HW 1 (EE19MTECH01008)

Tuesday, 24 September 2019 8:04 PM

① Derive expression for optimal decorrelating transform for a set of observations $X \in \mathbb{R}^{d \times N}$ where each row is assumed to be zero mean.

Ans:- $x \in \mathbb{R}^{d \times N}$ each row is zero mean

P is a linear transform such that,

$$Y = Px$$

$$\underbrace{\mathrm{Cor}}_{\text{Correlation matrix}}(Y) = \text{diagonal matrix}$$

$$C_x = \frac{1}{N}(xx^T) \quad \text{(Covariance matrix of } X\text{)}$$

Now, $Y = Px$

$$C_y = \frac{1}{N}\left[Px\,x^T P^T\right]$$

$$C_y = P C_x P^T$$

$\because C_x$ is PSD matrix

$$\therefore C_x = E_x \Sigma_x E_x^T \quad \text{(Eigen decomposition)}$$

$$\& \quad E_x E_x^T = E_x^T E_x = I$$

$$C_y = P E_x \Sigma_x E_x^T P^T$$

$$= (P E_x) \Sigma_x (E_x^T P^T)$$

If $P = E_x^T$

$C_y = \Sigma_x$ results in initial criteria.

$\therefore \boxed{P = E_x^T}$ is the transform applied on X such that Px is decorrelated.

② Derive expressions for partial derivatives of the log likelihood function of GMM w.r.t each parameters. Set these derivatives to zero and find expression for "locally optimal" parameters in terms of posterior probabilities & the observations.

set these derivatives to zero and find expression for "locally optimal" parameters in terms of posterior probabilities & the observations.

Ans:- $p(\underline{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i ; \mu_k, \Sigma_k)$

π_k is mixing parameter

$\pi_k \in (0,1]$ & $\sum \pi_k = 1$

$\underline{z}$ is latent variable with joint pdf $p(\underline{x}, \underline{z})$

$$p(x) = \sum_{\underline{z}} p(\underline{x}, \underline{z})$$

$$= \sum_{\underline{z}} p(\underline{z})\, p(\underline{x} \mid \underline{z})$$

$\underline{z}$ is one-hot vector with k^{th} index 1.

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

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

$$\Rightarrow p(x) = \sum_{\underline{z}} \left(\prod_{k=1}^{K} \pi_k^{z_k} \right) \left(p(\underline{x} \mid \underline{z}) \right)$$

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

Now,

$$L(x, \theta) = \prod_{i=1}^{N} p(\underline{x}_i, \pi, \mu, \Sigma)$$

$$\log L = \sum_{i=1}^{N} \log \left[p(\underline{x}_i, \pi, \mu, \Sigma) \right]$$

$$= \sum_{i=1}^{N} \log \left[\sum_{j=1}^{N} \pi_j \mathcal{N}(\underline{x}_i, \mu, \Sigma) \right]$$

Performing partial derivatives wrt parameters,

$$\frac{\partial \log L}{\partial \pi_j} = \sum_{i=1}^{N} \left[\frac{\mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j)}{\sum_{j=1}^{N} \pi_j \mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j)} \right] \quad \text{— } (1)$$

$$\frac{\partial \log L}{\partial \mu_j} = \sum_{i=1}^{N} \left[\frac{\pi_k \frac{\partial}{\partial \mu_j} \mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j)} \right] \quad - (2)$$

$$\frac{\partial \log L}{\partial \Sigma_j} = \sum_{i=1}^{N} \left(\frac{\pi_k \frac{\partial}{\partial \Sigma_j} \mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j)} \right) \quad - (3)$$

Multivariate Gaussian is given by

$$\mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j) = |2\pi\Sigma|^{1/2} \exp\left[-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right] \quad - (4)$$

$$\frac{\partial}{\partial \mu_j} \mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j) = |2\pi\Sigma|^{1/2} \exp\left[-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right] \left[2 (x-\mu)^T \Sigma^{-1} \right] \quad - (5)$$

posterior probabilities are given by

$p(z_k = 1 | \underline{x}_i) \rightarrow$ prob $\underline{x}_i$ come from K^{th} Gaussian $\because \underline{z}$ is one-hot vector

$p(\underline{x}_i | z_k = 1) \Rightarrow$ prob. $\underline{x}_i$ comes from $\mathcal{N}(\mu_k, \Sigma_k)$

$$\therefore p(\underline{x}_i | z_k = 1) = \mathcal{N}(\underline{x}_i, \mu_k, \Sigma_k)$$

