterazione se ne uniscono due, riducendo di uno il numero comlessivo. Si termina quando c'è un solo cluster.

9. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X; Y) = 0$?

   (a) Che le due variabili sono indipendenti: la conoscenza dell'esito di X non ci dice nulla sull'esito di Y.
   (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, non è necessaria alcuna informazione aggiuntiva per conoscere l'esito di Y.
   (c) Che non sappiamo nulla sulla possibile dipendenza fra le due variabili.

Significa che l'entropia di X non varia se la si condiziona alla conoscenza di Y.

10. Date due variabili casuali discrete X e Y, che cosa possiamo dire se la loro informazione mutua vale $I(X;Y) = 1$?

    (a) Che le due variabili sono dipendenti: la conoscenza dell'esito di X riduce l'informazione necessaria a comunicare l'esito di Y.

    (b) Che le due variabili sono completamente dipendenti: se conosciamo l'esito di X, questo ci basta per determinare l'esito di Y.

    (c) Che esiste una forte dipendenza lineare fra le due variabili.

    L'informazione mutua rappresenta la diminuzione dell'entropia di X quando si conosce Y. In questo caso la diminuzione c'è. L'entropia non misura dipendenze lineari. Si osservi che, dato che l'entropia può assumere qualunque valore positivo, una diminuzione pari a 1 non rappresenta necessariamente una dipendenza completa.

Griglie di soluzione

Elenco delle risposte corrette per il terzo esercizio.

