FS Solution to Ex 8

Freitag, 16. Juni 2023 20:34

Task 1)

ii) $\boxed{\text{max wrt. } \pi_k}$

We have that $\sum_{k=1}^{K} \pi_k = 1$

and π_k only in term 1 of $Q(\theta, \theta^{old})$, we then optimize

$$\tilde{Q} = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{1k}^n) \ln \pi_k$$
$$+ \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

$$\frac{\partial \tilde{Q}}{\partial \pi_k} = \sum_{n=1}^{N} \gamma(z_{1k}^n) \frac{1}{\pi_k} + \lambda \overset{!}{=} 0$$

$$\Rightarrow \lambda = - \sum_{n=1}^{N} \sum_{j=1}^{K} \gamma(z_{1j}^n)$$

same trick as in ex7 ...

substituting back gives

$$\sum_{n=1}^{N} \gamma(z_{1k}^n) \frac{1}{\pi_k} - \sum_{n=1}^{N} \sum_{j=1}^{K} \gamma(z_{1j}^n) = 0$$

$$\Rightarrow \quad \pi_k = \frac{\sum_{n=1}^{N} \gamma(z_{1k}^n)}{\sum_{n=1}^{N} \sum_{j=1}^{K} \gamma(z_{1j}^n)}$$

$\boxed{\text{max wrt. } A_{ij}}$

constraint is $\sum_{k=1}^{K} A_{jk} = 1$,

we then optimize

$$\hat{Q} = \sum_{n=1}^{N} \sum_{t=2}^{T} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{(t-1)j}^n, z_{tk}) \ln A_{jk}$$
$$+ \sum_{j=1}^{K} \lambda_j \left(\sum_{k=1}^{K} A_{jk} - 1 \right)$$

$\partial \hat{Q}$ N T

$$\sum_{j=1} \Big(\sum_{k=1} \quad \Big)$$

$$\frac{\partial \hat{Q}}{\partial A_{jk}} = \sum_{n=1}^{N} \sum_{t=2}^{T} \xi\left(z_{(t-1)j}, z_{tk}\right) \frac{1}{A_{jk}} + \lambda_j \overset{!}{=} 0$$

$$\Rightarrow \quad \lambda_j = -\sum_{n=1}^{N} \sum_{t=2}^{T} \sum_{l=1}^{K} \xi\left(z_{(t-1)j}, z_{tl}\right)$$

$$\Rightarrow \quad A_{jk} = \frac{\sum_{n=1}^{N} \sum_{t=2}^{T} \xi\left(z_{(t-1)j}, z_{tk}\right)}{\sum_{n=1}^{N} \sum_{t=2}^{T} \sum_{l=1}^{K} \xi\left(z_{(t-1)j}, z_{tl}\right)}$$

iii)

$\boxed{\text{max wrt. } \mu_{ki}}$

constraint is $\sum_{i=1}^{D} \mu_{ki} = 1$,

there we optimize

$$\bar{Q} = \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma(z_{tk}^{n}) \sum_{i=1}^{D} x_{ti}^{n} \ln \mu_{ki}$$

$$+ \lambda \left(\sum_{i=1}^{D} \mu_{ki} - 1 \right)$$

$$\frac{\partial \bar{Q}}{\partial \mu_{ki}} = \sum_{n=1}^{N} \sum_{t=1}^{T} \gamma(z_{tk}^{n})\, x_{ti}^{n} \frac{1}{\mu_{ki}} + \lambda \overset{!}{=} 0$$

$$\Rightarrow \quad \lambda = -\sum_{n=1}^{N} \sum_{t=1}^{T} \underbrace{\sum_{j=1}^{D} \gamma(z_{tk}^{n})\, x_{tj}^{n}}_{= \gamma(z_{tk}^{n})}$$

$\rightsquigarrow$ inserted into above gives

$$\mu_{ki} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T} \gamma(z_{tk}^{n})\, x_{ti}^{n}}{\sum_{n=1}^{N} \sum_{t=1}^{T} \gamma(z_{tk}^{n})}$$

iv)

$$\mathbb{E}_Z \left[\ln p(X, Z \mid \theta) \right]$$

$$= \mathbb{E}_Z \left[\sum_{n=1}^{N} \ln p(x^n, z^n \mid \theta) \right]$$ by linearity of Expectation

$$= \mathbb{E}_Z\left[\sum_{n=1}^{N} \ln p(x^n, z^n|\theta)\right] \quad \text{by linearity of Expectation}$$

$$= \sum_{n=1}^{N} \mathbb{E}_Z\left[\ln p(x^n, z^n|\theta)\right]$$

$$= \sum_{n=1}^{N} \sum_{z^1} \cdots \sum_{z^N} \prod_{m=1}^{N} p(z^m|x^m, \theta^{old}) \ln p(x^n, z^n|\theta)$$

$$= \sum_{n=2}^{N} \sum_{z^1} \cdots \sum_{z^N} \prod_{m=1}^{N} p(z^m|x^m, \theta^{old}) \ln p(x^n, z^n|\theta)$$

$$+ \sum_{z^1} \cdots \sum_{z^N} \prod_{m=1}^{N} p(z^m|x^m, \theta^{old}) \ln p(x^1, z^1|\theta)$$

$$= \sum_{n=2}^{N} \sum_{z^1} \cdots \sum_{z^N} \prod_{m=1}^{N} p(z^m|x^m, \theta^{old}) \ln p(x^n, z^n|\theta)$$

$$+ \sum_{z^1} p(z^1|x^1, \theta^{old}) \ln p(x^1, z^1|\theta) \underbrace{\sum_{z^2} p(z^2|x^2, \theta^{old}) \cdots \sum_{z^N} p(z^N|x^N, \theta^{old})}_{= 1 \cdots 1 = 1}$$

$$= \sum_{n=1}^{N} \sum_{z^n} p(z^n|x^n, \theta^{old}) \ln p(x^n, z^n|\theta)$$

$\rightsquigarrow$ Thus the expression is just Equation 13.17 from Bishop summed over all sequences (with different indices for sequences and timesteps ...)

$\blacksquare$