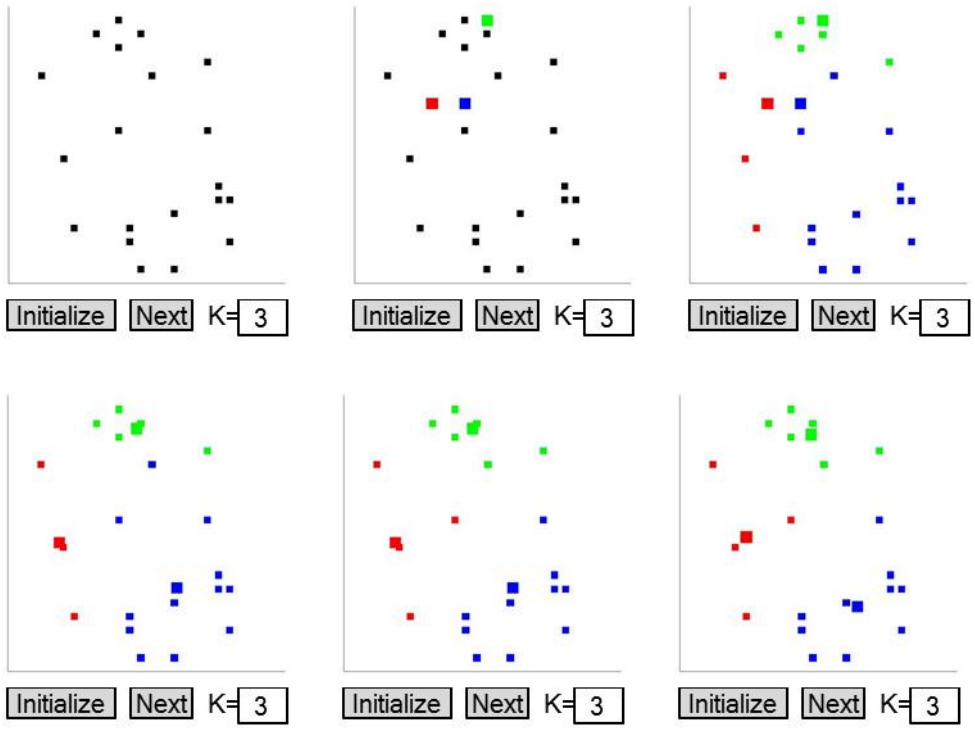
COMP 4420 Introduction to Data Mining
Project#2 –k-Means Animation
Due Date: 10:00 AM, 05/06/16

Overview:

For this project, you will develop a program in Java to animate the *k*-means clustering algorithm. You will assume a two-dimensional space to represent the data points, i.e., each data point will be made up of two values (X-axis and Y-axis). The program will generate 20 random data points to begin with, each with an X-axis and a Y-axis value. You will assume a 20x20 grid where the values along each dimension for a data point has to be between 1 and 20 (inclusive). Make sure that the same data point is not generated twice. Once the data points are generated, the program should display them in black color. Then depending on the number of clusters specified by the user, the program will allow the user to step through the clustering process. Each iteration should consist of assigning each data point to its nearest cluster and recalculating each cluster mean. The mean and the points in each cluster need to be displayed with a unique color. The user can specify anywhere between 2-5 clusters. At the beginning, all the cluster means should be randomly initialized. Each cluster mean should be displayed a little bigger than the individual data points. The following sequence shows the tentative screenshots of the program for some randomly generated data points.



Deliverables:

You will need to turn in your Java source file(s) Blackboard by May 6th.