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

An "easy" distribution $\underline{Q(X)}$

$$KL(Q \| P) = \sum_x Q(x) \log \frac{Q(x)}{P(X)}$$

$$= \sum_x Q(x) \log \frac{Q(X)}{\tilde{P}(X)} + \log Z$$

$$= -\sum_x Q(x) \log \frac{\tilde{P}(x)}{Q(x)} + \log Z$$

$$\underbrace{\sum_x Q(x) \log \frac{\tilde{P}(x)}{Q(x)}}_{L(Q,\tilde{P})} = \log Z - KL(Q \| P)$$

variational lower bound

1. $L(Q, \tilde{P}) \leqslant \log Z$
2. $Q^* = \underset{Q}{\operatorname{argmin}}\ KL(Q \| P) = \underset{Q}{\operatorname{argmax}}\ L(Q, \tilde{P})$

$$P(x) = \frac{1}{Z} e^{F(x)} \quad , \quad F(x) = \sum_{i=1}^{n} F_i(X_i) + \sum_{(i,j)\in \mathcal{E}} F_{ij}(X_i, X_j)$$

$$-L(Q,\tilde{P}) = \sum_i \sum_{X_i} Q_i(X_i) \log Q_i(X_i) - \sum_{i=1}^{n} Q_i(X_i) F_i(X_i) - \sum_{(i,j)\in\mathcal{E}} Q_i(X_i) Q_j(X_j) F_{ij}(X_i, X_j)$$

$$L(Q,\tilde{P}) = -\sum_i \sum_{X_i} Q_i(X_i) \log Q_i(X_i) + \sum_{i=1}^{n} \sum_{X_i} Q_i(X_i) F_i(X_i) + \sum_{(i,j)\in\mathcal{E}} \sum_{X_i} \sum_{X_j} Q_i(X_i) Q_j(X_j) F_{ij}(X_i, X_j)$$

$$Q(X) = Q(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} Q_i(X_i)$$ → mean-field

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

$$\mathcal{L}(Q,\tilde{P}) = -\sum_{i=1}^{n}\sum_{X_i} \underline{Q_i(X_i)} \log Q_i(X_i) + \sum_{i=1}^{n}\sum_{X_i} Q_i(X_i) F_i(X_i)$$
$$+ \sum_{(i,j)\in\mathcal{E}} \sum_{X_i}\sum_{X_j} Q_i(X_i) Q_j(X_j) F_{ij}(X_i, X_j)$$

$$\max_{Q} L(Q,\tilde{P}) = \max_{Q_1, Q_2, \dots, Q_n} L(Q,\tilde{P})$$

Discrete Case

$$Q_i(X_i = k) = q_{ik} \qquad Q_i \xrightarrow{\text{parameterize}} (q_{i1}, q_{i2}, \dots, q_{iL})$$
$$k = 1, 2, \dots, L \qquad \sum_{k=1}^{L} q_{ik} = 1$$

$$\mathcal{L}(Q,\tilde{P}) = -\sum_{i=1}^{n}\sum_{k=1}^{L} q_{ik}\log q_{ik} + \sum_{i=1}^{n}\sum_{k=1}^{L} q_{ik} F_i(k)$$
$$+ \sum_{i,j\in\mathcal{E}} \sum_{k=1}^{L}\sum_{k'=1}^{L} q_{ik}\, q_{jk'} F_{ij}(k,k')$$

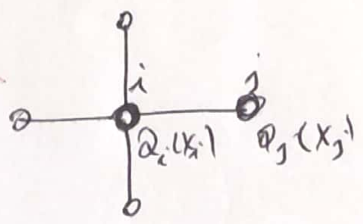
$$\max_{\substack{Q\\ q_{ik}}} \mathcal{L}(Q,P) \quad \text{subject to} \quad \sum_{k=1}^{L} q_{ik} = 1 \quad \text{for } i = 1 \cdots n$$

