Clustering Problems

1. **Cluster the following data using k-means clustering algorithm . Consider k=2 . Dataset={2,4,10,12,3,20,30,11,25}**

   Solution:

   1. Randomly assign means: $m_1 = 3$, $m_2 = 4$
   2. The numbers which are close to mean $m_1 = 3$ are grouped into cluster K_1 an numbers which are close to mean $m_2 = 4$ are grouped into cluster K_2.
   3. Again calculate the new mean for new cluster groups.
   4. $K_1=\{2,3\}$, $K_2=\{4,10,12,20,30,11,25\}$, $m_1=2.5, m_2=16$
   5. $K_1=\{2,3,4\}, K_2=\{10,12,20,30,11,25\}$, $m_1=3, m_2=18$
   6. $K_1=\{2,3,4,10\}, K_2=\{12,20,30,11,25\}$, $m_1=4.75, m_2=19.6$
   7. $K_1=\{2,3,4,10,11,12\}, K_2=\{20,30,25\}$, $m_1=7, m_2=25$
   8. $K_1=\{2,3,4,10,11,12\}, K_2=\{20,30,25\}$
   9. Stop as the clusters with these means (in step 7 and 8) are the same. The clust in the last two groups are identical.
   10. So the final answer is $K_1=\{2,3,4,10,11,12\}, K_2=\{20,30,25\}$

2. **For the following distance matrix , draw single link and complete link dendogram.**

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

Solution:

Single Link:

Step 1 :

$$\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & & & & \\ 2 & 2 & 0 & & & \\ 3 & 6 & 3 & 0 & & \\ 4 & 10 & 9 & 7 & 0 & \\ 5 & 9 & 8 & 5 & 4 & 0 \end{array} \Rightarrow \begin{array}{c|cccc} & (1,2) & 3 & 4 & 5 \\ (1,2) & 0 & & & \\ 3 & 3 & 0 & & \\ 4 & 9 & 7 & 0 & \\ 5 & 8 & 5 & 4 & 0 \end{array}$$

$d_{(1,2)3} = \min \{d_{1,3}, d_{2,3}\} = \min \{6, 3\} = 3$

$d_{(1,2)4} = \min \{d_{1,4}, d_{2,4}\} = \min \{10, 9\} = 9$

$d_{(1,2)5} = \min \{d_{1,5}, d_{2,5}\} = \min \{9, 8\} = 8$

$$\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & & & & \\ 2 & 2 & 0 & & & \\ 3 & 6 & 3 & 0 & & \\ 4 & 10 & 9 & 7 & 0 & \\ 5 & 9 & 8 & 5 & 4 & 0 \end{array} \Rightarrow \begin{array}{c|cccc} & (1,2) & 3 & 4 & 5 \\ (1,2) & 0 & & & \\ 3 & 3 & 0 & & \\ 4 & 9 & 7 & 0 & \\ 5 & 8 & 5 & 4 & 0 \end{array} \Rightarrow \begin{array}{c|ccc} & (1,2,3) & 4 & 5 \\ (1,2,3) & 0 & & \\ 4 & 7 & 0 & \\ 5 & 5 & 4 & 0 \end{array}$$

$d_{(1,2,3),4} = \min \{d_{(1,2),4}, d_{3,4}\} = \min \{9, 7\} = 7$

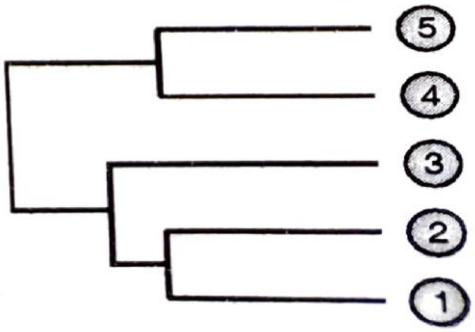
$d_{(1,2,3)5} = \min \{d_{(1,2),5}, d_{3,5}\} = \min \{8, 5\} = 5$

