Problem 2 :-

We know; $L = \sum_{j=1}^{k} \sum_{x_i \in S_j} ||x_i - \mu_j||^2$

$x_1, x_2, x_3, \ldots, x_n$ are sample points
$\mu_1, \mu_2, \mu_3, \ldots, \mu_k$ are centers

Lets Minimize L with the help of batch gradient descent with S_j fixed.
So,

$$\frac{dL}{d\mu_1} = \frac{d}{d\mu_1} \sum_{x_i \in S_1} (x_i - \mu_1)^T (x_i - \mu_1)$$

$$= \sum_{x_i \in S_1} 2(\mu_1 - x_i)$$

∴ We have;

$$\mu_1 \leftarrow \mu_1 + \epsilon \sum_{x_i \in S_1} (x_i - \mu_i)$$

lets use stochastic gradient descent to derive the updated formula for μ_1

$\mu_1 \leftarrow \mu_1 + \epsilon(x_i - \mu_i)$ if $x_i \in S_i$ otherwise there will be no change.
ϵ is learning rate.

$$\mu_1 \leftarrow \sum_{x_i \in S_1} \frac{1}{|S_1|} x_i$$

Comparing the answer with the one in the first;

$$\sum_{x_i \in S_1} \frac{1}{|S_1|} x_i = \mu_1 + \epsilon \sum_{x_i \in S_1} (x_i - \mu_1)$$

$$\sum_{x_i \in S_1} \frac{1}{|S_1|} x_i - \sum_{x_i \in S_1} \frac{1}{|S_1|} \mu_1 = \epsilon \sum_{x_i \in S_1} (x_i - \mu_1)$$

$$\sum_{x \in S_1} \frac{1}{|S_1|} (x_1 - \mu_1) = \epsilon \sum_{x_i \in S_1} (x_i - \mu_1)$$

$$\therefore \epsilon = \frac{1}{|S_1|}$$

Problem 3:-

Latent Variable is z

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

where $\{\pi_k\}$ satisfies: $0 \leq \pi_k \leq 1$ & $\sum_{k=1}^{k} \pi_k = 1$

Conditional distribution of x for given value of z is

$$p(x|z_k=1) = N(x|\mu_k, \Sigma_k)$$

$p(z)$ is the prior

$$\therefore p(z_k=1) = \pi_k$$

Eg:- $P(z=(0,0,1)) = P(z_3=1) = \pi_3$

$$\therefore \pi_1^{z_1=0} \cdot \pi_2^{z_2=0} \cdot \pi_3^{z_3=1} = \pi_3$$

$$\therefore P(z) = \prod_{k=1}^{k} \pi_k^{z_k} \quad \text{—①}$$

Similarly,

$$p(x|z) = (0,0,0,\cdots,1,0,0) = N(x|\mu_k, \Sigma_k)$$

$$\therefore p(x|z) = \prod_{k=1}^{k} N(x|\mu_k, \Sigma_k)^{z_k} \quad \text{—②}$$

Let model be gaussian. Hence it must contain some marginal probability for x observation.

$\therefore$ Let $P(x)$ be marginal probability

From the marginal distribution property,

$$P_x(x_1) = \sum_j P(x_i, y_j)$$

∴ Normal distribution is;

$$P(x) = \sum_{k=1}^{k} P(x, z=k)$$

$$= \sum_{k=1}^{k} P(z=k)\, P(x \mid z=k)$$

From ① & ②, we get

$$P(x) = \sum_{k=1}^{k} \pi_k \, \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

We use GMM algorithm

- <u>K-Means</u>:- hard assignments. Each point is associated uniquely with one cluster.

- <u>GMM</u>:- soft assignments. Based on posterior probabilities.

Problem 4 :-

Bayesian Network

$$\begin{tikzcd} A_1 \arrow[r] \arrow[d] & B_2 & \\ B_1 \arrow[r] & A_2 \arrow[u] \arrow[r] & B_3 \end{tikzcd}$$

Joint Distribution of $P(A_1, A_2, B_1, B_2, B_3)$

$$= P(A_1) \times P(A_2 | B_1) \times P(B_1 | A_1) \times P(B_2 | A_1, A_2) \times P(B_3 | A_2)$$

$\therefore$ No. of independent parameters needed to fully specify the joint distribution is;

$$2^0 + 2^1 + 2^1 + 2^2 + 2^1 = 11$$

Factorizations of $P(A_1, A_2, B_1, B_2, B_3)$

$$\therefore P(A_1 | A_2, B_1, B_2, B_3) . P(A_2, B_1, B_2, B_3)$$

$$= P(A_1 | A_2, B_1, B_2, B_3) . P(A_2 | B_1, B_2, B_3) . P(B_1, B_2, B_3)$$

$$= P(A_1 | A_2, B_1, B_2, B_3) . P(A_2 | B_1, B_2, B_3) . P(B_1 | B_2, B_3) . P(B_2, B_3)$$

$$= P(A_1 | A_2, B_1, B_2, B_3) . P(A_2 | B_1, B_2, B_3) . P(B_1 | B_2, B_3) . P(B_2 | B_3) . P(B_3)$$

$$\therefore 2^4 + 2^3 + 2^2 + 2^1 + 2^0 = \underline{\underline{31}}$$

31 Independent Parameters.