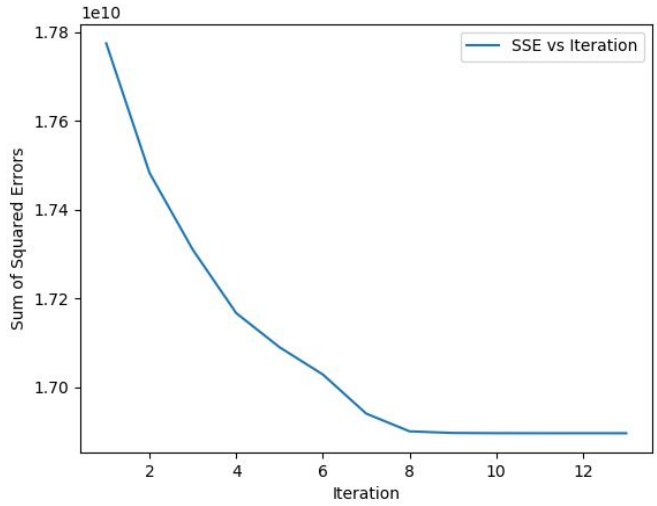
David Sahni
Burton Jaursch
Chase McWhirt
CS 434: Machine Learning and Data Mining
May 28th, 2019
Word Count: 389

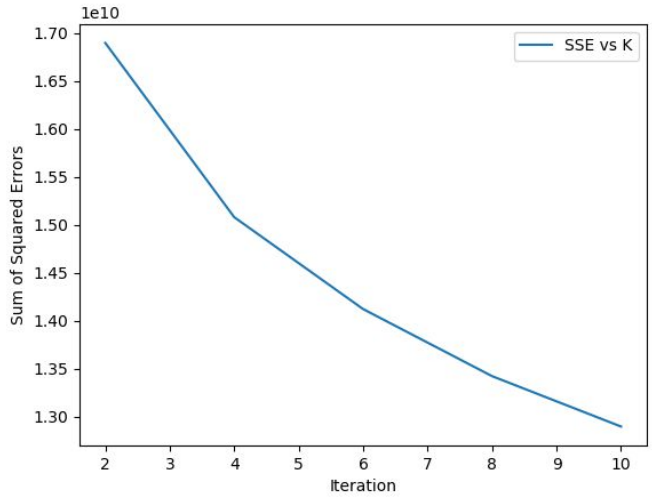
Implementation 4 Write Up

K-Means Clustering

Section 1: SSE vs Iterations w/ k = 2



Section 2: SSE vs K-Value



Note: the X-axis should be labeled as K-Value, not Iteration

While SSE has only decreased as we introduce higher k-values, we believe that the proper k-value should be 10, considering the context is that we are looking at images of 10 different handwritten digits. While increasing the k-value beyond this amount would certainly improve the SSE, we don't believe it would be beneficial for the problem at hand.

