**CS 6350 – Big Data Analytics and Management**

**Fall 2014**

**Assignment# 4**

**Due: 11/17/2014 by 11.55 p.m. (Part I)**

**12/01/2014 by 11.55 p.m. (Part II)**

**Part I:**

**Question A**

Consider the following eight points in 2-dimensional space: (2,10); (2,5); (8,4); (5,8); (7,5); (6,4); (1,2); (4,9). Suppose we plan to use the Euclidean distance metric and that we are interested in clustering these points into 3 clusters.

(i).Plot the data to see what might be appropriate clusters.

(ii) Beginning with the points (2,10), (5,8) and (1,2) as initial cluster centers, form the three initial clusters.

(iii) Use the 3-means clustering algorithm to get the final three clusters. What are the resulting centers and resulting clusters?

**B Clustering**

Please do the followings from **Chapter 8 (Page 563-567: Tan, Steinbach, and Kumar’s book):**

* Problem#16
* Problem#17
* Problem# 23

**Part II:**

C. **Classification**

Please do the followings from **Chapter 5 (Page 318-322: Tan, Steinbach, and Kumar’s book):**

* Problem#7
* Problem#8

D. Please see the following problem:

