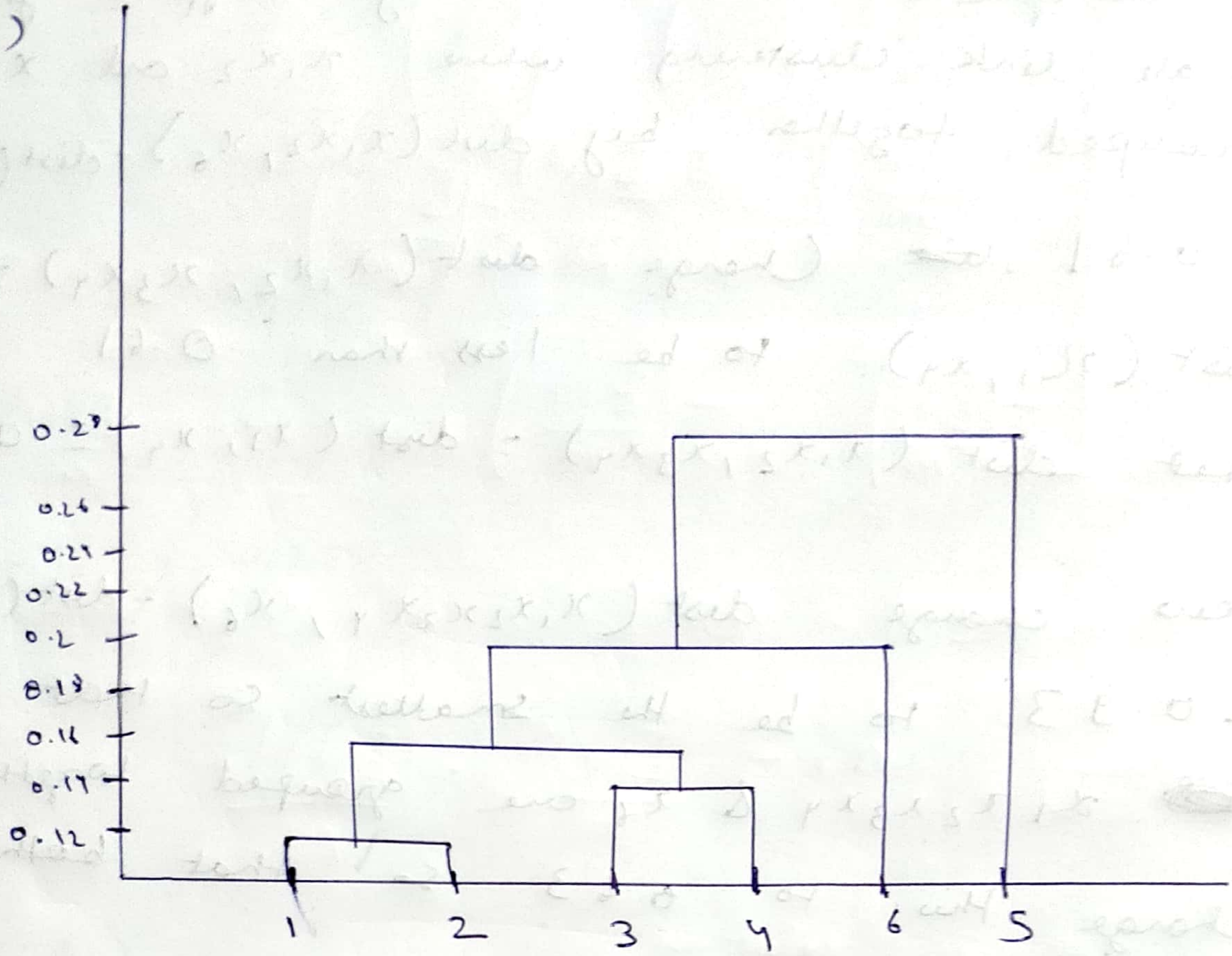
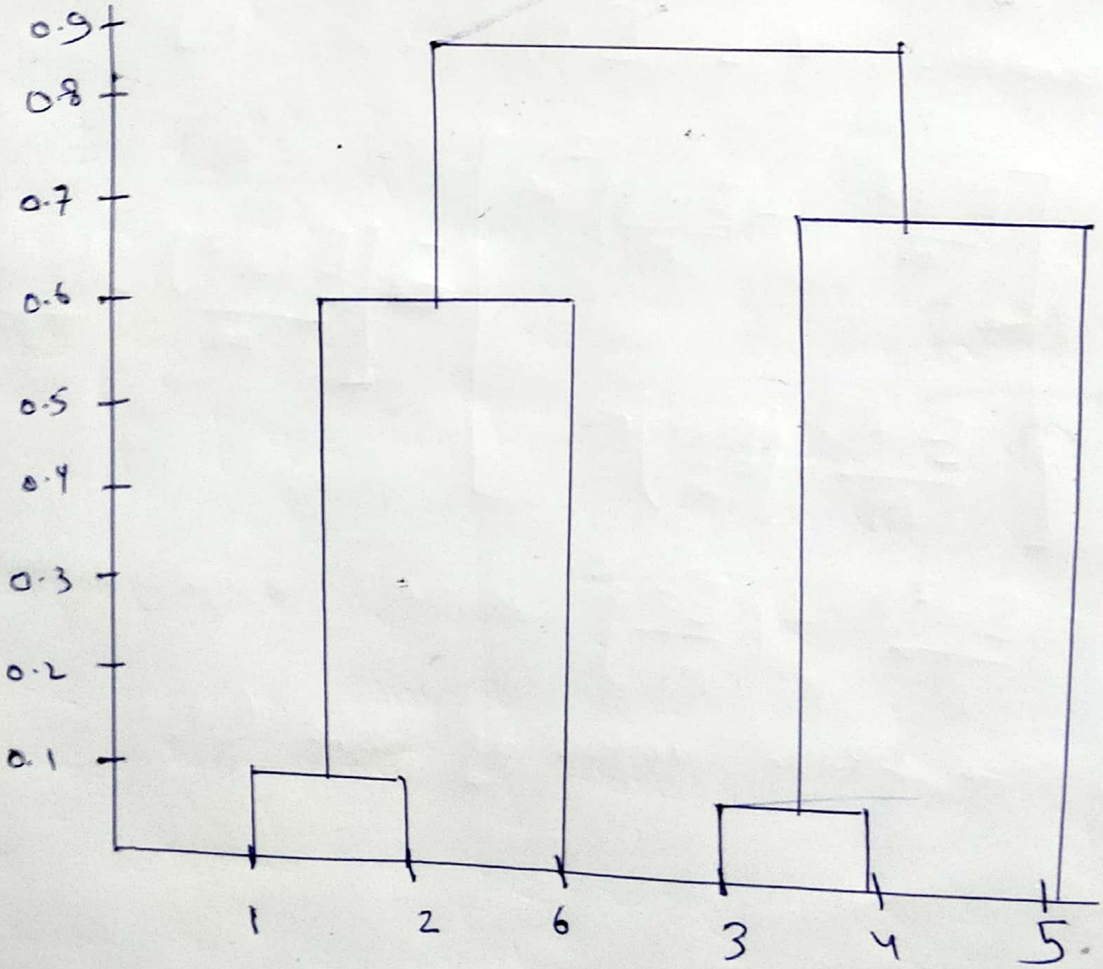
1 (a)


