

School of Information Technology and Electrical Engineering
INFS4203/7203 – Data Mining

Tutorial 5-6 Clustering

Question 1.

Suppose that the task is to cluster the following eight points (with (x,y) representing location) into three clusters.

A1: (2, 10), A2: (2, 5), A3: (8, 4),
B1: (5, 8), B2: (7, 5), B3: (6, 4),
C1: (1, 2), C2: (4, 9)

The distance function is Euclidean distance. Suppose initially we assign A1, B2, and C1 as the center of each cluster, respectively. Use the *k-means* algorithm to show:

- a) The three cluster centres after the first round execution, and
- b) The final three clusters.

Question 2. (Read Chapter 8.3 of the Textbook, pp515-526)

Consider a set of six objects: {A, B, C, D, E, F}. Let the following be a dissimilarity matrix between these objects.

	A	B	C	D	E	F
A	0.0	1.0	5.0	9.0	10.0	2.0
B	1.0	0.0	3.5	8.0	7.0	5.5
C	5.0	3.5	0.0	3.0	4.0	6.5
D	9.0	8.0	3.0	0.0	0.5	4.5
E	10.0	7.0	4.0	0.5	0.0	2.5
F	2.0	5.5	6.5	4.5	2.5	0.0

Construct agglomerative clustering hierarchies for these objects using

- a) The single linkage,
- b) The complete linkage, and
- c) The average linkage method.

Are the constructed hierarchies similar or different to each other?
Are some of them more reasonable than the others?

Question 3. (From the Textbook Exercise 8-3)

Many partitional clustering algorithms that automatically determine the number of clusters claim that this is an advantage. List two situations in which this is not the case.

Question 4. (From the Textbook Exercise 8-32)

In Figure 8.9, match the similarity matrices, which are sorted according to cluster labels, with the sets of points. Differences in shading and marker shape distinguish between clusters, and each set of points contains 100 points and three clusters. In the set of points labelled 2, there are three very tight, equal-sized clusters.

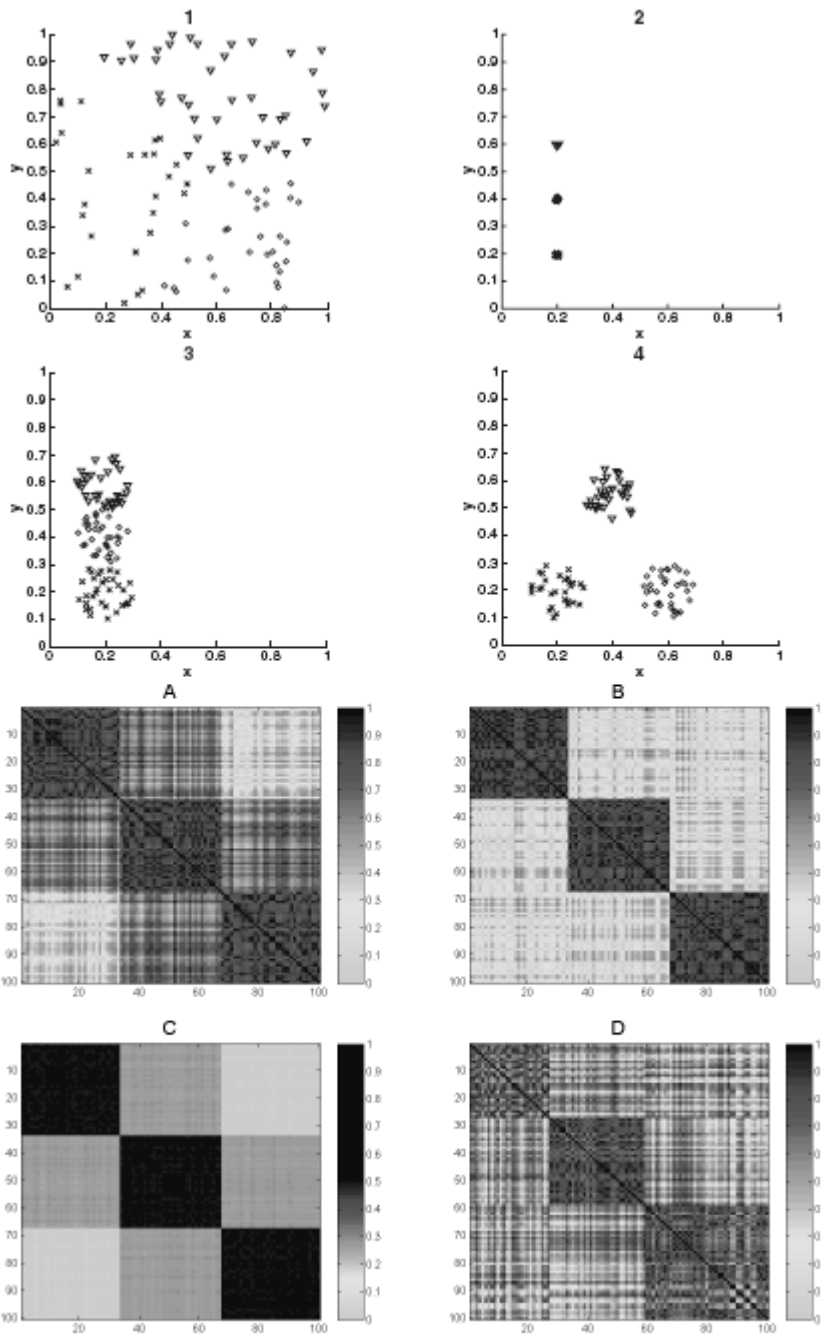


Figure 8.9. Points and similarity matrices for Exercise 32.