Q.1) Apply the K-means clustering for following data with $K=2$. State difference between K-means clustering and hierarchical clustering.

	x_1	x_2
A	2	3
B	6	1
C	1	2
D	3	0

→

Here, we have given $K=2$ Therefore we use A and D are two initial centroids.

$\therefore C_1(2,3)$ & $C_2(3,0)$.

Proximity matrix using Euclidean distance is

$$d(P_1,P_2) = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$$

Data Points			C_1	C_2	Closest Cluster
			2 | 3	3 | 0	
A	2	3	0	3.16	C_1
B	6	1	4.47	3.16	C_2
C	1	2	1.41	2.82	C_1
D	3	0	3.16	0	C_2

New centroids will be,

$C_1\left(\frac{3}{2}, \frac{5}{2}\right)$ & $C_2\left(\frac{9}{2}, \frac{1}{2}\right)$

New centroids will be

$C_1\left(\frac{2+1}{2}, \frac{3+2}{2}\right)$, $C_2\left(\frac{6+3}{2}, \frac{1+0}{2}\right)$

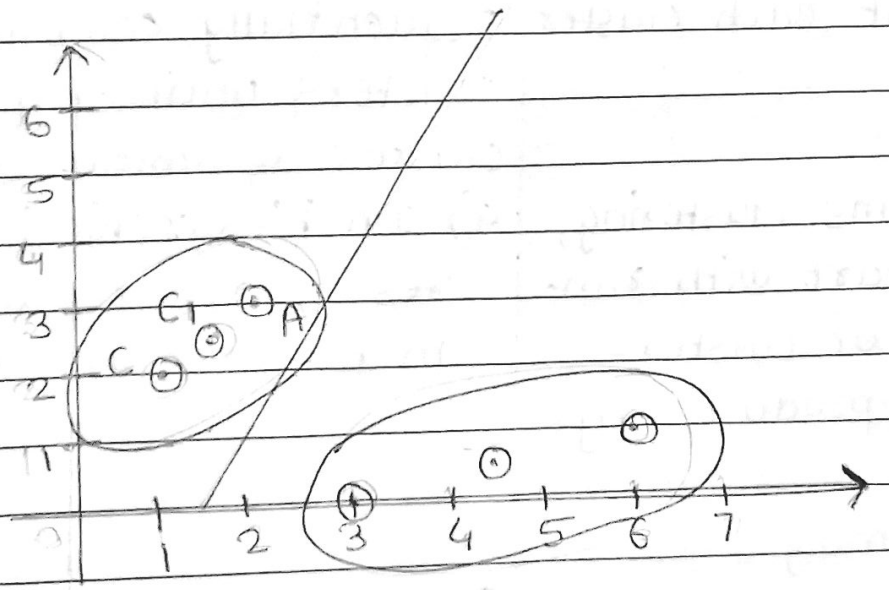
i.e. $C_1(3/2, 5/2)$, $C_2(9/2, 1/2)$.

Data Points		C_1		C_2		Closest cluster
		3/2	5/2	9/2	1/2	
A	2 3	0.71		3.54		C_1
B	6 1	4.74		1.58		C_2
C	1 2	0.71		3.80		C_1
D	3 0	2.92		1.58		C_2

This cluster is similar to previous cluster. it can not be further divided. Therefore New cluster will be,

$C_1\left(\frac{2+1}{2}, \frac{3+2}{2}\right)$, $C_2\left(\frac{6+3}{2}, \frac{1+0}{2}\right)$

i.e. $C_1(3/2, 5/2)$, $C_2(9/2, 1/2)$.



K means clustering	Hierarchical clustering.
1) K-means using a pre-specified number of clusters the method assigns records to each cluster to find the mutually exclusive cluster of spherical shape based on distance.	1) Hierarchical methods can be either divisive or agglomerative.
2) K means clustering needed advance knowledge of K. i.e. no. of clusters one want to divide your data.	2) In hierarchical clustering one can stop at any number of clusters, one find appropriate by interpreting the dendrogram
3) One can use median or mean as a cluster center to represent each cluster	3) Agglomerative methods begin with 'n' clusters and sequentially combine similar clusters until only one cluster is obtained
4) In K means clustering, since one start with random choice of clusters. the results produced by running the algorithm many times may differ.	4) In hierarchical clustering results are reproducible in hierarchical clustering.
5) Better performance when dealing with convex clusters.	5) Generate better result when dealing with non convex clusters.

Q.2) Draw a dendrogram for the following proximity matrix and find the number of clusters that we get.

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	(2)	8	0

Pair (3,5) → (2)

	1	2	(3,5)	4
1	0			
2	9	0		
(3,5)	(3)	7	0	
4	6	5	8	0

$|(3,5) \rightarrow 2| = \min[(3,2),(5,2)]$

$= \min[7,10] = 7$

Pair (3,5) and (1) (3) →

	((3,5) & (1))	2	4
((3,5) & (1))	0		
2	7	0	
4	6	(5)	0

pair (3,5), (1), (2)

(5) → (3,5), (1), (2)

4

(3,5), (1), (2) 4

0

5 0.

Dendrogram will be as follow,