Raktim Gautam Goswami
EE17BTECH11051

Representation Learning

HW 1

Q1)

Sol^n Let us be given data $X \in \mathbb{R}^{d \times N}$

We know, $C_{XX} = \frac{1}{N} X . X^T$ (C_{XX} is covariance matrix)

Our goal is to find $Y = PX$,

such that $C_{YY} = \frac{1}{N} Y . Y^T$ is minimized.

This happens when C_{YY} is a diagonal matrix.

$$C_{YY} = \frac{1}{N} PX . (PX)^T$$

$$= \frac{1}{N} P X X^T P^T$$

$$= P \frac{XX^T}{N} P^T$$

$$= P C_{XX} P^T$$

From spectral decomposition theorem, we know that,

$C_{XX} = E D E^T$ ($\because C_{XX}$ is symmetric)

E is orthonormal matrix

& D is diagonal.

$\therefore C_{YY} = P E D E^T P^T$

$\therefore$ it is clear that C_{YY} is diagonal if $P = E^T$

$\therefore Y = E^T X$.

Q2)

Sol: For GMM,

$$p(\underline{x}; \mu_k, \Sigma_k) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\underline{x}; \underline{\mu}_k, \Sigma_k)$$

Also, $\sum \pi_k = 1$

$0 \le \pi_k \le 1$
$\forall k$

Let us define $z := [0 \ldots 1 \ldots 0]^T$

Let $p(z_k = 1) = \pi_k$

$$\therefore p(\underline{z}) = \prod_{k=1}^{K} \{\pi_k^{z_k}$$

$$p(\underline{x} \mid z_k = 1) = \mathcal{N}(\underline{x}; \mu_k, \Sigma_k)$$

$$\Rightarrow p(\underline{x} \mid \underline{z}) = \prod_{k=1}^{K} \mathcal{N}(\underline{x}, \mu_k, \Sigma_k)^{z_k}$$

We now define,

prior prob $\rightarrow p(z_k = 1) = \pi_k$, $p(\underline{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$

Posterior Prob $\Rightarrow$

$$p(z_k = 1 \mid \underline{x}) = \frac{p(\underline{x} \mid z_k = 1)\, p(z_k = 1)}{p(\underline{x})}$$

$$p(z_k = 1 \mid \underline{x}) = \frac{\pi_k \mathcal{N}(\underline{x}, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_k \mathcal{N}(\underline{x}, \underline{\mu}_j, \Sigma_j)}$$

$$= \gamma(z_k)$$

Similarily

$$\gamma(z_{nk}) = p(z_k = 1 \mid \underline{x}_n) = \frac{\pi_k \mathcal{N}(\underline{x}_n; \mu_k, \Sigma_k}{\sum_{j=1}^{K} \pi_k \mathcal{N}(\underline{x}_n; \mu_j, \Sigma_j}$$

$$\log(L(X,\theta)) = \sum_{n=1}^{N} \log\left[\sum_{k=1}^{K} \pi_k \mathcal{N}(\underline{x}_n, \underline{\mu}_k ; \Sigma_k)\right]$$

$$\mathcal{N}(\underline{x}_n, \mu_j, \Sigma_j) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} \exp\left[-\frac{1}{2}(\underline{x}-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)\right]$$

For simplicity, we assume Σ to be a diagonal matrix

Now, for finding optimal parameters,

$$\frac{\partial}{\partial \mu_k} \log L(x; \underline{\mu}_k, \underline{\Sigma}_k) = \sum_{n=1}^{N} \frac{\frac{\partial}{\partial \mu_k}\left[\pi_k \mathcal{N}(\underline{x}_n, \underline{\mu}_k, \Sigma_k)\right)}{\sum_{i=1}^{K} \pi_i \mathcal{N}(\underline{x}_n, \underline{\mu}_i, \Sigma_i)}$$

$$\frac{\partial}{\partial \mu_k}(\mathcal{N}(x_n, \underline{\mu}_k, \Sigma_k)) = \mathcal{N}(x_n, \mu_k, \Sigma_k) \frac{\partial}{\partial \mu_k}\left[-\frac{1}{2}\sum_i \frac{(x_n - \mu_k)}{\sigma_{ki}}\right]$$

$$= \mathcal{N}(\underline{x}_n, \underline{\mu}_k, \Sigma_k) . \Sigma^{-1} (\underline{x}_n - \underline{\mu}_k)$$