Tema 1

Esercizio 3: `1.b 2.c 3.b 4.b 5.b 6.c 7.a 8.b 9.b 10.c`

Tema 2

Esercizio 3: `1.c 2.a 3.b 4.a 5.b 6.b 7.a 8.b 9.c 10.a`

Tema 3

Esercizio 3: `1.b 2.c 3.a 4.a 5.b 6.c 7.b 8.a 9.a 10.a`

Tema 4

Esercizio 3: `1.c 2.c 3.c 4.b 5.b 6.b 7.b 8.a 9.b 10.c`

Tema 5

Esercizio 3: `1.b 2.c 3.b 4.c 5.b 6.b 7.c 8.a 9.b 10.c`

Tema 6

Esercizio 3: `1.a 2.a 3.b 4.a 5.b 6.c 7.b 8.a 9.a 10.a`

Tema 7

Esercizio 3: `1.a 2.c 3.a 4.a 5.a 6.b 7.b 8.a 9.b 10.c`

Tema 8

Esercizio 3: `1.a 2.c 3.a 4.b 5.b 6.c 7.c 8.b 9.a 10.a`

Tema 9

Esercizio 3: `1.b 2.b 3.b 4.c 5.b 6.b 7.a 8.b 9.c 10.a`

Tema 10

Esercizio 3: `1.b 2.a 3.a 4.a 5.c 6.b 7.b 8.c 9.b 10.c`

Tema 11

Esercizio 3: `1.b 2.c 3.a 4.a 5.b 6.b 7.a 8.b 9.b 10.b`

Tema 12

Esercizio 3: `1.c 2.a 3.b 4.b 5.a 6.b 7.c 8.b 9.a 10.a`

Tema 13

Esercizio 3: `1.c 2.b 3.c 4.c 5.b 6.c 7.c 8.b 9.b 10.b`

Tema 14

Esercizio 3: `1.b 2.b 3.c 4.b 5.a 6.c 7.b 8.b 9.b 10.a`

Tema 15

Esercizio 3: `1.a 2.b 3.c 4.a 5.a 6.c 7.a 8.b 9.c 10.a`

Tema 16

Esercizio 3: `1.c 2.b 3.b 4.b 5.a 6.c 7.b 8.b 9.c 10.b`

Tema 17

Esercizio 3: `1.a 2.b 3.c 4.b 5.a 6.c 7.c 8.a 9.a 10.b`

Tema 18
Esercizio 3: `1.c 2.c 3.a 4.b 5.c 6.b 7.a 8.a 9.a 10.c`

Tema 19
Esercizio 3: `1.b 2.b 3.c 4.c 5.c 6.a 7.c 8.a 9.c 10.a`

Tema 20
Esercizio 3: `1.a 2.a 3.b 4.c 5.b 6.b 7.c 8.c 9.b 10.b`

Tema 21
Esercizio 3: `1.a 2.c 3.a 4.a 5.a 6.c 7.b 8.b 9.a 10.b`

Tema 22
Esercizio 3: `1.b 2.c 3.b 4.a 5.b 6.c 7.b 8.a 9.b 10.c`

Tema 23
Esercizio 3: `1.c 2.a 3.c 4.c 5.c 6.a 7.c 8.a 9.b 10.c`

Tema 24
Esercizio 3: `1.c 2.a 3.b 4.b 5.b 6.b 7.b 8.c 9.b 10.c`

Tema 25
Esercizio 3: `1.b 2.b 3.b 4.b 5.b 6.c 7.b 8.a 9.c 10.a`

Tema 26
Esercizio 3: `1.b 2.a 3.c 4.a 5.c 6.b 7.c 8.c 9.b 10.a`

Tema 27
Esercizio 3: `1.c 2.a 3.c 4.c 5.b 6.b 7.a 8.c 9.b 10.c`

Tema 28
Esercizio 3: `1.a 2.a 3.a 4.b 5.a 6.a 7.a 8.c 9.b 10.c`

Tema 29
Esercizio 3: `1.b 2.c 3.a 4.c 5.c 6.c 7.a 8.a 9.c 10.a`

Tema 30
Esercizio 3: `1.a 2.b 3.c 4.a 5.b 6.a 7.b 8.c 9.a 10.a`

Tema 31
Esercizio 3: `1.a 2.b 3.c 4.a 5.a 6.a 7.c 8.a 9.c 10.c`

Tema 32
Esercizio 3: `1.c 2.c 3.c 4.b 5.a 6.c 7.c 8.b 9.a 10.a`

