Variational Inference

pgm 27 (I)

$P(X_t | X_e = x_e) \Longrightarrow \boxed{P(X_c)}$ marginalization $\boxed{\phi_c(X_c)}$

$x_1^*, \dots, x_n^* = \arg\max_{x_1 \dots x_n} P(X_1, X_2, \dots, X_n)$

$$P(X_c) = \sum_{X \setminus X_c} P(X_1, \dots, X_n)$$

How to solve this using an optimization problem?

Latent variable models

$P_\theta(X, Z) = \checkmark$

data $x^1, x^2, \dots, x^n$

$p(Z|X) = \checkmark$ Hard to compute

$$\boxed{P_\theta(X) = \frac{1}{Z} \underbrace{\tilde{P}(X)}_{\text{product of factors}} = \frac{1}{Z} e^{F(X)}} \quad F(X) \rightarrow \text{sum of}$$

MRF

BN or MRF with evidence

$P(X) = P(X_t, \underbrace{X_e}_{\text{evidence (known)}}) \Longrightarrow$ need $p(X_t | X_e) = \dfrac{P(X_t, x_e)}{\sum_{X_t} P(X_t, x_e)}$ $\quad Z(X_e)$

$P(X) = P(Y, \underbrace{Z}_{\text{latent}})$ $\quad p(Z|Y) = \dfrac{p(Z,Y)}{\sum_Z p(Z,Y)}$ $\rightarrow Z(Y)$

$P(X) = \frac{1}{Z} \tilde{P}(X)$

complex

(Hard to do inference on)

find $Q(X)$ $\begin{cases} Q(X) \text{ is easy to handle} \\ Q(X) \text{ is close to } P(X) \end{cases}$

Measure distance between $Q(X), P(X)$

KL-divergence

$$KL(Q \| P) = \sum_X Q(X) \log \frac{Q(X)}{P(X)} = E_Q\left\{\log \frac{Q(X)}{P(X)}\right\}$$

$$= \sum_X Q(X) \left[\log Q(X) - \log P(X) \right]$$

$P(X) = Q(X)$ for all $X \Longrightarrow KL(Q \| P) = 0$

$KL(Q \| P) = 0 \Longrightarrow P(X) = Q(X)$ for all X

almost all (continuous)

$KL(Q \| P) \geq 0$

$KL(Q \| P) \neq KL(P \| Q)$ in general

$$KL(Q \| P) = \sum_X Q(X) \log Q(X) - \sum_X Q(X) \log P(X)$$

what choice of Q maximizes $\sum_X Q(X) \log P(X)$?

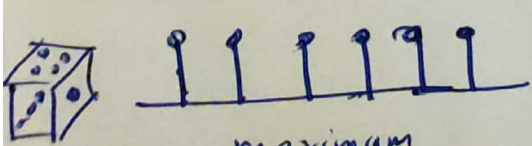
$\sum_X Q(X) = 1$

$\int_X Q(X) = 1$

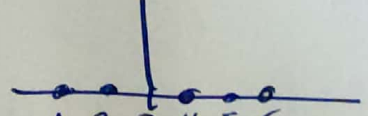
$$\text{Entropy} = \sum_X Q(X) \log \frac{1}{Q(X)} = -\sum_X Q(X) \log Q(X) = H(Q)$$

$$KL(Q \| P) = -H(Q) - E_Q\{\log P(X)\}$$

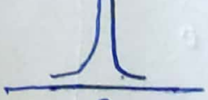
minimizing $KL(Q \| P) \equiv$ maximizing $H(Q) + E_Q\{\log P(X)\}$



maximum entropy



minimum entropy



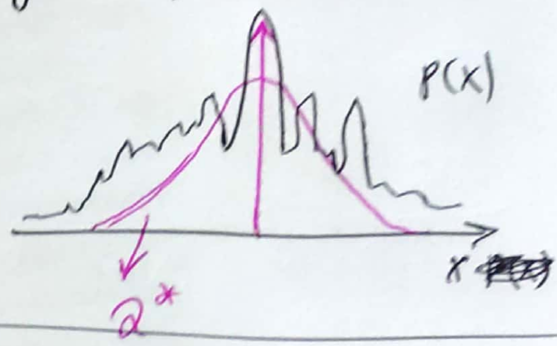
Q_1

Q_2

$H(Q_1) < H(Q_2)$

$$KL(Q\|P) = -H(Q) - E_Q\{\log P(X)\}$$

$$\underset{Q\in \bar{Q}_{ee}}{\text{minimize}}\; KL(Q\|P)$$

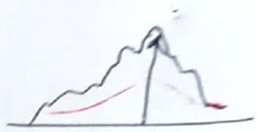


Simple case: $Q(X) = Q(X_1, X_2, \dots, X_n) = Q_1(X_1)\, Q_2(X_2) \cdots Q_n(X_n)$

