HW-1

(1) Optimal Decorrelating Transform :- AKA (PCA)

Let $x \in \mathbb{R}^{d \times N}$ data where each row is zero centred.

Let P be a linear transformation such that

$$Y = PX$$

Which results that $Cor(Y)$ = diagonal matrix

$Cor(Y)$ is the correlation matrix

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

If $Y = PX$

$$C_Y = \frac{1}{N}\left[PXX^TP^T\right]$$

$$C_Y = PC_XP^T$$

Now C_X is a symmetric and positive semi definite matrix

$$\Rightarrow C_X = E_X \Lambda_X E_X^T \quad \{\text{Eigen decomposition}\}.$$

$$\& \; E_X E_X^T = E_X^T E_X = I$$

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

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

Now if $P = E_X^T$

$$C_Y = \Lambda_X$$

which meets our initial criterion

$\therefore$ $P = E_X^T$ is the linear transformation

we can apply on our data X such that

PX is decorrelated.

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

where w_k is called as the mixing parameter

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

Let $\vec{z}$ be a latent variable and has a joint

distribution $p(\vec{x}, \vec{z})$ with $\vec{x}$

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

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

If $\vec{z}$ is a one-hot vector & $p(z_k = 1) = w_k$

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

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

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

Reason :- given $\vec{z}$, we know which gaussian has the most contribution. and z_k is either very close to zero or 1. Hence, raising $\mathcal{N}(\mu_k, \Sigma_k)$ to z_k gives the conditional.

Now let $\mathcal{L}(x, \theta) = \prod_{i=1}^{N} p(\vec{x}^{(i)}; w, \mu, \Sigma)$

$$\log \mathcal{L} = \sum_{i=1}^{N} \log \left[p(\vec{x}^{(i)}; w, \mu, \Sigma) \right]$$

$$= \cancel{\sum_{i=1}^{N} \log \left[\sum_{j=1}^{K} p(\vec{x}^{(i)}; w_j, \mu_j, \Sigma_j) \right]}$$

$$= \sum_{i=1}^{N} \log \left[\sum_{j=1}^{K} w_j \mathcal{N}(x^{(i)}, \mu_j, \Sigma_j) \right]$$

$\log p(x^{(i)}; \mu_j, \Sigma_j, \omega_j)$ → this term can't be broken down further or derivative can't be computed easily.

$$\frac{\partial \log L}{\partial \omega_j} = \sum_{i=1}^{N} \left[\frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{j=1}^{K} \omega_j \mathcal{N}(x^{(i)}, \mu_j, \Sigma_j)} \right]$$

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

$$\frac{\partial \log L}{\partial \sigma_j} = \sum_{i=1}^{N} \left[\frac{\omega_k \frac{\partial}{\partial \sigma_j} \mathcal{N}(x^{(i)}, \mu_j, \sigma_j)}{\sum_{j=1}^{K} \omega_j \mathcal{N}(x^{(i)}, \mu_j, \Sigma_j)} \right]$$

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

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

$$\Rightarrow \frac{\partial \log L}{\partial \mu_j} = 0$$

Consider the following posterior probabilities

$p(z_k=1 | x^{(i)})$ tells the probability that $x^{(i)}$ came from k^{th} gaussian

$p(x^{(i)} | z_k=1)$ tells the probability that $x^{(i)}$ comes from.

$$\mathcal{N}(\mu_k, \Sigma_k)$$

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

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

& $$p(z_k=1 | x^{(i)}) = \frac{p(x^{(i)} | z_k=1)\, p(z_k=1)}{p(x^{(i)})}$$ (Bayesian Rule)

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

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

$\rightarrow \gamma(z_k^{(i)})$ tells the probability that

$x^{(i)}$ came from distribution denoted

by z_k (k-s) $\mathcal{N}(\mu_k, \Sigma_k)$

Coming back to $\frac{\partial \log L}{\partial \mu_j}$

$$\frac{\partial \log L}{\partial \mu_j} = \sum_{i=1}^{N} \left(\frac{w_k \mathcal{N}(x^{(i)}, \mu_k, \Sigma_k)\, 2(x-\mu_k)^T \Sigma^{-1}}{\sum_{j=1}^{K} w_j \mathcal{N}(x^{(i)}, \mu_j \Sigma_j)} \right)$$

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

$$\Rightarrow \sum_{i=1}^{N} \gamma(z_k^{(i)}) [x^{(i)}]^T = \mu^T \sum_{i=1}^{N} \gamma(z_k^{(i)})$$

$$\Rightarrow \mu = \frac{\sum_{i=1}^{N} \gamma(z_k^{(i)}) [x^{(i)}]}{\sum_{i=1}^{N} \gamma(z_k^{(i)})}$$