Tema 33
Esercizio 3: `1.a 2.a 3.c 4.c 5.c 6.b 7.a 8.b 9.c 10.b`

Tema 34
Esercizio 3: `1.c 2.b 3.c 4.a 5.b 6.c 7.a 8.b 9.b 10.a`

Tema 35
Esercizio 3: `1.c 2.b 3.c 4.a 5.b 6.c 7.a 8.a 9.b 10.b`

Tema 36
Esercizio 3: 1.a 2.c 3.b 4.a 5.a 6.a 7.b 8.c 9.a 10.a

Tema 37
Esercizio 3: 1.b 2.c 3.b 4.b 5.b 6.b 7.c 8.c 9.b 10.b

Tema 38
Esercizio 3: 1.b 2.c 3.b 4.c 5.b 6.c 7.b 8.c 9.a 10.b

Tema 39
Esercizio 3: 1.b 2.a 3.b 4.c 5.c 6.c 7.c 8.c 9.c 10.c

Tema 40
Esercizio 3: 1.c 2.a 3.a 4.b 5.b 6.b 7.a 8.b 9.b 10.b

Tema 41
Esercizio 3: 1.b 2.a 3.b 4.c 5.c 6.c 7.b 8.a 9.b 10.c

Tema 42
Esercizio 3: 1.b 2.c 3.c 4.a 5.b 6.c 7.c 8.b 9.b 10.a

Tema 43
Esercizio 3: 1.b 2.c 3.c 4.b 5.a 6.a 7.b 8.b 9.c 10.a

Tema 44
Esercizio 3: 1.a 2.b 3.c 4.b 5.b 6.b 7.a 8.c 9.b 10.c

Tema 45
Esercizio 3: 1.c 2.a 3.b 4.c 5.c 6.c 7.c 8.a 9.a 10.a

Tema 46
Esercizio 3: 1.c 2.b 3.c 4.a 5.b 6.c 7.c 8.b 9.b 10.a

Tema 47
Esercizio 3: 1.a 2.b 3.b 4.b 5.c 6.a 7.a 8.c 9.c 10.b

Tema 48
Esercizio 3: 1.c 2.c 3.a 4.a 5.c 6.c 7.c 8.c 9.a 10.c

Tema 49
Esercizio 3: 1.a 2.b 3.b 4.b 5.a 6.a 7.a 8.c 9.a 10.c

Tema 50
Esercizio 3: 1.a 2.b 3.a 4.c 5.c 6.b 7.b 8.b 9.b 10.a

Tema 51
Esercizio 3: 1.a 2.b 3.a 4.b 5.b 6.b 7.b 8.b 9.b 10.b

Tema 52
Esercizio 3: 1.c 2.a 3.a 4.b 5.b 6.a 7.c 8.c 9.b 10.c

Tema 53
Esercizio 3: 1.c 2.a 3.a 4.b 5.b 6.a 7.a 8.c 9.b 10.c

Tema 54
Esercizio 3: 1.a 2.b 3.b 4.b 5.c 6.b 7.a 8.a 9.a 10.c

Tema 55
Esercizio 3: 1.c 2.a 3.b 4.c 5.a 6.b 7.c 8.a 9.b 10.b

Tema 56
Esercizio 3: 1.c 2.a 3.c 4.b 5.a 6.b 7.a 8.a 9.b 10.b

Tema 57
Esercizio 3: 1.b 2.c 3.c 4.c 5.a 6.b 7.a 8.b 9.c 10.a

Tema 58
Esercizio 3: 1.a 2.b 3.c 4.c 5.b 6.a 7.a 8.b 9.c 10.b

Tema 59
Esercizio 3: 1.c 2.a 3.c 4.c 5.b 6.b 7.c 8.a 9.b 10.c

Tema 60
Esercizio 3: 1.a 2.a 3.b 4.c 5.c 6.a 7.b 8.c 9.b 10.b

Tema 61
Esercizio 3: 1.a 2.a 3.c 4.b 5.a 6.a 7.c 8.a 9.a 10.b

Tema 62
Esercizio 3: 1.c 2.c 3.b 4.c 5.b 6.b 7.a 8.c 9.c 10.c

Tema 63
Esercizio 3: 1.b 2.b 3.a 4.c 5.c 6.a 7.a 8.b 9.c 10.b

Tema 64
Esercizio 3: 1.c 2.a 3.b 4.c 5.a 6.a 7.a 8.b 9.b 10.a

Tema 65
Esercizio 3: 1.a 2.c 3.c 4.b 5.b 6.c 7.a 8.a 9.a 10.a

