CS215 Assignment2 Problem 4

Parth Pujari and Anish Kulkarni

October 2022

1 Principal Component Analysis

The aim is to analyse a data base of 60,000 images of handwritten digits from 1 to 9 and apply Principal Component Analysis on each of the images to give rise to principal or significant modes of variation along certain eigen vectors.

For each number we computed the:

- Mean
- Covariance Matrix
- The principal mode of variation

The mean vector was visualised for the number 2:

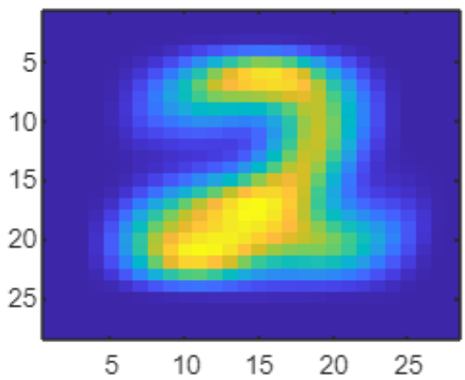

Figure 1: Mean Number

2 Principal Modes

For each number, the principal modes, or the significantly large eigen values were analysed. Graphs showed that only a couple of the eignen values were significantly large.
The significant modes signify the eigen values along which the variance of the data set is the largest.
Given that most of the images in the data set were identical to some extent, the maximum variance was to be expected across only a few nodes, else it would imply lack of similarity between the images labelled say 1. The the plots for the eigen values are below

3 Images around the mean

The images corresponding to the vectors of the mean and that of the mean plus and minus the principal mode give us an idea of how the images vary accross the entire set.
For example, for 1, the mean is slightly tilted to the right, the right image is almost straight while the leftmost image is heavily tilted to the right, which shows that the variance is caused due to the deviation from the slightly tilted right (mean) and that people generally write ones tilted.
Similarly the curve at the bottom of 2 is heavily shown in the image to the right of 2, not so much in the mean and not at all in the image to its left.

The images comparing each digit are shown below.

