Esercizi sull'uso delle probabilità

Esercizio 1

Si consideri un problema medico in cui bisogna modellare le relazioni causali (cioè di causa/effetto) fra fumo, neoplasia ai polmoni, tosse e risultato di una biopsia ai polmoni. Indichiamo la distribuzione di probabilità congiunta delle variabili in gioco con $\boldsymbol{P}(S, L, C, B)$, dove $S = fumare$, $L =$ *neoplasia ai polmoni*, $C =$ *tosse*, $B =$ *risultato biopsia*. Tutte le variabili sono booleane, nessuna relazione è deterministica, e si considera conoscenza medica relativa ad una popolazione di ex-fumatori anziani.

Si considerino le seguenti tre fattorizzazioni della distribuzione di probabilità congiunta:

1. $\boldsymbol{P}(S, L, C, B) = \boldsymbol{P}(S)\boldsymbol{P}(L|S)\boldsymbol{P}(C|S, L)\boldsymbol{P}(B|L)$
2. $\boldsymbol{P}(S, L, C, B) = \boldsymbol{P}(S)\boldsymbol{P}(B)\boldsymbol{P}(C)\boldsymbol{P}(L|S, B, C)$
3. $\boldsymbol{P}(S, L, C, B) = \boldsymbol{P}(S)\boldsymbol{P}(C)\boldsymbol{P}(L|S, C)\boldsymbol{P}(B|S, L)$

a) dire quale fattorizzazione codifica meglio la conoscenza medica corrente

b) quale fattorizzazione utilizza meno parametri ?

c) usando la fattorizzazione (1) si derivi una espresssione simbolica per $\boldsymbol{P}(B|S)$ che utilizzi le probabilità condizionali disponibili nella fattorizzazione.

d) usando la fattorizzazione (1) si derivi una espresssione simbolica per $\boldsymbol{P}(L|B)$ che utilizzi le probabilità condizionali disponibili nella fattorizzazione.

Soluzione:

a) la fattorizzazione (1)

b) riferendoci in ordine alle variabili S, C, L, B:

   - la fattorizzazione (1) ha 1+4+2+2=9 parametri;
   - la fattorizzazione (2) ha 1+1+1+8=11 parametri;
   - la fattorizzazione (3) ha 1+1+4+4=10 parametri.

   Quindi la risposta è: la fattorizzazione (1).

c) ricordando che nella rappresentazione simbolica i prodotti e le divisioni si riferiscono a singole coimponenti delle distribuzioni, abbiamo

$$\begin{aligned}
\boldsymbol{P}(B|S) &= \frac{\boldsymbol{P}(B,S)}{\boldsymbol{P}(S)} = \frac{1}{\boldsymbol{P}(S)} \sum_{C=c,L=l} \boldsymbol{P}(B,S,C,L) \\
&= \frac{1}{\boldsymbol{P}(S)} \sum_{C=c,L=l} \boldsymbol{P}(C|S,L)\boldsymbol{P}(B|L)\boldsymbol{P}(L|S)\boldsymbol{P}(S) \\
&= \frac{1}{\boldsymbol{P}(S)} \boldsymbol{P}(S) \sum_{L=l} \boldsymbol{P}(B|L)\boldsymbol{P}(L|S) \sum_{C=c} \boldsymbol{P}(C|S,L)
\end{aligned}$$

   e poiché $\sum_{C=c} \boldsymbol{P}(C|S,L) = 1$, si ha $\boldsymbol{P}(B|S) = \sum_{L=l} \boldsymbol{P}(B|L)\boldsymbol{P}(L|S)$.

d) usando la regola di Bayes: $\boldsymbol{P}(L|B) = \dfrac{\boldsymbol{P}(B|L)\boldsymbol{P}(L)}{\boldsymbol{P}(B)} = \alpha \boldsymbol{P}(B|L)\boldsymbol{P}(L)$.
   Rimane da calcolare

$$\begin{aligned}
\boldsymbol{P}(L) &= \sum_{S=s,C=c,B=b} \boldsymbol{P}(B,S,C,L) \\
&= \sum_{S=s,C=c,B=b} \boldsymbol{P}(C|S,L)\boldsymbol{P}(B|L)\boldsymbol{P}(L|S)\boldsymbol{P}(S) \\
&= \sum_{S=s} \boldsymbol{P}(L|S)\boldsymbol{P}(S) \underbrace{\sum_{B=b} \boldsymbol{P}(B|L)}_{=1} \underbrace{\sum_{C=c} \boldsymbol{P}(C|S,L)}_{=1}
\end{aligned}$$

   e quindi

$$\boldsymbol{P}(L|B) = \frac{1}{\boldsymbol{P}(B)} \boldsymbol{P}(B|L) \sum_{S=s} \boldsymbol{P}(L|S)\boldsymbol{P}(S) = \alpha \boldsymbol{P}(B|L) \sum_{S=s} \boldsymbol{P}(L|S)\boldsymbol{P}(S)$$

