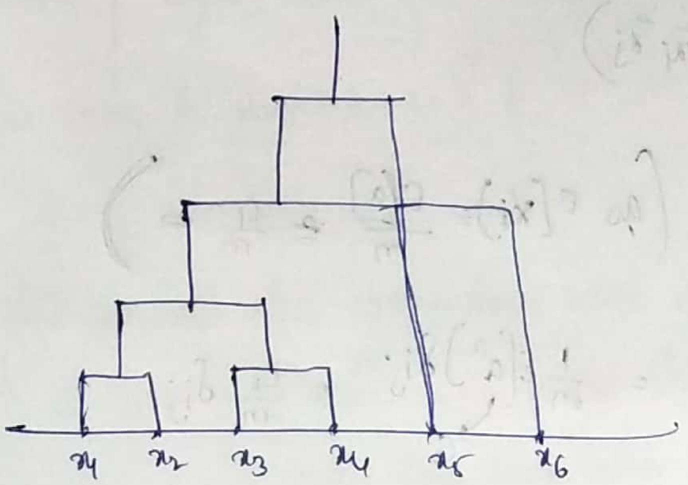
Assignment - 4

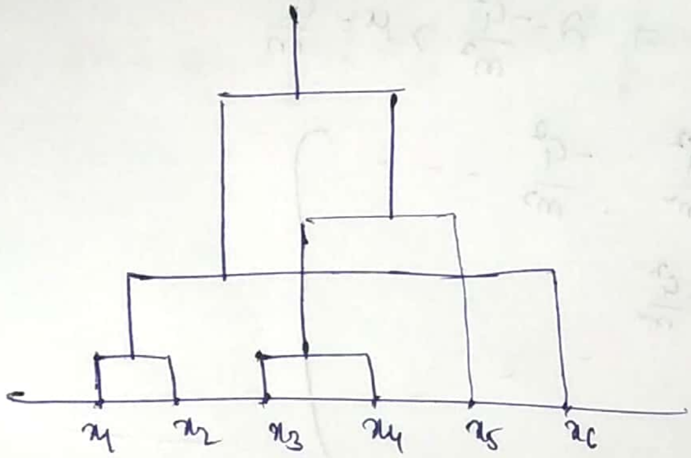
G. Deepika
CS18B16EHC0015

(1)

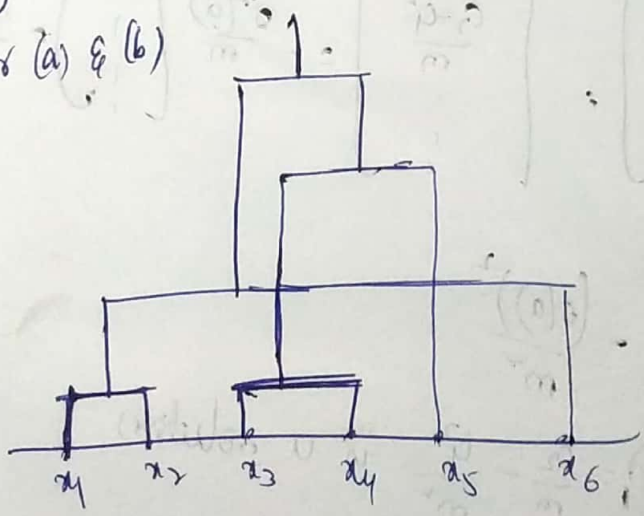
(a)



(b)



(c) On changing (x_6, x_1) to 0.15 and (x_3, x_5) to 0.15, we get same dendograms for (a) & (b)



(2)

(a) $C_{ij} = E[(x_i - \bar{x}_i)(x_j - \bar{x}_j)]$

$= E[x_i x_j - x_i \bar{x}_j - \bar{x}_i x_j + \bar{x}_i \bar{x}_j]$

$\Rightarrow C_{ij} = E[x_i x_j] - \left(\frac{C_1}{m}\right)^2 \quad \left(\text{as } E[x_i] = \frac{E[a]}{m} = \frac{C_1}{m}\right)$

$E[x_i x_j] = 0$ if $i \neq j$ $\quad = \frac{1}{m} E[a^2]\delta_{ij} = \frac{C_2}{m}\delta_{ij}$

$\frac{E[a^2]}{m}$ if $i = j$ $\qquad E[a^2] \to C_2$

$C_{ij} = \frac{C_2 \delta_{ij}}{m} - \frac{C_1^2}{m^2} \quad \Rightarrow \lambda = -\frac{C_1^2}{m^2}, \ \mu = \frac{C_2}{m}$

$$\Rightarrow C = \begin{bmatrix} \frac{C_2}{m} - \frac{C_1^2}{m^2} & -\frac{C_1^2}{m^2} & -\frac{C_1^2}{m^2} & \cdots \\ & \frac{C_2}{m} - \frac{C_1^2}{m^2} & & \\ & & \ddots & \\ & & & \end{bmatrix}$$

(b) let $u = [1\ 1\ 1 \cdots 1]$

$$\Rightarrow Cu = \begin{bmatrix} \frac{C_2}{m} - (m)\frac{C_1^2}{m^2} \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \frac{C_2 - C_1^2}{m} \\ \vdots \\ \vdots \end{bmatrix} = \frac{\sigma^2(a)}{m}\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} = \frac{\sigma_a^2}{m} \cdot u$$

clearly; Eigen value $= \frac{E[a^2]}{m} - \frac{(E(a))^2}{m^2}$

Consider $|C - \lambda_1 I|$; $\lambda_1 = \frac{C_2}{m} - \frac{C_1^2}{m^2}$ is a solution

$\Rightarrow rank(|C-\lambda_2 I|) = 1$

$\therefore$ Number of eigen vectors $= m - rank(|C-\lambda_2 I|) = n-1$

$\therefore$ One vector is $u = (1\ 1\ 1\ 1\ 1)^T$ & rest have same eigen value.

(C) In this case, all eigen vectors have same value (except one). So it is not possible to ignore any vectors as it ~~might~~ leads to loss of variance.
Ideal condition to use PCA: When eigen values are in decreasing order
Therefore, PCA is not a good way to select features for this problem.