Step 3 :

$$\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & & & & \\ 2 & 2 & 0 & & & \\ 3 & 6 & 3 & 0 & & \\ 4 & 10 & 9 & 7 & 0 & \\ 5 & 9 & 8 & 5 & 4 & 0 \end{array} \Rightarrow \begin{array}{c|cccc} & (1,2) & 3 & 4 & 5 \\ (1,2) & 0 & & & \\ 3 & 3 & 0 & & \\ 4 & 9 & 7 & 0 & \\ 5 & 8 & 5 & 4 & 0 \end{array} \Rightarrow \begin{array}{c|ccc} & (1,2,3) & 4 & 5 \\ (1,2,3) & 0 & & \\ 4 & 7 & 0 & \\ 5 & 5 & 4 & 0 \end{array}$$

$$\begin{array}{c|cc} & (1,2,3) & 4,5 \\ (1,2,3) & 0 & \\ (4,5) & 5 & 0 \end{array}$$

$d_{(1,2,3),(4,5)} = \min \{d_{(1,2,3),4}, d_{(1,2,3),5}\} = \min \{7, 5\} = 5$



Complete Link :

Step 1 :

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ 2 & 0 & & & \\ 6 & 3 & 0 & & \\ 10 & 9 & 7 & 0 & \\ 9 & 8 & 5 & 4 & 0 \end{bmatrix} \end{array} \Rightarrow \begin{array}{c} \\ (1,2) \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{cccc} (1,2) & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & \\ 6 & 0 & & \\ 10 & 7 & 0 & \\ 9 & 5 & 4 & 0 \end{bmatrix} \end{array}$$

$$d_{(1,2),3} = \max\{d_{1,3}, d_{2,3}\} = \max\{6, 3\} = 6$$

$$d_{(1,2)4} = \max\{d_{,1,4}, d_{2,4}\} = \max\{10, 9\} = 10$$

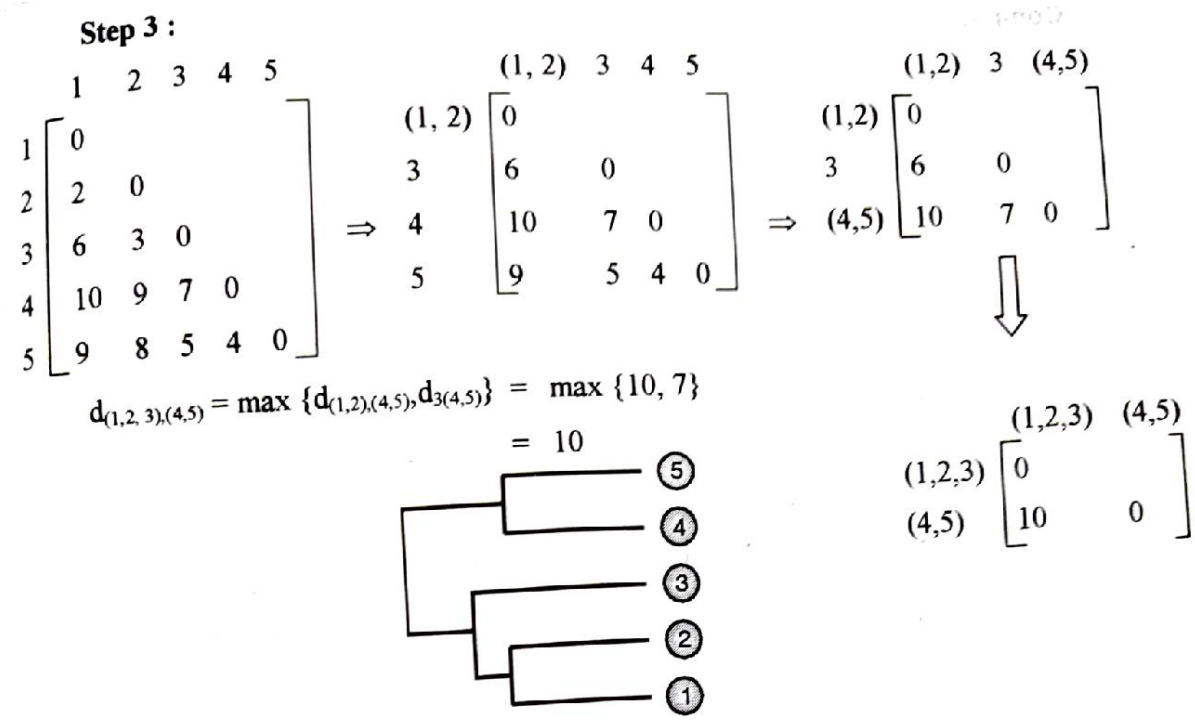
$$d_{(1,2)5} = \max\{d_{,1,5}, d_{2,5}\} = \max\{9, 8\} = 9$$

