Homework 3 Solution: Mathematics for AI
Due date: November 9th, 11:59 pm.
Note: Submit your solution file to TA, HyeongJoo Hwang (hjhwang@ai.kaist.ac.kr), by email.
For questions, contact TA as well.

---

1. Consider the graphical model in Figure 1 which defines the following:

$$p\left(x_n \mid \theta\right)=\sum_{j=1}^m p_j\left[\sum_{k=1}^l q_k N\left(x_n \mid \mu_j, \sigma_k^2\right)\right],$$

where $\theta=\{p_1,...,p_m,\mu_1,...,\mu_m,q_1,...,q_l,\sigma_1^2,...,\sigma_l^2\}$ are all the parameters. Here $p_j \triangleq P(J_n=j)$ and $q_k \triangleq P(K_n=k)$ are the equivalent of mixture weights. We can think of this as a mixture of m non-Gaussian components, where each component distribution is a scale mixture, $p(x|j;\theta)=\sum_{k=1}^l q_k N(x;\mu_j,\sigma_k^2)$, combining Gaussians with different variances (scales).

For this model, we will now derive a generalized EM algorithm where we do a partial update in the M step instead of finding the exact maximum.

(a) Derive an expression for the responsibilities, $P(J_n=j,K_n=k|x_n,\theta)$, needed for the E step.

Solution : $r_{njk}=P\left(J_n=j, K_n=k \mid x_n, \boldsymbol{\theta}\right)=\frac{P\left(J_n=j, K_n=k, x_n \mid \boldsymbol{\theta}\right)}{p\left(x_n \mid \boldsymbol{\theta}\right)}$

$$=\frac{P\left(J_n=j\right) P\left(K_n=k\right) P\left(x_n \mid \boldsymbol{\theta}\right)}{\sum_{j^{\prime}} p_{j^{\prime}}\left[\sum_{k^{\prime}} q_{k^{\prime}} \mathcal{N}\left(x_n \mid \mu_{j^{\prime}}, \sigma_{k^{\prime}}^2\right)\right]}=\frac{p_j q_k \mathcal{N}\left(x_n \mid \mu_j, \sigma_k^2\right)}{\sum_{j^{\prime}} p_{j^{\prime}}\left[\sum_{k^{\prime}} q_{k^{\prime}} \mathcal{N}\left(x_n \mid \mu_{j^{\prime}}, \sigma_{k^{\prime}}^2\right)\right]}$$

(b) Write out a full expression for the expected complete log-likelihood

$$Q\left(\theta^{\text{new}}, \theta^{\text{old}}\right)=E_{\theta^{\text{old}}} \sum_{n=1}^N \log P\left(J_n, K_n, x_n \mid \theta^{\text{new}}\right)$$

Solution : $Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}_{t-1}\right)=\mathbb{E}_{\theta_{l-1}}\left[\sum_n \log p\left(x_n, J_n, K_n \mid \boldsymbol{\theta}\right)\right]$

$$=\sum_n \mathbb{E}\left[\log \prod_j \prod_k p\left(x_n, J_n, K_n \mid \boldsymbol{\theta}\right)^{\mathbf{1}_{J_n=j, K_n=k}}\right]$$

$$=\sum_n \sum_j \sum_k r_{njk}\left(\log p_j+\log q_k+\log \mathcal{N}\left(x_n \mid \mu_j, \sigma_k^2\right)\right)$$

(c) Solving the M-step would require us to jointly optimize the means $\mu_1,...,\mu_m$ and the variances $\sigma_1^2,...,\sigma_l^2$. It will turn out to be simpler to first solve for the μ_j's given fixed σ_j^2's, and subsequently solve for σ_j^2's given the new values of μ_j's. We will just do the first part. Derive an expression for the maximizing μ_j's given fixed $\sigma_{1:l}^2$, i.e., solve $\frac{\partial Q}{\partial \mu^{\text{new}}}=0$.

