HW1

Q.1) consider set of observations $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\}$ where $x \in \mathbb{R}^d$

Given: Data is zero mean.

X is the original dataset matrix (N×d).

Let Y be a matrix (N×d) related by linear transformation P.

$$Y = PX$$

Covariance Matrix

$$C_X = \frac{1}{N} X \cdot X^T$$

$$C_{X_{ii}} = \frac{1}{N} \bar{X}_i^T \cdot \bar{X}_i$$

- Diagonal elements are the variances of corresponding features.
- Off-Diagonal elements are covariance b/w features.

$$C_Y = \frac{1}{N} Y \cdot Y^T$$

$C_Y \rightarrow$ covariance matrix of Y (optimized)

Goal: All off-diagonal elements should be zero i.e. Y is decorrelated

Each successive dimension in Y should be rank-ordered according to variance.

$C_Y = \frac{1}{N}(PX)(PX)^T$ $\quad (\because Y = PX$

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

$= P\left(\frac{X X^T}{N}\right) P^T$

$$\boxed{C_Y = P C_X P^T}$$

C_X is positive semidefinite & symmetric.

$\therefore C_X = E_X \Lambda_X E_X^T$ (SVD)

E_X - matrix of Eigen vectors of C_X

Λ_X - Diagonal matrix of eigen values of C_X.

$E_X E_X^T = I$ & $E_X^T E_X = I$

$C_Y = P C_X P^T$

$= P E_X \Lambda_X E_X^T P^T$

IF $P = E_X^T$

$C_Y = \cancel{P}\, E_X^T E_X \Lambda_X E_X^T E_X$

$$\boxed{C_Y = \Lambda_X}$$

So, choice of P diagonalises C_Y.

Q.2) Gaussian Mixture distribution can be written as linear superposition of gaussian.

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

π_k - mixing coefficient

$$0 \leq \pi_k \leq 1$$

$$\sum_{k=1}^{K} \pi_k = 1$$

If z is discrete latent variable and has a joint distribution $p(x,z)$ with x then $p(\bar{x})$ is marginal of $p(x,z)$

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

$$= \sum_{z} p(z) . p(x|z)$$

If $\bar{z}$ is a one hot vector of dimension k i.e. $z_K \in \{0,1\}$ with $p(z_k = 1) = \pi_k$

$$p(\bar{z}) = \prod_{k=1}^{K} \pi_k^{z_k} \qquad \pi_k \neq 0$$

Assuming, data points are drawn independently from distribution, then log likelihood is given by

$$L(\theta, x) = \ln p(x| \pi_k, \mu_k, \Sigma_k)$$

$$= \sum_{i=1}^{N} \ln \left[\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k) \right]$$

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

$$= \frac{p(z_k = 1, \bar{x}^{(i)})}{p(\bar{x}^{(i)})}$$

$$= \frac{p(\bar{x}^{(i)} \mid z_k = 1)\, p(z_k = 1)}{\sum_{j=1}^{K} p(\bar{x}^{(i)}, z_j)}$$

$$= \frac{p(\bar{x}^{(i)} \mid z_k = 1)\, p(z_k = 1)}{\sum_{j=1}^{K} p(\bar{x}^{(i)} \mid z_j = 1)\, p(z_j = 1)}$$

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

At local optimum,

$$\frac{\partial L(\theta, \bar{x})}{\partial \mu_k} = 0 \qquad \frac{\partial L(\theta, \bar{x})}{\partial \Sigma_k} = 0 \qquad \frac{\partial L(\theta, \bar{x})}{\partial \pi_k} = 0$$

for multivariate Gaussian,

$$\mathcal{N}(\bar{x}; \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\bar{x} - \bar{\mu})^T \Sigma^{-1} (\bar{x} - \bar{\mu})\right]$$

$$L(\theta; \bar{x}) = \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k \left[\frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\bar{x} - \bar{\mu})^T \Sigma^{-1} (x - \bar{\mu})\right)\right] \quad \text{—— (1)}$$

$$\frac{\partial L(\theta; \bar{x})}{\partial \mu_k} = 0 \qquad \left(\text{differentiating eqn (1) w.r.t } \mu_k\right)$$

$$0 = \sum_{i=1}^{N} \left[\frac{\pi_k \cdot \mathcal{N}(x; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(x; \mu_j, \Sigma_j)} \frac{(x-\mu_k)}{|\Sigma|} \right]$$

$$0 = \sum_{i=1}^{N} \left[\frac{\gamma(z_{nk}) \cdot (x - \mu_k)}{\Sigma} \right]$$

$$\sum_{i=1}^{N} \gamma(z_{nk}) \cdot x_n = \sum_{i=1}^{N} \gamma(z_{nk}) \mu_k$$

$$\mu_k = \frac{\sum_{i=1}^{N} \gamma(z_{nk}) x_n}{\sum_{i=1}^{N} \gamma(z_{nk})}$$

$$\boxed{\mu_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{nk}) x_n} \quad \left(N_k = \sum_{i=1}^{N} \gamma(z_{nk}) \right)$$

for π_k,

we know $\sum_{i=1}^{K} \pi_k = 1$, ~~and~~ Add Langrange Multiplier to eqn (1),

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

differentiating w.r.t π_k.

$$0 = \sum_{i=1}^{N} \left[\frac{\mathcal{N}(x; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x; \mu_j, \Sigma_j)} \right] + \lambda \quad \text{--- (2)}$$

Multiplying both sides by π_k and summing over K.

$$0 = \sum_{i=1}^{N} \left[\frac{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j)} \right] + \lambda \sum_{i=1}^{K} \pi_k$$

$$0 = N + \lambda \qquad \left(\because \sum_{i=1}^{K} \pi_k = 1 \right)$$

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

substitute value of λ in eqn ②, and multiply both sides by π_k.

$$0 = \sum_{i=1}^{N} \left[\frac{\pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j)} \right] - N\pi_k$$

$$0 = \sum_{i=1}^{N} \gamma(z_{nk}) - N\pi_k$$

$$N\pi_k = N_k \qquad \left[\because N_k = \sum_{i=1}^{N} \gamma(z_{nk}) \right]$$

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

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

$$0 = \frac{1}{\sqrt{(2\pi)^D}} \left[\frac{-1}{2|\Sigma|^{3/2}} |\Sigma| \Sigma^{-1} \exp\left(\frac{-1}{2} (x-\mu_k) \Sigma^{-1} (x-\mu_k)^T \right) + \frac{1}{|\Sigma|^{1/2}} \exp\left(\frac{-1}{2} (x-\mu_k) \Sigma^{-1} (x-\mu_k)^T \right) \cdot \left(\frac{-1}{2} \Sigma^{-1} (x-\mu_k)(x-\mu_k)^T \right) \right]$$

Rearranging the above eqn,

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

Scanned by CamScanner