1. Majority vote approach:

P(Class is Red | X) =

$$\underbrace{0.1 \quad 0.15 \quad 0.2 \quad 0.2}_{4} \quad \underbrace{0.55 \quad 0.6 \quad 0.6 \quad 0.65 \quad 0.7 \quad 0.75}_{6}$$

$$4 < 6$$

Red #

Average probability

$$\frac{\sum P(\text{Class is Red} \mid X)}{10} = 0.45$$

Green #

2. In order to show that the regularized LDA can be calculated based on kernels, it suffices to prove that LD function

$$\delta_k(x) = \phi(x)^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k, \quad k=1,2, \quad \cdots ①$$

can be computed only through kernels.

$$\begin{aligned} \hat{\Sigma} &= W_\phi + \gamma \mathbb{I} \\ &= \frac{1}{N-2} \sum_{k=1}^{2} \sum_{y_i=k} (\phi(x_i) - \hat{\mu}_k)(\phi(x_i) - \hat{\mu}_k)^T + \gamma \mathbb{I} \\ &= \frac{1}{N-2} (\Phi - YD^{-1}Y^T\Phi)^T (\Phi - YD^{-1}Y^T\Phi) + \gamma \mathbb{I} \\ &= \gamma \mathbb{I} + \Phi^T R \Phi, \end{aligned}$$

where $D = \mathrm{diag}(N_1, N_2)$ and $R = (I - YD^{-1}Y^T)^T(I - YD^{-1}Y^T)$.

$$\hat{\mu}_k = \frac{1}{N_k} \sum_{y_i=k} \phi(x_i) = \frac{\Phi^T y_k}{N_k}, \quad \cdots ②$$

where y_k is the kth column vector of Y.

Subtitute ② into ①,

$$\begin{aligned} \delta_k(x) &= \phi(x)^T \Sigma^{-1} \frac{\Phi^T y_k}{N_k} - \frac{1}{2} \left(\frac{\Phi^T y_k}{N_k}\right)^T \Sigma^{-1} \frac{\Phi^T y_k}{N_k} + \log \pi_k \\ &= \frac{1}{N_k} \phi(x)^T \Sigma^{-1} \Phi^T y_k - \frac{1}{2N_k^2} y_k^T \Phi^T \Sigma^{-1} \Phi^T y_k + \log \pi_k. \quad \cdots ③ \end{aligned}$$

Claim that $\Sigma^{-1}\Phi^T = \Phi^T(R\underbrace{\Phi\Phi^T}_{K} + \gamma \mathbb{I})^{-1} = \Phi^T(RK + \gamma \mathbb{I})^{-1}$,

and then subtitute it into ③, we get

$$\delta_k(x) = \frac{1}{N_k} k^T (RK + \gamma \mathbb{I})^{-1} y_k - \frac{1}{2N_k^2} y_k^T K (RK + \gamma \mathbb{I})^{-1} y_k + \log \pi_k,$$

which depends on kernels.

3. (a)

$$K(x, x') = \langle (x, x^2), (x', x'^2) \rangle$$

$$= xx' + x^2x'^2 = xx'(1 + xx')$$ #

(b)