Solution :

$$\frac{\partial Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}_{t-1}\right)}{\partial \mu_j}=\sum_n \sum_k r_{njk}\left(\frac{x_n-\mu_j}{\sigma_k^2}\right)=0$$

$$\text{Thus, } \mu_j=\frac{\sum_n \sum_k \frac{r_{njk}\left(x_n-\mu_j\right)}{\sigma_k^2}}{\sum_n \sum_k \frac{r_{njk}}{\sigma_k^2}}$$

2. Derive the residual error for PCA by following the steps below:

(a) Based on the fact that (1) $\mathbf{v}_j^T\mathbf{v}_j = 1$ and $\mathbf{v}_j^T\mathbf{v}_k = 0$ for $k \neq j$ and (2) $z_{ij} = \mathbf{x}_i^T\mathbf{v}_j$, show:

$$\left\| \mathbf{x}_i - \sum_{j=1}^{K} z_{ij}\mathbf{v}_j \right\|^2 = \mathbf{x}_i^T\mathbf{x}_i - \sum_{j=1}^{K} \mathbf{v}_j^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{v}_j$$

Hint: Begin with the case $K = 2$.

Solution :

$$\begin{aligned}\left\| \mathbf{x}_i - \sum_j z_{ij}\mathbf{v}_j \right\|^2 &= \left(\mathbf{x}_i - \sum_j z_{ij}\mathbf{v}_j\right)^T \left(\mathbf{x}_i - \sum_j z_{ij}\mathbf{v}_j\right) \\ &= \mathbf{x}_i^T\mathbf{x}_i - 2\sum_j z_{ij}\mathbf{x}_i^T\mathbf{v}_j + \sum_j z_{ij}^2\mathbf{v}_j^T\mathbf{v}_j + \sum_j\sum_k z_{ij}z_{ik}\mathbf{v}_j^T\mathbf{v}_k \\ &= \mathbf{x}_i^T\mathbf{x}_i - 2\sum_j \mathbf{v}_j^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{v}_j + \sum_j z_{ij}^2 = \mathbf{x}_i^T\mathbf{x}_i - \sum_j \mathbf{v}_j^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{v}_j\end{aligned}$$

(b) Based on the fact that $\mathbf{v}_j^T\mathbf{C}\mathbf{v}_j = \lambda_j\mathbf{v}_j^T\mathbf{v}_j = \lambda_j$, prove the following equality:

$$J_K \triangleq \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T\mathbf{x}_i - \sum_{j=1}^{K}\mathbf{v}_j^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{v}_j\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i - \sum_{j=1}^{K}\lambda_j$$

Solution :

$$\begin{aligned}\frac{1}{n}\sum_i\left(\mathbf{x}_i^T\mathbf{x}_i - \sum_j\mathbf{v}_j^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{v}_j\right) &= \frac{1}{n}\sum_i\mathbf{x}_i^T\mathbf{x}_i - \frac{1}{n}\sum_i\sum_j\mathbf{v}_j^T\mathbf{C}\mathbf{v}_j \\ &= \frac{1}{n}\sum_i\mathbf{x}_i^T\mathbf{x}_i - \frac{1}{n}\sum_i\sum_j\mathbf{v}_j^T\lambda_j\mathbf{v}_j \\ &= \frac{1}{n}\sum_i\mathbf{x}_i^T\mathbf{x}_i - \frac{1}{n}\sum_i\sum_j\lambda_j = \frac{1}{n}\sum_i\mathbf{x}_i^T\mathbf{x}_i - \sum_j\lambda_j\end{aligned}$$

(c) If $K = d$ there is no truncation, so $J_d = 0$. Use this to show that the error from only using $K < d$ terms is given by

$$J_K = \sum_{j=K+1}^{d}\lambda_j$$

Hint: partition the sum $\sum_{j=1}^{d}\lambda_j$ into $\sum_{j=1}^{K}\lambda_j$ and $\sum_{j=K+1}^{d}\lambda_j$.

Solution : Since $J_d = 0$,

$$\frac{1}{n}\sum_i\mathbf{x}_i^T\mathbf{x}_i = \sum_{j=1}^{d}\lambda_j$$

$$\text{Thus, } J_k = \sum_{j=1}^{d}\lambda_j - \sum_{j=1}^{K}\lambda_j = \sum_{j=K+1}^{d}\lambda_j$$