James May

Garrett Bauer

Peter Dorich

Assignment 4 Report

K-Mean

|  |
| --- |
| **Figure 1:** SSE is plotted as a function of iterations, where k = 2, and the number of points in each cluster stops changing at the 12th iteration. |

The sum of squares error (SSE) is shown to converge in figure 1, roughly around iteration 7, but precisely at iteration 12. The algorithm was run with a large range starting seed positions and this was found to have the highest iteration count and therefore chosen to show convergence of the SSE.

|  |
| --- |
| **Figure 2:** SSE is plotted as a function of k. Each k is run 10 times and the lowest final SSE value is plotted for each k. |

As seen in figure 2, the SSE drops dramatically after iteration 6 and appears to start converging at iteration 8. Since we know the SSE will continue to decrease with an increasing k value, we can use the dramatic decrease in slope, after k=8, as an indicator that there are likely only 8 groupings in this data.

Principal Component Analysis

|  |
| --- |
| **Figure 3**: Error Rate is plotted as a function of depth, where blue is training error, and orange is testing error. The data points are recorded in the table below. |