Principal Component Analysis

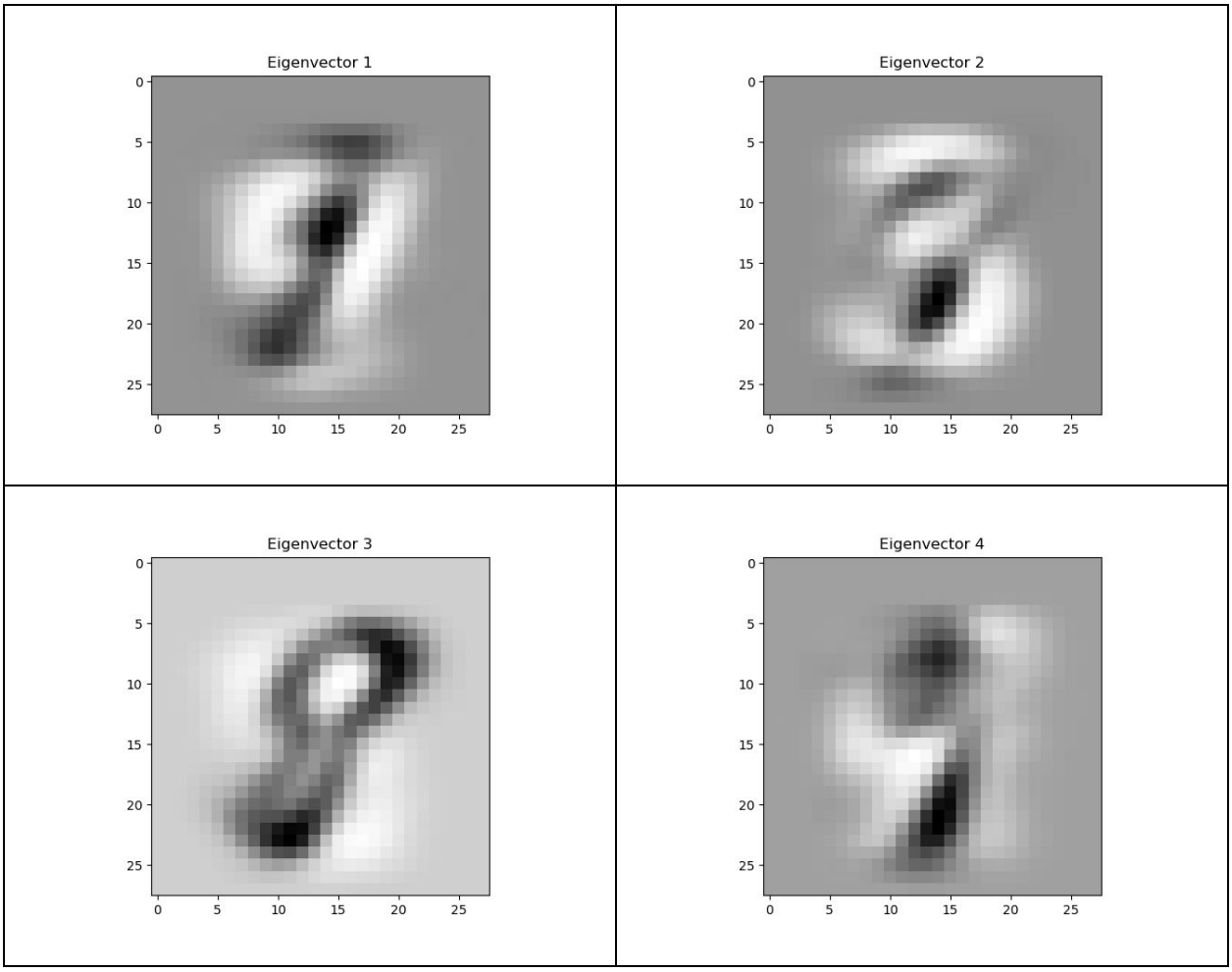
Section 1: Eigen-values

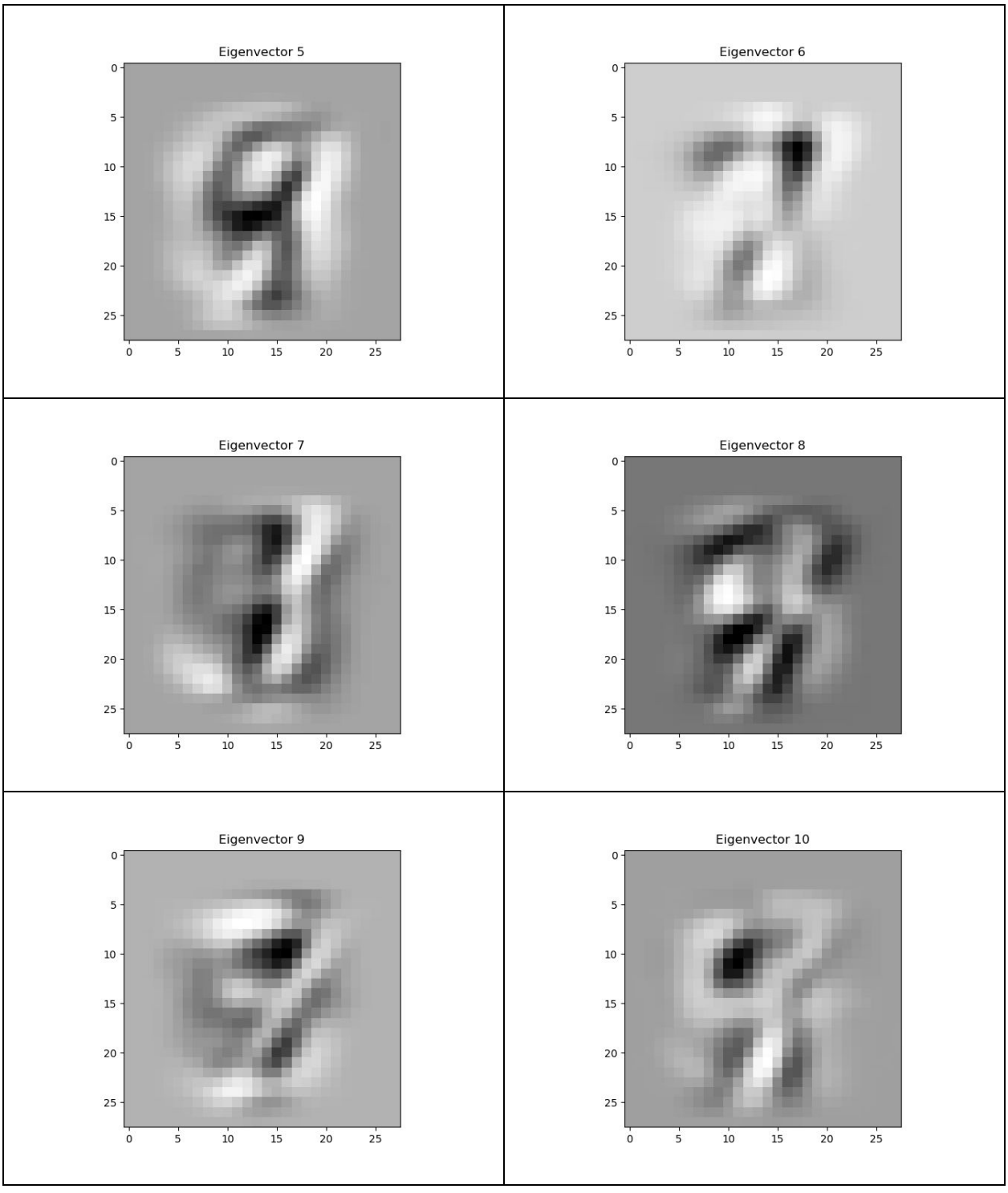
List of the top ten largest eigen-values:

1. 352,868.69
2. 267,895.87
3. 227,632.70
4. 174,703.49
5. 130,486.76
6. 115,542.50
7. 99,726.44
8. 90,576.06
9. 85,326.54
10. 71,547.97