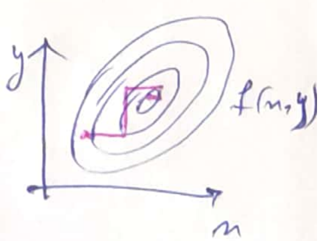
coordinate ascent

Loop

for $i = 1 \cdots n$

$\max_{q_i:}$





$$\max_{q_{i1}, q_{i2}, \dots, q_{iL}} L(Q,P) \Longrightarrow \frac{\partial}{q_{il}} L(Q,P) + \lambda\left(\sum_{k=1}^{L} q_{ik} - 1\right)$$

$$= -\log q_{il} - 1 + F_i(l) + \sum_{j\in N_i}\sum_{k'=1}^{L} q_{jk'} F_{ij}(l,k') + \underline{\lambda} = 0$$

for $l = 1, 2, \dots, L$

$$\sum_{k=1}^{l} q_{ik} = 1$$

$$\Rightarrow \log q_{il} = F_i(l) + \sum_{j \in N_i} \sum_{k=1}^{L} q_{jk} F_{ij}(l,k) + \lambda - 1$$

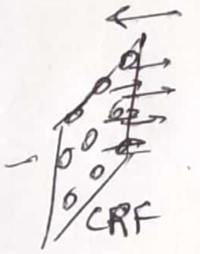
$$q_{il} = \exp\left(F_i(l) + \sum_{j \in N_i} \sum_{k=1}^{L} q_{jk} F_{ij}(l,k)\right) e^{\lambda - 1}$$

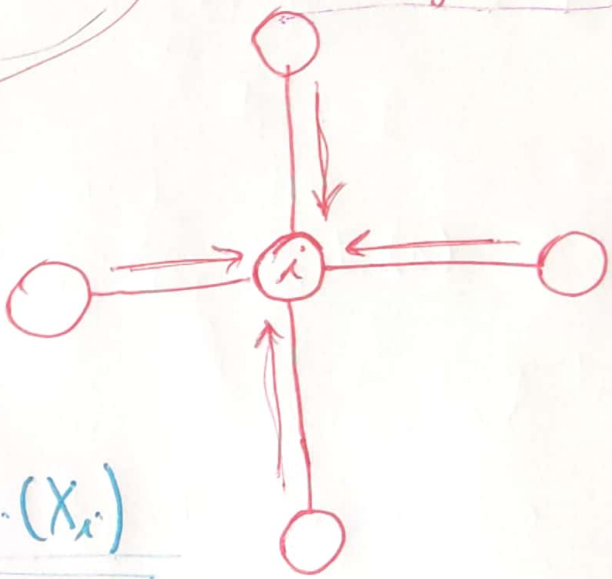
$l = 1, 2, \dots, L$ $\qquad \sum_{k=1}^{L} q_{il} = 1$

$$q'_{il} = \exp\left(F_i(l) + \sum_{j = N_i} \sum_{k=1}^{L} q_{jk} F_{ij}(l,k)\right)$$

$$q_{il} \leftarrow \frac{q'_{il}}{\sum_{k=1}^{L} q'_{ik}}$$

→ mean-field message passing





Case 2: Continuous $Q_i(X_i)$

Among all $Q_i(X)$ with $Q_i(X) \geq 0$, $\int Q_i(X) dx = 1$
$i = 1, \dots, n$

find the one that maximizes $\mathcal{L}(Q, \tilde{P})$

minimizes $KL(Q \| P)$

Variational Calculus $\quad f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is a function

optimize $J(f)$ s.t. so some constraint

IV

Variational Learng

$$P_\theta(X) = \frac{1}{Z(\theta)} \tilde{P}_\theta(X) = \frac{1}{Z(\theta)} e^{F_\theta(X)}$$

Latent Variable Models

$$P_\theta(x) = \int P_\theta(x, z)\, dz$$

Data $\underbrace{x^1, x^2, \dots, x^m}$

$$\max_\theta \sum_i \log P_\theta(x^i)$$

$$\max_\theta ll(\theta) = \max_\theta \sum_{i=1}^{m} \log P_\theta(X^i)$$

$$= \max \sum_{i=1}^{m} \log \int P_\theta(x^i; z)\, dz$$

Solution 1

$$\frac{\partial ll(\theta)}{\partial \theta_k} = \frac{\partial}{\partial \theta_k} \sum_{i=1}^{m} \log \underbrace{\int P_\theta(x^i; z)\, dz}_{\text{hard to compute}}$$

$$= \sum_{i=1}^{m} \frac{\int \frac{\partial}{\partial \theta_k} P_\theta(x^i; z)\, dz}{\int \underbrace{P_\theta(x^i; z)\, dz}}$$

Solution 2: Expectation-Maximization

$\Longrightarrow$ Need posterior $\underbrace{P_\theta(z \mid x^i)}$ for $i = 1, \dots, m$

$\rightarrow$ hard to compute