18. Probabilistički grafički modeli II

Strojno učenje 1, UNIZG FER, ak. god. 2021./2022.

Jan Šnajder, natuknice s predavanja, v1.4

1 Zaključivanje

- Primjer trave i prskalice (v. predavanje 17):

$$p(c, s, r, w) = p(c)p(s|c)p(r|c)p(w|s, r)$$

- Upit: Ako je trava mokra ($w = 1$), koja je vjerojatnost kiše (r) i prskalice (s)?

$$\begin{aligned}
P(s = 1|w = 1) &= \frac{P(s = 1, w = 1)}{P(w = 1)} \\
&= \frac{\sum_{c,r} P(c, s = 1, r, w = 1)}{\sum_{c,r,s} P(c, s, r, w = 1)} = 0.2781/0.6471 = 0.43 \\
P(r = 1|w = 1) &= \frac{P(r = 1, w = 1)}{P(w = 1)} \\
&= \frac{\sum_{c,s} P(c, s, r = 1, w = 1)}{\sum_{c,r,s} P(c, s, r, w = 1)} = 0.4851/0.6471 = 0.708
\end{aligned}$$

  gdje je $P(w = 1)$ **vjerojatnost dokaza**

- Dvije vrste upita: (1) posteriorni upiti i (2) MAP-upiti
- **Posteriorni upit** je izračun uvjetne vjerojatnosti:

$$p(\mathbf{x}_q|\mathbf{x}_o) = \frac{\sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)}{p(\mathbf{x}_o)} = \frac{\sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)}{\sum_{\mathbf{x}'_n, \mathbf{x}'_q} p(\mathbf{x}'_q, \mathbf{x}_o, \mathbf{x}'_n)}$$

  gdje su $\mathbf{x}_q$ varijable upita, $\mathbf{x}_o$ su opažene varijable, a $\mathbf{x}_n$ varijable smetnje (*nuisance*)

- **MAP-upiti** – najvjerojatnija vrijednost varijabli upita:

$$\mathbf{x}^*_q = \underset{\mathbf{x}_q}{\mathrm{argmax}} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$$

2 Zaključivanje: eliminacija varijabli

- Odgovaranje upita iziskuje konstrukciju $p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ pa marginalizaciju/normalizaciju
- Velik $n \Rightarrow$ **kombinatorna eksplozija** $\Rightarrow$ NP-složen problem
- Poništava prednost Bayesove mreže (sažet zapis zajedničke distribucije)
- Alternative: **egzaktno zaključivanje** i **približno zaključivanje**
- **Eliminacija varijabli** – egzaktno zaključivanje pomoću dinamičkog programiranja
- **Eliminacija varijabli zbroj-umnožak** – potiskivanje marginalizacije što dublje:

$$\begin{aligned}
p(w) &= \sum_c \sum_s \sum_r p(c,s,r,w) \\
&= \sum_c \sum_s \sum_r p(c)p(s|c)p(r|c)p(w|s,r) \\
&= \sum_s \sum_r p(w|s,r) \underbrace{\sum_c p(c)p(s|c)p(r|c)}_{t_1(s,r)} \\
&= \sum_s \underbrace{\sum_r p(w|s,r)t_1(s,r)}_{t_2(s,w)} \\
&= \sum_s t_2(s,w) \\
&= t_3(w)
\end{aligned}$$

- Varijante algoritma za skriveni Markovljev model (HMM):
  - elimacija varijabli $\Rightarrow$ **algoritam naprijed nazad** (*forward-backward algoritham*)
  - MAP-upiti $\Rightarrow$ **Viterbijev algoritam**
- Za općenite Bayesove mreže eliminacija varijabli je presložena
- Alternativa: približno zaključivanje – **propagacijski algoritmi** i **metode uzorkovanja**

3 Zaključivanje: metode uzorkovanja

- Ideja: procjena distribucije na temelju uzorka
- Ako uzorkujemo uzorke $\mathbf{x} \sim P(\mathbf{x})$, očekivanje je:

$$P(\mathbf{x} = x) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{\mathbf{x} = x\}$$

- Najjednostavnija metoda: **unaprijedno uzorkovanje** (*forward sampling*)
  - Uzorkovanje varijable za varijablom, prema topološkom uređaju mreže

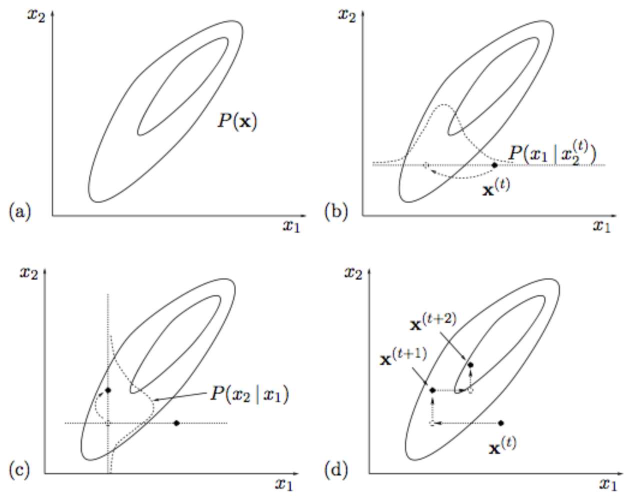
- Problem: želimo uzorkovati iz uvjetne vjerojatnosti $P(\mathbf{x}_q|\mathbf{x}_o)$