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

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

$$\frac{\partial}{\partial \Sigma_k} \mathcal{N} = \frac{\partial}{\partial \Sigma_k} \left[|2\pi \Sigma_k|^{-1/2} \right] \exp\left[-\frac{1}{2} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \right]$$

$$+ |2\pi \Sigma_k|^{-1/2} \frac{\partial}{\partial \Sigma_k} \exp\left(-\frac{1}{2} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) \right)$$

Note:-
Σ is a symmetric matrix
$\rightarrow \frac{\partial}{\partial X} |X^k| = k|X^k| X^{-T}$
$= k|X^k| X^{-1}$
i.e, $\Sigma^{-T} = \Sigma^{-1}$
$\Sigma^T = \Sigma$

$$= \left(\cancel{|2\pi|^{-1/2}} \left(-\frac{1}{2} |\Sigma_k|^{-1/2} \Sigma_k^{-1} \right) \mathcal{N}(x^{(i)}, \mu_k, \Sigma_k) \right) + \left(|2\pi \Sigma_k|^{-1/2} \mathcal{N}(x^{(i)} \mu_k \Sigma_k) \cdot \frac{1}{2} \Sigma^{-1} (x^{(i)} - \mu) (x^{(i)} - \mu)^T \Sigma^{-1} \right)$$

$$= -\frac{1}{2} \left(\Sigma_k^{-1} - \Sigma_k^{-1} (x^{(i)} - \mu_k) (x^{(i)} - \mu_k)^T \Sigma_k^{-1} \right) \mathcal{N}(x^{(i)}, \mu_k \Sigma_k)$$

$$\therefore \frac{\partial \log L}{\partial \Sigma_k} = 0 \Rightarrow \sum_{i=1}^{N} \gamma(z_k^{(i)}) \left[\cancel{\Sigma_k^{-1}} - \Sigma_k^{-1} (x^{(i)} - \mu_k) (x^{(i)} - \mu_k)^T \Sigma_k^{-1} \right] = 0$$

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

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

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

for w_k

$$\frac{\partial \log L}{\partial w_k} = 0 \quad \& \quad \sum_{i=1}^{K} w_k = 1$$

$$\Rightarrow \sum_{i=1}^{N} \frac{\mathcal{N}(x^{(i)}, \mu_k, \Sigma_k)}{\sum_{j=1}^{K} w_k \mathcal{N}(x^{(i)}, \mu_k, \Sigma_k)} = 0$$

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

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

$$\Rightarrow \left[\sum_{k=1}^{K} w_k - 1 \right] = 0$$

$\Rightarrow$ By using Lagrange multipliers

$$\cancel{\sum_{i=1}^{N} \frac{\mathcal{N}(x^{(i)}, \mu_k, \Sigma_k)}{\sum_{k=1}^{K} w_k \mathcal{N}(x^{(i)}, \mu_k, \Sigma_k)} + \lambda \left[\sum_{k=1}^{K} w_k - 1 \right] = 0}$$

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

applying $\frac{\partial}{\partial w_k}$ on both sides yields

$$\sum_{i=1}^{N} \frac{\mathcal{N}\left(x^{(i)}, \mu_k, \Sigma_k\right)}{\sum_{j=1}^{K} w_j \mathcal{N}\left(x^{(i)}, \mu_j, \Sigma_j\right)} + \lambda = 0 \quad \text{— (1)}$$

Multiply by w_k on b.s

$$\Rightarrow \lambda w_k + \sum_{i=1}^{N} \frac{w_k \mathcal{N}\left(x^{(i)}, \mu_k, \Sigma_k\right)}{\sum_{j=1}^{K} w_j \mathcal{N}\left(x^{(i)}, \mu_j, \Sigma_j\right)} = 0.$$

Summing over k on both sides

$$\lambda \sum_{k=1}^{K} w_k + \sum_{i=1}^{N} \frac{\sum_{k=1}^{K} w_k \mathcal{N}_k\left(x^{(i)}, \mu_k, \Sigma_k\right)}{\sum_{j=1}^{K} w_k \mathcal{N}\left(x^{(i)}, \mu_k, \Sigma_k\right)} = 0$$

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

$$\boxed{\lambda = -N} \quad \text{— (2)}$$

Substituting this back in (1) gives

$$\sum_{i=1}^{N} \frac{\mathcal{N}\left(x^{(i)}, \mu_k; \Sigma_k\right)}{\sum_{j=1}^{K} w_j \mathcal{N}\left(x^{(i)}, \mu_j, \Sigma_j\right)} - N = 0$$

$$\Rightarrow \sum_{i=1}^{N} \frac{\gamma\left(z_k^{(i)}\right)}{w_k} = N \quad \Rightarrow \quad w_k = \frac{\sum_{i=1}^{N} \gamma(z_k^{i})}{N}$$

$$\Rightarrow \boxed{w_k = \frac{N_k}{N}}$$