Eric Zounes, Ian Fridge

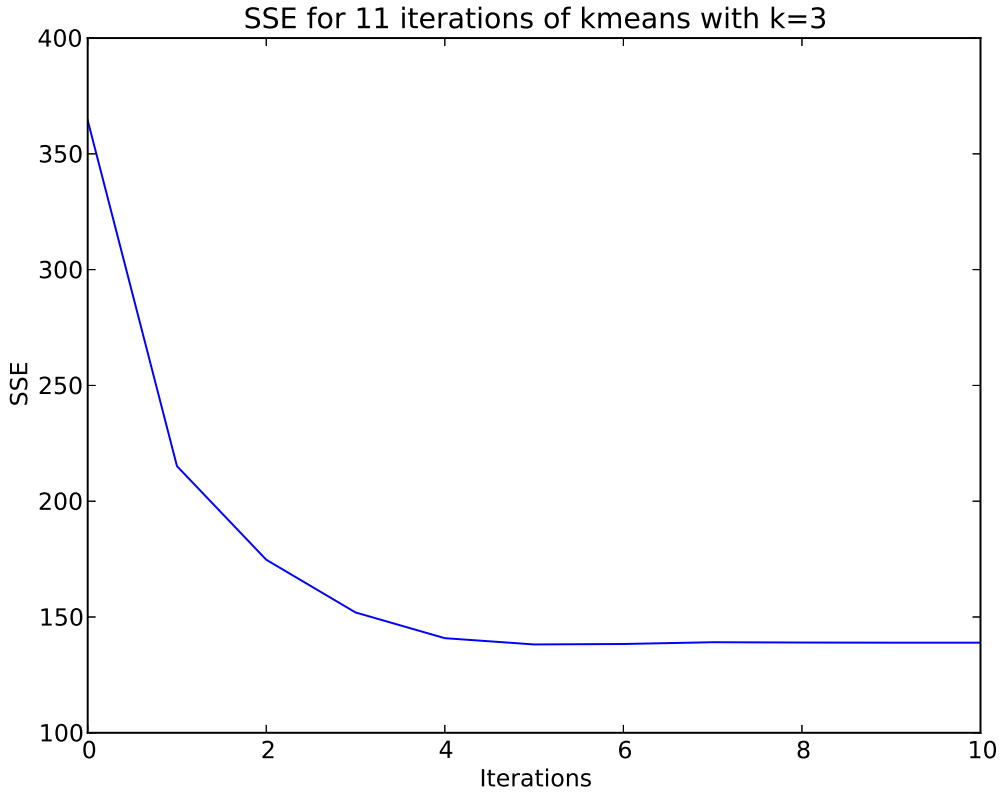
November 2, 2012

CS434: Assignment 3

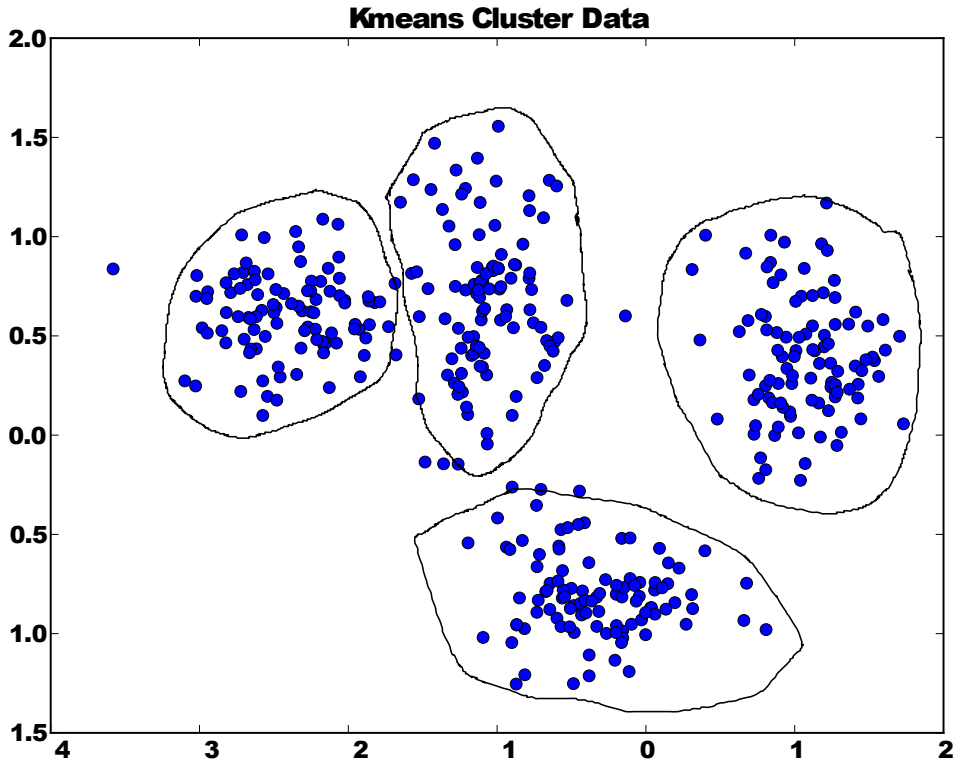
1 Implementation Assignment

1. Implementing K-Means

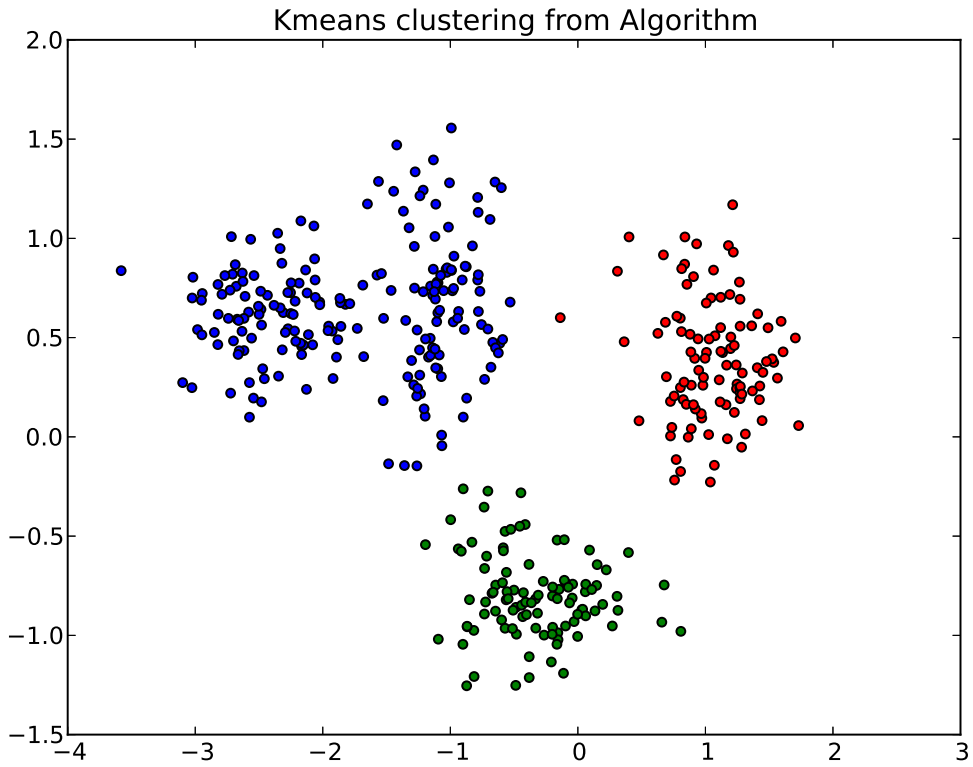
   (a) Run your kmeans algorithm with $k = 3$. To verify that your algorithmn actually converges, please plot the objective of the Kmeans algorithm (SSE) as a function of the iterations. From one run to another run, this curve may look different. Just present the results of a typical run.



(b) Plot the scatter plot of the given data, and inspect the scatter plot visually. How many clusters do you see in this data?



I see 4 clusters.



The above graph is the kmeans clustering determined by our algorithm. Each color is associated with a cluster.

(c) Now apply your kmeans implementaion to this dat with differnet values of k $(2, 3, \ldots, 6$. For each value of k, please run your algorithm 10 times, each time with a different random initialization, record the lowest SSE value achieved in these 10 repeats for each value of k. Plot the recorded SSE values against the changing k value. Does the curve confirm your belief about the k value? Why or why not?

It shows that as we increase the value of k we have less and less SSE. This is because the clusters are smaller and smaller and thus the error in each is decreasing

in size. The large clusters will have a larger average distance from the center of the cluster. Each point is the the minimum sum of the SSE for each iteartion for each value of k. We ran it 10 times at each value of k and summed the SSE of each iteration and plotted that value on the y axis with the value of k on the x axis.