∵ support vectors $(-2, 4), (-1, 1)$

$$w_1 Y_1 + w_2 Y_2 + c = 0$$

$$\Rightarrow (-2-(-1))Y_1 + (4-1)Y_2 + c = 0$$

$$\Rightarrow -Y_1 + 3Y_2 + c = 0$$

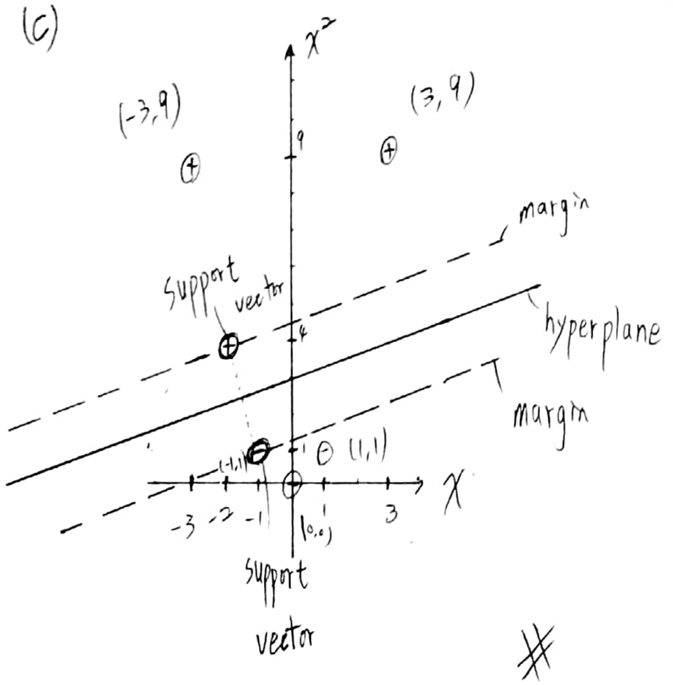
Substitute $\frac{(-2,4)+(-1,1)}{2}$ into it

$$\Rightarrow -Y_1 + 3Y_2 - 9 = 0$$

$$\text{margin} = \sqrt{(-2-(-1))^2 + (4-1)^2} / 2 = \frac{\sqrt{10}}{2}$$

$w_1 = -1$, $w_2 = 3$, $c = -9$, margin $= \frac{\sqrt{10}}{2}$ #

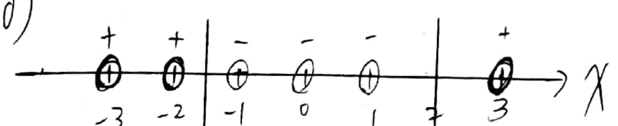
(c)



#

(e) This question is quite difficult to solve manually (needs computer program to solve). So, full points (8%) will award to everyone.

(d)



#

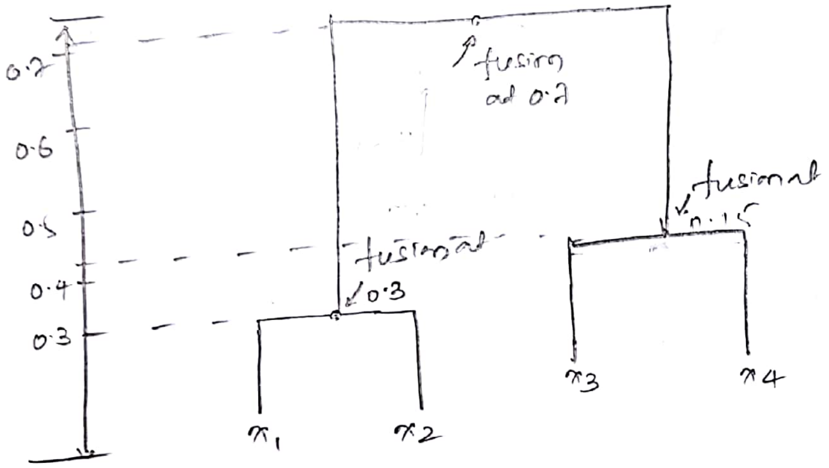
Scanned by CamScanner

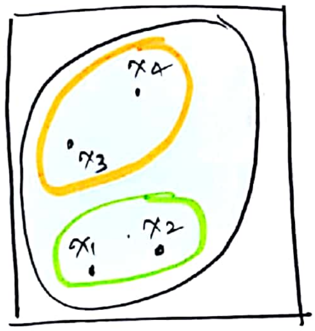
Q.4 Let x_1, x_2, x_3, x_4 are observation with dissimilarity matrix as

$$\begin{bmatrix} & \overset{(x_1,x_2)}{0.3} & \overset{(x_1,x_3)}{0.4} & \overset{(x_1,x_4)}{0.7} \\ 0.3 & & \overset{(x_2,x_3)}{0.5} & \overset{(x_2,x_4)}{0.6} \\ 0.4 & 0.5 & & \overset{(x_3,x_4)}{0.45} \\ 0.7 & 0.6 & 0.45 & \end{bmatrix}$$

