ML

Page No. ______
Date

Ans 1:

(a).
$$p(x) = \sum_{k=1}^{K} \pi_k N(x \mid \mu_k, \Sigma_k)$$

let there be K-dimensional binary random variable z having 1 of K representat' as 1 and all other 0.
that is $z_k \in \{0, 1\}$

let $p(z_k \overset{=1}{}) = \pi_k$

\# $\sum_k \pi_k = 1$

$0 \leq \pi_k \leq 1$

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

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

$$\gamma(z_k) = p(z_k \overset{=1}{} / x) = \frac{p(z_k = 1)\ p(x / z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)\, p(x / z_j = 1)}$$

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

Now, log likelihood:

$$\ln p(X \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln\left[\sum_{k=1}^{K} \pi_k N(x_n \mid \mu_k, \Sigma_k)\right] \quad (1)$$

* Differentiating ① wrt μ_k;

$$\sum_{n=1}^{N} \frac{\pi_k \, N(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \, N(x_n \mid \mu_j, \Sigma_j)} \Sigma^{-1} (x_n - \mu_k)$$

$$= \sum_{n=1}^{N} \gamma_{nk} \Sigma^{-1} (x_n - \mu_k) = 0.$$

* $$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} x_n$$

$$N_k = \sum_{n=1}^{N} \gamma_{nk}$$

* Differentiating ① wrt Σ_k; $\frac{\partial}{\partial \Sigma_k} \left[\sum_{n=1}^{N} \ln \left[\sum_{k=1}^{K} \pi_k \, N(x_n \mid \mu_k, \Sigma_k) \right] \right] =$

* $$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} (x_n - \mu_k)(x_n - \mu_k)^T$$

$$N_k = \sum_{n=1}^{N} \gamma_{nk}$$

* For mixing coefficients :

We use Lagrange multiplier :

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

Differentiating wrt π_k;

$$\sum_{n=1}^{N} \frac{N(x_n \mid \mu_k \, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \, N(x_n \mid \mu_j, \Sigma_j)} + \lambda = 0 \qquad ②$$

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

$$\sum_{n=1}^{N} \cancel{\sum_{k}} \frac{\sum_{k=1}^{K} \pi_k \, N(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \, N(x_n \mid \mu_j, \Sigma_j)} + \lambda \sum_{k=1}^{K} \pi_k = 0$$

$$\sum_{n=1}^{N} (1) + \lambda = 0$$

$$\lambda = -N$$

\# $$y_{nk} = \frac{\pi_k \, N(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \, N(x_n \mid \mu_j, \Sigma_j)} = \frac{y_{nk}}{\pi_k}$$

$$= \frac{N(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \, N(x_n \mid \mu_j, \Sigma_j)}$$

$$N_k = \sum_{n=1}^{N} y_{nk}$$

So, $\displaystyle\sum_{n=1}^{N} \frac{y_{nk}}{\pi_k} - N = 0$ From 2

$$\frac{1}{\pi_k} N_k - N = 0$$

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

So, we observe that when we maximise log likelihood of a function, we get mean and variances which are fitted independently to corresponding group of data

Page No.
Date

~~Ans 7~~ points. Mixing coefficient is given by fraction of points in each group.

(b) Assume each Gaussian is a GMM
~~Let~~ Let covariance matrices of mixture components is given by εI, where ε is variance parameter ~~shared~~ shared by all components, and I is identity matrix;

$$\text{so } P(x \mid \mu_k, \Sigma_k) = \frac{1}{\sqrt{2\pi\varepsilon}} e^{-\left[\frac{\|x-\mu_k\|^2}{2\varepsilon}\right]}$$

$$E_Z[\ln p(X, Z \mid \mu, \Sigma, \pi)] \to -\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|x_n - \mu_k\|^2 + \text{const}$$

$$\gamma(z_{nk}) = \frac{\pi_k e^{-\left[\frac{\|x_n-\mu_k\|^2}{2\varepsilon}\right]}}{\sum_j \pi_j e^{-\left[\frac{\|x_n-\mu_j\|^2}{2\varepsilon}\right]}}$$

$\varepsilon \to 0$, $\|x_n - \mu_j\|^2$ is smallest and will go to 0
so $\gamma(z_{nk})$ will be 0 for all x_n except x term j.

Therefore, it is hard clustering as each point goes only in one cluster.
A point cannot belong to multiple clusters.

Page No. ______
Date

Ans 2:- $p(X, Z \mid \mu, \Sigma, \pi) = \prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k)^{z_{nk}} \left[N(x_n \mid \mu_k, \Sigma_k)\right]^{z_{nk}}$

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

Using Lagrange Multipliers, we get;

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

$$P(Z \mid X, \mu, \Sigma, \pi) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} \left[\pi_k N(x_n \mid \mu_k, \Sigma_k)\right]^{z_{nk}}$$

- $E[z_{nk}] = \dfrac{\sum_{z_{nk}} z_{nk} \left[\pi_k N(x_n \mid \mu_k, \Sigma_k)\right]^{z_{nk}}}{\sum_{z_{nj}} \left[\pi_j N(x_n \mid \mu_j, \Sigma_j)\right]^{z_{nj}}}$

$$= \frac{\pi_k N(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n \mid \mu_j, \Sigma_j)} = \gamma(z_{nk})$$

- $E[\ln p(X, Z \mid \mu, \Sigma, \pi) = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left[\ln \pi_k + \ln N(x_n \mid \mu_k, \Sigma_k)\right]$

Gaussian mixture model in which variance matrix of mixture is given by ϵI; $\epsilon \rightarrow$ variance parameter shared by all components.

Date

- $p(x \mid \mu_k, \Sigma_k) = \dfrac{1}{\sqrt{2\pi\varepsilon}} e^{-\frac{(x-\mu)^2}{2\varepsilon}}$

- $\gamma(z_{nk}) = \dfrac{\pi_k \, e^{\left[-\frac{\|x_n - \mu_k\|^2}{2\varepsilon}\right]}}{\sum_j \pi_j \, e^{\left[-\frac{\|x_n - \mu_j\|^2}{2\varepsilon}\right]}}$

- $\cancel{E_z[\ln p(X, Z \mid \mu, \Sigma, \pi) \to -\frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2 + \text{const}}$

If we consider the limit $\varepsilon \to 0$, we see that denominator term for $\|x_n - \mu_j\|^2$ is smallest and will go to 0 ~~at~~ most slowly

$\gamma(z_{nk})$ becomes 0 for all terms except j.

so, ~~r_{nk}~~ $\gamma(z_{nk}) \to r_{nk}$; r_{nk} is true if n^{th} point belongs to k^{th} cluster.

$$E_z[\ln p(X, Z \mid \mu, \Sigma, \pi)] \to -\frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2 + \text{const.} \quad (1)$$

So maximising data log ~~likel~~ likelihood for this model is equivalent to minimising distortion measure for K-means algorithm given by

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2$$

as in expression it is $-ve$, so it has to be minimised to make it maximum.