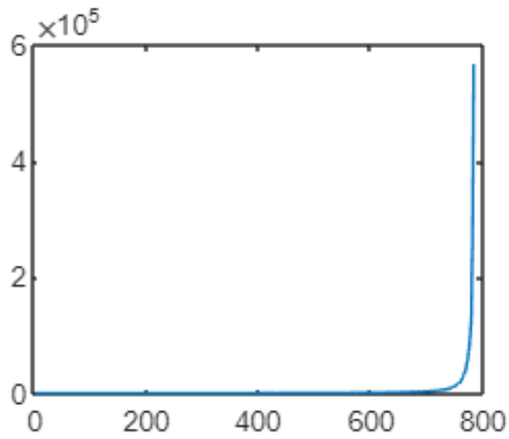

Figure 2: Eigen Plot for 0

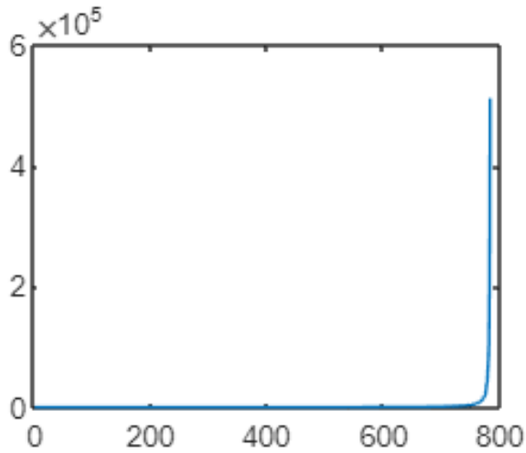

Figure 3: Eigen Plot for 1

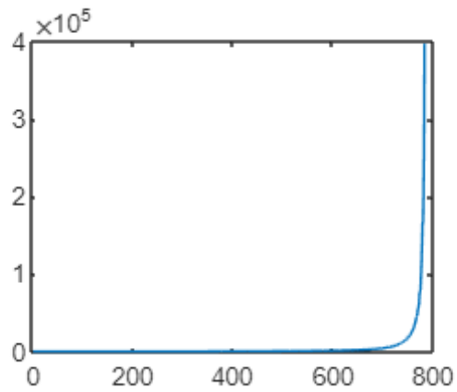

Figure 4: Eigen Plot for 2

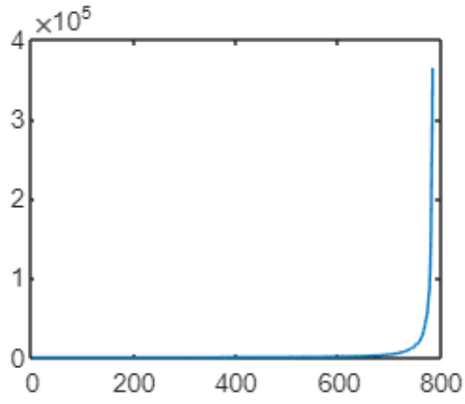

Figure 5: Eigen Plot for 3

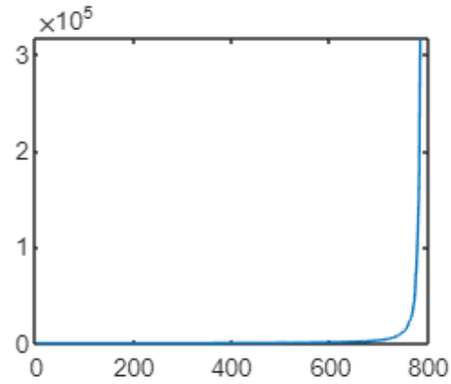

Figure 6: Eigen Plot for 4

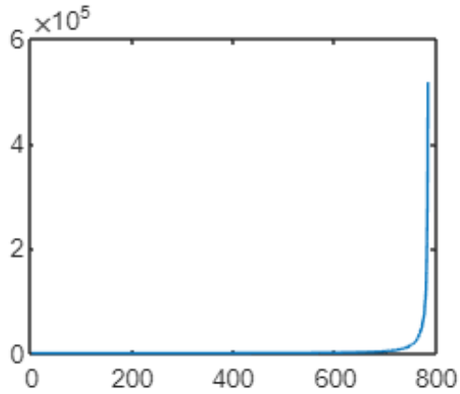

Figure 7: Eigen Plot for 5

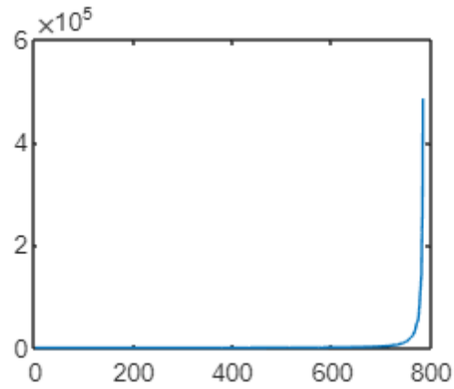

Figure 8: Eigen Plot for 6

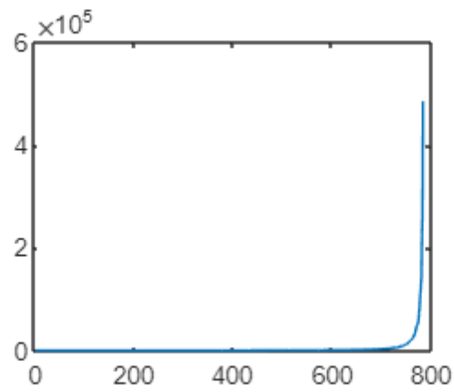

Figure 9: Eigen Plot for 7

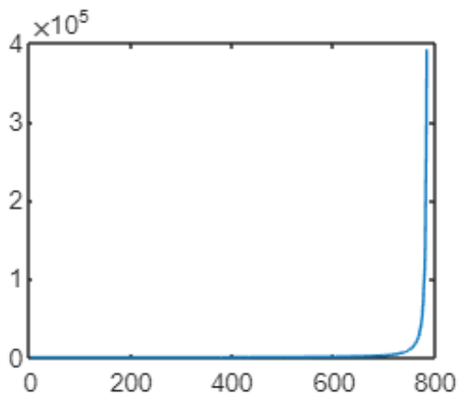

Figure 10: Eigen Plot for 8

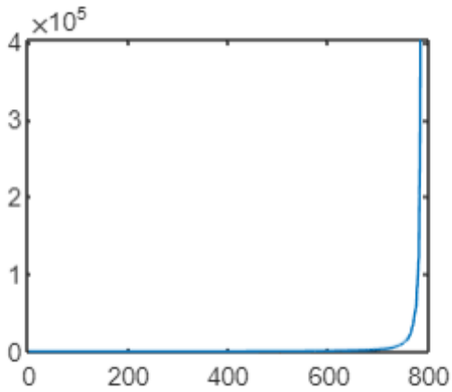

Figure 11: Eigen Plot for 9

