Q.1) Assume that the data set D is given by the table below. Follow complete linkage technique to find clusters in D. Use Euclidean distance measure.

Points	A	B
P_1	2	4
P_2	8	2
P_3	9	3
P_4	1	5
P_5	8.5	1

Soln: P_1, P_3, P_5 are center of each cluster
Number of Clusters K = 3
Initial cluster ~~Centers~~ Centres be $C_1 = P_1(2,4)$
$C_2 = P_3(9,3)$ $\quad$ $C_3 = P_5(8.5, 1)$
Euclidian distance formula
$$\text{Distance}[(x,y),(a,b)] = \sqrt{(x-a)^2 + (y-b)^2}$$

Iteration 1:
$P_1(2,4)$

$\text{Distance}[(2,4),(2,4)] = \sqrt{(2-2)^2 + (4-4)^2} = 0 \rightarrow$ small
$\text{Distance}[(2,4),(9,3)] = \sqrt{(2-9)^2 + (4-3)^2} = 7.07$
$\text{Distance}[(2,4),(8.5,1)] = \sqrt{(2-8.5)^2 + (4-1)^2} = 7.15$

$P_1(2,4)$ belongs to cluster C1.

$P_2(8,2)$

$\text{Distance}[(8,2),(2,4)] = \sqrt{(8-2)^2 + (2-4)^2} = 6.32$
$\text{Distance}[(8,2),(9,3)] = \sqrt{(8-9)^2 + (2-3)^2} = 1.41$
$\text{Distance}[(8,2),(8.5,1)] = \sqrt{(8-8.5)^2 + (2-1)^2} = 1.11 \rightarrow$ small

P_2 (8,2) belongs to cluster C_3.

P_3 (9,3)

Distance [(9,3), (2,4)] = $\sqrt{(9-2)^2 + (3-4)^2}$ = 7.07

Distance [(9,3), (9,3)] = $\sqrt{(9-9)^2 + (3-3)^2}$ = 0 → small

Distance [(9,3), (8.5,1)] = $\sqrt{(9-8.5)^2 + (3-1)^2}$ = 2.06

P_3 (9,3) belongs to cluster C_2.

P_4 (1,5)

Distance [(1,5), (2,4)] = $\sqrt{(1-2)^2 + (5-4)^2}$ = 1.414 ← small

Distance [(1,5), (9,3)] = $\sqrt{(1-9)^2 + (5-3)^2}$ = 8.24

Distance [(1,5), (8.5,1)] = $\sqrt{(1-8.5)^2 + (5-1)^2}$ = 8.5

P_4 (1,5) belongs to cluster C_1.

P_5 (8.5,1)

Distance [(8.5,1), (2,4)] = $\sqrt{(8.5-2)^2 + (1-4)^2}$ = 7.15

Distance [(8.5,1), (9,3)] = $\sqrt{(8.5-9)^2 + (1-3)^2}$ = 2.06

Distance [(8.5,1), (8.5,1)] = $\sqrt{(8.5-8.5)^2 + (1-1)^2}$ = 0 → small

P_5 (8.5,1) belongs to cluster C_3.

After iteration 1

cluster C_1 = [P_1 (2,4), P_4 (1,5)]

cluster C_2 = [P_3 (9,3)]

cluster C_3 = [P_2 (8,2), P_5 (8.5,1)]

Iteration 2

Centers of new clusters

cluster C_1 = $\left(\frac{2+1}{2}, \frac{4+5}{2}\right)$ = (1.5, 4.5)

cluster C_2 = (9,3)

cluster C_3 = $\left[\frac{8+8.5}{2}, \frac{2+1}{2}\right]$ = (8.25, 1.5)

$P_1(2,4)$

Distance$[(2,4),(1.5,4.5)] = \sqrt{(2-1.5)^2+(4-4.5)^2} = 0.707 \rightarrow$ small

Distance$[(2,4),(9,3)] = \sqrt{(2-9)^2+(4-3)^2} = 7.07$

Distance$[(2,4),(8.25,1.5)] = \sqrt{(2-8.25)^2+(4-1.5)^2} = 6.73$

$P_1(2,4)$ belongs to cluster C_1

$P_2(8,2)$

Distance$[(8,2),(1.5,4.5)] = \sqrt{(8-1.5)^2+(2-4.5)^2} = 6.96$

Distance$[(8,2),(9,3)] = \sqrt{(8-9)^2+(2-3)^2} = 1.414$

Distance$[(8,2),(8.25,1.5)] = \sqrt{(8-8.25)^2+(2-1.5)^2} = 0.55 \rightarrow$ small

