Solution to Ex7

Freitag, 9. Juni 2023 16:25

Task 1:

i) opt 9.55 – M step wrt π_k
show that we get 9.60

9.55: $\mathbb{E}_Z[\ln p(X,Z|\mu,\pi)]$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \Big\{ \ln \pi_k + \sum_{i=1}^{D}[x_{ni}\ln\mu_{ki} + (1-x_{ni})\ln(1-\mu_{ki})]\Big\}$$

9.60 $\pi_k = \frac{N_k}{N}$, $N_k = \sum_{n=1}^{N}\gamma(z_{nk})$

⇒ Add Lagrange multiplier term to (9.40) to enforce the constraint $\sum_{k=1}^{K}\pi_k = 1$:

$$J := \mathbb{E}_Z[\ln p(X,Z|\mu,\pi)] + \lambda\Big(\sum_{k=1}^{K}\pi_k - 1\Big)$$

$$\frac{\partial J}{\partial \pi_k} = \sum_{n=1}^{N}\gamma(z_{nk})\frac{1}{\pi_k} + \lambda$$

$$= \frac{N_k}{\pi_k} + \lambda \overset{!}{=} 0$$

$$\Rightarrow \quad N_k = -\lambda\pi_k \quad \text{*}$$

Summing both sides over k

$$\sum_k N_k = -\lambda\sum_k \pi_k$$

$$\sum_{n=1}^{N}\underbrace{\sum_k \gamma(z_{nk})}_{=1} = -\lambda \underbrace{\sum_k \pi_k}_{=1}$$

$$\Rightarrow \quad N = -\lambda$$

$$\text{*} \quad N_k = N\pi_k \Rightarrow \pi_k = \frac{N_k}{N}$$

Task 2:

i) Show that

$$\ln p(X|\theta) = \mathcal{L}(q,\theta) + KL(q||p)$$

for $\mathcal{L}(q,\theta) = \sum_z q(z) \ln \left\{ \frac{p(X,z|\theta)}{q(z)} \right\}$

and $KL(q||p) = -\sum_z q(z) \ln \left\{ \frac{p(z|X,\theta)}{q(z)} \right\}$

$\rightsquigarrow$

$$\mathcal{L}(q,\theta) = \sum_z q(z) \ln \left\{ \frac{p(X,z|\theta)}{q(z)} \right\}$$

$$= \sum_z q(z) \ln \left\{ \frac{p(z|X,\theta)\, p(X|\theta)}{q(z)} \right\}$$

$$= \sum_z q(z) \left\{ \ln \left\{ \frac{p(z|X,\theta)}{q(z)} \right\} + \ln p(X|\theta) \right\}$$

$$= \sum_z q(z) \ln \frac{p(z|X,\theta)}{q(z)} + \sum_z q(z) \ln p(X|\theta)$$

$$= -KL(q||p) + \ln p(X|\theta) \qquad \Big| + KL(q||p)$$

$$\Rightarrow \quad \ln p(X|\theta) = \mathcal{L}(q,\theta) + KL(q||p)$$

ii) This follows directly from the fact that $KL(q||p) \geq 0$

iii) Jensen's inequality for concave f:

$$f(\mathbb{E}[x]) \geq \mathbb{E}[f(x)]$$

$$\ln p(X|\theta) = \ln \sum_z p(X,z|\theta)$$

$$= \ln \sum_z \frac{q(z)}{q(z)} p(X,z|\theta)$$

$$\geq \sum_z q(z) \ln \left\{ \frac{p(X,z|\theta)}{q(z)} \right\}$$

$$= \mathcal{L}(q,\theta)$$

[note that $\mathbb{E}_{p(x)}[X] = \sum_x p(x) \cdot X$
for discrete random variables]

iv) For $q(z) = p(z|X, \theta^{old})$
we have

$$\nabla_\theta \ln p(X|\theta) =$$

$$= \nabla_\theta \mathcal{L}(p(z|X,\theta^{old}), \theta)$$

$$+ \nabla_\theta KL(p(z|X,\theta^{old}) \| p(z|X,\theta))$$

Since we know that the KL-divergence
$K(p(z|X,\theta^{old}) \| p(z|X,\theta))$
equals its <u>minimum</u> $(=0)$
for $\theta = \theta^{old}$, it
follows that $\nabla_\theta KL(p(z|X,\theta^{old}) \| p(z|X,\theta))$
is 0 at $\theta = \theta^{old}$.

v) Show that

$$\boxed{\frac{\prod_{n=1}^{N} p(x_n, z_n|\theta)}{\sum_{z} \prod_{n=1}^{N} p(x_n, z_n|\theta)}} = \prod_{n=1}^{N} p(z_n|x_n, \theta)$$

$= *$

We need to show that

$$\boxed{\sum_{z} \prod_{n=1}^{N} p(x_n, z_n|\theta) = \prod_{n=1}^{N} \sum_{z_n} p(x_n, z_n|\theta)}$$

[note: I missed the "n" here during class!]

since then

$$* = \prod_{n=1}^{N} \frac{p(x_n, z_n|\theta)}{p(x_n|\theta)} = \prod_{n=1}^{N} p(z_n|x_n, \theta)$$

$$\sum_{z} \prod_{n=1}^{N} p(x_n, z_n|\theta) = \sum_{z_1} \cdots \sum_{z_N} \left[p(x_1, z_1|\theta) \cdots p(x_N, z_N|\theta)\right]$$

$$= \sum_{z_1}\left[p(x_1, z_1|\theta)\right] \cdots \sum_{z_N}\left[p(x_N, z_N|\theta)\right]$$

$$= \prod_{n=1}^{N} \sum_{z_n} p(x_n, z_n|\theta)$$

[note that $\sum_{z_n}$ means $\sum_{z_n \in \{\text{"all possible values of } z_n\text{"}\}}$]