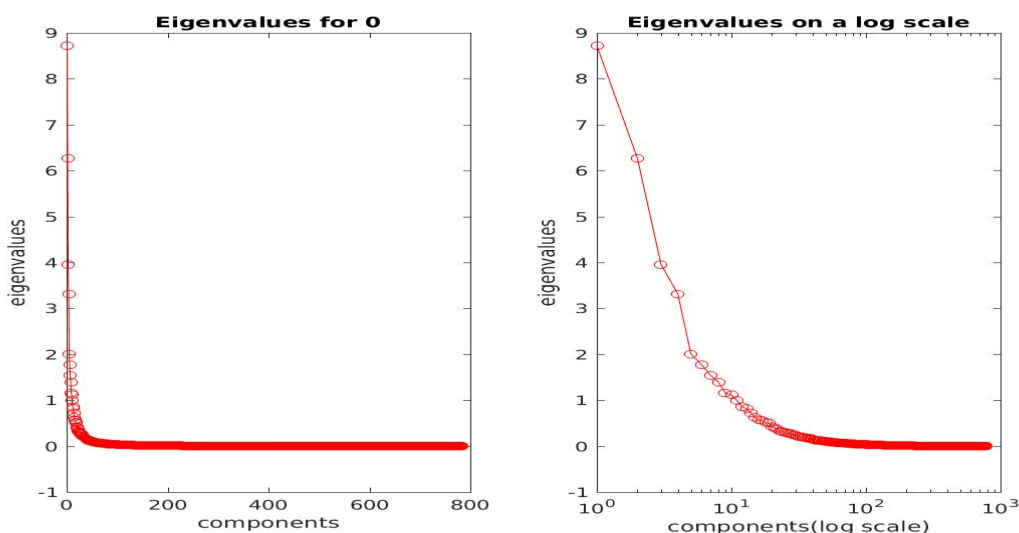# Question 4

4.1 : Part A

For each digit, the Mean Vectors, Covariance Matrices, eigenvalues, and corresponding first Modes of Variations have been calculated and stored in 'results.mat' (present in the results directory). More on it's loading in Sub-Section 4.4 (Usage of code).
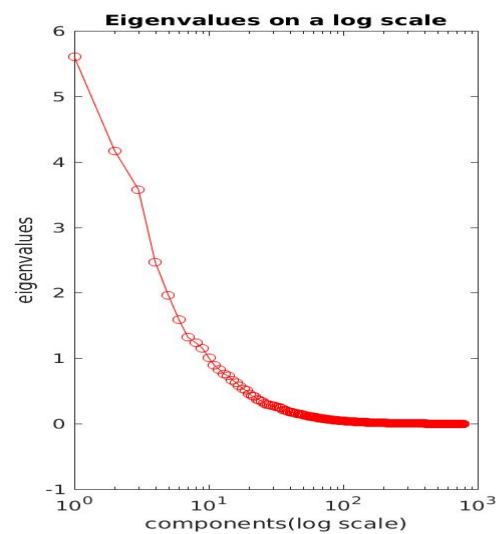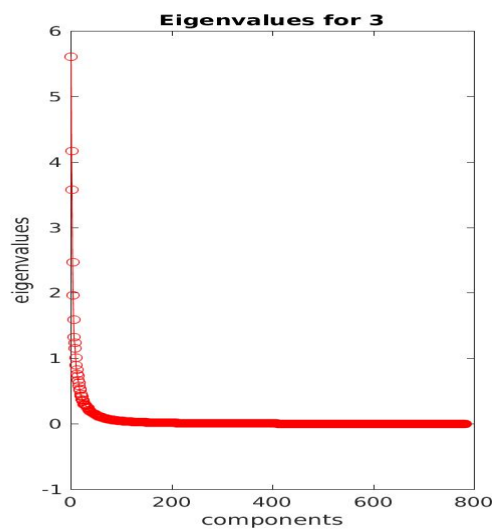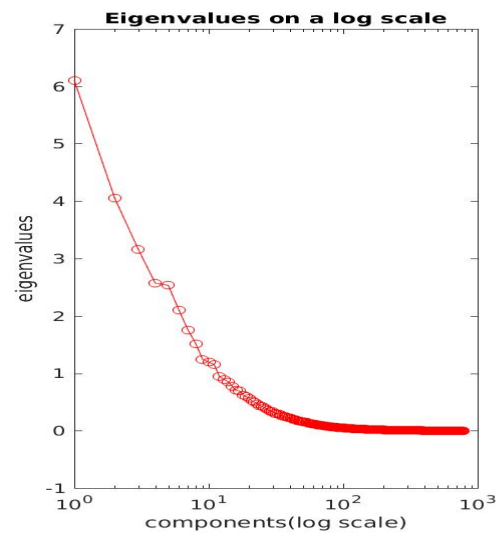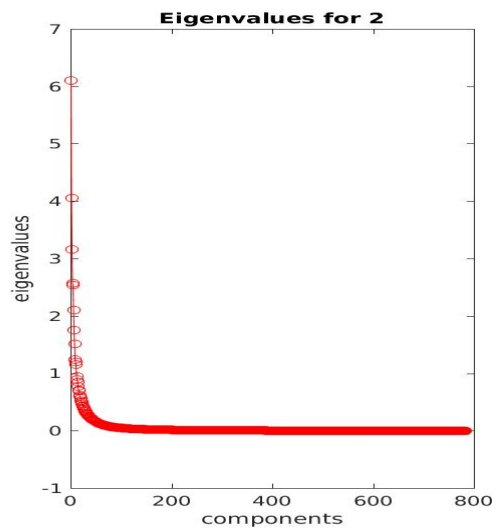
4.2 : Part B