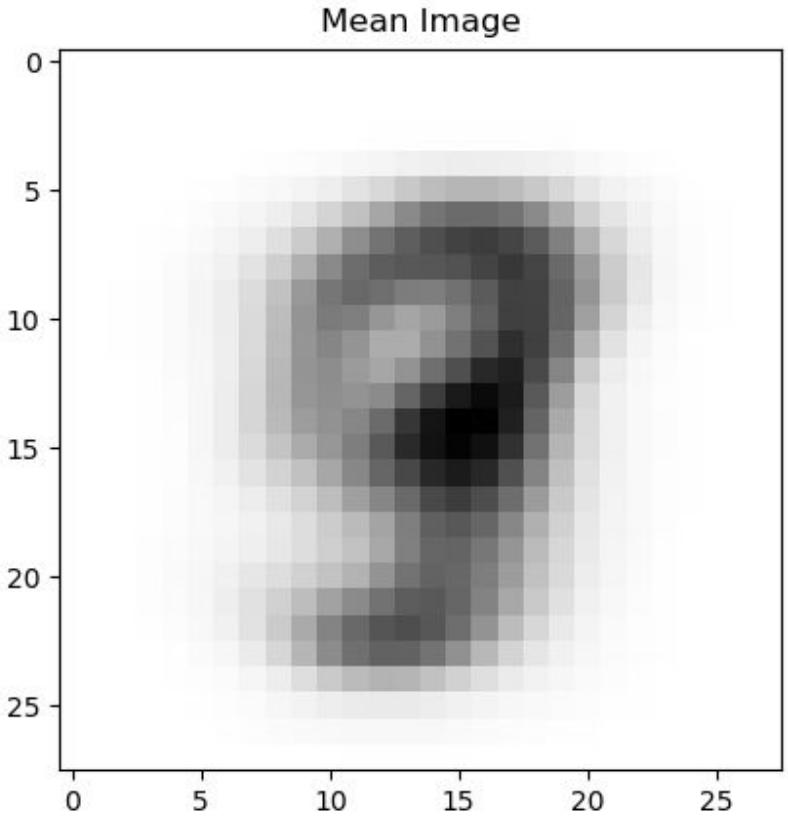
Section 2: Eigen-vectors and Mean Images

Eigen-vectors will be in the order related to the top ten largest eigen-values above.





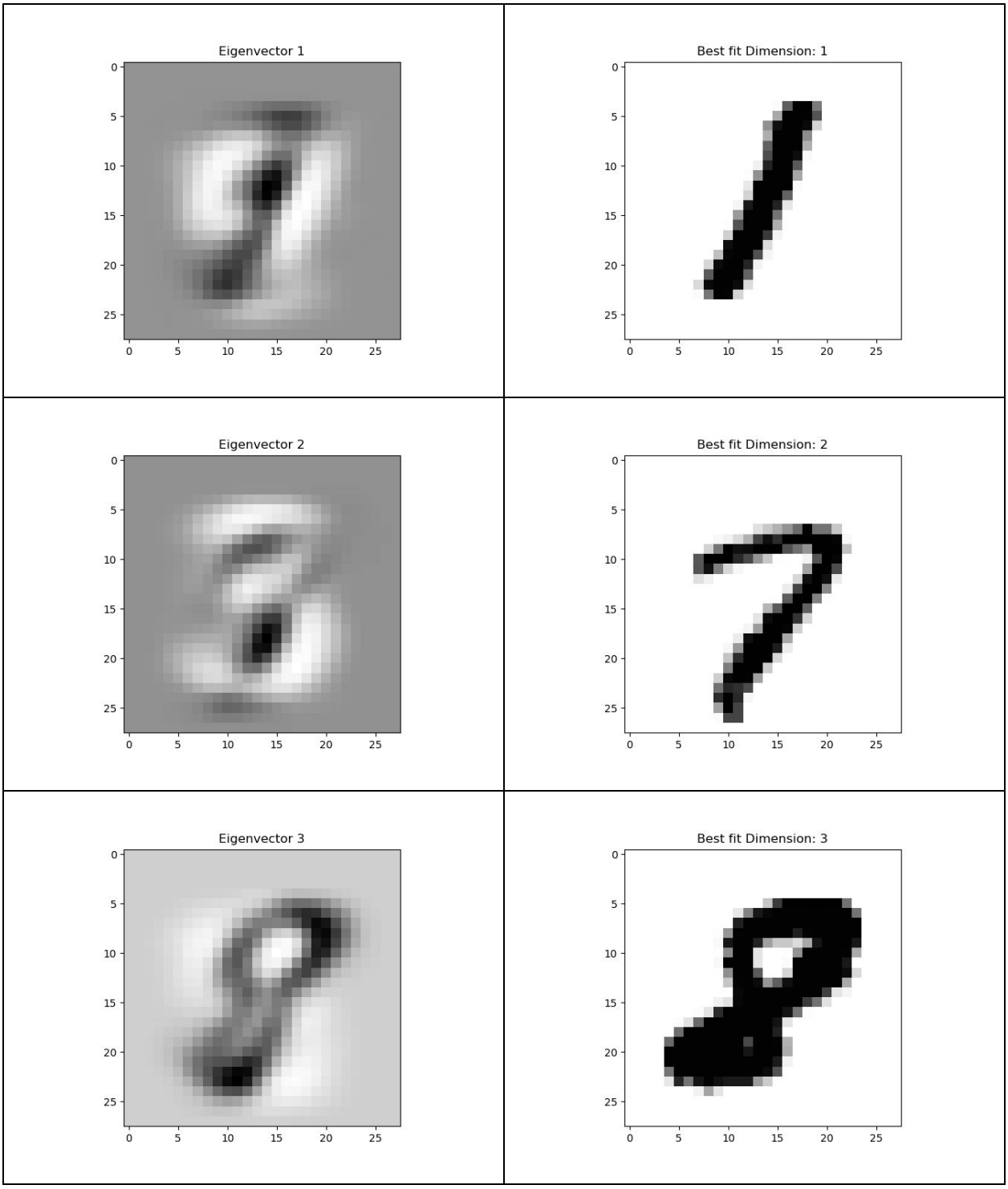
Given these images, some digits are fairly visible. It's strange how some are more visible as the lighter pixels in the plot. Based on the most prominent features, they appear to be a 7, 3, 8, 9, 9, 7, 3, 4, 7, 9.