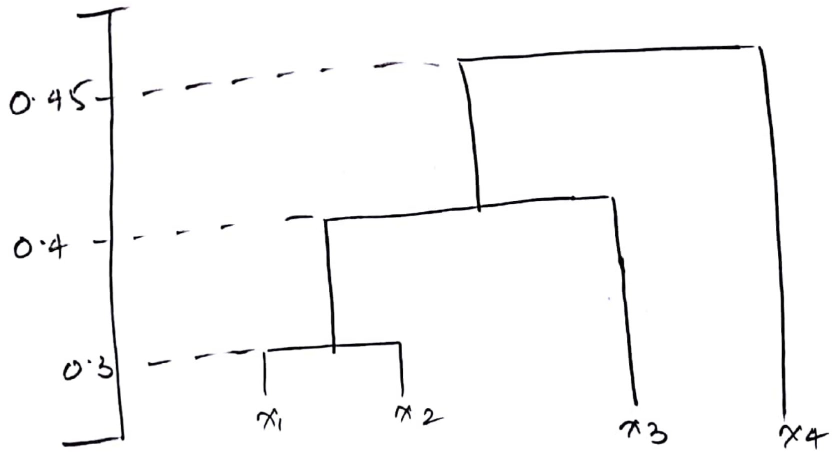
from above we can show observations in $\mathbb{R}^2$ as

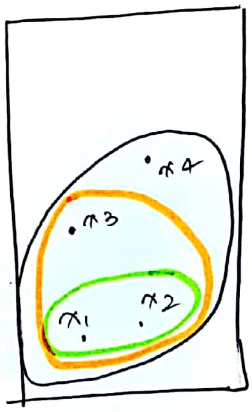
a) Complete linkage based clustering





b) Single linkage based clustering



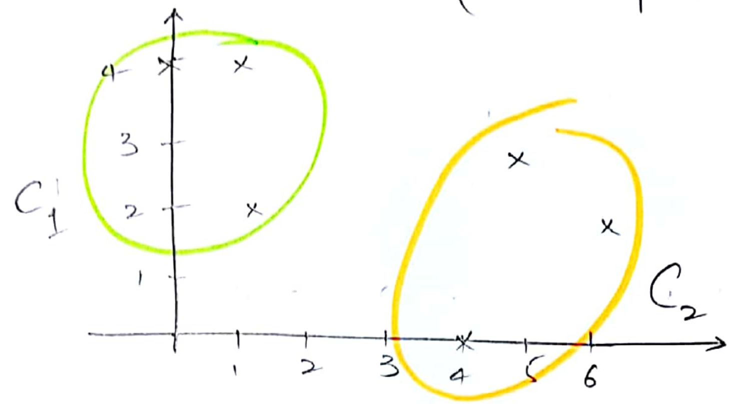


c) $x_1\, x_2 \rightarrow$ cluster 1 ; $x_3\, x_4 \rightarrow$ cluster 2

d) $x_1, x_2\, x_3 \rightarrow$ cluster 1 ; $x_4 \rightarrow$ cluster 2

Q.5 Given data (Initialize) (Let 6 points are $p_1, p_2, p_3, p_4, p_5, p_6$)

Step 1:

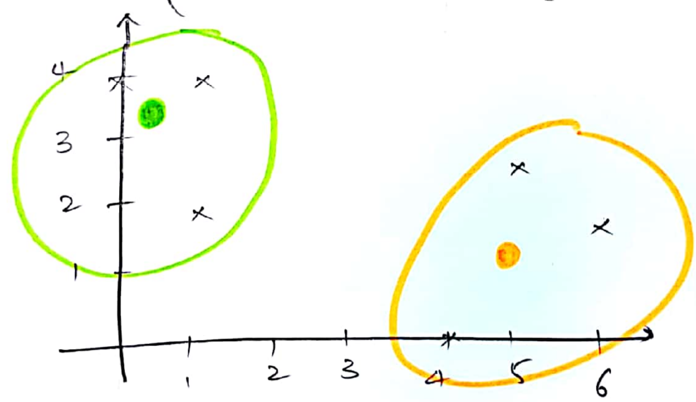


Iteration 1 (step 2a)

