Assignment 3

Question 1

$$d(x,y) = |x-y|^2$$

(i) $d(x,y) \in [0,\infty)$ & $d(x,x) = 0$

$\Rightarrow$ satisfies positive definite

(ii) $d(x,y) = d(y,x)$ $\because$ $|x-y|^2 = |y-x|^2$

$\Rightarrow$ satisfies symmetry

(iii) To check : $d(x,z) \leq d(x,y) + d(y,z)$

$$|x-z|^2 \leq |x-y|^2 + |y-z|^2$$

Without loss of generality, say, $x \geq y \geq z$

$$(x-z)^2 \leq (x-y)^2 + (y-z)^2$$

$$\cancel{x^2} + \cancel{z^2} - 2xz \leq \cancel{x^2} + y^2 - 2xy + y^2 + \cancel{z^2} - 2yz$$

$$2y(y-x) - 2z(y-x) \geq 0$$

$$2(y-z)(y-x) \geq 0$$

$$\underbrace{(x-y)}_{\geq 0}\ \underbrace{(z-y)}_{\leq 0} \geq 0$$

$\because$ LHS ≤ 0 & RHS ≥ 0 which contradicts

$\therefore$ $d(x,y) = |x-y|^2$ fails to qualify Triangle inequality

$\Rightarrow$ NOT a valid distance metric

Question 2

$x = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$

x

$x_i \in C$	$d(x, x_i)$
x_1	5.1478
x_4	4.9244
x_7	4.1231
x_2	5
x_5	3.6056
x_8	2.6926
x_3	4.1608
x_6	5.0249

$$D_{min}(x,C) = \min_{V \in C} \{\delta(x,V)\} = 2.6926$$

$$D_{max}(x,C) = \max_{V \in C} \{\delta(x,V)\} = 5.1478$$

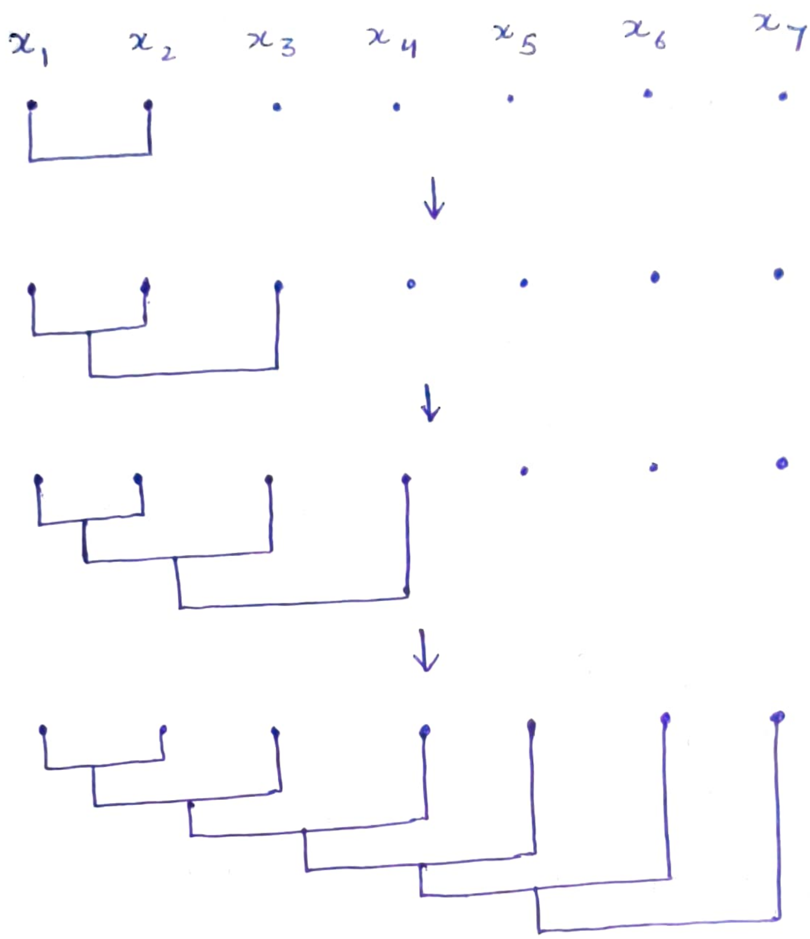
$$D_{avg}(x,C) = \langle \delta(x,V) \rangle = 4.3349$$

Question 3

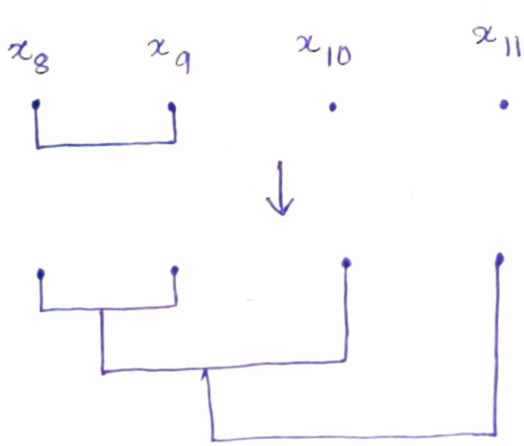
Single Linkage Algorithm

$$D_{min} = \min_{\substack{u \in C_1 \\ v \in C_2}} \{ \delta(u,v) \}$$

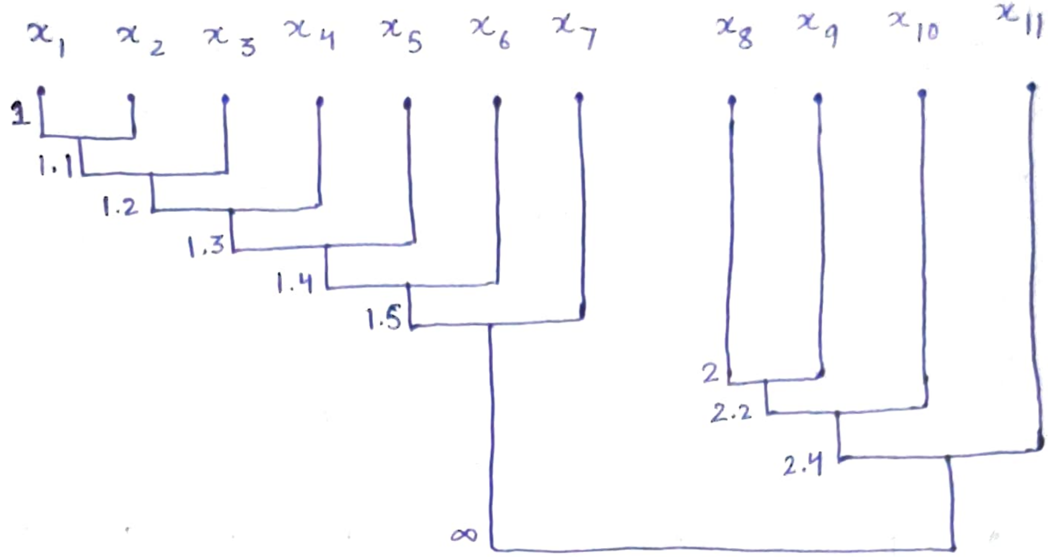
→ Drawing dendogram for elongated clusters x_1 to x_7



