Q1) (a) $\ln(p(X, Z \mid \mu, \pi, \Sigma) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\{\ln \pi_k + \ln N(x_n \mid \mu_k, \Sigma_k)\}$

We also have $\sum_{k=1}^{K}\sum_{n=1}^{N} z_{nk} = N$, and $\sum_{k=1}^{K} \pi_k = 1$

→ Consider the function $F = \ln(p(X, Z \mid \mu, \pi, \Sigma) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)$. Maximing this gives:

$$\Rightarrow \frac{\partial F}{\partial \pi_k} = 0 \Rightarrow \frac{\partial}{\partial \pi_k} \sum_{n=1}^{N}\sum_{k=1}^{K} (z_{nk} \ln \pi_k + z_{nk} N(x_n \mid \mu_k, \Sigma_k)) + \lambda\left(\sum \pi_k - 1\right) = 0$$

$$\Rightarrow \sum_{n=1}^{N} \cancel{\sum_{k=1}^{K}} \frac{z_{nk}}{\pi_k} + 0 + \lambda = 0,$$ as $z_{nk} N(x_n \mid \mu_k, \Sigma_k)$ is independent of π_k, and remaining terms are also independent of π_k

$$\Rightarrow \lambda = -\sum_{n=1}^{N} \cancel{\sum_{k=1}^{K}} \frac{z_{nk}}{\pi_k}$$

Multiplying by π_k on both sides, and summing over k

Using $\sum_{k=1}^{K} \pi_k = 1$, we get

$$\lambda = -N$$

→ Substituting back, we get:

$$\cancel{\sum_{k=1}^{K}} \sum_{n=1}^{N} \frac{z_{nk}}{\pi_k} = N$$ or

$$\boxed{\pi_k = \frac{\sum_{n=1}^{N} z_{nk}}{N}}$$

Or, the mixing coefficients are given by the fraction of points assigned to the corresponding components.

Scanned by CamScanner

Now, differentiating w.r.t to μ_k,

$$\frac{\partial F}{\partial \mu_k} = \frac{\partial}{\partial \mu_k}\left(\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk} \ln \pi_k + z_{nk} \ln\left(\frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|} e^{-1/2(x_n-\mu_k)\ldots}\right)\right) = 0$$

Now, $\log(ab) = \log a + \log b$

$$\Rightarrow \frac{\partial F}{\partial \mu_k} = \frac{\partial}{\partial \mu_k}\left(\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk} \ln \pi_k + z_{nk} \ln c - \frac{z_{nk}}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right) = 0$$

$$\Rightarrow + \sum_{n=1}^{N} z_{nk} \Sigma_k^{-1}(x_n-\mu_k) = 0 \quad (\text{All } z_{nk'}, k' \neq k \text{ become } 0)$$

$$\Rightarrow \sum_{n=1}^{N} (z_{nk} x_n - z_{nk}\mu_k) = 0 \quad (\text{Multiplying by } \Sigma_k)$$

$$\text{or} \quad \boxed{\mu_k = \frac{\sum_{n=1}^{N} z_{nk} x_n}{\sum_{n=1}^{N} z_{nk}}}$$

Differentiating w.r.t to Σ_k

$$\frac{\partial F}{\partial \Sigma_k} = 0 = \frac{\partial}{\partial \Sigma_k}\sum^{N}\sum^{K}\left(z_{nk}\ln \pi_k + z_{nk}\left(-\frac{1}{2}\right)\ln|\Sigma| + \frac{1}{2}\ln(x_n-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$\Rightarrow \sum_{n=1}^{N} z_{nk}\left[-\Sigma_k^{-1} + \Sigma_k^{-1}(x-\mu)(x-\mu)^T\Sigma_k^{-1}\right] = 0$$

Multiplying Σ_k,

$$\sum_{n=1}^{N} z_{nk}\left[(x-\mu)(x-\mu)^T\Sigma_k^{-1} - I\right] = 0$$

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

i(b) For hard clustering, the cluster with the highest posterior probability should be assigned to the point, as it has the maximum responsibility of describing the point.

Date ___/___/
Page ___

(Q2) $E_Z[\ln p(X, Z|\mu, \Sigma, \pi)] =$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \ln \pi_k + \gamma z_{nk} \ln N(\ldots$$

$\cancel{N} = N(x_n | \mu_k, \Sigma_k) = \dfrac{\cancel{\xi}\, e^{-1/2 (x_n-\mu_k)^T \Sigma^{-1} (x_n - \mu_k)}}{\sqrt{(2\pi)^D |\Sigma|}}$

If $\Sigma_k = \epsilon I$, $|\Sigma_k| = |\epsilon I| = \epsilon |I| = \epsilon$

and $\Sigma_k^{-1} = \dfrac{I}{\epsilon}$ as $\Sigma_k \cdot \Sigma_k^{-1} = \dfrac{I}{\epsilon} \times \epsilon I = I^2 = I$

$\Rightarrow \lim_{\epsilon \to 0} E_Z[\ln p(X, Z | \mu, \pi, \Sigma)]$

$= \lim_{\epsilon \to 0} \cancel{\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma z_{nk} (\ln \pi_k + \ln \left(\frac{1}{\sqrt{(2\pi)^D \epsilon}}\right.}$

$= \lim_{\epsilon \to 0} \sum^{N}\sum^{K} \gamma(z_{nk}) \cdot [\ln \pi_k + \ln N(x_n | \mu_k, \Sigma_k)]$

Now, $\gamma(z_{nk}) = \dfrac{\pi_k \, e^{-\frac{\|x_n - \mu_k\|^2}{2\epsilon}}}{\sum_{j=1} \pi_j \, e^{-\frac{\|x_n - \mu_j\|^2}{2\epsilon}}}$

When $\epsilon \to 0$, the term having least L_2 norm goes to 0 at the least rate and hence $\gamma(z_{nk}) \to r_{nk}$ as $\epsilon \to 0$ (remaining go to 0.

And hence,

$\lim_{\epsilon \to 0} E_Z[\ln(p(X, Z | \mu, \Sigma, \pi))] \to -\frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n \ldots$

+ constant