ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Università di Bologna - Scuola di Scienze

Esame scritto di Calcolo delle Probabilità e Statistica
5 settembre 2018

Esercizio 1

Una tecnologia *antispam* è rappresentata dai **filtri bayesiani**. Essi determinano l'*indice di spamicity* di una e-mail, ovvero la probabilità che sia spam sapendo che contiene determinati *elementi distintivi* (come ad esempio l'espressione *"free money"* oppure *"$$$"*). Se l'indice di spamicity supera una certa soglia (ad esempio 90%), l'e-mail è classificata come spam.

Supponiamo di conoscere le seguenti probabilità (riguardanti ad esempio le e-mail che arrivano alla nostra casella di posta personale):

(a) la probabilità che una *qualsiasi e-mail* (*spam* oppure *non spam*) sia *spam* è pari al 50%;

(b) la probabilità che una *e-mail spam* contenga l'espressione *"free money"* è pari al 70%;

(c) la probabilità che una *e-mail spam* contenga la stringa *"$$$"* è pari all'80%;

(d) la probabilità che una *e-mail spam* contenga entrambe le espressioni *"free money"* e *"$$$"* è pari al 60%;

(e) la probabilità che una *e-mail non spam* contenga l'espressione *"free money"* è pari al 15%;

(f) la probabilità che una *e-mail non spam* contenga la stringa *"$$$"* è pari al 10%;

(g) la probabilità che una *e-mail non spam* contenga entrambe le espressioni *"free money"* e *"$$$"* è pari al 5%.

Scegliamo una e-mail a caso tra quelle che abbiamo ricevuto.

1) Qual è la probabilità che *non sia spam*?

2) Qual è la probabilità che contenga l'espressione *"free money"*?

3) Qual è la probabilità che *non sia spam* sapendo che contiene l'espressione *"free money"*?

4) Si calcoli la probabilità che sia *spam* sapendo che contiene entrambe le espressioni *"free money"* e *"$$$"*? (Tale probabilità è l'*indice di spamicity* della e-mail calcolato da un filtro bayesiano che tiene conto solo delle due espressioni *"free money"* e *"$$$"*.)

Soluzione

Introduciamo gli eventi:

$$\begin{aligned} S &= \text{“è una } \textit{e-mail spam}\text{”} \\ S^c &= \text{“è una } \textit{e-mail non spam}\text{”} \\ E_{fm} &= \text{“contiene l'espressione } \textit{free money}\text{”} \\ E_{\$\$\$} &= \text{“contiene la stringa } \textit{\$\$\$}\text{”} \end{aligned}$$

Dal testo dell'esercizio, sappiamo che:

(a) $P(S) = 0.5$;

(b) $P(E_{fm}|S) = 0.7$;

(c) $P(E_{\$\$\$}|S) = 0.8$;

(d) $P(E_{fm} \cap E_{\$\$\$}|S) = 0.6$;

(e) $P(E_{fm}|S^c) = 0.15$;

(f) $P(E_{\$\$\$}|S^c) = 0.1$;

(g) $P(E_{fm} \cap E_{\$\$\$}|S^c) = 0.05$.

1) $P(S^c) = 1 - P(S) = 50\%$.

2) Utilizzando la formula delle probabilità totali, si ottiene

$$P(E_{fm}) = P(E_{fm}|S)P(S) + P(E_{fm}|S^c)P(S^c) = 42.5\%.$$

3) Utilizzando la formula di Bayes, si ottiene

$$P(S^c|E_{fm}) = \frac{P(E_{fm}|S^c)P(S^c)}{P(E_{fm})} = \frac{0.15 \cdot 0.5}{0.425} = 17.65\%.$$

4) Dalla formula delle probabilità totali, si ha che

$$P(E_{fm} \cap E_{\$\$\$}) = P(E_{fm} \cap E_{\$\$\$}|S)P(S) + P(E_{fm} \cap E_{\$\$\$}|S^c)P(S^c) = 32.5\%.$$

Utilizzando infine la formula di Bayes, si ottiene

$$P(S|E_{fm} \cap E_{\$\$\$}) = \frac{P(E_{fm} \cap E_{\$\$\$}|S)P(S)}{P(E_{fm} \cap E_{\$\$\$})} = \frac{0.6 \cdot 0.5}{0.325} \simeq 92.31\%.$$

Esercizio 2

Un dado non truccato a *quattro* facce viene lanciato due volte. Si considerino le variabili aleatorie X ed Y date da

$$\begin{aligned} X &= \text{“prodotto dei valori apparsi nei due lanci”} \\ Y &= \text{“valore massimo che appare nei due lanci”} \end{aligned}$$

1) Determinare supporto e densità discreta di X.