0.28
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
1
2
3
4
6
5

(b)


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1
2
6
3
4
5.

(c) Complete link clustering differs from single link clustering where $x_1 x_2$ and x_6 are grouped together by $dist(x_1 x_2, x_6) = dist(x_2, x_6)$

$= 0.61$. ~~the~~ Change $dist(x_1 x_2, x_3 x_4) =$ $dist(x_1, x_4)$ to be less than 0.61

Let $dist(x_1 x_2, x_3 x_4) = dist(x_1, x_4) = 0.55$

Now change $dist(x_1 x_2 x_3 x_4, x_6) = dist(x_3 x_6)$ $= 0.93$ to be the smallest so that $x_1 x_2 x_3 x_4$ & x_6 are grouped together.

Change this to 0.63 so that both dendograms become same.

2 (a)

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

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

$$E(x_i) = \frac{E(a)}{m}$$

$$C_{ij} = E(x_i x_j) - \left(\frac{E(a)}{m}\right)^2 - \left(\frac{E(a)}{m}\right)^2 + \left(\frac{E(a)}{m}\right)^2$$

$$C_{ij} = E(x_i x_j) - \left(\frac{E(a)}{m}\right)^2$$

$$E(x_i x_j) = \begin{cases} 0 & , \ i \neq j \\ \dfrac{E(a^2)}{m} & \end{cases} \Rightarrow E(x_i x_j) = \frac{1}{m} E(a^2)\delta_{ij}$$

$$C_{ij} = \frac{E(a)^2}{m} \delta_{ij} - \left(\frac{E(a)}{m}\right)^2$$

$$\lambda = -\left(\frac{E(a)}{m}\right)^2 , \quad \mu = \frac{E(a^2)}{m}$$

(b)

$$C = \begin{bmatrix} \frac{E(a^2)}{m} - \left(\frac{E(a)}{m}\right)^2 & -\left(\frac{E(a)}{m}\right)^2 & \cdots & -\left(\frac{E(a)}{m}\right)^2 \\ -\left(\frac{E(a)}{m}\right)^2 & \frac{E(a^2)}{m} - \left(\frac{E(a)}{m}\right)^2 & \cdots & -\left(\frac{E(a)}{m}\right)^2 \\ \vdots & & & \\ -\left(\frac{E(a)}{m}\right)^2 & -\left(\frac{E(a)}{m}\right)^2 & \cdots & \frac{E(a^2)}{m} - \left(\frac{E(a)}{m}\right)^2 \end{bmatrix}$$

$|C - \lambda I| = 0$, λ is eigen value

Clearly $\lambda_i = \frac{E(a)^2}{m} - \frac{(E(a))^2}{m}$

is a solution

Also consider, $v = [1, 1, \ldots, 1]^T$

$$Cv = \begin{bmatrix} \frac{E(a)^2}{m} - \left(\frac{E(a)}{m}\right)^2 M \\ \vdots \\ \frac{E(a^2)}{m} - \left(\frac{E(a)}{m}\right)^2 m \end{bmatrix} = \left[\frac{E(a^2)}{m} - \frac{(E(a))^2}{M}\right] v$$

$\lambda_2 = \frac{E(a^2)}{M} - \left(\frac{E(a)}{M}\right)^2$ is an eigen value

and has eigen vector of form $(1, 1, \ldots, 1)$

Consider λ_i

$\text{rank}(|C - \lambda_i I|) = 1$

No. of eigen vectors of $\lambda_2 = m - \text{rank}(C - \lambda_i I)$
$= m - 1$

$\Rightarrow$ One eigen vector is of the form $(1, 1, \ldots, 1)$ and all other eigen vectors have same value which is $\frac{E(a^2)}{m} - \left(\frac{E(a)}{m}\right)^2$

(c) PCA will not work well as all eigen values are same, thereby removing dimensions will result in a large loss in variance explained by PCA.