Tema 66
Esercizio 3: 1.c 2.b 3.a 4.b 5.c 6.b 7.a 8.c 9.c 10.b

Tema 67
Esercizio 3: 1.b 2.c 3.a 4.a 5.c 6.c 7.b 8.b 9.c 10.c

Tema 68
Esercizio 3: 1.a 2.a 3.c 4.b 5.a 6.a 7.a 8.a 9.b 10.b

Tema 69
Esercizio 3: 1.b 2.c 3.b 4.b 5.c 6.b 7.c 8.a 9.c 10.b

Tema 70
Esercizio 3: 1.b 2.c 3.b 4.c 5.a 6.c 7.a 8.a 9.b 10.a

Tema 71
Esercizio 3: 1.b 2.a 3.a 4.b 5.c 6.b 7.b 8.a 9.c 10.b

Tema 72
Esercizio 3: 1.c 2.b 3.a 4.b 5.a 6.c 7.a 8.c 9.c 10.b

Tema 73
Esercizio 3: 1.b 2.c 3.b 4.c 5.a 6.c 7.c 8.a 9.c 10.a

Tema 74
Esercizio 3: 1.c 2.c 3.b 4.b 5.b 6.a 7.b 8.b 9.c 10.c

Tema 75
Esercizio 3: 1.b 2.a 3.c 4.c 5.b 6.a 7.c 8.c 9.c 10.c

Tema 76
Esercizio 3: 1.b 2.a 3.c 4.a 5.c 6.a 7.c 8.c 9.a 10.b

Tema 77
Esercizio 3: 1.a 2.b 3.a 4.c 5.b 6.a 7.b 8.c 9.a 10.a

Tema 78
Esercizio 3: 1.c 2.c 3.b 4.c 5.c 6.b 7.b 8.c 9.c 10.b

Tema 79
Esercizio 3: 1.c 2.a 3.b 4.b 5.a 6.a 7.c 8.a 9.c 10.a

Tema 80
Esercizio 3: 1.a 2.a 3.c 4.a 5.a 6.b 7.a 8.b 9.c 10.a

Tema 81
Esercizio 3: 1.b 2.b 3.c 4.b 5.b 6.b 7.c 8.b 9.a 10.c

Tema 82
Esercizio 3: 1.a 2.b 3.c 4.c 5.a 6.c 7.b 8.a 9.b 10.a

Tema 83
Esercizio 3: 1.c 2.b 3.a 4.b 5.b 6.a 7.b 8.c 9.a 10.b

Tema 84
Esercizio 3: 1.c 2.b 3.b 4.a 5.c 6.c 7.b 8.c 9.c 10.c

Tema 85
Esercizio 3: 1.a 2.a 3.a 4.b 5.a 6.b 7.c 8.b 9.b 10.a

Tema 86
Esercizio 3: 1.b 2.a 3.c 4.a 5.c 6.c 7.a 8.c 9.b 10.a

Tema 87
Esercizio 3: 1.a 2.b 3.b 4.c 5.a 6.c 7.c 8.c 9.a 10.c

Tema 88
Esercizio 3: 1.c 2.b 3.b 4.c 5.b 6.b 7.a 8.c 9.c 10.c

Tema 89
Esercizio 3: 1.c 2.a 3.a 4.c 5.b 6.c 7.a 8.a 9.b 10.b

Tema 90
Esercizio 3: 1.c 2.b 3.c 4.b 5.b 6.c 7.b 8.c 9.c 10.b

Tema 91
Esercizio 3: 1.a 2.a 3.b 4.a 5.c 6.c 7.a 8.b 9.a 10.c

Tema 92
Esercizio 3: 1.c 2.b 3.a 4.b 5.a 6.a 7.a 8.b 9.b 10.c

Tema 93
Esercizio 3: 1.a 2.b 3.a 4.b 5.b 6.a 7.b 8.c 9.c 10.b

Tema 94
Esercizio 3: 1.a 2.c 3.c 4.b 5.a 6.c 7.b 8.b 9.c 10.a

Tema 95
Esercizio 3: 1.b 2.b 3.a 4.a 5.a 6.c 7.c 8.c 9.a 10.a

Tema 96
Esercizio 3: 1.c 2.b 3.b 4.a 5.a 6.c 7.a 8.b 9.c 10.c

Tema 97
Esercizio 3: 1.b 2.a 3.b 4.b 5.c 6.b 7.a 8.a 9.b 10.a

Tema 98
Esercizio 3: 1.a 2.c 3.c 4.a 5.a 6.b 7.b 8.a 9.a 10.c

