Data Mining: Assignment 2

Clayton Mercer

I have written a python script to read raw data from a csv file pertaining to students. It normalizes their ages and gpas and makes a secondary csv file with the normalized data. From that point it reads that new data and runs a k-means algorithm to cluster the information therein. The output from the finished product will return 3 files being a report on the cluster membership and the centroid tied to each and every point, a SSE text file that shows the error for each cluster, and a data plot in the form of a pdf. It is also dynamic and will accept input from the user as to how many clusters the data should be divided into. All in all it took a great deal of time and I think it turned out wonderfully.