EE 511 Simulation Methods for Stochastic Systems
Project #3: Clusters and Mixtures

Problem 1 [Testing Faith]
Download the "old faithful" data set from blackboard. This contains samples of a 2-D random variable: the first dimension is the duration of the old faithful geyser eruptions. The second is the waiting time between eruptions. Generate a 2-D scatter plot of the data. Run a k-means clustering routine on the data for k=2. Show the two clusters in a scatterplot.

Problem 2 [Generating Mixed Samples]
Implement a random number generator for a random variable with the following mixture distribution:

$$f(x) = 0.4N(-1,1) + 0.6N(1,1)$$

Generate N=1000 samples and histogram them. Try out a k-means clustering routine (k=2) on the data.

Turn in:

- A summary of your experiments including any relevant plots
- A brief discussions of the results
- Your code.