→ Drawing dendogram for $x_8 \rightarrow x_{11}$

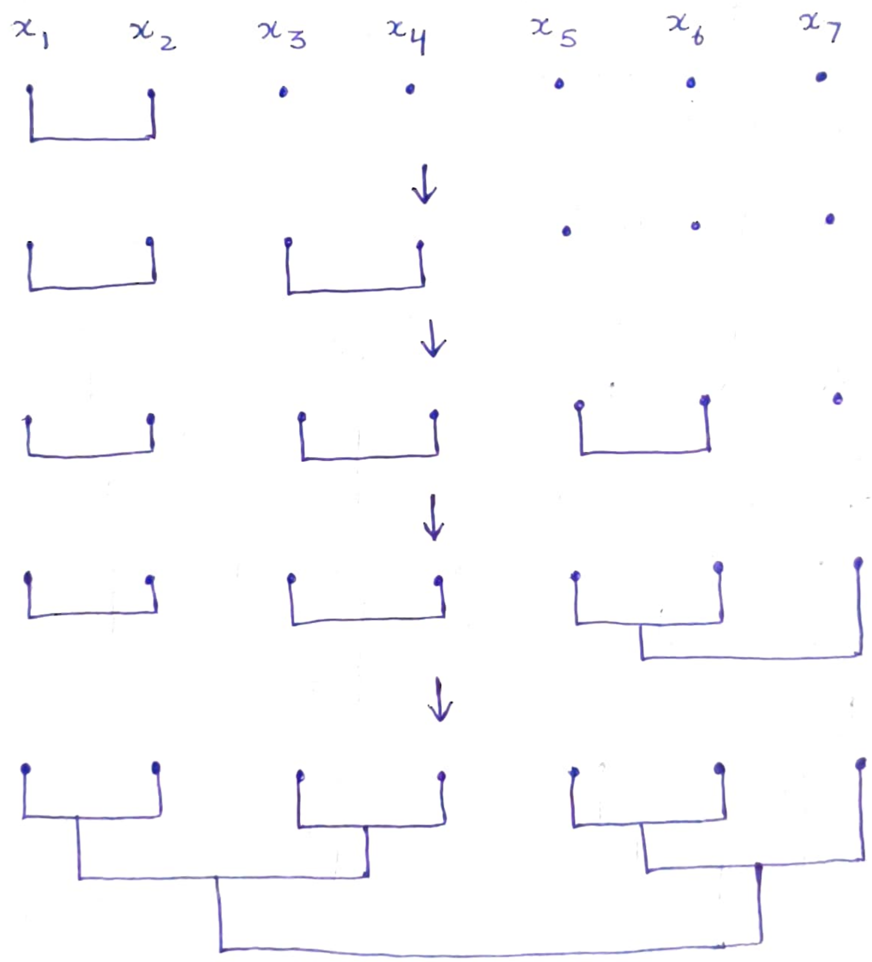


→ Merging both

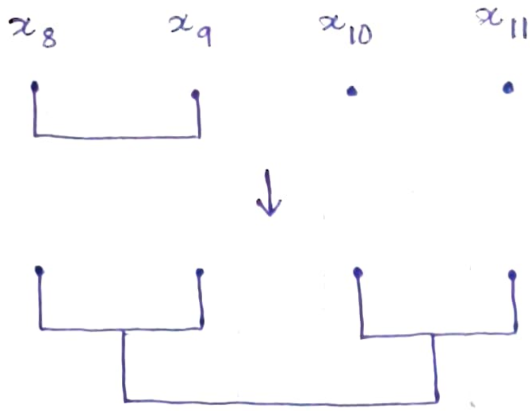


Complete Linkage Algorithm

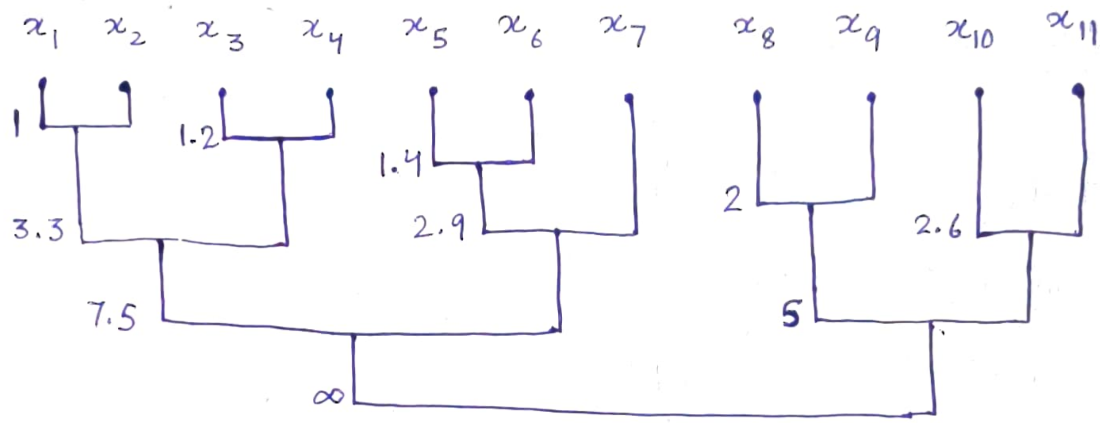
→ Drawing dendogram for clusters $x_1 \to x_7$