Centroid of "C_1" $\Rightarrow \left(\frac{1+1+0}{3}, \frac{4+2+4}{3}\right) = \left(2/3, 10/3\right)$

$= (0.67, 3.33)$

Centroid of "C_2" $\Rightarrow \left(\frac{5+6+4}{3}, \frac{3+2}{3}\right) = (5, 1.67)$



Iteration 1, step 2b)

Observations	distance to centroid C_1	C_2
(1, 4)	0.75	4.62
(1, 2)	1.3703	4.013
(0, 4)	0.94	5.5162
(5, 3)	4.3426	1.33
(6, 2)	5.49	1.053
(4, 0)	4.7093	1.9468

from above table, we can see the euclidean distance between cluster data to its centroid doesn't change

$\therefore$ the previous clustering remains same $\times$

Q.6 The likelihood function is given by

$$L = \prod_{i=1}^{n} \underbrace{\sum_{k=1}^{K} \pi_k N(x_i; \mu_k, \Sigma)}_{p(x)},$$

loglikelihood function is

$$\ln L = \sum_{i=1}^{n} \ln \sum_{k=1}^{K} \pi_k N(x_i; \mu_k, \Sigma)$$

$\therefore$ Q-function is given by

$$Q = E_{z|x, \theta^{old}}[\ln p(x, z|\theta)]$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \ln[p(z_k|\theta)\, p(x_i|z_k; \theta)\, p(z_k|x_i; \theta^{old})]$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \ln[\pi_k N(x_i; \mu_k, \Sigma)]\, \gamma_k^{(t)}(x_i)$$

(where $\gamma_k^{(t)}(x_i) = p(z_k|x_i; \theta^{old}) = \dfrac{p(x_i|z_k; \theta^{old})\, p(z_k|\theta^{old})}{\sum_{k=1}^{K} p(x_i|z_k; \theta^{old})\, p(z_k|\theta^{old})}$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \left[\ln \pi_k - \frac{1}{2}\ln|2\pi\Sigma| - \frac{1}{2}(\mu_k - x_i)^T \Sigma^{-1} (\mu_k - x_i)\right] \gamma_{ik}$$

$$\Rightarrow \sum_{i=1}^{n} \sum_{k=1}^{K} \left[\ln \pi_k - \frac{1}{2}(\mu_k - x_i)^T \Sigma^{-1} (\mu_k - x_i)\right] - \frac{KN}{2}\ln|2\pi\Sigma| \quad (*)$$

$$\frac{dQ}{d\mu_k} = 0 \Rightarrow \sum_{i=1}^{n} \Sigma^{-1}(\mu_k - x_i)\, \gamma_k^{(t)}(x_i) = 0$$

and

$$\frac{dQ}{d\Sigma} = 0 \Rightarrow \sum_{i=1}^{n} \sum_{k=1}^{K} \left[-\Sigma + \Sigma^{-1}(\mu_k - x_k)(\mu_k - x_i)^T \Sigma^{-1} \right] \gamma_k^{(t)}(x_i) = 0$$

$$\therefore \hat{\mu}_k = \frac{1}{n_k^{(t)}} \sum_{i=1}^{n} \gamma_k^{(t)}(x_i)\, x_i \; ; \quad \boxed{n_k^{(t)} \triangleq \sum_{i=1}^{n} \gamma^{(t)}(x_i)}$$

(i.e average # of points in cluster k)

$$\hat{\Sigma} = \frac{1}{\sum_{k=1}^{K} n_k^{(t)}} \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_k^{(t)}(x_i)\,(x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T$$