CS 307 Lecture 27

A training set $S = \{x^{(1)}, x^{(2)} \ldots x^{(n)}\}$

Learn is distribution from which S is drawn.

The distribution of drawing x depends on a latent variable z.

Assume: Latent variable z take one of $\{1, 2 \ldots k\}$.

Let θ be the parameters of the distribution.

$$L(\theta) = \log \prod_{i=1}^{n} \underbrace{p_\theta(x^{(i)})}_{}$$

Depends on latent variable $z \in \{1, \ldots k\}$.

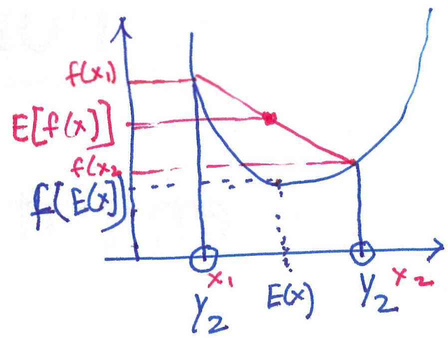
$$= \log \prod_{i=1}^{n} \sum_{z=1}^{k} p_\theta(x^{(i)}, z)$$

$$= \sum_{i=1}^{n} \log \sum_{z=1}^{k} p_\theta(x^{(i)}, z) \quad \leftarrow$$

Jensen's Inequality: Let $f: \mathbb{R} \to \mathbb{R}$ be ~~concave~~ convex.

$f''(x) \geq 0 \ \forall x$.

$$f(E(X)) \leq E[f(X)]$$



Let $f: \mathbb{R} \to \mathbb{R}$ be concave then

$f''(x) \leq 0 \ \forall x$.

$$f(E[X]) \geq E[f(X)].$$

Example: $f(x) = \log x$.

f is concave $\quad f''(x) = -\frac{1}{x^2} \leq 0 \ . \forall x$.

$f(x) = -\log x$ is convex.

Find θ that maximizes

$$L(\theta) = \sum_{i=1}^{n} \log \sum_{z=1}^{k} p_\theta(x^{(i)}, z).$$

$$= \sum_{i=1}^{n} \log \sum_{z=1}^{k} Q_i(z) \frac{p_\theta(x^{(i)}, z)}{Q_i(z)}.$$

Suppose we pick Q_i such that $\forall i$.

$$\sum_{z=1}^{k} Q_i(z) = 1 \quad \text{and} \quad Q_i(z) \geq 0 \ \forall z.$$

$f(z)$, $z \in \{1, \ldots k\}$.

$E[f(z)] = \sum_{i=1}^{k} p(z=i) \cdot f(i)$

$$\sum_{z=1}^{k} Q_i(z) \frac{p_\theta(x^{(i)}, z)}{Q_i(z)} = \mathop{E}_{Q_i}\left[\frac{p_\theta(x^{(i)}, z)}{Q_i(z)}\right]$$

$$L(\theta) = \sum_{i=1}^{n} \log E_{z \sim Q_i}\left[\frac{p_\theta(x^{(i)}, z)}{Q_i(z)}\right]$$

$$\geq \sum_{i=1}^{n} E_{z \sim Q_i}\left[\log \frac{p_\theta(x^{(i)}, z)}{Q_i(z)}\right] \quad \text{(by Jensen's ineq)}$$

$$= \sum_{i=1}^{n} \sum_{z=1}^{k} Q_i(z)\left[\log \frac{p_\theta(x^{(i)}, z)}{Q_i(z)}\right]. \quad = F(\theta, Q)$$

Proposition: $\forall \theta$. $L(\theta) \geq F(\theta, Q) = \sum_{i=1}^{n} \sum_{z=1}^{k} Q_i(z)\left[\log \frac{p_\theta(x^{(i)}, z)}{Q_i(z)}\right]$

Expection-step: $Q^{(t+1)} = \arg\max_Q F(\theta^t, Q)$

Maximization Step: $\theta^{(t+1)} = \arg\max_\theta F(\theta, Q^{(t+1)})$

Assumption: ~~Computing θ that maximizes $F(\theta, Q)$~~

For any Q, the problem of computing θ that maximizes $F(\theta, Q)$ is computationally tractable.

$$L(\theta) = \sum_{i=1}^{n} \log \sum_{z=1}^{k} p_\theta(x^{(i)}, z) \geq \sum_{i=1}^{n} \sum_{z=1}^{k} Q_i(z) \log \frac{p_\theta(x^{(i)}, z)}{Q_i(z)}$$
$$= F(\theta, Q).$$

Take $Q_i(z) = p(Z=z \mid x^{(i)}; \theta) = p_\theta(Z=z \mid x^{(i)})$

In this case,

$$F(\theta, Q) = \sum_{i=1}^{n} \sum_{z=1}^{k} p_\theta(z \mid x^{(i)}) \log \underbrace{\frac{p_\theta(x^{(i)}, z)}{p_\theta(z \mid x^{(i)})}}_{\frac{p_\theta(x^{(i)}) \cdot p_\theta(z \mid x^{(i)})}{p_\theta(z \mid x^{(i)})} = p_\theta(x^{(i)})}$$

$$= \sum_{i=1}^{n} \sum_{z=1}^{k} p_\theta(z \mid x^{(i)}) \log p_\theta(x^{(i)})$$

$$= \sum_{i=1}^{n} \log p_\theta(x^{(i)}) \Big(\underbrace{\sum_{z=1}^{k} \underbrace{p_\theta(z \mid x^{(i)})}_{Q_i(z)}}_{=1} \Big)$$

$$= \sum_{i=1}^{n} \log p_\theta(x^{(i)}) = L(\theta)$$

Proposition: For any θ, $\arg\max_Q F(\theta, Q)$ is the distributions

$Q_i(z) = p_\theta(z \mid x^{(i)})$ and

$\max_Q F(\theta, Q) = L(\theta)$.

E-M Algo

Repeat.

E-step: $Q^{(t+1)} = \underset{Q}{\arg\max}\ F(Q, \theta^{(t)})$

M-step: $\theta^{(t+1)} = \underset{\theta}{\arg\max}\ F(Q^{(t+1)}, \theta)$

Until converge.

Proposition: $L(\theta^{(t)}) \le L(\theta^{(t+1)})$

$$L(\theta^{(t)}) = F(Q^{(t+1)}, \theta^{(t)}) \underbrace{\le F(Q^{(t+1)}}_{\theta^{(t+1)} \text{ maximizes } F \text{ at } Q^{(t+1)}.}, \theta^{(t+1)}) \le F(Q^{(t+2)}, \theta^{(t+1)}) = L(\theta^{(t+1)})$$