→ Drawing dendogram for clusters $x_8 \to x_{11}$

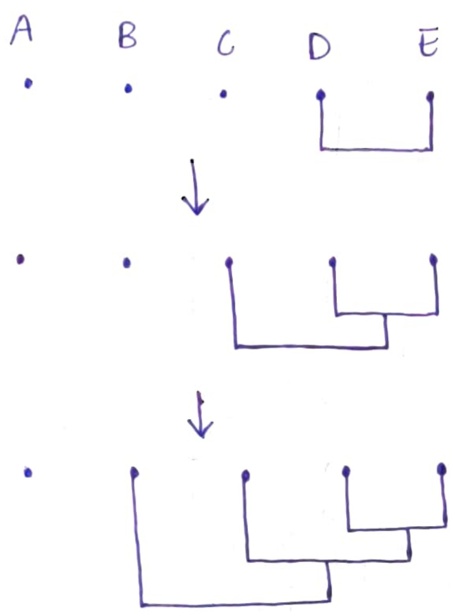


→ Merging both



Question 4

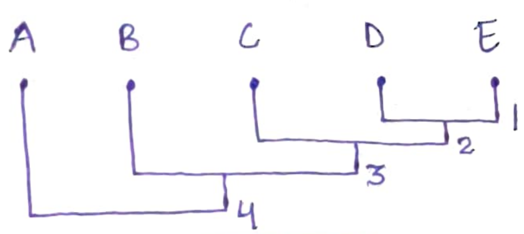
Single Linkage Algorithm



$$\Rightarrow \begin{bmatrix} 0 & 4 & 9 & 5 \\ 4 & 0 & 3 & 7 \\ 9 & 3 & 0 & 2 \\ 5 & 7 & 2 & 0 \end{bmatrix}$$

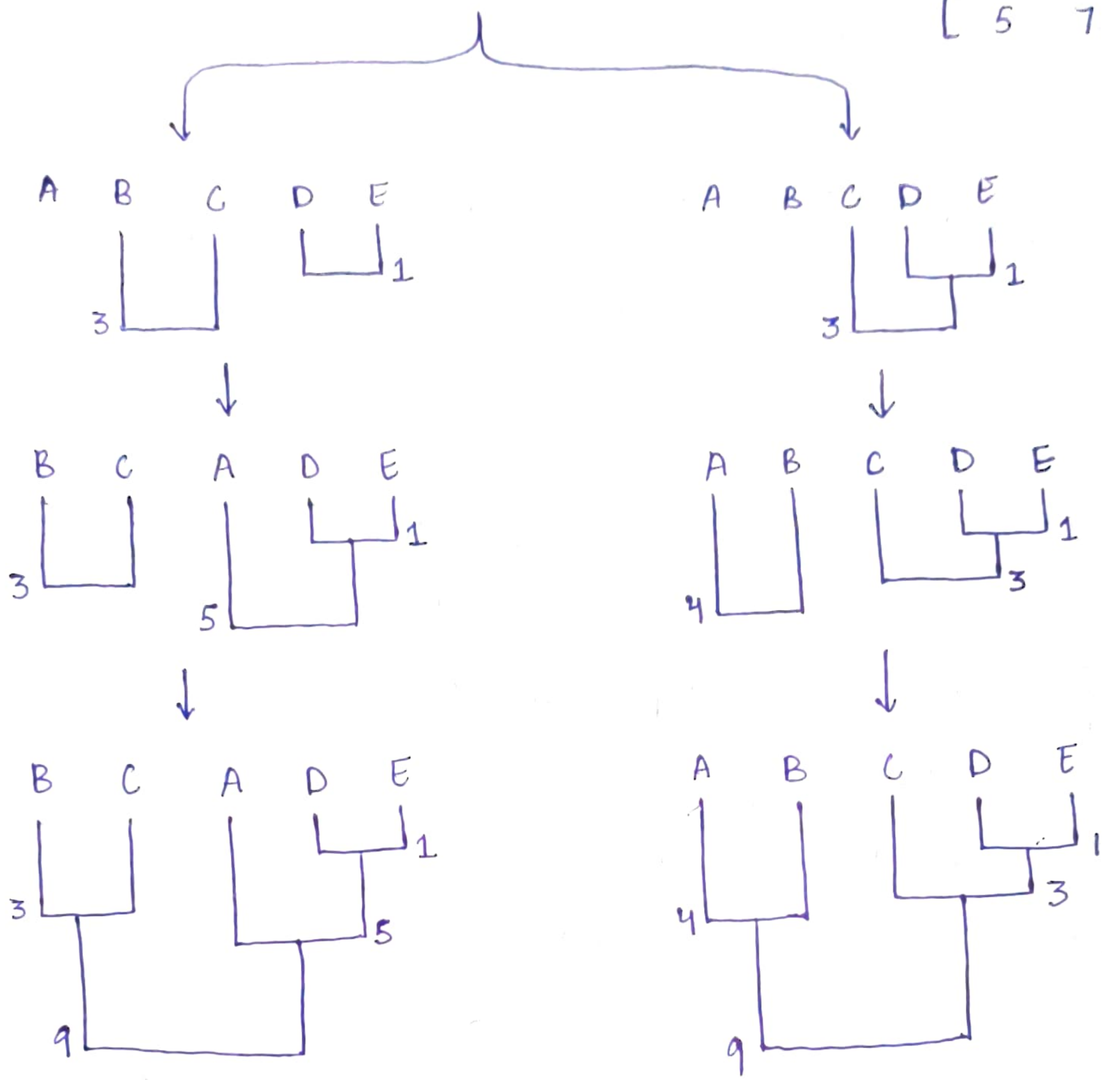
$$\Rightarrow \begin{bmatrix} 0 & 4 & 5 \\ 4 & 0 & 3 \\ 9 & 3 & 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix}$$