Tema 99
Esercizio 3: 1.c 2.b 3.c 4.a 5.c 6.b 7.a 8.c 9.a 10.a

Tema 100
Esercizio 3: 1.b 2.c 3.a 4.c 5.a 6.b 7.a 8.b 9.a 10.a

Tema 101
Esercizio 3: 1.b 2.a 3.b 4.c 5.c 6.b 7.b 8.b 9.c 10.c

Tema 102
Esercizio 3: 1.a 2.a 3.b 4.a 5.a 6.a 7.b 8.c 9.b 10.b

Tema 103
Esercizio 3: 1.b 2.b 3.b 4.a 5.a 6.a 7.c 8.c 9.c 10.b

Tema 104
Esercizio 3: 1.c 2.c 3.b 4.b 5.c 6.a 7.a 8.c 9.c 10.c

Tema 105
Esercizio 3: 1.c 2.b 3.a 4.a 5.b 6.b 7.b 8.a 9.c 10.b

Tema 106
Esercizio 3: 1.b 2.c 3.a 4.a 5.c 6.c 7.a 8.c 9.b 10.a

Tema 107
Esercizio 3: 1.c 2.a 3.a 4.a 5.c 6.c 7.b 8.b 9.b 10.b

Tema 108
Esercizio 3: 1.c 2.b 3.c 4.a 5.b 6.b 7.b 8.a 9.b 10.b

Tema 109
Esercizio 3: 1.b 2.c 3.b 4.c 5.c 6.a 7.a 8.c 9.a 10.c

Tema 110
Esercizio 3: 1.a 2.c 3.c 4.c 5.a 6.b 7.a 8.c 9.a 10.b

Tema 111
Esercizio 3: 1.b 2.a 3.c 4.c 5.c 6.c 7.c 8.b 9.b 10.a

Tema 112
Esercizio 3: 1.b 2.a 3.c 4.c 5.a 6.a 7.a 8.a 9.c 10.a

Tema 113
Esercizio 3: 1.a 2.c 3.c 4.b 5.a 6.c 7.a 8.a 9.c 10.b

Tema 114
Esercizio 3: 1.b 2.b 3.b 4.c 5.a 6.b 7.b 8.a 9.a 10.a

Tema 115
Esercizio 3: 1.b 2.a 3.a 4.b 5.c 6.a 7.b 8.a 9.c 10.a

Tema 116
Esercizio 3: 1.b 2.b 3.a 4.c 5.a 6.b 7.c 8.a 9.a 10.b

Tema 117
Esercizio 3: 1.b 2.c 3.a 4.c 5.c 6.c 7.a 8.b 9.c 10.c

Tema 118
Esercizio 3: 1.c 2.b 3.b 4.c 5.b 6.c 7.a 8.b 9.a 10.b

Tema 119
Esercizio 3: 1.a 2.c 3.c 4.b 5.b 6.c 7.c 8.c 9.b 10.b

Tema 120
Esercizio 3: 1.a 2.c 3.c 4.c 5.a 6.b 7.c 8.b 9.a 10.a

Tema 121
Esercizio 3: 1.a 2.a 3.a 4.b 5.a 6.a 7.b 8.a 9.a 10.b

Tema 122
Esercizio 3: 1.c 2.c 3.b 4.a 5.c 6.c 7.c 8.b 9.c 10.a

Tema 123
Esercizio 3: 1.a 2.c 3.b 4.a 5.a 6.a 7.c 8.c 9.a 10.b

Tema 124
Esercizio 3: 1.a 2.a 3.c 4.c 5.c 6.c 7.b 8.c 9.b 10.b

Tema 125
Esercizio 3: 1.b 2.c 3.a 4.a 5.c 6.c 7.b 8.c 9.c 10.b

Tema 126

Esercizio 3: `1.a 2.c 3.a 4.c 5.a 6.b 7.b 8.c 9.b 10.c`

Tema 127

Esercizio 3: `1.a 2.c 3.a 4.a 5.c 6.c 7.c 8.b 9.a 10.a`

Tema 128

Esercizio 3: `1.a 2.a 3.a 4.c 5.a 6.b 7.c 8.b 9.a 10.c`

Tema 129

Esercizio 3: `1.a 2.a 3.a 4.a 5.b 6.c 7.b 8.b 9.a 10.b`

Tema 130

Esercizio 3: `1.a 2.c 3.a 4.a 5.b 6.c 7.c 8.a 9.a 10.a`