Date: __.__.__

Q2 GMM distribution :-

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

$$\theta = \{\pi_k, \underline{\mu}_k, \Sigma_k\}_{k=1}^{K}$$

Also $\sum \pi_k = 1$ $\quad 0 \leq \pi_k \leq 1 \;\forall k$

$\Sigma_k$ is a diagnal matrix

Define $Z = [\underbrace{0 \,-\!-\, 1 \,-\!-\, 0}_{K}]^T$ (one hot vector)

let $P(z_k = 1) = \pi_k$

$$P(z) = \prod_{k=1}^{K} (\pi_k)^{z_k}$$

$$P(\cancel{x} / z_k = 1) = \mathcal{N}(\underline{x}; \underline{\mu}_k, \Sigma_k)$$

$$P(x/z) = \prod_{k=1}^{K} \mathcal{N}(\underline{x}, \underline{\mu}_k, \Sigma_k)^{z_k}$$

Prior Probability $= P(z_k = 1) = \pi_k$, $P(z) = \prod_{k=1}^{K} \pi_k^{z_k}$

apsara

Date: __.__.__

Posterior Prob:- $P(z_k=1/\underline{x}) = \dfrac{P(\underline{x}/z_k=1)\,P(z_k=1)}{P(x)}$

~~$P(z_k=1/x)$~~ → (Bayes Theorem)

$$P(z_k=1/x) = \frac{\pi_k \mathcal{N}(\underline{x}; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\underline{x}; \mu_j, \Sigma_j)}$$

$$= \gamma(z_k)$$

$$\gamma(z_{nk}) = P(z_k=1/x_n)$$

$$= \frac{\pi_k \mathcal{N}(x_n; \mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n; \mu_j, \Sigma_j)}$$

$$\log(\mathcal{L}(x,\theta)) = \sum_{n=1}^{N} \log\left[\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n, \mu_k, \Sigma_k)\right]$$

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

as $\Sigma_j$ is diagnol matrix.

apsara

Date:__.__.__

$$-\frac{1}{2}\left[(x_n-\mu_j)^T \Sigma_j^{-1}(x_n-\mu_j)\right]$$

$$= -\frac{1}{2}\sum_{i=1}^{d}\frac{(x_{ni}-\mu_{ji})^2}{\sigma_{ji}^2}$$

$$\Rightarrow \frac{\partial}{\partial \mu_j}\left[-\frac{1}{2}(x_n-\mu_j)^T\Sigma_j^{-1}(x_n-\mu_j)\right] = \Sigma_j^{-1}(x_n-\mu_j)$$

$$\frac{\partial}{\partial u_k}\log(L(x;\theta)) = \sum_{n=1}^{N}\frac{\frac{\partial}{\partial u_k}\left(\pi_k N(x_n;\mu_k,\Sigma_k)\right)}{\sum_{i=1}^{K}\pi_i N(x_n;u_i,\Sigma_i)}$$

$$\frac{\partial}{\partial u_k}\left(N(x_n;\mu_k,\Sigma_k)\right) = N(x_n;u_k,\Sigma_k)\frac{\partial}{\partial u_k}\left[-\frac{1}{2}\sum_i\frac{(x_n-u_i)^2}{\sigma_{ki}}\right]$$

$$= N(x_n;\mu_k,\Sigma_k)\,\Sigma^{-1}(x_n-\mu_k)$$

$$\frac{\partial}{\partial u_k}\log(L(x;\theta)) = \sum_{n=1}^{N}\left[\frac{\pi_k N(x_n,\mu_k,\Sigma_k)\,\Sigma^{-1}(x_n-\mu_k)}{\sum_{i=1}^{K}\pi_i N(x_n;u_i,\Sigma_i)}\right]$$

$$\frac{\partial \log L}{\partial \mu_k} = \sum_{n=1}^{N} \gamma(z_{nk}) \Sigma^{-1} (\underline{x_n} - \underline{\mu_k}) = 0$$

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

$$\mu_k = \frac{\sum_{n=1}^{N} [\gamma(z_{nk}) \underline{x_n}]}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

Define $\rightarrow \sum_{n=1}^{N} \gamma(z_{nk}) = N_k$

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

Now

$$\frac{\partial \log(L(x;\theta))}{\partial \Sigma_k} = \frac{\partial}{\partial \Sigma_k} \left[ \sum_{n=1}^{N} \log \left[ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n, \mu_k, \Sigma_k) \right] \right]$$

$$= \sum_{n=1}^{N} \gamma(z_{nk}) \left[ \frac{-1}{\sigma_{kj}^2} + \frac{(x_{nj} - \mu_{kj})^2}{\sigma_{kj}^3} \right]$$

For optimal $\Sigma_{ki}$ putting $\frac{\partial \log(L)}{\partial \Sigma} = 0$

$$\sum_{n=1}^{N} \gamma(z_{nk}) \left[ \frac{-1}{\sigma_{kj}} + \frac{(x_{nj} - \mu_{kj})^2}{\sigma_{kj}^3} \right] = 0$$

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

$$= N_k$$

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

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

we know that $\sum_{k=1}^{K} \pi_k = 1$

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

which gives
$$0 = \frac{\sum_{n=1}^{N} N(x_n, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} N(x_n; \mu_j, \Sigma_j)} + c$$

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

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

$$\boxed{c = -N}$$

$$0 = \sum_{n=1}^{N} \frac{\mathcal{N}(x_n, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \mathcal{N}(x_n, \mu_j, \Sigma_j)} - N$$

$$N = \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n, \mu_k, \Sigma_k)}{\sum_{j} \mathcal{N}(x_n; \mu_j, \Sigma_j)} \left(\frac{1}{\pi_k}\right)$$

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

apsara

Date: __.__.____

HW0

Q1

$X \in \mathbb{R}^{d \times N}$, each row has a zero mean

Lets assume there is a matrix $P \in \mathbb{R}^{d \times d}$ such that $Y = PX$ gives diagnol covariance matrix i.e

$$C_{YY} = \frac{1}{N}(PX)(PX)^T \quad \text{is diagnol}$$

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

$$C_{YY} = \frac{1}{N} P C_{XX} P^T \qquad \boxed{XX^T = C_{XX}}$$

Since $C_{XX}$ is symmetric, we can write

$$C_{XX} = EDE^T \quad - (2)$$

where

$D$ is diagnol matrix and $E$ is orthogonal matrix

$$D = \begin{bmatrix} d_1 & & \\ & d_2 & 0 \\ 0 & & \ddots d_n \end{bmatrix} \qquad E = \begin{bmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_n \\ | & | & & | \end{bmatrix}$$

$$EE^T = I$$

So

$$C_{YY} = \frac{1}{N} PEDE^T P^T$$

apsara

Date: ___.___.___

if $P = E^T$

$$\boxed{C_{yy} = D}$$

hence $Y = E^T X$ gives us optimal decorrelated form of data contained in X.