For each digit, the eigenvalues were calculated and plotted vs the component numbers, linearly and on a semilog axis and the following inference was made-

- **Very few prominent eigenvalues** were found for every digit. Not more than 30 values were of a significant magnitude, which is far far less than 784.

- One can justify this by the following argument. Every image has 784 paramenters(or perhaps variables). But **not all 784** of these are **significant**. **Many** of these are **redundant** and don't contribute in distinguishing between different digits.
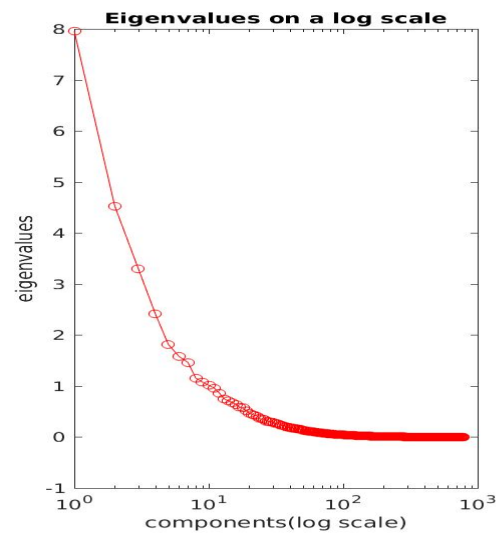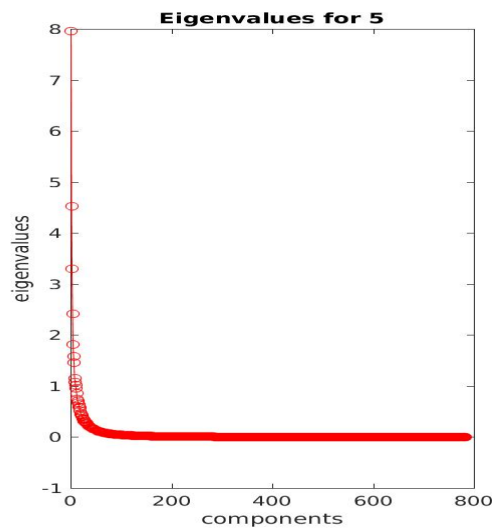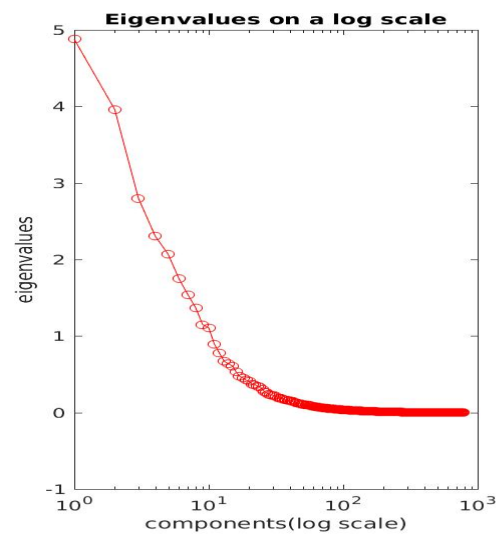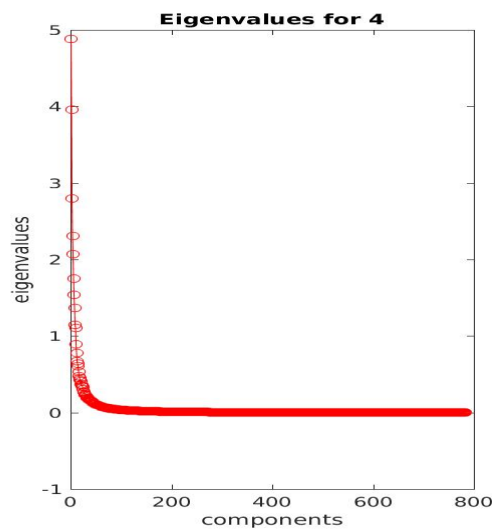
- Since eigenvalues of a matrix correspond to the change in magnitude of input vectors when multiplied with that matrix, we can say that every eigenvector corresponds to a feature. And since we concluded that many features will be redundant, then many eigenvalues too have to be insignificant.