$P_2(8,2)$ belongs to Cluster C_3

$P_3(9,3)$

Distance$[(9,3),(1.5,4.5)] = \sqrt{(9-1.5)^2+(3-4.5)^2} = 7.64$

Distance$[(9,3),(9,3)] = \sqrt{(9-9)^2+(3-3)^2} = 0 \rightarrow$ small

Distance$[(9,3),(8.25,1.5)] = \sqrt{(9-8.25)^2+(3-1.5)^2} = 1.677$

$P_3(9,3)$ belongs to cluster C_2

$P_4(1,5)$

Distance$[(1,5),(1.5,4.5)] = \sqrt{(1-1.5)^2+(5-4.5)^2} = 0.707 \rightarrow$ small

Distance$[(1,5),(9,3)] = \sqrt{(1-9)^2+(5-3)^2} = 8.24$

Distance$[(1,5),(8.25,1.5)] = \sqrt{(1-8.25)^2+(5-1.5)^2} = 8.05$

$P_4(1,5)$ belongs to cluster C_1

$P_5(8.5,1)$

Distance$[(8.5,1),(1.5,4.5)] = \sqrt{(8.5-1.5)^2+(1-4.5)^2} = 7.8$

Distance$[(8.5,1),(9,3)] = \sqrt{(8.5-9)^{22}+(1-3)^2} = 2.06$

Distance$[(8.5,1),(8.25,1.5)] = \sqrt{(8.5-8.25)^2+(1-1.5)^2} = 0.55 \rightarrow$ small

$P_5(8.5,1)$ belongs to cluster C_3

After Iteration 2

Cluster C1 = [P_1(2,4) ; P_4(1,5)]

Cluster C2 = [P_3(9,3)]

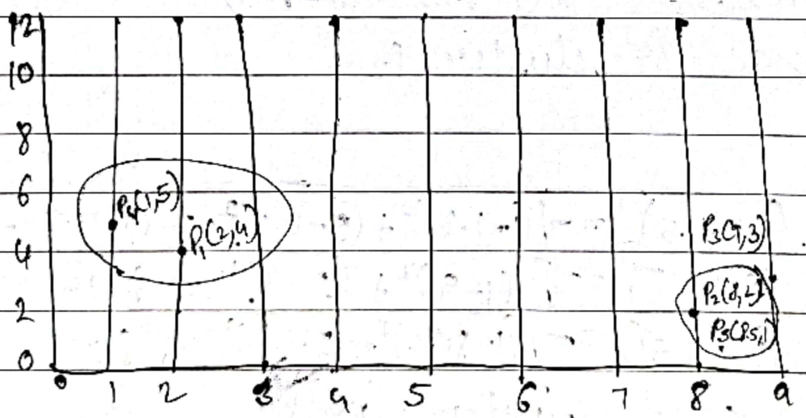
Cluster C3 = [P_2(8,2) ; P_5(8.5,1)]

Comparing the clustering of iteration 1 and iteration 4, we find that objects does not move cluster anymore. Thus, the computation of the K-Mean Clustering has reached its stability and no more iteration is needed. So the final clusters are:

Cluster C_1 = { P_1(2,4) ; P_4(1,5) }

Cluster C_2 = { P_3(9,3) }

Cluster C_3 = { P_2(8,2) ; P_5(8.5,1) }



Clustering after last iteration

Q.2) Co-ordinates of objects are given below. Apply K medoid (PAM) to cluster the coordinates into two clusters

Objects	X	Y
0	8	7
1	3	7
2	4	9
3	9	6
4	8	5
5	5	8
6	7	3
7	8	4
8	7	5
9	4	5

Soln:- Step 1:- We select 2 random representative objects

(object) i	x	y	c		Distance / cost = C
0	8	7	4	9	$\|8-4\| + \|7-9\| = 6$
1	3	7	4	9	$\|3-4\| + \|7-9\| = 3$
3	9	6	4	9	$\|9-4\| + \|6-9\| = 8$
4	8	5	4	9	$\|8-4\| + \|5-9\| = 8$
5	5	8	4	9	$\|5-4\| + \|8-9\| = 2$
6	7	3	4	9	$\|7-4\| + \|3-9\| = 9$
8	7	5	4	9	$\|7-4\| + \|5-9\| = 7$
9	4	5	4	9	$\|4-4\| + \|5-9\| = 4$

Object(i)	x	y	C_2		Distance/cost	c
0	8	7	8	4	$\|8-8\| + \|7-4\|$	3
1	3	7	8	4	$\|3-8\| + \|7-4\|$	8
3	9	6	8	4	$\|9-8\| + \|6-4\|$	3
4	8	5	8	4	$\|8-8\| + \|5-4\|$	1
5	5	8	8	4	$\|5-8\| + \|8-4\|$	7
6	7	3	8	4	$\|7-8\| + \|3-4\|$	2
8	7	5	8	4	$\|7-8\| + \|5-4\|$	2
9	4	5	8	4	$\|4-8\| + \|5-4\|$	2

Compare cost of cost(C_1) and cost(C_2) for every i and select the minimum one.

