Homework 3: 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.

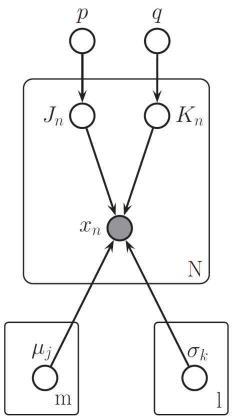
---

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).

Figure 1: A mixture of Gaussians with two discrete latent indicators. $J_n$ specifies which mean to use, and $K_n$ specifies which variance to use.




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.

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

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

(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 $\mu_j$'s given fixed $\sigma_j^2$'s, and subsequently solve for $\sigma_j^2$'s given the new values of $\mu_j$'s. We will just do the first part. Derive an expression for the maximizing $\mu_j$'s given fixed $\sigma_{1:l}^2$, i.e., solve $\frac{\partial Q}{\partial \mu^{\text{new}}}=0$.

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 $K = 2$.

(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:

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

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