DSAA 5002 - Data Mining and Knowledge Discovery in Data Science

Q1 [15 Marks]

(Fall Semester 2023)

Homework 3

Deadline: 22 Nov 2023 11:59pm

(Please hand in via Canvas.) Full Mark: 100 Marks

Q1 [20 Marks]

Apply the agglomerative hierarchical clustering algorithm with the following distance matrix and the single linkage. Plot the cluster tree and mark out all the merging levels.

	1	2	3	4
2	2.33			
3	3.15	1.30		
4	1.90	1.50	3.70	
5	3.01	0.47	1.40	1.82

Table 1： distance matrix

Q2 [20 Marks]

Use the similarity matrix in Table 2 to perform single-link hierarchical clustering. Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the clusters are merged.

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

Table 2: Similarity matrix for Q2

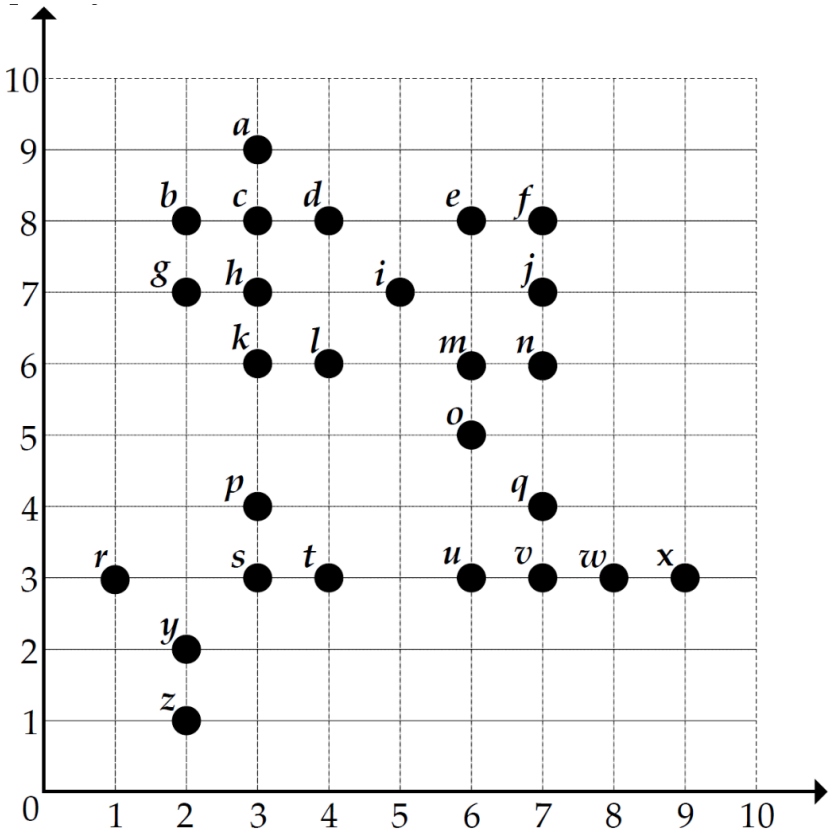
Q3 [30 Marks]

Apply DBSCAN with parameters MinPts=4 and Eps = $\sqrt{2}$ to get clustering results.

First, for every data point, answer if it is a core, a border, or an outlier.

Second, for data points that are not outliers, show the clusters detected.

Third, show your detailed steps of DBSCAN process, including the content of the queue you maintain, whenever a new core is found.



Q4 [20 Marks] Fuzzy Cluster

Assume there are 2 clusters in which the data is to be divided, initializing the data point randomly. Each data point lies in both clusters with some membership value which can be assumed anything in the initial state.

The table below represents the values of the data points along with their membership (gamma) in each cluster.

Cluster	(1,3)	(2,5)	(4,8)	(7,9)	(9,12)
1)	0.8	0.7	0.5	0.3	0.1
2)	0.2	0.3	0.5	0.7	0.9

Please work out the centroids, the distance of each point from centroid, and the cluster membership value.