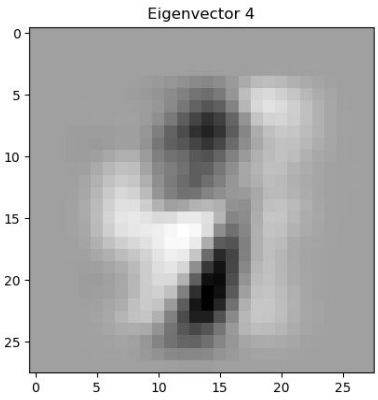


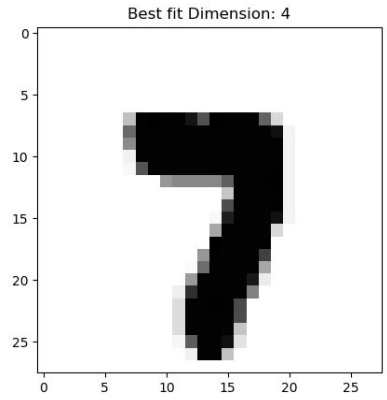
Finally, we can see the mean image of across all of the data. Looks roughly like a lowercase “g” which is fairly close to a 9. Based on the predictions made on the most prominent features, this shape makes sense. It also makes sense since Many numbers share a similar shape to 9 (2, 3, 5, 6, 8, and 0). Those that don’t (1, 4, 7) are in the minority. Though, that being said, they are similar to one another.

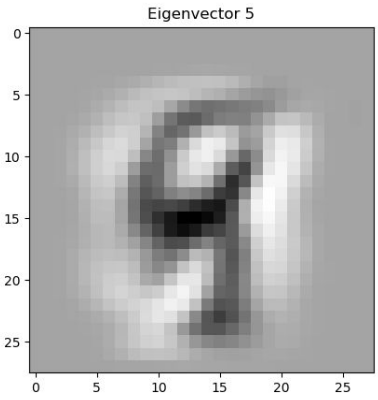
Section 3: 10 Dimensional PCA Plots

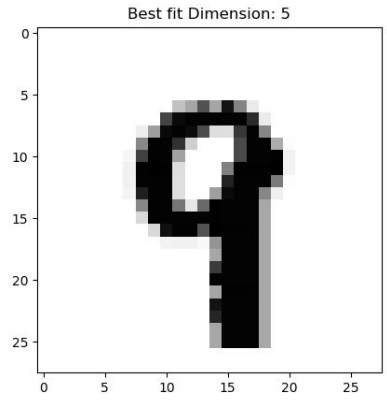
This list will be in the same order as the previous two lists (corresponding to the ten largest eigen-values in descending order). Eigen-vectors will be on the left for comparison.