Step 2 :

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ 2 & 0 & & & \\ 6 & 3 & 0 & & \\ 10 & 9 & 7 & 0 & \\ 9 & 8 & 5 & 4 & 0 \end{bmatrix} \end{array} \Rightarrow \begin{array}{c} \\ (1,2) \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{cccc} (1,2) & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & \\ 6 & 0 & & \\ 10 & 7 & 0 & \\ 9 & 5 & 4 & 0 \end{bmatrix} \end{array} \Rightarrow \begin{array}{c} \\ (1,2) \\ 3 \\ (4,5) \end{array}\begin{array}{c} \begin{array}{ccc} (1,2) & 3 & (4,5) \end{array} \\ \begin{bmatrix} 0 & & \\ 6 & 0 & \\ 10 & 7 & 0 \end{bmatrix} \end{array}$$

$$d_{(1,2),(4,5)} = \max\{d_{(1,2),4}, d_{(1,2),5}\} = \max\{10, 9\} = 10$$

$$d_{3,(4,5)} = \max\{d_{,3,4}, d_{3,5}\} = \max\{7, 5\} = 7$$



3. **Apply DBSCAN clustering algorithm to cluster following dataset. Consider Eps=1.9 and Minpts=4**

Point	X	Y
P1	7	4
P2	6	4
P3	5	6
P4	4	2
P5	6	3
P6	5	2
P7	3	3
P8	4	5
P9	6	5
P10	3	6
P11	4	4
P12	8	2

Solution:

Step1: Find distance matrix using Euclidean distance

P1	0.00											
P2	1.00	0.00										
P3	2.83	2.24	0.00									
P4	3.61	2.83	4.12	0.00								
P5	1.41	1.00	3.16	2.24	0.00							
P6	2.83	2.24	4.00	1.00	1.41	0.00						
P7	4.12	3.16	3.61	1.41	3.00	2.24	0.00					
P8	3.16	2.24	1.41	3.00	2.83	3.16	2.24	0.00				
P9	1.41	1.00	1.41	3.61	2.00	3.16	3.61	2.00	0.00			
P10	4.47	3.61	2.00	4.12	4.24	4.47	3.00	1.41	3.16	0.00		
P11	3.00	2.00	2.24	2.00	2.24	2.24	1.41	1.00	2.24	2.24	0.00	
P12	2.24	2.83	5.00	4.00	2.24	3.00	5.10	5.00	3.61	6.40	4.47	0.00
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12

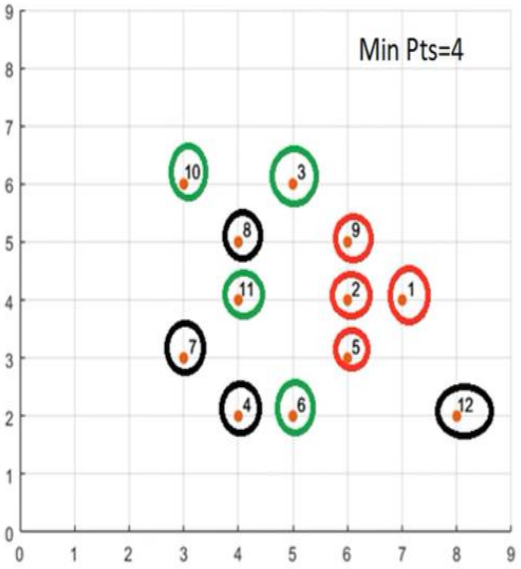
Eps=1.9

MinPts=4

Step 2: Find Є-neighborhood of each data point

P1: P2,P5, P9	**P2:P1, P5, P9**	P3:P8, P9	P4: P6, P7
P5:P1,P2,P6	P6:P4, P5	P7: P4, P11	**P8:P3, P10, P11**
P9:P1, P2, P3	P10:P8	P11: P7, P8	P12:

Step3: Identify core points, border points and Noise points



Point	Status	
P1	Core	
P2	Core	
P3	Noise	Border
P4	Noise	
P5	Core	
P6	Noise	Border
P7	Noise	
P8	Core	
P9	Core	
P10	Noise	Border
P11	Noise	Border
P12	Noise	

https://www.youtube.com/watch?v=S5OvKmWIdZA