Esercizio 2

Siete in vacanza ad Atene, e durante una delle vostre notti brave, in una strada poco illuminata, siete testimoni dell'investimento di un pedone da parte di un taxi, che invece di fermarsi, fugge via. Voi pensate di aver riconosciuto che il colore del taxi fosse blu. Sapendo che ad Atene 1 taxi su 10 è blu ed i restanti sono verdi, e che in condizioni di scarsa illuminazione il blu e il verde si discriminano nel 75% dei casi, calcolare la probabilità che il taxi fosse realmente di colore blu. (Aiuto: differenziare la sentenza "il taxi era blu" dalla sentenza "il taxi sembrava blu").

Soluzione:

Sia B la variabile booleana associata alla sentenza "il taxi era blu" e SB la variabile booleana associata alla sentenza "il taxi sembrava blu".

Il fatto che in condizioni di scarsa illuminazione il blu e il verde si discriminano nel 75% dei casi, ci permette di dire che

$$P(SB|B) = 0.75 \quad P(\neg SB|\neg B) = 0.75$$

Inoltre, il fatto che ad Atene 1 taxi su 10 è blu ed i restanti sono verdi, ci permette di dire che $P(B) = 0.1$.

Usando la regola di Bayes abbiamo:

$$P(B|SB) = \alpha P(SB|B)P(B) = \alpha\ 0.75 \times 0.1 = \alpha\ 0.075$$

$$P(\neg B|SB) = \alpha P(SB|\neg B)P(\neg B) = \alpha 0.25 \times 0.9 = \alpha\ 0.225$$

e quindi, normalizzando,

$$P(B|SB) = \frac{0.075}{0.075 + 0.225} = 0.25 \quad P(\neg B|SB) = \frac{0.225}{0.075 + 0.225} = 0.75$$

Esercizio 3

Si consideri la seguente fattorizzazione che codifica conoscenza riguardante la probabilità per un politico di essere eletto:

$$\boldsymbol{P}(H, S, P, E) = \boldsymbol{P}(H)\boldsymbol{P}(S|H)\boldsymbol{P}(P|H, S)\boldsymbol{P}(E|P)$$

dove $H = onesto$, $S = abile$, $P = popolare$, $E = eletto$ e

P(H)
0.1

H	P(S| H)
T	0.3
F	0.9

H	S	P(P| H,S)
T	T	0.9
T	F	0.5
F	T	0.8
F	F	0.2

P	P(E|P)
T	0.6
F	0.1

a) calcolare $P(h, s, \neg p, \neg e)$

b) dato un politico onesto, calcolare la distribuzione di probabilità per la sua elezione (variabile E)

Soluzione:

a) usando la fattorizzazione data abbiamo:

$$\begin{aligned} P(h, s, \neg p, \neg e) &= P(h)P(s|h)P(\neg p|h, s)P(\neg e|\neg p) \\ &= 0.1 \times 0.3 \times 0.1 \times 0.9 = 0.00027 \end{aligned}$$

b) usando l'algoritmo enumerativo:

$$
\begin{aligned}
\boldsymbol{P}(E|h) &= \alpha \sum_{S=s,P=p} \boldsymbol{P}(h,S,P,E) \\
&= \alpha \sum_{S=s,P=p} P(h)\boldsymbol{P}(S|h)\boldsymbol{P}(P|h,S)\boldsymbol{P}(E|P) \\
&= \alpha P(h) \sum_{S=s} \boldsymbol{P}(S|h) \sum_{P=p} \boldsymbol{P}(P|h,S)\boldsymbol{P}(E|P) \\
&= \alpha 0.1[0.3 \times (0.9 < 0.6, 0.4 > +0.1 < 0.1, 0.9 >) + \\
&\quad 0.7 \times (0.5 < 0.6, 0.4 > +0.5 < 0.1, 0.9 >)] =< 0.41, 0.59 >
\end{aligned}
$$