CSE597 Fall 2017 Assignment 4

Assigned: Monday, November 06, 2017

Due: Monday, November 13, 2017 (Typed hardcopy at the beginning of class)

Maximum: 100 point

Note: This assignment is to be done by an individual student, no team work allowed.

1. (25%) Perform K-means on the data shown in Table 1. Use K=2 and suppose A and C are randomly selected as initial means. Please show the clustering results of nodes of each round until converged.

Table 1

i	X_1	X_2
A	1	1
B	1	0
C	0	2
D	2	4
E	3	5

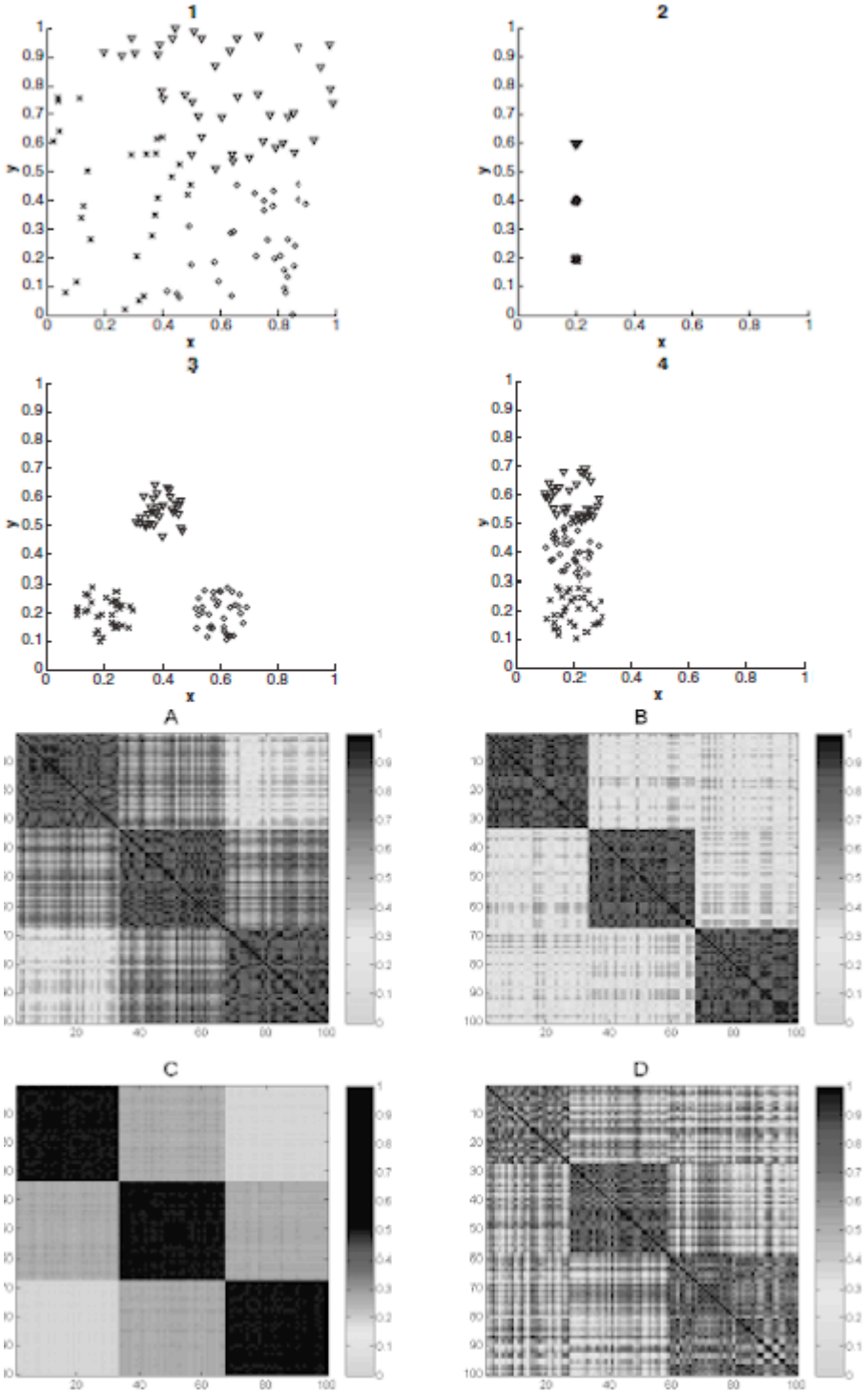
2. (25%) Use the similarity matrix in Table 2 to perform MIN and MAX similarity hierarchical clustering (i.e., MIN: the similarity of two groups of nodes is the MIN similarity among all possible pairs of nodes. MAX: the similarity of two groups of nodes is the MAX similarity among all possible pairs of nodes). Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged.

Table 2

	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~~

3. (25%) In Figure 1, 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 labeled 2, there are three very tight, equal-sized clusters.

Figure 1



4. (25%) One way to sparsify a proximity matrix is the following: For each object (row in the matrix), set all entries to 0 except for those corresponding to the objects k -nearest neighbors. However, the sparsified proximity matrix is typically not symmetric.
    a. If object a is among the k -nearest neighbors of object b , why is b not guaranteed to be among the k -nearest neighbors of a ?
    b. Suggest at least two approaches that could be used to make the sparsified proximity matrix symmetric.