HW1 - Representation Learning

1) $C_x = \frac{1}{n} XX^T$

$\because$ C_x is a symmetric matrix,

$C_x = EDE^T$ where D is a digonal matrix and E is a matrix of eigen vectors of C_x.

We have to find an orthonormal matrix P in $Y = PX$ such that $C_Y = \frac{1}{n} YY^T$ is a diggonal matrix.

$\rightarrow C_Y = \frac{1}{n} YY^T$

$\Rightarrow C_Y = \frac{1}{n} PX X^T P^T$

$\Rightarrow C_Y = P(\frac{1}{n} XX^T)P^T$

$\Rightarrow C_Y = P C_x P^T$

$\Rightarrow C_Y = P(EDE^T)P^T$

$C_Y = (PE)D(E^T P^T)$

Since C_Y is a diagonal matrix, $PE = I = E^T P^T$

$\Rightarrow$ $\underline{P = E^T}$

2) Log likelihood function of a GMM:

$$\mathcal{L}(x;\theta) = \sum_{n=1}^{N} \log\left[\sum_{j=1}^{k} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)\right]$$

$$\frac{\partial}{\partial \mu_k} \mathcal{L}(x;\theta) = \sum_{n=1}^{N} \frac{\pi_k \frac{d}{d\mu_k}[\mathcal{N}(x_n, \mu_k, \Sigma_k)]}{\sum_{j=1}^{k} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)} \quad -①$$

$$\frac{d}{d\mu_k}(\mathcal{N}(x_n, \mu_k, \Sigma_k)) = \frac{1}{\sqrt{(2\pi)^d \Sigma_k}} e^{-\frac{1}{2}[(x_n-\mu_k)^T \Sigma_k^{-1} (x_n-\mu_k)]} \cdot (-1) \Sigma_k^{-1}(x_n-\mu_k)$$

$$= \mathcal{N}(x_n, \mu_k, \Sigma_k) \cdot (-1) \cdot \Sigma_k^{-1}(x_n-\mu_k) \quad -②$$

Substituting ② in ①

$$\frac{\partial}{\partial \mu_k} \mathcal{L}(x;\theta) = \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n, \mu_k, \Sigma_k)}{\sum_{j=1}^{k} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)} \cdot \Sigma_k^{-1}(x_n-\mu_k) = 0$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k) = 0$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(z_{nk}) x_n - \mu_k \sum_{n=1}^{N} \gamma(z_{nk}) = 0$$

$$\Rightarrow \mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})} = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{N_k}$$

$$\frac{\partial}{\partial \Sigma_k} l(x;\theta) = \sum_{n=1}^{N} \frac{\pi_k \frac{d}{d\Sigma_k}[\mathcal{N}(x_n, \mu_k, \Sigma_k)]}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)}$$

$$\frac{d}{d\Sigma_{kj}} = \frac{1}{\sqrt{(2\pi)^d}\,\sigma_{kj}} e^{-\left(\frac{\sum_{j=1}^{d}(x_{nj}-\mu_{kj})^2}{2\sigma_{kj}^2}\right)} \cdot \left[\frac{-1}{\sigma_{kj}} + \frac{(x_{nj}-\mu_{kj})^2}{\sigma_{kj}^3}\right]$$

$$\frac{\partial}{\partial \Sigma_{kj}} = \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)} \left[\frac{-1}{\sigma_{kj}} + \frac{(x_{nj}-\mu_{kj})^2}{\sigma_{kj}^3}\right] = 0$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(z_{nk})\left(-\sigma_{kj}^2 + (x_{nj}-\mu_{kj})^2\right) = 0$$

$$\Rightarrow \sigma_{kj}^2 = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_{nj}-\mu_{kj})^2}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\Rightarrow \Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\Rightarrow \Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T}{N_k}$$

For finding the value of π_k, we need to take into account the constraint, $\sum_{k=1}^{K} \pi_k = 1$

We need to maximize $\ell(x;\theta) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)$

$$\Rightarrow \sum_{n=1}^{N} \left[\frac{\mathcal{N}(x_n, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)} \right] + \lambda = 0$$

$$\Rightarrow \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)} + \lambda \pi_k = 0$$

$$\Rightarrow \sum_{n=1}^{N} \frac{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)} + \lambda \sum_{k=1}^{K} \pi_k = 0$$

$$\Rightarrow \sum_{n=1}^{N} 1 + \lambda(1) = 0$$

$$\Rightarrow \underline{\underline{\lambda = -N}}$$

$$\Rightarrow N\pi_k = \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n, \mu_j, \Sigma_j)}$$

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

$$\Rightarrow \underline{\underline{\pi_k = \frac{N_k}{N}}}$$