Data Mining; Assignemt 6

Jesse Annan | ID: 002708111

March 29, 2023

Find step by step procedure of 1-5 in attached excel file.

1. K-means algorithm
   Starting with k=4 clusters. I selected *(0,7), (4,4), (0,0), and (9,9)* as my
   initial cluster centers.
   d(x,y) = Manhattan distance.

First Iteration:

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

Second Iteration:

	Cluster centers		
(1.5, 7.1$\dot{6}$)	(5.9, 3.7)	(2, 0.5)	(8, 8.5)
cluster 1	cluster 2	cluster 3	cluster 4
(0,7)	(5,3)	(1,1)	(7,8)
(1,6)	(6,2)	(3,0)	(9,9)
(1,8)	(6,4)		
(2,5)	(7,2)		
(2,7)	(7,3)		
(2,8)	(7,5)		
(3,6)	(8,3)		
(3,7)	(8,4)		

Third Iteration:

| | Cluster centers | | |
| (1.75, 6.75) | (6.75, 3.25) | (2, 0.5) | (8, 8.5) |
cluster 1	cluster 2	cluster 3	cluster 4
(0,7)	(5,3)	(1,1)	(7,8)
(1,6)	(6,2)	(3,0)	(9,9)
(1,8)	(6,4)		
(2,5)	(7,2)		
(2,7)	(7,3)		
(2,8)	(7,5)		
(3,6)	(8,3)		
(3,7)	(8,4)		

Algorithm terminated since the clusters where the same.

2. K-medoid algorithm
   Starting with k=4 clusters. I selected *(1,8), (1,1), (2,5), and (7,8)* as my initial cluster centers.
   d(x,y) = Manhattan distance.

First Iteration:

Cluster centers			
(1, 8) cluster 1	(1, 1) cluster 2	(2, 5) cluster 3	(7, 8) cluster 4
(0,7)	(1,1)	(1,6)	(7,2)
(1,8)	(3,0)	(2,5)	(7,3)
(2,7)	(6,2)	(3,6)	(7,5)
(2,8)		(5,3)	(7,8)
(3,7)		(6,4)	(8,3)
			(8,4)
			(9,9)
Cost = 59			

Second Iteration:

Cluster centers			
(1, 8) cluster 1	(1, 1) cluster 2	(5, 3) cluster 3	(7, 8) cluster 4
(0,7)	(1,1)	(5,3)	(7,5)
(1,6)	(3,0)	(6,2)	(7,8)
(1,8)		(6,4)	(9,9)
(2,5)		(7,2)	
(2,7)		(7,3)	
(2,8)		(8,3)	
(3,6)		(8,4)	
(3,7)			
Cost = 43			

4

Third Iteration:

| | Cluster centers | | |
| (1, 8) | (1, 1) | (7, 3) | (7, 8) |
cluster 1	cluster 2	cluster 3	cluster 4
(0,7)	(1,1)	(5,3)	(7,8)
(1,6)	(3,0)	(6,2)	(9,9)
(1,8)		(6,4)	
(2,5)		(7,2)	
(2,7)		(7,3)	
(2,8)		(7,5)	
(3,6)		(8,3)	
(3,7)		(8,4)	
Cost = 36			

Forth Iteration:

| | Cluster centers | | |
| (1, 8) | (1, 1) | (7, 3) | (7, 8) |
cluster 1	cluster 2	cluster 3	cluster 4
(0,7)	(1,1)	(5,3)	(7,8)
(1,6)	(3,0)	(6,2)	(9,9)
(1,8)		(6,4)	
(2,5)		(7,2)	
(2,7)		(7,3)	
(2,8)		(7,5)	
(3,6)		(8,3)	
(3,7)		(8,4)	
Cost = 30			

Algorithm terminated because of the reduced cost and unchanged clusters.

3. AGNES algorithm

   d(x,y) = Manhattan distance.
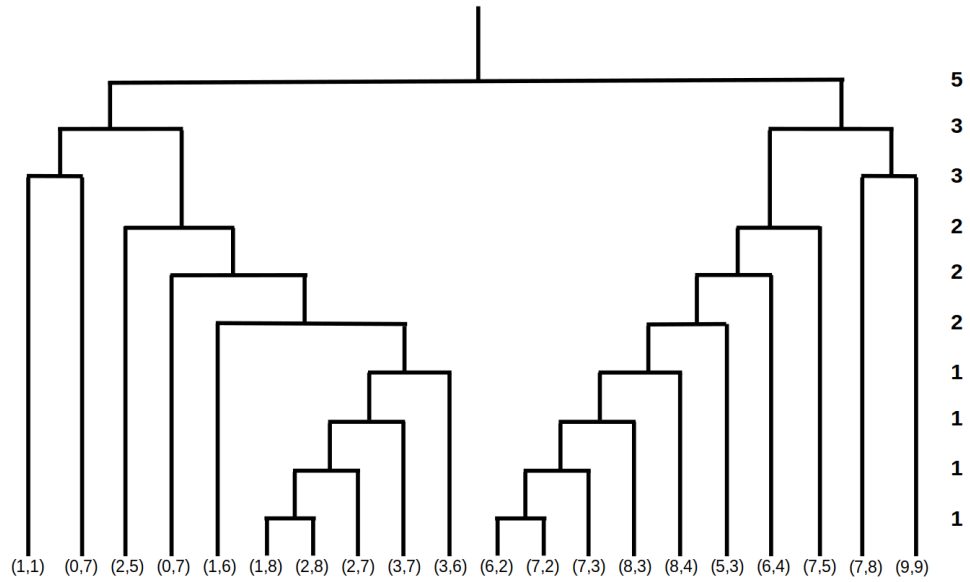
Algorithm terminated because every single cluster joined the hierarchical tree.

4. DBSCAN algorithm
   d(x,y) = Euclidean distance.

Core points					
(1,6)	(1,8)	(2,7)	(2,8)	(3,6)	(3,7)
(6,2)	(6,4)	(7,2)	(7,3)	(8,3)	(8,4)
Border points					
	(0,7)	(2,5)	(5,3)	(7,5)	

Clustering core points.

	Core points					
Cluster 1	(1,6)	(1,8)	(2,7)	(2,8)	(3,6)	(3,7)
Cluster 2	(6,2)	(6,4)	(7,2)	(7,3)	(8,3)	(8,4)

Adding border points to the clusters

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

5. OPTICS algorithm
   d(x,y) = Manhattan distance.

Optics clustering graph
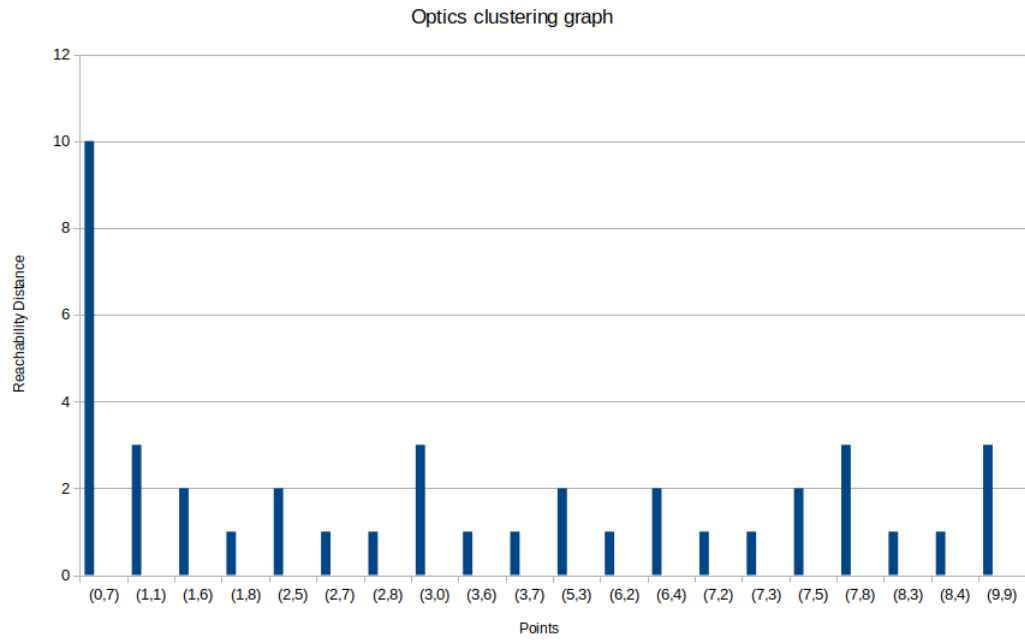
6. Spectral clustering algorithm
   Kindly check attached jupyter notebook file.