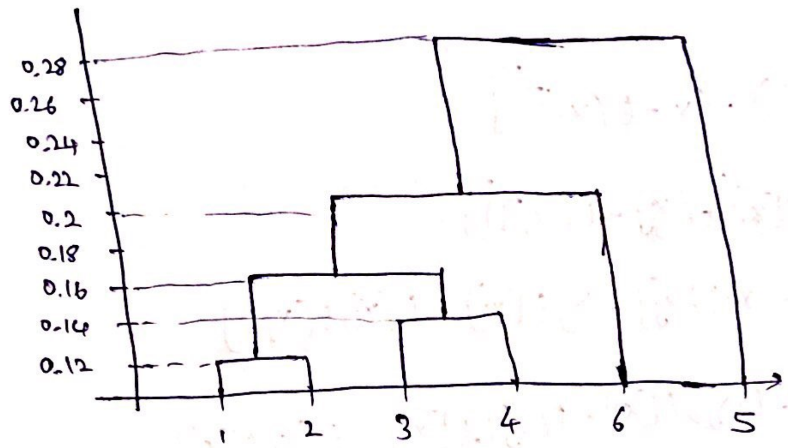
Assignment - 4
AML

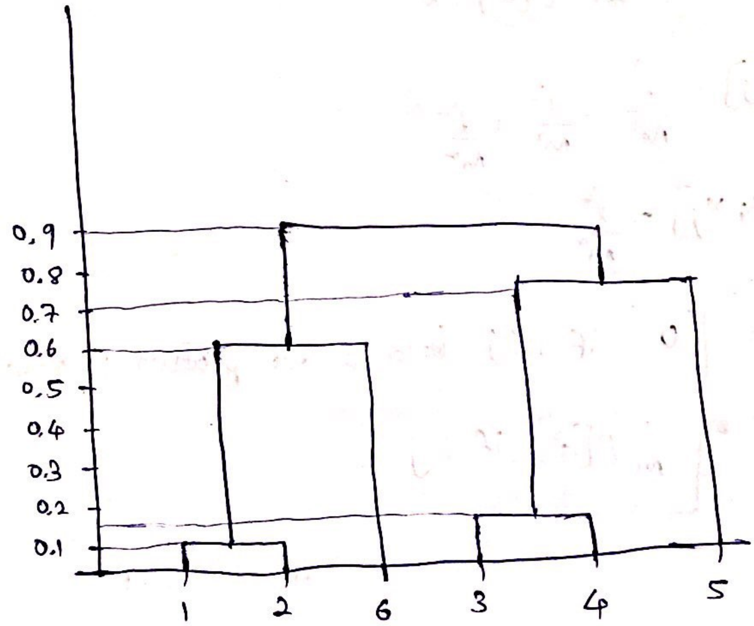
C.Saurav vara prasad
ES16BTECH11007

1A) a)



heirarchical clustering dendogram with single link

b)



heirarchical clustering with complete link dendogram

c) the first step that the complete link clustering differs with single link clustering is where AB and F are grouped together by dist (AB, F) = dist (B,F) = 0.61. We'd want dist (AB, CD) = dist(A,D), to be smaller than its value, such as 0.57. Then we want ~~dist (AB,CD)~~

dist (ABCD,F) = dist(C,F) = 0.93 to be the smallest so that ABCD and F are grouped together. We set this value to 0.63. After these changes

2A) Given

$$C = E\left[(x - E(x)) \cdot (x - E(x))^T\right]$$

$$C_{ij} = E\left[(x_i - E[x_i]) \cdot (x_j - E[x_j])\right]$$

$$= E\left[x_i x_j - x_i E[x_j] - x_j E[x_i] + E[x_i] E[x_j]\right]$$

$$= E[x_i x_j] - E[x_j] E[x_i] - E[x_j] \cdot E[x_i] + E[x_i] E[x_j]$$

we know that

$$E[x_i] = \frac{1}{M} E[a] \quad , \quad \text{let } E[a] = l$$

So, $$C_{ij} = E[x_i x_j] - \frac{l^2}{M^2} - \frac{l^2}{M^2} + \frac{l^2}{M^2}$$

$$= E[x_i x_j] - \frac{l^2}{M^2}.$$

Now, $$E[x_i x_j] = \begin{cases} 0 & \text{if } i \neq j \text{ because dot product is zero vector} \\ \frac{1}{M} E[a^2] & \text{if } i = j \end{cases}$$

let $E[a^2] = p$.

So, $$C_{ij} = \frac{p}{M} - \frac{l^2}{M^2} \quad \text{if } i \neq j$$

So, the matrix would be

$$C_{ij} = \begin{cases} -\frac{l^2}{M^2} & \text{if } i \neq j \\ \frac{p}{M} - \frac{l^2}{M^2} & \text{if } i = j \end{cases}$$

Covariance matrix.

b) Vector $v = [1, 1, \_\_, 1]$

Consider

$$Cv = \begin{bmatrix} \frac{p}{m} - \frac{l^2}{m^2} & -\frac{l^2}{m^2} & \_\_ \\ -\frac{l^2}{m^2} & \frac{p}{m} - \frac{l^2}{m^2} & \_\_ \\ & & \ddots \\ & & & \frac{p}{m} - \frac{l^2}{m^2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} p/m - \frac{m l^2}{m^2} \\ \frac{p}{m} - \frac{m l^2}{m^2} \end{bmatrix} = \left[\frac{p}{m} - \frac{l^2}{m}\right] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \lambda v$$

So, $v = [1, 1, \_\_, 1]$ is an eigen vector

Consider

$$C - \lambda I = \begin{bmatrix} \frac{p}{m} - \frac{l^2}{m^2} - \lambda & -\frac{l^2}{m^2} & \_\_ & -\frac{l^2}{m^2} \\ -\frac{l^2}{m^2} & \frac{p}{m} - \frac{l^2}{m^2} - \lambda & \_\_ \\ \_\_ & & \frac{p}{m} - \frac{l^2}{m^2} - \lambda \end{bmatrix}$$

when $\lambda = \left[\frac{p}{m} - \frac{l^2}{m^2}\right]$

$(C - \lambda I)$, then diagonal elements become zero

then matrix becomes $\underbrace{-\frac{l^2}{m^2}\begin{bmatrix} 0 & 1 & \_\_ & \_\_ \\ 1 & 0 \\ 1 & & \ddots \end{bmatrix}}_{\text{rank} = 1}$

So, this matrix has $(m-1)$ eigen vectors whose value is $\left[\frac{p}{m} - \frac{l^2}{m^2}\right]$

and m^{th} eigen vector is $[1, 1, 1, \_\_, 1]$

c) No, PCA is not a good way because all the features are equally important as all the (m-1) eigen values are same. So PCA generally ignores the features which are less important, but in this case all (m-1) values are same by discarding all of them leads to so much of the data loss.

So, PCA is not prefered to this kind of problems.