Complete Linkage Algorithm

$$\Rightarrow \begin{bmatrix} 0 & 4 & 9 & 5 \\ 4 & 0 & 3 & 7 \\ 9 & 3 & 0 & 3 \\ 5 & 7 & 3 & 0 \end{bmatrix}$$



Question 5

$N = 5$

5 points mapped to $K = 3$ clusters, namely,

$C_1 = (2,3)$, $C_2 = (5,8)$, $C_3 = (9,4)$

$d(x,y)$ is (Euclidean distance b/w x, y points)^2

= distortion function

Point p	$d(p, C_1)$	$d(p, C_2)$	$d(p, C_3)$	Nearest centroid
(1, 2)	2	52	68	C_1
(3, 4)	2	20	36	C_1
(6, 7)	32	2	18	C_2
(8, 3)	36	34	2	C_3
(5, 5)	13	9	17	C_2

(a) $C_{1,\text{new}} = \left(\frac{1+3}{2}, \frac{2+4}{2}\right) = (2, 3)$

$C_{2,\text{new}} = \left(\frac{6+5}{2}, \frac{7+5}{2}\right) = (5.5, 6)$

$C_{3,\text{new}} = (8, 3)$

(b) Distortion old $= 2 + 2 + 2 + 2 + 9 = 17$

Point p	$d(p, C_{1\text{new}})$	$d(p, C_{2\text{new}})$	$d(p, C_{3\text{new}})$	Nearest centroid
(1, 2)	2	36.25	50	C_1
(3, 4)	2	10.25	26	C_1
(6, 7)	32	1.25	20	C_2
(8, 3)	36	15.25	0	C_3
(5, 5)	13	1.25	13	C_2

Distortion new $= 2 + 2 + 1.25 + 0 + 1.25 = 6.5$

Distortion decreases after this iteration

Question 6

$$E_Z\left[\ln p(X, Z \mid \mu, \Sigma, \pi)\right] = \sum_{n=1}^{N} \sum_{k=1}^{K} \underbrace{\gamma(z_{nk})}_{\text{responsibility (fixed, constant)}} \left(\ln \pi_k + \ln N(x_n \mid \mu_k, \Sigma_k)\right)$$

$$N(x_n | \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)}$$

$$\ln N(x_n | \mu_k, \Sigma_k) = -\frac{D}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_k| - \frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)$$

$$\mathbb{E}_Z[\ln p(x, z | \mu, \Sigma, \pi)] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \Big[\ln \pi_k - \frac{1}{2} \ln |\Sigma_k| - \frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\Big]$$

