**Data Mining (CS F415 / CS C415)**

**Assignment 2**

**Submission Date: April 25th**

Find a dataset that interests you at the [UCI Machine Learning Repository](http://www.ics.uci.edu/~mlearn/MLRepository.html). Make sure that it is a dataset that has a number of numeric features, since we will just use the numeric ones for this exercise. If the dataset already has a classification, make sure to ignore it and leave it out.

Create or make use of software that uses the k-means clustering algorithm to find clusters in the data. You should try and vary the number of clusters, and choose an appropriate number. Use two different techniques for picking initial seeds, and determine how it affects your results. Turn in on paper an explanation of the methodologies that you used, the results that you found, and a clear description of the final clusters that you discovered.