- **Uzorkovanje s odbijanjem** (*rejection sampling*)
  - Uzorkovanje unaprijed i odbijanje $\mathbf{x}$ za koje $\mathbf{x}_o$ nisu na željenim vrijednostima
  - Problem: neučinkovito, osobito ako je vjerojatnost dokaza $P(\mathbf{x}_o)$ malena

- **Uzorkovanje po važnosti** (*importance sampling*)
  - Postavljanje $\mathbf{x}_o$ na željene vrijednosti, unaprijedno uzorkovanje i korekcija očekivanja
  - Problem: loša kvaliteta procjene, pogotovo ako su $\mathbf{x}_o$ pri dnu Bayesove mreže

- **Gibbsovo uzokovanje** (*Gibbs sampling*)
  - Postupak iz porodice **Markov Chain Monte Carlo (MCMC)**
  - Krenuvši od slučajnog vektora $\mathbf{x}$, uzorkujemo ciklički varijablu po varijablu

$$
\begin{aligned}
&\mathbf{x}^0 \sim p(x_1^0, x_2^0, x_3^0) && \Rightarrow \text{početni vektor (npr., unaprijednim uzorkovanjem)}\\
&x_1^1 \sim p(x_1|x_2^0, x_3^0)\\
&x_2^1 \sim p(x_2|x_1^1, x_3^0)\\
&x_3^1 \sim p(x_3|x_1^1, x_2^1) && \Rightarrow \text{vektor } \mathbf{x}^1 = (x_1^1, x_2^1, x_3^1)\\
&x_1^2 \sim p(x_1|x_2^1, x_3^1)\\
&x_2^2 \sim p(x_2|x_1^2, x_3^1)\\
&x_3^2 \sim p(x_3|x_1^2, x_2^2) && \Rightarrow \text{vektor } \mathbf{x}^2 = (x_1^2, x_2^2, x_3^2)\\
&\vdots
\end{aligned}
$$



4 Učenje

- PGM-ovi su probabilistički modeli ⇒ učenje se svodi na **procjenu parametara $\boldsymbol{\theta}$**

- MLE, MAP ili bayesovska procjena

- Log-izglednost za općenitu Bayesovu mrežu:

$$\begin{aligned}
\ln \mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) &= \ln p(\mathcal{D}|\boldsymbol{\theta}) \\
&= \ln p(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}|\boldsymbol{\theta}) \\
&= \ln \prod_{i=1}^{N} p(\mathbf{x}^{(i)}|\boldsymbol{\theta}) \\
&= \ln \prod_{i=1}^{N} \prod_{k=1}^{n} p(x_k^{(i)}|\mathrm{pa}(x_k^{(i)}), \boldsymbol{\theta}_k) \\
&= \ln \prod_{k=1}^{n} \prod_{i=1}^{N} p(x_k^{(i)}|\mathrm{pa}(x_k^{(i)}), \boldsymbol{\theta}_k) \\
&= \sum_{k=1}^{n} \sum_{i=1}^{N} \ln p(x_k^{(i)}|\mathrm{pa}(x_k^{(i)}), \boldsymbol{\theta}_k)
\end{aligned}$$

- MLE procjena za k-ti čvor:

$$\boldsymbol{\theta}_k^* = \underset{\boldsymbol{\theta}_k}{\mathrm{argmax}} \sum_{i=1}^{N} \ln p(x_k^{(i)}|\mathrm{pa}(x_k^{(i)}), \boldsymbol{\theta}_k)$$

- MAP procjena za k-ti čvor:

$$\boldsymbol{\theta}_k^* = \underset{\boldsymbol{\theta}_k}{\mathrm{argmax}} \Big(\sum_{i=1}^{N} \ln p(x_k^{(i)}|\mathrm{pa}(x_k^{(i)}), \boldsymbol{\theta}_k) + \ln p(\boldsymbol{\theta}_k) \Big)$$

- MAP-procjena za kategorijsku razdiobu (Dirichlet-kategorijski model uz $\alpha = 2$):

$$\begin{aligned}
\hat{\mu}_{k,j,l} &= \frac{N_{kjl} + 1}{N_{kj} + K_k} \\
N_{kjl} &= \sum_{i=1}^{N} \mathbf{1}\{\mathbf{x}^{(i)}_{\mathrm{pa}(x_k)} = j \wedge x_k^{(i)} = l\} \\
N_{kj} &= \sum_{l} N_{kjl}
\end{aligned}$$

  gdje je K_k broj mogućih vrijednosti varijable x_k

- Primjer: MAP procjena za čvor w u mreži s travom i prskalicom (v. predavanje 17):

$$P(w|s,r) = \frac{\sum_{i=1}^{N} \mathbf{1}\{x_s^{(i)} = s \wedge x_r^{(i)} = r \wedge x_w^{(i)} = w\} + 1}{\sum_{i=1}^{N} \mathbf{1}\{x_s^{(i)} = s \wedge x_r^{(i)} = r\} + 2}$$

- Modeli sa skrivenim varijablama (npr., HMM, GMM) $\Rightarrow$ tzv. **nepotpuni podatci**
  - Log-izglednost se ne dekomponira po strukturi grafa
  - $\Rightarrow$ MLE nema rješenje u zatvorenoj formi
  - Učenje pomoću **algoritma maksimizacije očekivanja** ili **gradijentnim usponom**

- **Učenje strukture mreže**:
  - Polunaivan Bayesov klasifikator: v. 4.3.2–4.3.4 u skripti
  - Učenje općenite strukture Bayesove mreže: **algoritam K2**