Theory

(1)

$$Cov(x) = \frac{1}{N}(x x^T) \quad (\because E(x)=0)$$

Now take the transformation

$$Y = PX$$

We have to get maximum decorrelation i.e the Covariance matrix of Y has to be diagonal

$$cov(Y) = \frac{1}{N}(Y Y^T)$$

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

$$= P\left[\frac{XX^T}{N}\right) P^T$$

$$= P\,Cov(X)\,P^T$$

$$Cov(X) = E_x \lambda E_x^T \quad \text{(Eigen decomposition)}$$

$$E_x E_x^T = I$$

$$Cov(Y) = P\,E_x\,\lambda\,E_x^T P^T$$

If $P = E_x^T$,

Then $cov(Y)$ will be diagonal

since

$$Cov(Y) = (E_x^T E_x)\,\lambda\,(E_x^T E_x) = \lambda \rightarrow \text{diagonal matrix}$$

Thus

$$\boxed{P = E_x^T}$$

$$\boxed{Y = E_x^T X}$$

(2)

The MLE of θ for a GMM given $X = \{\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}\}$

$$L(\theta;x) = \prod_{i=1}^{N} P(\vec{x}^{(i)}; \vec{N}; \vec{\mu}, \Sigma)$$

$$\log(L(\theta;x)) = \log\left[\prod_{i=1}^{N} P(\vec{x}^{(i)}; \vec{N}; \vec{\mu}, \Sigma)\right]$$

$$= \sum_{i=1}^{N} \log\left[\sum_{k=1}^{K} w_K \mathcal{N}(\vec{x}^{(i)}; \vec{\mu}_k, \Sigma_k)\right]$$

First defining the posterior probabilities,

$$\gamma(z_k^{(i)}) = P(z_k = 1 \mid x^{(i)}; \theta)$$

$$\gamma(z_k^{(i)}) = P(z_k = 1 \mid x^{(i)}; \theta)$$

$$\gamma(z_k^{(i)}) = \frac{P(z_k = 1, x^{(i)}; \theta)}{P(x^{(i)}; \theta)}$$

$$\boxed{\gamma(z_k^{(i)}) = \frac{w_k \mathcal{N}(x^{(i)}, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} w_j \mathcal{N}(x^{(i)}; \mu_j, \Sigma_j)}}$$

coming to question,

taking partial derivatives of the log likelihood function of GMM with respect to each of the parameter and finding the local optimal solution.

$$\frac{\partial [\log L(\theta; x))]}{\partial \mu_k} = 0$$

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

$$= \sum_{i=1}^{N} \underbrace{\left[\frac{\omega_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \omega_j \mathcal{N}(x^{(i)}; \mu_j, \Sigma_j)} \right]}_{\gamma(z_k^{(i)})} \Sigma_k^{-1}(x_i - \mu_k) -$$

$$= \sum_{i=1}^{N} \gamma(z_k^{(i)}) \Sigma_k (x_i - \mu_k) = 0 .$$

$$\left[\sum_{i=1}^{N} \gamma(z_k^{(i)}) x_i \right] \Sigma_k^{-1} = \mu_k \left[\sum_{i=1}^{N} \gamma(z_k^{(i)}) \right] \Sigma_k^{-1}$$

multiply by $\underline{\Sigma_k}$ $[\Sigma_k \Sigma_k^{-1} = I]$

$$\sum_{i=1}^{N} \gamma(z_k^{(i)}) \mu_k = \sum_{i=1}^{N} \gamma(z_k^{(i)}) x_i$$

$$\boxed{\mu_k = \frac{\sum_{i=1}^{N} \gamma(z_k^{(i)}) x^{(i)}}{\sum_{i=1}^{N} \gamma(z_k^{(i)})}}$$

$$\Rightarrow \mu_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_k^{(i)}) x^{(i)} \qquad \boxed{N_k = \sum_{i=1}^{N} \gamma(z_k^{(i)})}$$

$$\frac{\partial \log(L(\theta;x))}{\partial \Sigma_1} = 0$$

$$\frac{\partial \log(L(\theta;x))}{\partial \Sigma_k} = \frac{\sum_{i=1}^{N}\left[\left(w_k \left(\frac{\partial N_k}{\partial \Sigma_k}\right)\right) \Big/ \left[\sum_{j=1}^{k} w_j \, N(x^{(i)}; \mu_j, \Sigma_j)\right]\right]}{L} = 0 \quad (1)$$

Now finding

$$\frac{\partial N_k}{\partial \Sigma_k} = \frac{\partial}{\partial \Sigma_k}\left[\sqrt{(2\pi)}^{k} \, |\Sigma_k|^{-1/2} \, e^{-\frac{(x^{(i)}-\mu_k)^T \Sigma_k^{-1} (x^{(i)}-\mu_k)}{2}}\right]$$

ignoring $\sqrt{(2\pi)}^{k}$ since its a constant

$$= \left(\Sigma_k^{-1}\right)^T \left(-\frac{1}{2}\right)\left(|\Sigma_k|^{-1/2}\right) e^{-\left(\frac{(x^i-\mu)^T \Sigma_k^{-1} (x^i-\mu_k)}{2}\right)}$$

$$+ \frac{1}{2} |\Sigma_k|^{-1/2} e^{-\frac{(x^i-\mu_k)^T \Sigma_k^{-1} (x^i-\mu_k)}{2}} \cdot (\Sigma_k^{-1})^T (x-\mu)(x-\mu)^T (\Sigma_k^{-1})^T$$

$$= \frac{(\Sigma_k^{-1})^T |\Sigma_k|^{-1/2}}{2}\left[-e^{-\frac{(x^i-\mu_k)^T \Sigma_k^{-1} (x^i-\mu_k)}{2}}\right]\left[I - (x^i-\mu_k)(x^i-\mu_k)^T \Sigma_k^{-1}\right]$$

substituting back in (1)

$$\frac{(\Sigma_k^{-1})^T |\Sigma_k|^{-1/2}}{2} \sum_{i=1}^{N}\left(\frac{w_k \, N(x^{(i)}; \mu_k; \Sigma_k)}{\sum_{j=1}^{k} w_j \, N(x^{(i)}; \mu_j, \Sigma_j)}\right)\left[I - (x^{(i)}-\mu_k)^T (x^{(i)}-\mu_k) \Sigma_k^{-1}\right] = 0$$

$$\sum_{i=1}^{N} \gamma(z_k^{(i)}) \left[I - (x^{(i)}-\mu_k)(x^{(i)}-\mu_k)^T \Sigma_k^{-1}\right] = 0 .$$

on solving

$$(\Sigma_k) \sum_{i=1}^{N} \gamma(z_k^{(i)}) = \sum_{i=1}^{N} \gamma(z_k^{(i)}) (x^{(i)}-\mu_k)(x^{(i)}-\mu_k)^T$$

$$\Rightarrow \Sigma_k = \frac{\sum_{i=1}^{N} \gamma(z_k^{(i)}) (\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T}{\sum_{i=1}^{N} \gamma(z_k^{(i)})}$$

Scanned by CamScanner

Sol

$N \rightarrow$ we derive for the optimal ω_K by using Lagrangian multipliers and maximising, with the condition

$\sum_{i:}$

$$\boxed{\sum_{k=1}^{K} \omega_K = 1}$$

ie maximising

$$Q(\theta) = \sum_{i=1}^{N} \log\left[\left(\sum_{k=1}^{K} \omega_K \mathcal{N}(x^{(i)}; \mu_K, \Sigma_K)\right)\right] + \lambda\left(\sum_{k=1}^{K} \omega_K - 1\right)$$

$$\frac{\partial Q(\theta)}{\partial \omega_K} = 0$$

$$\frac{\partial Q(\theta)}{\partial \omega_K} = \sum_{p=1}^{N} \frac{\mathcal{N}(x^{(i)}; \mu_K, \Sigma_K)}{\sum_{j=1}^{K} \omega_j \mathcal{N}(x^{(i)}; \mu_j, \Sigma_j)} + \lambda = 0 \quad \text{--} (2)$$

multiply both sides by ω_K, $\boxed{\omega_K \neq 0 \quad \forall K}$

Summing over K making use of the constraint $\sum_{k=1}^{K} \omega_k = 1$

$$\sum_{i=1}^{N} \left[\frac{\sum_{k=1}^{K} \omega_k \mathcal{N}(x^{(i)}; \mu_K, \Sigma_K)}{\sum_{j=1}^{K} \omega_j (\mathcal{N}(x^{i}; \mu_j, \Sigma_j)}\right] = -\lambda \sum_{i=1}^{K} \omega_K$$

Scanned by CamScanner

$$\sum_{i=1}^{N} (1) = -\lambda(1)$$

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

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

from ②

Substitute $\lambda = -N$ and multiply both sides by ω_k.

$$\sum_{i=1}^{N} \left[\frac{\omega_k \mathcal{N}(x^{(i)}; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \omega_j \mathcal{N}(x^{(i)}; \mu_j, \Sigma_j)} \right] = N\omega_k$$

$$\sum_{i=1}^{N} \gamma(z_k^{(i)}) = N\omega_k$$

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

where $\boxed{\sum_{i=1}^{N} \gamma(z_k^{(i)}) = N_k}$

$$\rightarrow \mu_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_k^{(i)}) \vec{x}^{(i)}$$

$$\boxed{N_k = \sum_{n=1}^{N} \gamma(z_k^{(n)})}$$

$$\rightarrow \Sigma_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_k^{(i)}) (\vec{x}^{(i)} - \vec{\mu}_k)(\vec{x}^{(i)} - \vec{\mu}_k)^T$$

$$\rightarrow \omega_k = \frac{N_k}{N}$$