$+ \; C \rightarrow$ independent of Σ_k, π_k

(i) Diff. w.r.t Σ_k & equating to 0

$$-\frac{1}{2} \sum_{n=1}^{N} \gamma(z_{nk}) \left[\Sigma_k^{-1} - \Sigma_k^{-1} (x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1}\right] = 0$$

Multiplying Σ_k on both sides

$$\sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T = N_k \Sigma_k \qquad \left[N_k = \sum_{n=1}^{N} \gamma(z_{nk})\right]$$

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

(ii) Term with π_k is $\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \ln \pi_k$

Adding the constraint $\left(\sum_{k=1}^{K} \pi_k = 1\right)$

Using Lagrange Multiplier method,

$$L = \sum_{k=1}^{K} \sum_{n=1}^{N} \gamma(z_{nk}) \ln \pi_k + \lambda \left(\sum_{k=1}^{K} \pi_k - 1\right)$$

Diff. w.r.t π_k & equating to 0

$$\sum_{n=1}^{N} \frac{\gamma(z_{nk})}{\pi_k} + \lambda = 0 \Rightarrow \pi_k = -\frac{N_k}{\lambda} \quad \text{—(1)}$$

Summing over all K

$$\sum_{k=1}^{K} \pi_k = -\frac{1}{\lambda} \sum_{k=1}^{K} \gamma(z_{nk}) = -\frac{1}{\lambda} \cdot N = 1 \Rightarrow \lambda = -N$$

Putting in (1) $\Rightarrow \pi_k = -\frac{N_k}{-N} \Rightarrow \pi_k = \frac{N_k}{N}$

Question 7

Probability density function of the mixture distribution is given by, $p(x) = \sum_{k=1}^{K} \pi_k \, p(x|k)$

↳ mixing coefficient

$$p(x_b | x_a) = \frac{p(x_b, x_a)}{p(x_a)}$$

$$p(x_b|x_a) = \frac{\sum_{k=1}^{K} \pi_k \, p(x_b, x_a | k)}{\sum_{k=1}^{K} \pi_k \, p(x_a|k)}$$

$$= \sum_{k=1}^{K} \underbrace{\frac{\pi_k \, p(x_a|k)}{p(x_a)}}_{\bar{\pi}_k \rightarrow \text{updated mixing coefficient}} \; p(x_b | x_a, k)$$

$$= \sum_{k=1}^{K} \bar{\pi}_k \, p(x_b | x_a, k)$$

Question 8

(a) Likelihood func^n, $p(x|\theta) = \prod_{n=1}^{N} p(x_n|\theta)$

$$= \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \, N(x_n | \mu_k, \Sigma_k)$$

Log likelihood func^n, $L(\theta) = \ln p(x|\theta)$

$$= \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k \, N(x_n | \mu_k, \Sigma_k)$$

This is complete as it sums over all components. If we assume that the component responsible for generating each data point is known, we introduce latent variable z_n, where z_n is a one-hot encoded indicator for component k

$$z_{nk} = \begin{cases} 1 & , \text{if } x_n \text{ was generated by component } k \\ 0 & , \text{otherwise} \end{cases}$$

Complete log likelihood func^n,

$$L(\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \ln\left(\pi_k \, N(x_n | \mu_k, \Sigma_k)\right)$$

(b) (i) Update Rule for π_k

Need to maximize $\sum_{n=1}^{N} z_{nk} \ln \pi_k$ subject to the constraint $\sum_{k=1}^{K} \pi_k = 1$ to get MLE estimate for π_k

Using the lagrange multiplier method, we get

$$\pi_k = \frac{N_k}{N} \qquad \text{where } N_k = \sum_{n=1}^{N} z_{nk}$$

(ii) Update rules for μ_k

Maximize the complete log likelihood w.r.t μ_k, we differentiate $\sum_{n=1}^{N} z_{nk} \ln N(x_n | \mu_k, \Sigma_k)$

Taking derivative w.r.t μ_k & equating to 0 gives

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

(iii) Update rule for Σ_k

Maximizing w.r.t Σ_k,

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