Putting,

$$\therefore \frac{\partial \log L}{\partial \mu_k} = 0.$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(z_{nk}) \Sigma^{-1} (\underline{x}_n - \underline{\mu}_k) = 0$$

$$\Rightarrow \underline{\mu}_k = \frac{\sum_{n=1}^{N} [\gamma(z_{nk}) \underline{x}_n]}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\therefore \boxed{\underline{\mu}_k = \frac{\sum_{n=1}^{N} [\gamma(z_{nk}) \underline{x}_n]}{N_k}} \qquad N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

Now, for Σ_k (optimal)

$$\frac{\partial \log L}{\partial \Sigma_k} = \sum_{n=1}^{N} \frac{\frac{\partial}{\partial \Sigma_k}(\pi_k \mathcal{N}(\underline{x}_n, \underline{\mu}_k, \Sigma_k))}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x_n, \mu_l, \Sigma_l)}$$

$$\mathcal{N}(\underline{x}_n, \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^d}\sqrt{|\Sigma_k|}} \exp\left[-\frac{1}{2}(\underline{x}_n - \underline{\mu}_k)^T \Sigma_k^{-1} (\underline{x}_n - \underline{\mu}_k)\right]$$

$$= \frac{1}{\sqrt{(2\pi)^d}} \frac{1}{\prod_{j=1} \sigma_{kj}} \exp\left[-\frac{1}{2} \sum_{j=1}^{d} \frac{(x_{nj} - \mu_{kj})^2}{\sigma_{kj}^2} \right]$$

Setting this to 0,

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

$$\therefore \boxed{\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\underline{x}_n - \underline{\mu}_k)(\underline{x}_n - \underline{\mu}_k)^T}$$

Now, for minimizing w.r.t. π_k

$\therefore$ we will maximize

$$\log L(x, \theta) + d \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

We are using this because $\sum_{k=1}^{K} \pi_k = 1$

$$\therefore \sum_{n=1}^{N} \frac{N(\underline{x}_n ; \underline{\mu}_k, \Sigma_k)}{\sum_{j=1}^{K} N(\underline{x}_n, \underline{\mu}_j, \Sigma_j)} + d = 0 \quad \text{---} (1)$$

$$\Rightarrow \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{\pi_k N(\underline{x}_n, \underline{\mu}_k, \Sigma_k)}{\sum_{j=1}^{K} N(x_n, \underline{\mu}_j, \Sigma_j)} + \sum_{k=1}^{K} d\, \pi_k = 0$$

$$\Rightarrow \sum_{n=1}^{N} 1 + d \sum_{k=1}^{K} \pi_k = 0$$

$$\Rightarrow N + d = 0$$

$$\Rightarrow d = -N$$

From (1), we have,

$$\sum_{n=1}^{N} \frac{\mathcal{N}(\underline{x}_n, \underline{\mu}_k, \Sigma_k)}{\sum_{j=1} \mathcal{N}(\underline{x}_n, \underline{\mu}_j, \Sigma_j)} - N$$

$$\Rightarrow N = \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(\underline{x}_n, \underline{\mu}_k, \Sigma_k)}{\sum_j \mathcal{N}(\underline{x}_n, \underline{\mu}_j, \Sigma_j)} \frac{1}{\pi_k}$$

$$\Rightarrow N = N_k \frac{1}{\pi_k}$$

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