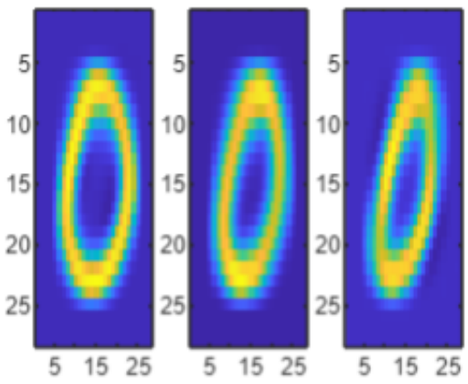

Figure 12: Comparison for 0

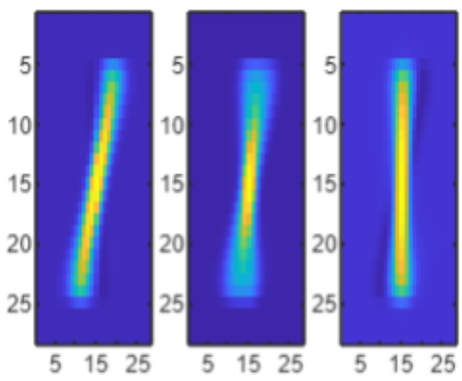

Figure 13: Comparison for 1

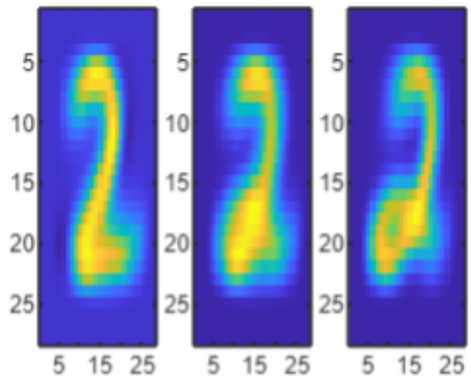

Figure 14: Comparison for 2

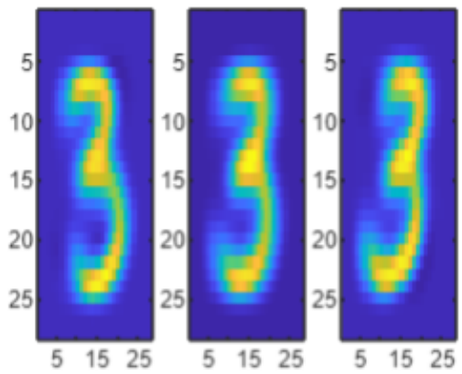

Figure 15: Comparison for 3

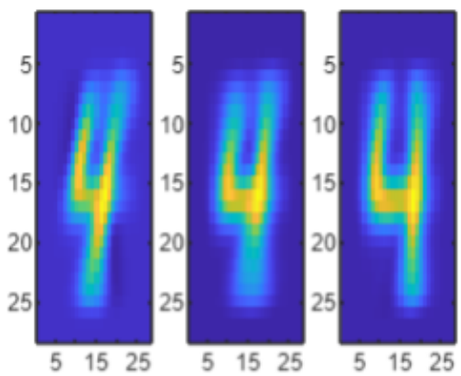

Figure 16: Comparison for 4

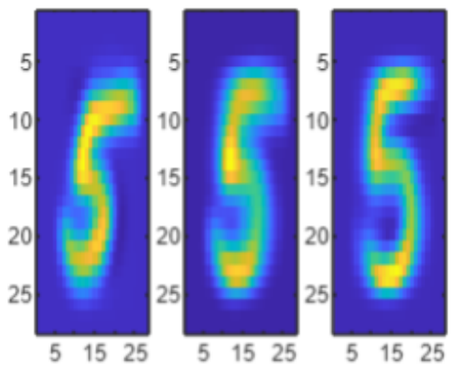

Figure 17: Comparison for 5

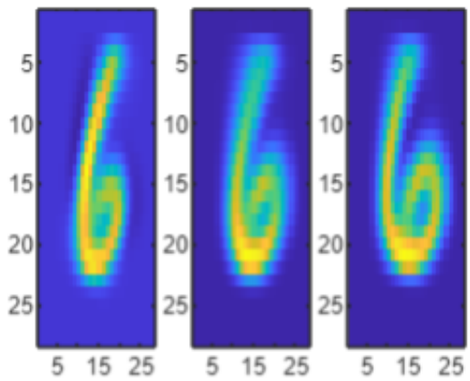

Figure 18: Comparison for 6

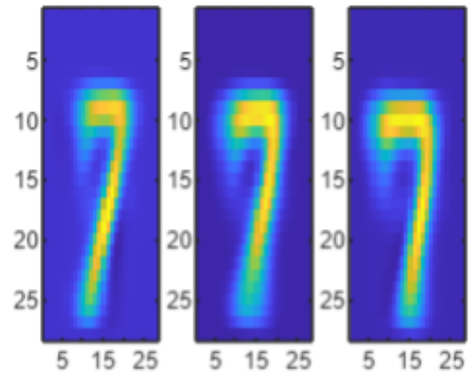

Figure 19: Comparison for 7

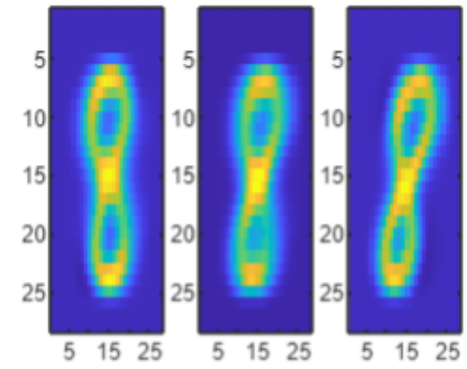

Figure 20: Comparison for 8

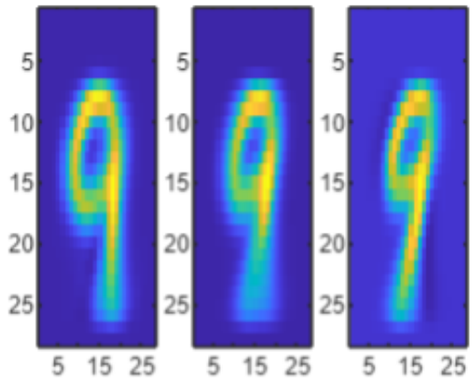

Figure 21: Comparison for 9