2) Quanto vale $\mathbb{P}(X > 6)$?

3) Trovare la densità discreta congiunta di X e Y.

4) X e Y sono indipendenti?

5) Calcolare $\mathbb{E}\left[\frac{X}{Y}\right]$.

Soluzione

1) $\mathcal{S}_X = \{1,2,3,4,6,8,9,12,16\}$

X	1	2	3	4	6	8	9	12	16
p_X	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

2) $\mathbb{P}(X > 6) = \frac{3}{8} = 37.5\%$.

3) La densità discreta congiunta di X e Y (insieme anche alle rispettive densità marginali) è data da:

$X \backslash Y$	1	2	3	4	p_X
1	$\frac{1}{16}$	0	0	0	$\frac{1}{16}$
2	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
3	0	0	$\frac{1}{8}$	0	$\frac{1}{8}$
4	0	$\frac{1}{16}$	0	$\frac{1}{8}$	$\frac{3}{16}$
6	0	0	$\frac{1}{8}$	0	$\frac{1}{8}$
8	0	0	0	$\frac{1}{8}$	$\frac{1}{8}$
9	0	0	$\frac{1}{16}$	0	$\frac{1}{16}$
12	0	0	0	$\frac{1}{8}$	$\frac{1}{8}$
16	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
p_Y	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	1

4) No, infatti ad esempio $p_{(X,Y)}(1,1) \neq p_X(1)\, p_Y(1)$.

5)

$$
\begin{aligned}
\mathbb{E}\left[\frac{X}{Y}\right] &= \sum_{\substack{i=1,2,3,4,6,8,9,12,16 \\ j=1,2,3,4}} \frac{i}{j}\, p_{(X,Y)}(i,j) \\
&= \frac{1}{1}\, p_{(X,Y)}(1,1) + \frac{2}{2}\, p_{(X,Y)}(2,2) + \frac{3}{3}\, p_{(X,Y)}(3,3) \\
&\quad + \frac{4}{2}\, p_{(X,Y)}(4,2) + \frac{4}{4}\, p_{(X,Y)}(4,4) + \frac{6}{3}\, p_{(X,Y)}(6,3) \\
&\quad + \frac{8}{4}\, p_{(X,Y)}(8,4) + \frac{9}{3}\, p_{(X,Y)}(9,3) + \frac{12}{4}\, p_{(X,Y)}(12,4) \\
&\quad + \frac{16}{4}\, p_{(X,Y)}(16,4) \;=\; \frac{15}{8} \;=\; 1.875.
\end{aligned}
$$

Esercizio 3

Sia X una variabile aleatoria continua con densità

$$f_X(x) \;=\; \begin{cases} c\,x^4, & -1 \le x \le 1, \\ 0, & \text{altrimenti.} \end{cases}$$

1) Determinare il valore della costante c.

2) Determinare la funzione di ripartizione di X.

Un messaggio binario ha un *valore iniziale* sconosciuto, descritto dalla variabile aleatoria discreta Y_{in} (Y_{in} può assumere solo i valori "0" oppure "1"). La densità discreta di Y_{in} è nota ed è data dalla seguente tabella:

Y_{in}	0	1
$p_{Y_{\text{in}}}$	$\frac{1}{4}$	$\frac{3}{4}$

Quando il messaggio binario arriva a destinazione è affetto da un certo "rumore", che supponiamo sia descritto dalla variabile aleatoria X. Più precisamente, il *valore finale* del messaggio binario è

$$Y_{\text{fin}} \;=\; Y_{\text{in}} + X.$$

Supponiamo che le variabili aleatorie Y_{in} e X siano *indipendenti*.
Una volta ricevuto, il valore finale viene decodificato in un valore binario in base alla seguente regola:

- se $Y_{\text{fin}} \ge 0.5$, si interpreta come "1";
- se $Y_{\text{fin}} < 0.5$, si interpreta come "0".

3) Qual è il valore atteso di Y_{fin}?

4) Sapendo che il valore iniziale del messaggio è "1", qual è la probabilità che il valore finale sia interpretato come "1" (ovvero che sia interpretato correttamente)?

Soluzione

1) Notiamo che $f_X(x) \geq 0$ per ogni x se e solo se $c \geq 0$. Inoltre

$$1 \;=\; \int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x \;=\; \int_{-1}^{1} c\,x^4\,\mathrm{d}x \;=\; \frac{2}{5}\,c,$$

da cui si ottiene $c = 5/2$.

