STA 5106: Homework Assignment #7

(Thursday, October 24)
Due: Thursday, October 31

1. Write a Matlab program to implement K-means clustering algorithm. Download two datasets from the class website and test your program on them with k = 5. As part of your results: (i) display all observations along with their labels in a plot, (ii) plot the evolution of the sum of squared error versus the iteration index.

2. Use the above Matlab program to perform K-means clustering on the two datasets given in Problem 2 of Homework 6, for k = 2. Display your results as earlier. Estimate the mean, variance, and proportion for each class.

Recall that you have already used EM algorithm on these two datasets (in Homework 6) to estimate means μ, variances σ^2, and proportions α. Compare the results obtained by EM algorithm and K-means clustering.

3 (Optional). Use Python program to finish Problem 1.