fully factorized case

$$P(X) = \frac{1}{Z}\tilde{P}(X) = \frac{1}{Z} e^{F(X)}$$

$$\min_Q KL(Q\|P) = \sum_X Q(X)\log\frac{Q(X)}{P(X)}$$



$$= \sum_X P(X)\log P(X) - \sum_X P(X)\boxed{\log Q(X)}$$

why not minimize $KL(P\|Q) = \sum_X \underline{P(X)}\log\frac{P(X)}{Q(X)}$?

$$KL(Q\|P) = \sum_X Q(X)\left[\log Q(X) - \log P(X)\right]$$

$$= \sum_X Q(X)\left[\log Q(X) - \log \frac{1}{Z}\tilde{P}(X)\right]$$

$$= \sum_X Q(X)\log Q(X) - \sum_X Q(X)\log\tilde{P}(X) + \sum_X Q(X)\log Z$$

$$= \sum_X Q(X)\log Q(X) - \sum_X Q(X)\underbrace{\log\tilde{P}(X)}_{F(X)} + \log Z$$

$$KL(Q\|P) = \underbrace{\sum_X Q(X)\log\frac{Q(X)}{\tilde{P}(X)}}_{-L(Q)} + \log Z$$

$$\boxed{L(Q) = E_Q\left\{\frac{\tilde{P}(X)}{Q(X)}\right\}}$$

$$L(Q) = +\log Z - KL(Q\|P) \Rightarrow L(Q) \le \log(Z)$$

$KL \ge 0$ variational lower bound!

$$P(X) = \frac{1}{Z}\tilde{P}(X) = \frac{1}{Z} e^{F(X)} = \frac{1}{Z} e^{\sum_c F_c(X_c)}$$ pgm 27 IV

find Q minimize $KL(Q\|P) = \sum_X Q(X)\log\frac{Q(X)}{\tilde{P}(X)} + \log Z$

Meanfield inference $Q(X) = Q(X_1 - X_n) = \prod_{i=1}^{n} Q_i(X_i)$

Fully factorized

$$Q^* = \underset{Q}{\operatorname{argmin}} \sum_X Q(X)\log\frac{Q(X)}{\tilde{P}(X)} + \log Z$$

$$= \underset{Q}{\operatorname{argmin}} \sum_X Q(X)\log\frac{Q(X)}{\tilde{P}(X)}$$

$$\sum_X Q(X)\log\frac{Q(X)}{\tilde{P}(X)} = \boxed{\sum_X Q(X)\log Q(X)} - \sum_X Q(X)\underbrace{\log\tilde{P}(X)}_{F(X)}$$

$$\sum_X Q(X)\log Q(X) = \sum_{X_1}\sum_{X_2} - \sum_{X_n} Q_1(X_1)\,Q_2(X_2) - Q_n(X_n)\left[\sum_{i=1}^{n}\log Q_i(X_i)\right]$$

$$= \sum_{i=1}^{n} Q_i(X)\log Q_i(X)$$

$$\sum_X Q(X)\log\tilde{P}(X) = \sum_X Q(X)\,F(X) = \sum_X Q(X)\sum_c F_c(X_c)$$

$$= \sum_c\sum_X Q(X)\,F_c(X_c) = \sum_c\sum_{X_c}\left[\sum_{X\setminus X_c} Q(X_c, X\setminus X_c)\right]F_c(X_c)$$

$$= \sum_c\sum_{X_c} Q(X_c)\,F_c(X_c)$$

$$\sum_i Q_i(X_i)\log Q_i(X_i) - \sum_c\sum_{X_c} Q(X_c)\,F_c(X_c)$$

Scanned with CamScanner

pairwise MRF

$$F(X) = \sum_{i=1}^{n} F_i(X_i) + \sum_{(i,j)\in\mathcal{E}} F_{ij}(X_i, X_j)$$

- $\sum_i Q_i(X_i) \log Q_i(X_i) - \sum_{i=1}^{n} Q_i(X_i) F_i(X_i)$

$$- \sum_{(i,j)\in\mathcal{E}} Q_i(X_i) Q_j(X_j) F_{ij}(X_i, X_j)$$