Step 2:- the cluster are

Cluster 1: {(3,7), (4,9), (5,8), (4,5)}

Cluster 2: {(8,7), (9,6), (8,5), (7,3), (8,4), (7,5)}

calculate total cost

$$T\,cost(x,c) = \sum_{i=1} |x_i - c_i|$$

Total cost = {cost((8,4),(8,7)), cost((8,4),(9,6)), cost((8,4),(8,5)), cost((8,4),(7,5)), cost((4,9),(3,7)), cost((4,9),(5,8)), cost((4,9),(5,8)), cost((4,9),(4,5))}

$= (3+3+1+2+2) + (3+2+4)$

$= 11 + 9 = 20$

Step 3:- select one of non-medoids O

Let $O' = (8,5)$ i.e. object(4)

So now medoid are $C_1(4,9)$ and $O'(8,5)$.

Object(i)	x	y	O'		Distance/cost	C
0	8	7	8	5	|8-8| + |7-5|	2
1	3	7	8	5	|3-8| + |7-5|	7
3	9	6	8	5	|9-8| + |6-5|	2
5	5	8	8	5	|5-8| + |8-5|	6
6	7	3	8	5	|7-8| + |3-5|	3
7	8	4	8	5	|8-8| + |4-5|	1
8	7	5	8	5	|7-8| + |5-5|	1
9	4	5	8	5	|4-8| + |5-5|	4

Object(i)	x	y	C_1		Distance/cost	C
0	8	7	4	9	|8-4| + |7-9|	6
1	3	7	4	9	|3-4| + |7-9|	3
3	9	6	4	9	|9-4| + |6-9|	8
5	5	8	4	9	|5-4| + |8-9|	2
6	7	3	4	9	|7-4| + |3-9|	9
7	8	4	4	9	|8-4| + |4-9|	9
8	7	5	4	9	|7-4| + |5-9|	7
9	4	5	4	9	|4-4| + |5-9|	4

Compare the cost of cost(C_1) and cost(O') energy i and select the minimum one

again create the cluster.

Cluster 1: {(3,7), (5,8), (4,5), (4,9)}
Cluster 2: {(8,7), (9,6), (7,3), (8,4), (7,5), (8,5)}

Current total cost = (2+2+3+1+1) + (3+2+4)
= 9+9 = 18

Step 4:- So cost of swapping medoid from C_2 to O' is
S = current ~~tot~~ total cost - past total cost
= 18-20 = -2 < 0

So moving O' would be a good idea
Now, move cluster again
Select one of ~~non medi~~ non-medoids O''
Let O'' = (7,3) i.e. object (6)

So now medoid are C_1 (4,9) and O'' (7,3)

object(i)	x	y	C_1		Distance/cost	C
0	8	7	4	9	|8-4| + |7-9|	6
1	3	7	4	9	|3-4| + |7-9|	3
3	9	6	4	9	|9-4| + |6-9|	8
4	8	5	4	9	|8-4| + |5-9|	8
5	5	8	4	9	|5-4| + |8-9|	2
7	8	4	4	9	|8-4| + |4-9|	9
8	7	5	4	9	|7-4| + |5-9|	7
9	4	5	4	9	|4-4| + |5-9|	4

Object(i)	x	y	O''		Distance|cost	c
0	8	7	7	3	|8-7| + |7-3|	5
1	3	7	7	3	|3-7| + |7-3|	8
3	9	6	7	3	|9-7| + |6-3|	5
4	8	5	7	3	|8-7| + |5-3|	3
5	5	8	7	3	|5-7| + |8-3|	7
7	8	4	7	3	|8-7| + |4-3|	2
8	7	5	7	3	|7-7| + |5-3|	2
9	4	5	7	3	|4-7| + |5-3|	5

again create the cluster
Cluster 1: {(3,7), (5,8), (4,5), (4,9)}
Cluster 2: {(8,7), (9,6), (8,5), (8,4), (7,5), (7,3)}

current total cost = (3+2+4) + (5+5+3+2+2)
= 9 + 17
= 26

So cost of swapping medoid from O' to O'' is
S = current total ~~tot~~ cost − Past total cost
S = 26−18 = 8 > 0

So moving O' could be a bad idea so previous choice was good.