**ECE 595 Homework 8 Due: Nov. 26, 4 PM**

Consider the unlabeled dataset ***X*** from the attached file, *KmeansDat.mat*, that you use to apply K-means algorithm. Plot the data and observe that there are three distinct clusters that form the data in . Using K = 3, develop a MATLAB code to arrive at the three cluster centroids with initial centroids of . From these initial centroids, iteratively find the cluster assignment for each training example  , *i = 1, 2, …, m* by finding the minimum squared Euclidean distance (squared norm) between each example and the three centroids. That is, for each example , its cluster assignment  is given by

 such that is the minimum, where  is the index of the cluster centroid closest to  and  is the *jth* centroid.

After the cluster assignment to all *m* examples, recalculate the centroids based on the current clusters as

, where is the set of examples that are assigned to centroid *k*.

If, for example, , that is, are assigned to centroid *k = 2*, then the updated centroid is given by



You may terminate the iterative process when D, the sum of all the distances in each cluster from the examples that are assigned to that cluster, is a minimum (choose, for example, D = 0.05).  (a) Plot the cost function D after each epoch – cluster assignment and centroid modification – as a function of epoch. If the error does not converge after 500 epochs, try raising D to 0.5 or so. What are the final centroids? (b) Plot the final centroid locations along with the data points. How many iterations does it take to converge? (c) Modify and run your code for K = 4 clusters and show the plots as in (a) and (b) using a random set of four values from the dataset for initial cluster centroids. What are the final centroids?