

Assignment 5

CIS 453/553 Data Mining, Fall 2017

due 11:59 pm, May 19th, Friday (Hard Deadline)

1. (20 points) Suppose that the data mining task is to cluster the following nine points (with (x,y) representing location) into three clusters:

$$A_1(3,10), A_2(4,6), A_3(9,5), B_1(3,8), B_2(8,5), B_3(6,6), C_1(2,3), C_2(5,7), C_3(6,8)$$

Suppose initially we assign A_1 , B_1 and C_1 as the center of each cluster, respectively. Please use *k-means* algorithm and square-error criterion to show

- (a) The three cluster centers after the first round execution.
- (b) The final three clusters. You should write a small program to do that.

2. (10 points) Please compare the strengths and weaknesses of *k-means* and *k-medoids* algorithms. Also compare them with a hierarchical clustering scheme (such as AGNES or DIANA.)

3. (20 points) Use density-based clustering to process the same nine data points of Problem 1. Let $\varepsilon = 3$ and $MinPts = 3$, show the clustering results.

To turn in by paper version: Ask Cheri to put your answers to Prof. Dejing Dou's mailbox.

To turn in by emails: We prefer that you send in a pdf file. If you are using Word, you should be able to convert your word file to a pdf file.