EE5601. HW1

P. ARAVIND GANESH

EE16BTECH11026

1. Optimal Decorrelating Linear Transformations

$X \in \mathbb{R}^{d \times N}$

$Y = PX$ $P \to$ linear transformation

such that $\mathrm{Cor}(Y)$ {correlation matrix} is diagonal matrix.

$C_x = \frac{1}{N}(XX^T)$ $\to$ covariance matrix of data points.

$C_y = \frac{1}{N}(PXX^TP^T)$

$= PC_xP^T$

$C_x \to$ symmetric, positive semi-definite.

$\Rightarrow$ Eigen decomposition

$$C_x = E_x \Lambda_x E_x^T$$

$E_x E_x^T = I$ and Λ_x is a diagonal matrix with eigen values as diagonal elements.

$$\cancel{E_x E} \quad C_y = P E_x \Lambda_x E_x^T P^T$$

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

$$P = E_x^T \Rightarrow C_y = \Lambda_x$$

So linear transformation is $\cancel{E_x}$

$$Y = E_x^T X$$

$\Downarrow$

Y is decorrelated

---

2 $\quad p(\bar{x}) = \sum_{k=1}^{K} w_k \mathcal{N}(\bar{x}_i, \mu_k, \Sigma_k)$

$$\sum_{k=1}^{K} w_k = 1, \quad w_k > 0 \quad \forall k$$

Latent variable, $\vec{z}$ → one-hot vector

$\cancel{p(\vec{z}^i, x^i)}$.

$$p(x^i) = \sum_{z} p(\bar{x}, \vec{z}) = \sum_{z} p(\vec{z})\, p(\bar{x}/\bar{z})$$

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

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

$$p(x) = \sum_{\bar{z}} p(\bar{x}/\bar{z}_k) \prod_{k=1}^{K} w_k^{z_k}$$

$$p(\bar{x}/\bar{z}_k) = \prod_{k=1}^{K} \left[\mathcal{N}(\mu_k, \Sigma_k)\right]^{z_k}$$

$$L(\bar{x}, \theta) = \prod_{i=1}^{N} p(\bar{x}^{(i)}, w, \bar{\mu}, \Sigma)$$

$$\log L = \sum_{i=1}^{N} \log p(\bar{x}^i, w, \bar{\mu}, \Sigma)$$

$$= \sum_{i=1}^{N} \ln\left[\sum_{j=1}^{K} w_j \mathcal{N}(\bar{x}^{(i)}, \bar{\mu}_j, \Sigma_j)\right]$$

$$\frac{\partial \ln L}{\partial w_j} = \sum_{i=1}^{N} \frac{\mathcal{N}(\bar{x}_i, \bar{\mu}_j, \Sigma_j)}{\sum_{j=1}^{K} w_j \mathcal{N}(\bar{x}_i, \bar{\mu}_j, \Sigma_j)}$$

$$\frac{\partial \ln L}{\partial \bar{\mu}_j} = \sum_{i=1}^{N} \frac{w_k \frac{\partial}{\partial \bar{\mu}_j} \mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k)}{\sum_{k=1}^{K} w_k \mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k)}$$

$$\frac{\partial \ln L}{\partial \bar{\Sigma}_k} = \sum_{i=1}^{N} \frac{w_k \frac{\partial}{\partial \Sigma_k} \mathcal{N}(\bar{x}^{(i)}, \bar{\mu}_k, \Sigma_k)}{\sum_{k=1}^{K} w_k \mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k)}$$

$$\mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma_k)}}$$

$$\exp\left(\frac{-1}{2} (\bar{x}_i - \bar{\mu}_k)^T \Sigma_k^{-1} (\bar{x}_i - \bar{\mu}_k)\right)$$

$$\frac{\partial \ell'}{\partial \mu_k} \quad \frac{1}{\sqrt{\det(2\pi\Sigma_k)}} \cdot \exp\left[-\frac{1}{2}(\bar{x}_i - \mu_k)^T \Sigma_k^{-1} (\bar{x}_i - \mu_k)\right]$$