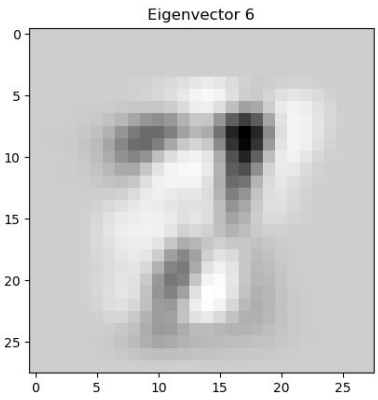


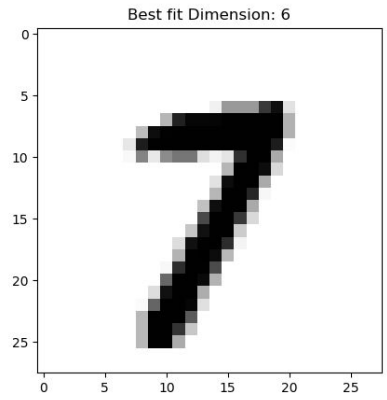

Eigenvector 4
0
5
10
15
20
25
0 5 10 15 20 25

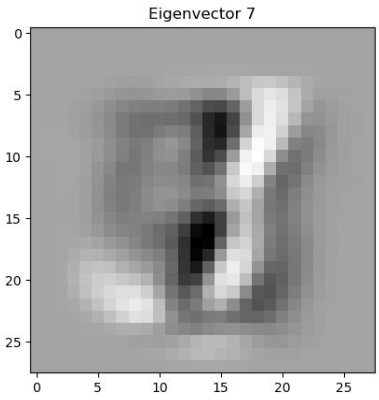

Best fit Dimension: 4
0
5
10
15
20
25
0 5 10 15 20 25

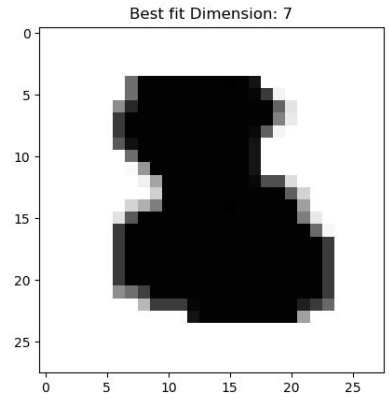

Eigenvector 5
0
5
10
15
20
25
0 5 10 15 20 25

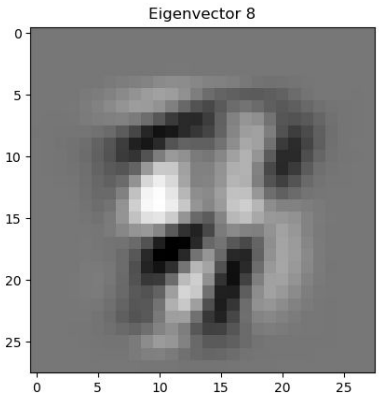

Best fit Dimension: 5
0
5
10
15
20
25
0 5 10 15 20 25

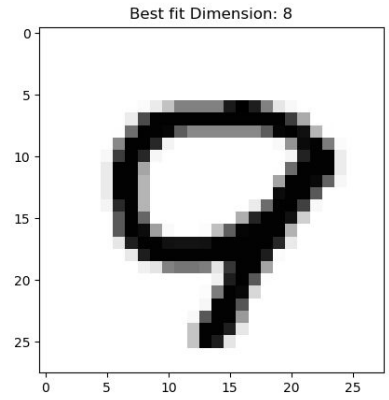

Eigenvector 6
0
5
10
15
20
25
0 5 10 15 20 25

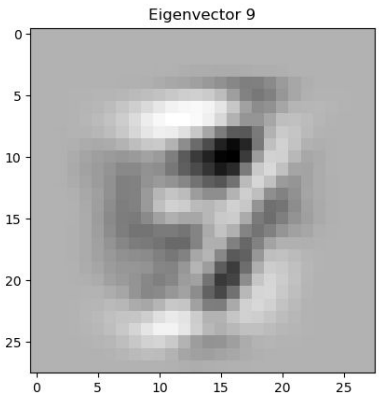

