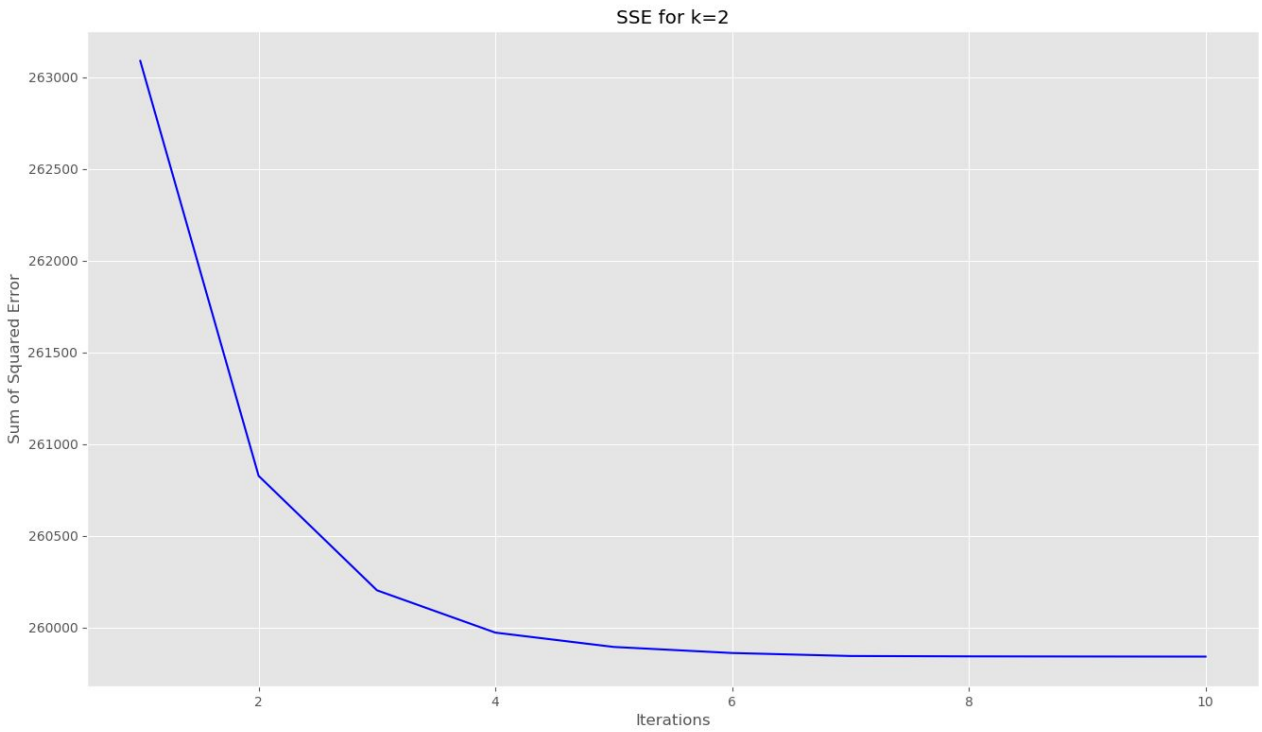
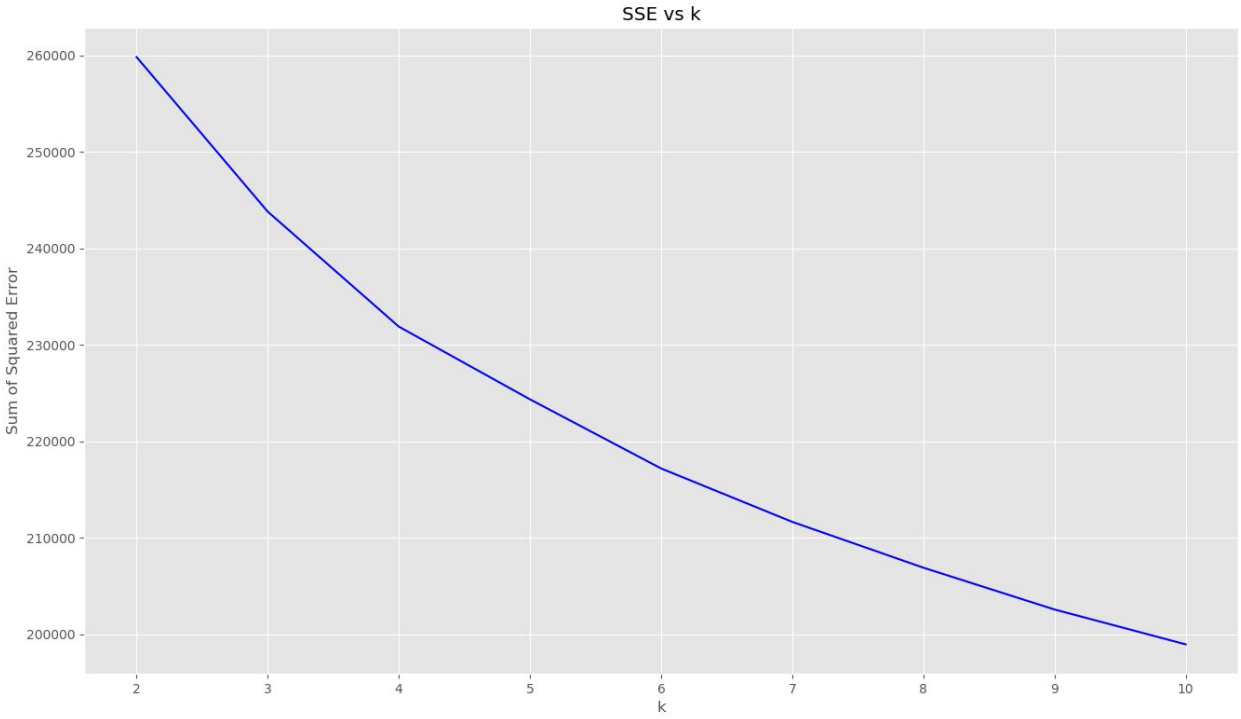
CS 434: Implementation 4 Report

Noah Berks
Jonathan Jones

2.1



2.2



In order to choose a proper k, we'll use the knee heuristic. By visual inspection, it appears that the knee occurs at k = 4. Thus by this heuristic we'll choose 4 as our optimal k value.

Principal Component Analysis

3.1

Top Ten eigenvalues
5.426661918583951
4.119890301738053
3.5006951056377984
2.6867126529514844
2.0067168375258646
1.7768935437414768
1.5336630023286901
1.3929420665662533
1.312211254259663
1.100314740676314

3.2

Mean image:

Eigen images:

The mean image represents the combined similarities between every digit. It appears as some sort of curly 9. Surprisingly it seems that the lower left quadrant of a number is least likely to appear in an image as this area is left almost completely blank in the mean image. Each of the eigen images represents the features with most variance of each digit in the dataset. These images are representations of what defines each number in an image form. It's hard to tell exactly what they mean as this is a machine learned reference image but it can be inferred that they are the most significant tells for each digit in the dataset.

3.3

Selected images by index: [94, 3319, 4466, 3148, 5700, 3508, 4548, 5499, 1858, 5369]

Above are the images that demonstrated the greatest values in each dimension from the 10 dimensionally compressed dataset. The black areas in the eigen images appear to correspond to the white areas in the selected images. Each of the eigen images represents a specific dimension in the 10 dimensional compressed dataset. The selected images showed the highest value for that specific dimension that the eigenvector helped to create. Meaning the features derived from this eigenvector compression resonate most with the selected image in the dataset. The images selected here are the images that represent the digits with most confidence.