$$2\,((\bar{x}_i - \mu_k)^T \Sigma^{-1}$$

$$= 0$$

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

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

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

$$= \frac{p(\bar{x}_i \mid z_k = 1)\, \omega_k}{\sum_{k=1}^{K} p(\bar{x}_i, z_k = 1)}$$

$$= \frac{p(\bar{x}_i \mid z_k = 1)\, \omega_k}{\sum_{k=1}^{K} p(\bar{x}_i \mid z_k = 1)\, p(z_k = 1)}$$

$$P(z_k = 1 \mid \bar{x}_i) = \gamma(z_k^{(i)})$$

$$= \frac{w_k \, \mathcal{N}(\bar{x}_i, \mu_k, \Sigma_k)}{\sum_{k=1}^{K} w_k \, \mathcal{N}(\bar{x}_i, \mu_k, \Sigma_k)}$$

$\gamma(z_k^{(i)}) \rightarrow$ prob that $\bar{x}_i$ came from $\mathcal{N}(\mu_k, \Sigma_k)$

$$\frac{\partial \ln L}{\partial \mu_k} = \cancel{\sum_{i=1}^{N} \gamma(z_k^{(i)}) (\bar{x}_i - \bar{\mu}_k)^T}$$

$$= \sum_{i=1}^{N} \gamma(z_k^{(i)}) (\bar{x}_i - \bar{\mu}_k)^T \, 2 \, \Sigma_k^{-1} = 0$$

$$\Rightarrow \quad \bar{\mu}_k = \frac{\sum \gamma(z_k^{(i)}) \, \bar{x}_i}{\sum_{i=1}^{N} \gamma(z_k^{(i)})}$$

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

$$\frac{\partial \cancel{\ln L}}{\partial \Sigma_k} \mathcal{N} = \frac{\partial}{\partial \Sigma_k} \frac{1}{\sqrt{\det(2\pi\Sigma_k)}} \exp(\cdots)$$

$$+ \frac{1}{\sqrt{\det(2\pi\Sigma_k)}} \frac{\partial}{\partial \Sigma_k} \exp\left(\frac{-1}{2}(\bar{x}_i - \bar{u}_k)^T \Sigma^{-1} (\bar{x}_i - \bar{u}_k)\right)$$

$$= \frac{-1}{2} \Sigma_k^{-1} - \Sigma_k^{-1} (\bar{x}_i - \bar{u}_k)(\bar{x}_i - \bar{u}_k)^T \Sigma_k^{-1} \cdot$$

$$\mathcal{N}(\bar{x}_i, \bar{u}_k, \Sigma_k) = 0$$

$$\Rightarrow \sum_{i=1}^{N} \gamma(z_k^{(i)}) \left[\Sigma_k^{-1} - \Sigma_k^{-1} (\bar{x}_i - \bar{u}_k)(\bar{x}_i - \bar{u}_k)^T \Sigma_k^{-1}\right] = 0$$

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

$$\frac{\partial \ln L}{\partial w_k} = 0 \quad \text{and} \quad \sum w_k = 1$$

$$\Rightarrow \sum_{i=1}^{N} \frac{\mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k)}{\sum_{k=1}^{K} w_k \mathcal{N}(\bar{x}_i, \mu_k, \Sigma_k)} = 0$$

$$\Rightarrow \sum_{i=1}^{N} \frac{\gamma(z_k^{(i)})}{w_k} = 0 \quad \text{and} \sum_{k=1}^{K} w_k = 1$$

Applying Lagrange multipliers:

$$\ln(p(X \mid \mu, \Sigma, w)) + \lambda\left(\sum_{k=1}^{K} w_k - 1\right) = 0$$

$$\frac{\partial}{\partial w_k}(\cdot).$$

$$\Rightarrow \sum_{i=1}^{N} \frac{\mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k)}{\sum_{k=1}^{K} w_k \mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k)} + \lambda = 0$$

$$\lambda w_k + \sum_{i=1}^{N} \frac{w_k \mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k)}{\sum_{k=1}^{K} w_k \mathcal{N}(\bar{x}_i, \bar{\mu}_k, \Sigma_k)} = 0$$

Summing over k = 1 to K

$$\lambda(1) + N = 0$$

$$\lambda = -N.$$

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

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