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

By bayes rule,

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

$$= \frac{p(\underline{x}_i | z_k = 1)\, p(z_k = 1)}{\sum_{i=1}^{K} p(x_i | z_k = 1)\, p(z_k)}$$

$$p(z_k = 1 | \underline{x}_i) = \frac{\pi_k \mathcal{N}(\underline{x}_i, \mu_k, \Sigma_k)}{} = \gamma(z_k)$$

$$p(z_k = 1 \mid \underline{x}_i) = \frac{\pi_k \, N(\underline{x}_i, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \, N(\underline{x}_i, \mu_j, \Sigma_j)} = \gamma(z_k) \quad (\text{let})$$

Putting,

$$\frac{\partial \log L}{\partial \mu_j} = 0 \quad (\text{from } 2)$$

$$\Rightarrow \sum_{i=1}^{N} \left[\frac{\pi_k \, N(\underline{x}_i, \mu_k, \Sigma_k) \, 2(x-\mu)^T \Sigma^{-1}}{\sum_{j=1}^{K} \pi_j \, N(\underline{x}_i, \mu_j, \Sigma_j)} \right] = 0$$

$$\Rightarrow \sum_{i=1}^{N} \gamma(z_{k(i)}) \left(2(x-\mu)^T \Sigma^T \right) = 0$$

$$\Rightarrow \boxed{\mu = \frac{\sum_{i=1}^{N} \gamma(z_{k(i)}) \, [\underline{x}_{(i)}]}{\sum_{i=1}^{N} \gamma(z_{k(i)})}}$$

$$\frac{\partial \log L}{\partial \Sigma_k} = 0 \quad \text{From } 3$$

$$\Rightarrow \sum_{i=1}^{K} \left[\frac{\pi_k \frac{\partial}{\partial \Sigma_k} N(\underline{x}_i, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \, N(\underline{x}_i, \mu_j, \Sigma_j)} \right]$$

Derivating 4 w.r.t. Σ_k

$$\therefore \frac{\partial N(\underline{x}_i, \mu_k, \Sigma_k)}{\partial \Sigma_k} = -\frac{1}{2} \left(\Sigma_k^{-1} - \Sigma_k^{-1} (\underline{x}_i - \mu_k)(x_i - \mu_k)^T \Sigma_k^{-1} \right) N(\underline{x}_i, \mu_k, \Sigma_k)$$

$$\frac{\partial L}{\partial \Sigma_k} = 0$$

$$\Rightarrow \sum_{i=1}^{N} \gamma(z_{k(i)}) \left[\Sigma_k^{-1} - \Sigma_k^{-1} (\underline{x}_i - \mu_k)(\underline{x}_i - \mu_k)^T \Sigma_k^{-1} \right] = 0$$

$$\Rightarrow \Sigma_k = \frac{\sum_{i=1}^{N} \gamma(z_{k(i)}) \, (\underline{x}_i - \mu_k)(\underline{x}_i - \mu_k)^T}{\sum \gamma(z \ldots}$$

$$\Rightarrow \boxed{\Sigma_K = \frac{\sum_{i=1}^{N} \gamma(z_{k(i)}) (\underline{x}_i - \mu_K)(\underline{x}_i - \mu_K)^T}{\sum_{i=1}^{N} \gamma(z_{k(i)})}}$$

From (3)

$$\frac{\partial L}{\partial \pi_K} = 0 \qquad \sum_{i=1}^{K} \pi_K = 1$$

$$\Rightarrow \sum_{i=1}^{N} \left(\frac{\mathcal{N}(\underline{x}_i, \mu_K, \Sigma_K)}{\sum_{j=1}^{K} \pi_K \mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j)} \right) = 0$$

$$\Rightarrow \sum_{i=1}^{N} \frac{\gamma z_{nk_i}}{\pi_K} = 0 \quad \& \quad \sum_{k=1}^{K} \pi_K = 1$$

$$\Rightarrow \sum_{K=1}^{K} \pi_K - 1 = 0$$

By method of Lagrange multipliers,

$$l(p(X|\mu, \Sigma, \pi)) + \lambda \left(\sum_{k=1}^{K} \pi_K - 1 \right) = 0 \quad - (6)$$

Derivating w.r.t. π_K

$$\Rightarrow \lambda \pi_K + \underbrace{\sum_{i=1}^{N} \frac{\pi_K \mathcal{N}(\underline{x}_i, \mu_K, \Sigma_K)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\underline{x}_i, \mu_j, \Sigma_j)}} = 0 \quad - (7)$$

Summing,

$$\Rightarrow \lambda = -N \text{ (say)}$$

Substituting this on (6),

$$\boxed{\pi_K = \frac{\sum_{i=1}^{N} \gamma z_{nk_i}}{N}}$$

where N is (7)