2)

$$F_X(t) \;=\; \int_{-\infty}^{t} f_X(x)\,\mathrm{d}x \;=\; \begin{cases} 0, & t \leq -1, \\ \dfrac{1+t^5}{2}, & -1 \leq t \leq 1, \\ 1, & t \geq 1. \end{cases}$$

3) Per la linearità del valore atteso (non si usa l'indipendenza in questo punto), si ha che

$$\mathbb{E}[Y_{\text{fin}}] \;=\; \mathbb{E}[Y_{\text{in}}] + \mathbb{E}[X] \;=\; \frac{3}{4} + \int_{-1}^{1} x\,\frac{5}{2}\,x^4\,\mathrm{d}x \;=\; \frac{3}{4}.$$

4)

$$\begin{aligned} P\big(Y_{\text{fin}} \geq 0.5 \,|\, Y_{\text{in}} = 1\big) \;&=\; \frac{P(\{Y_{\text{fin}} \geq 0.5\} \cap \{Y_{\text{in}} = 1\})}{P(Y_{\text{in}} = 1)} \;=\; \frac{P(\{1 + X \geq 0.5\} \cap \{Y_{\text{in}} = 1\})}{P(Y_{\text{in}} = 1)} \\ &=\; \frac{P(\{X \geq -0.5\} \cap \{Y_{\text{in}} = 1\})}{P(Y_{\text{in}} = 1)} \underset{\substack{\uparrow \\ \text{indip. di } X \text{ e } Y_{\text{in}}}}{=} \frac{P(X \geq -0.5) P(Y_{\text{in}} = 1)}{P(Y_{\text{in}} = 1)} \\ &=\; P(X \geq -0.5) \;=\; 1 - F_X(-0.5) \;\simeq\; 0.5156. \end{aligned}$$

Esercizio 4

Si consideri una catena di Markov $(X_n)_n$ con spazio degli stati $\mathcal{S} = \{0,1,2,3,4,5,6\}$ e matrice di transizione

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

1) Disegnare il grafo orientato associato alla catena di Markov.

2) Quali sono le classi comunicanti?

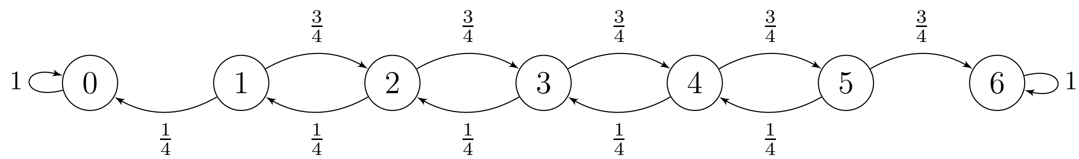
Consideriamo un giocatore che ad ogni giocata vinca 1 euro con probabilità $\frac{3}{4}$ e perda 1 euro con probabilità $\frac{1}{4}$. Se supponiamo che il giocatore smetta di giocare quando il suo capitale arriva a 0 oppure a 6, allora l'evoluzione nel tempo del capitale è descritta dalla catena di Markov $(X_n)_n$.

3) Sapendo che ad un certo istante il capitale del giocatore è pari a 3 euro, qual è la probabilità che dopo due giocate ("in due passi") il capitale ammonti a 5 euro?

4) Sapendo che ad un certo istante il capitale del giocatore è pari a 4 euro, qual è la probabilità che "in quattro passi" il capitale ammonti a 6 euro?

Soluzione

1)



2) $\{0\}$, $\{1, 2, 3, 4, 5\}$ e $\{6\}$.

3) Dobbiamo calcolare $\pi_{35}^{(2)}$. Si ha che

$$\pi_{35}^{(2)} = \underbrace{\frac{3}{4} \cdot \frac{3}{4}}_{\text{prob. cammino } 3 \to 4 \to 5} = \frac{9}{16} = 0.5625.$$

4) Dobbiamo calcolare $\pi_{46}^{(4)}$. Si ha che

$$\begin{aligned}\pi_{46}^{(4)} &= \underbrace{\frac{3}{4} \cdot \frac{3}{4} \cdot 1 \cdot 1}_{\text{prob. cammino } 4 \to 5 \to 6 \to 6 \to 6} + \underbrace{\frac{3}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4}}_{\text{prob. cammino } 4 \to 5 \to 4 \to 5 \to 6} \\ &+ \underbrace{\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4}}_{\text{prob. cammino } 4 \to 3 \to 4 \to 5 \to 6} = \frac{198}{256} \simeq 0.7734.\end{aligned}$$