

Assignment 5 (50 points only)
CIS 670 Data Science, Winter 2019
due 11:59 pm, Monday, March 4th

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

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

Suppose initially we assign A_1 , B_1 and C_1 as the center of each cluster, respectively. Please add a Map-reduce function for the K-means algorithm. Show the results for the first two iterations and explain how Map-reduce can help.

2. A database has six transactions. Let $\text{min_sup} = 50\%$.

TID	items_sold
T001	A, B, C, D, E, F
T002	B, H, E, C, F, T
T003	C, U, O, E, W, D
T004	W, A, B, C, F, X
T005	W, X, C, D, F, Y
T006	B, C, D, E, O, Z

Please add a Map-reduce function for the Apriori algorithm to generate all frequent itemsets. Show the results for each step and explain how Map-reduce can help.

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

To turn in by emails: Email your answers to dou@cs.uoregon.edu. 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.