Best fit Dimension: 6
0
5
10
15
20
25
0 5 10 15 20 25

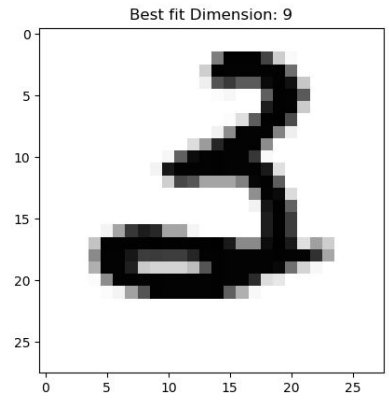

Eigenvector 7

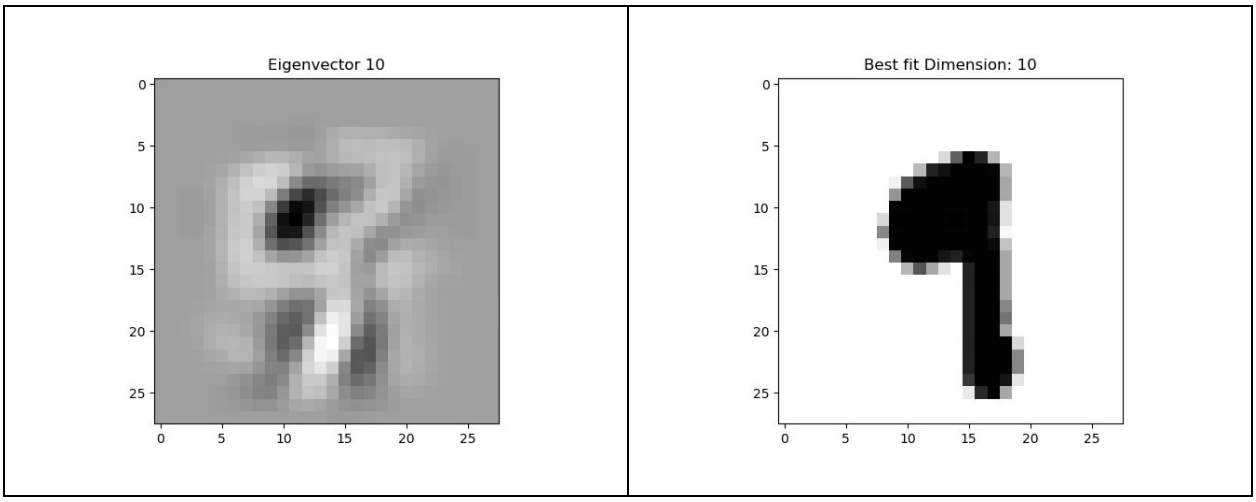

Best fit Dimension: 7


Eigenvector 8


Best fit Dimension: 8


Eigenvector 9


Best fit Dimension: 9



After taking the reduction step, we can see the results are still very convincing in that the PCA reduction images look like hand drawn digits. It appears that PCA reduction takes the darkest portions and uses them to create the hand drawn number. Thus, we end up with a list of 1, 7, 8, 7, 9, 7, 8, 9, 3, 9.