Bryan Anders
CSCD 429
Data Mining
Dr. Li

HW 4 Report

1) Running the program.

The directory in which this report resides (“hw4”) is also the top level directory for an eclipse project which contains all code needed to perform k-means clustering on the given data set. The project requires no runtime parameters.

2) Results.

Results are in a file named hw4_results.txt

3) Java generated images.

This project completes the extra credit by drawing the results of the k-means clustering in new images. Those images reside in the images directory. If an image is empty it means that it represents an empty cluster.

4) Distribution among clusters.

This implementation does not handle empty clusters. Initial centroids are randomly selected but there is no guarantee that all clusters will have members. I attempted to find a set of initial centroids that would prevent empty clusters but it was only a limited and ultimately unsuccessful effort.

5) RapidMiner Images

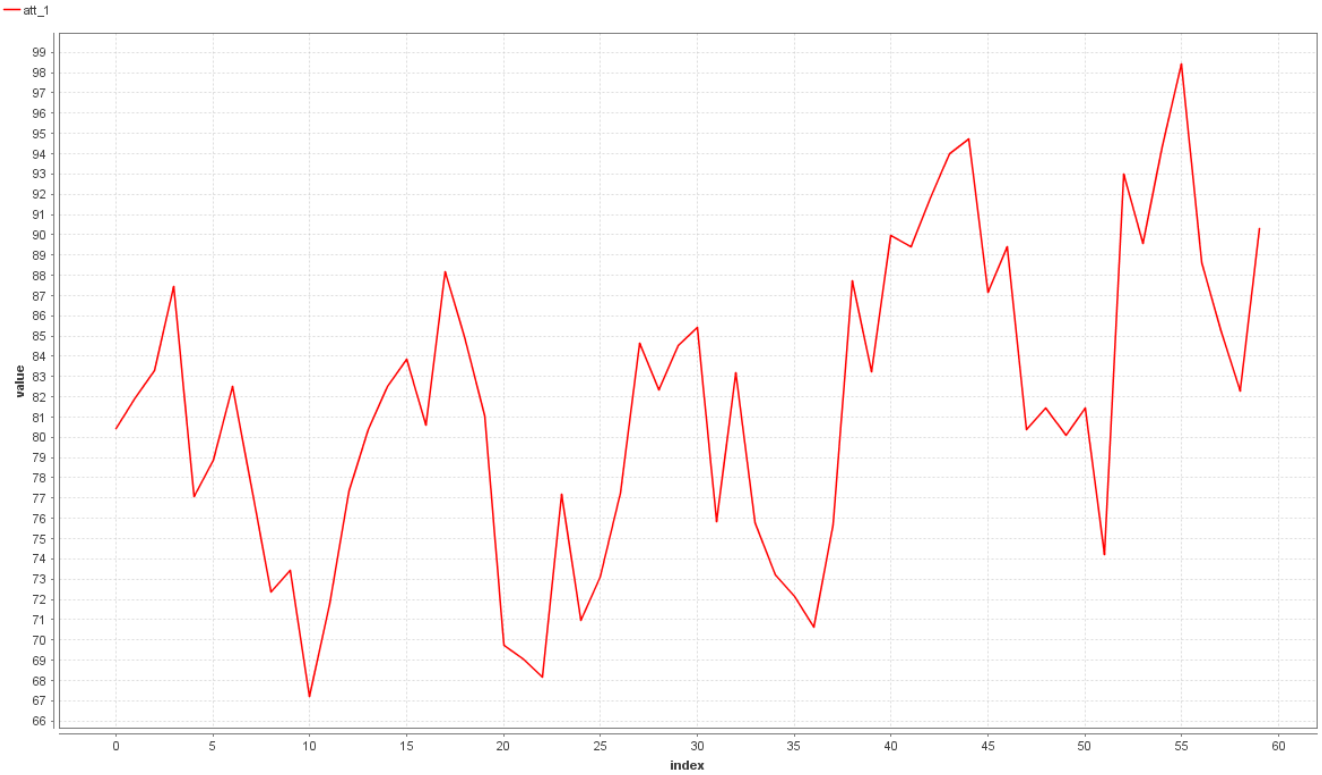
These images are from a sample run through and are not indicative of the exact results if the program was run again.

Cluster0: Empty

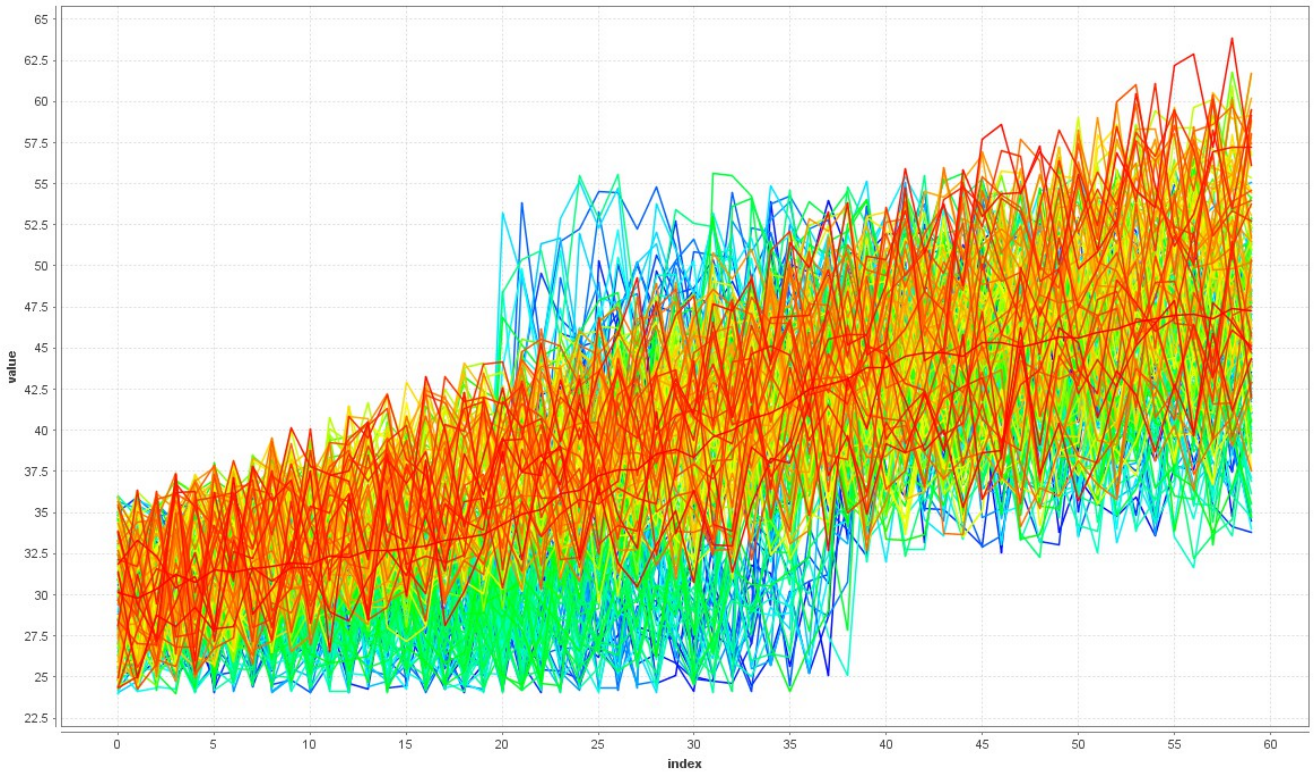
Cluster1: Empty

Cluster2: Empty

Cluster3:



Cluster4:



Cluster5: