

Bayesovská sieť

- Orientovaný acyklický graf
- Vrcholy: náhodné premenné $X_1 \dots X_n$
- Nech X_{S_i} sú predchodcovia premennej X_i
- Pre každé X_i tabuľka $\Pr(X_i|X_{S_i})$
- Sieť definuje celkové rozdelenie

$$\Pr(X_1 \dots X_n) = \prod_{i=1}^n \Pr(X_i|X_{S_i})$$

Inferencia v bayesovských sieťach

Rozdelíme premenné na tri typy:

- Vyfarbené: poznáme hodnotu
- Nevyfarbené: nepoznáme hodnotu, zaujíma nás
- Preškrtnuté: nepoznáme hodnotu, nezaujíma nás

Chceme

$$\Pr(\text{nevyfarbené}|\text{vyfarbené}) = \frac{\sum_{\text{preškrtnuté}} \Pr(\text{nevyfarbené}, \text{vyfarbené}, \text{preškrtnuté})}{\Pr(\text{vyfarbené})}$$

Eliminačný algoritmus

Udržuj vrece tabuliek.

- Vlož do vreca: $\Pr(X_i | X_{S_i})$ pre každé X_i a $\delta_i(x) = [x = x_i]$ pre každé vyfarbené X_i s hodnotou x_i
- Zorad' vyfarbené a zaškrtnuté premenné do vhodného poradia P
- Pre každú premennú X_i v poradí P :
nahrad' všetky tabuľky obsahujúce X_i jednou tabuľkou m_i , ktorá X_i neobsahuje
- Konečný výsledok: súčin zvyšných tabuliek po normalizácii
(tabuľka určujúca hodnotu pre každú konfiguráciu nevyfarbených premenných)

Eliminácia premennej X_i

- Vyber z vreca všetky tabuľky t_1, \dots, t_k zahrňajúce X_i
- Nech X_A sú všetky premenné v týchto tabuľkách okrem X_i
- Vytvor novú tabuľku m_i s premennými X_A :

$$m_i(X_A) = \sum_{x_i} \prod_{j=1}^k t_j(X_A)$$

- Vlož m_i do vreca

Bayesovská sieť

- Orientovaný acyklický graf
- Vrcholy: náhodné premenné $X_1 \dots X_n$
- Nech X_{S_j} sú predchodcovia premennej X_j
- Pre každé X_j tabuľka $\Pr(X_j | X_{S_j})$

Označenie $\Theta_{j,u,v} = \Pr(X_j = u | X_{S_j} = v)$

$$\sum_u \Theta_{j,u,v} = 1 \quad \forall j, v$$

- Sieť definuje celkové rozdelenie

$$\Pr(X_1 \dots X_n = x_1 \dots x_n) = \prod_{j=1}^n \Pr(X_j = x_j | X_{S_j} = x_{S_j}) = \prod_{j=1}^n \Theta_{j,x_j,x_{S_j}}$$

Učenie s neúplnými dátami

- $x^{(i)}$ - pozorované dáta vo vzorke i
- $X^{(i)}$ - pozorované premenné
- $y^{(i)}$ - chýbajúce dáta vo vzorke i
- $Y^{(i)}$ - chýbajúce premenné
- Cieľ: maximalizovať vierohodnosť

$$\Theta^* = \arg \max_{\Theta} \prod_{i=1}^t \Pr(X^{(i)} = x^{(i)} | \Theta)$$

EM algoritmus (Expectation maximization)

- Zvoľ počiatočné $\Theta^{(0)}$
- Iteruj kým nenastane konvergencia ($k = 0, 1, \dots$)
 - E-krok: spočítaj pre každé $i, y^{(i)}$:

$$Q_i^{(k)}(y^{(i)}) = \Pr(Y^{(i)} = y^{(i)} | X^{(i)} = x^{(i)}, \Theta^{(k)})$$

- M-krok: spočítaj

$$\Theta^{(k+1)} = \arg \max_{\Theta} \sum_{i=1}^t \sum_{y^{(i)}} Q_i^{(k)}(y^{(i)}) \log \Pr(X^{(i)} = x^{(i)}, Y^{(i)} = y^{(i)} | \Theta)$$

Maximalizujeme vierohodnosť dát doplnených vš. spôsobmi o chýbajúce hodnoty, váhované cez $Q^{(k)}$

M-krok EM algoritmu

Spočítaj

$$\Theta^{(k+1)} = \arg \max_{\Theta} \sum_{i=1}^t \sum_{\mathbf{y}^{(i)}} Q_i^{(k)}(\mathbf{y}^{(i)}) \log \Pr(X^{(i)} = \mathbf{x}^{(i)}, Y^{(i)} = \mathbf{y}^{(i)} | \Theta)$$

Maximalizujeme vierohodnosť dát doplnených vš. spôsobmi o chýbajúce hodnoty, váhované cez $Q^{(k)}$

Maximum pre pozorované frekvencie s váhovanými dátami:

$$\Theta_{j,u,v}^{(k+1)} = \frac{\#_{j,u,v}}{n_{j,v}}$$

- $\#_{j,u,v}$ je súčet váh doplnených vzoriek, v ktorých $X_j = u$ a $x_{S_j} = v$
- $n_{j,v}$ je súčet váh doplnených vzoriek, v ktorých $x_{S_j} = v$

EM algoritmus (Expectation maximization)

Lema: $L(\Theta^{(k+1)}|X) \geq L(\Theta^{(k)}|X)$, pričom rovnosť nastáva iba v lokálnom maxime alebo inflexnom bode $L(\Theta|X)$.

$$L(\Theta|X) = \prod_{i=1}^t \Pr(X^{(i)} = x^{(i)}|\Theta)$$

- Iterácie postupne zvyšujú vierohodnosť
- Môžu skončiť v lokálnom maxime
- Koľko iterácii treba?
- Koľko trvá jedna iterácia?