The plots of the eigenvalues follow:

Niraj Mahajan
Raaghav Raaj

Eigenvalues for 4

Eigenvalues on a log scale

Eigenvalues for 5

Eigenvalues on a log scale

Eigenvalues for 6

Eigenvalues on a log scale

Eigenvalues for 7

Eigenvalues on a log scale

Eigenvalues for 8

Eigenvalues on a log scale

Eigenvalues for 9

Eigenvalues on a log scale
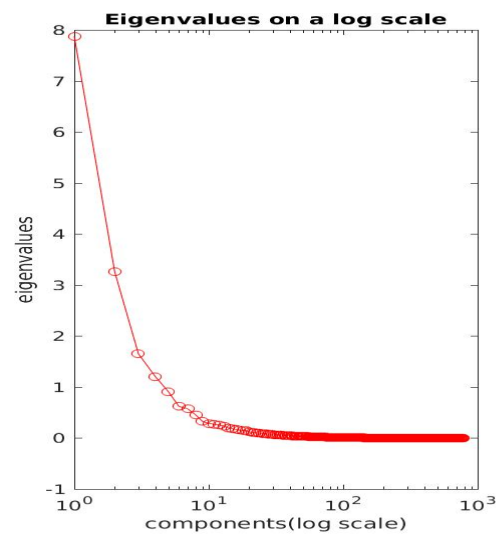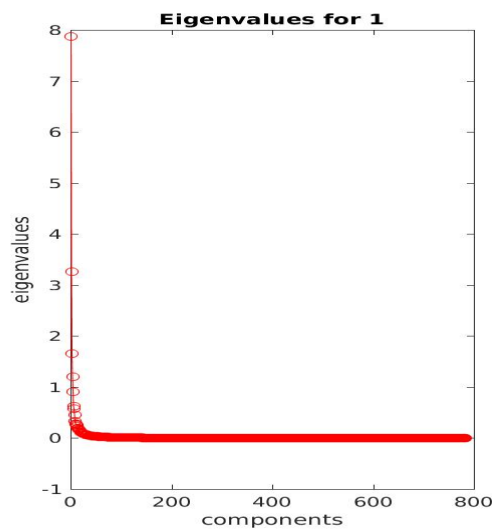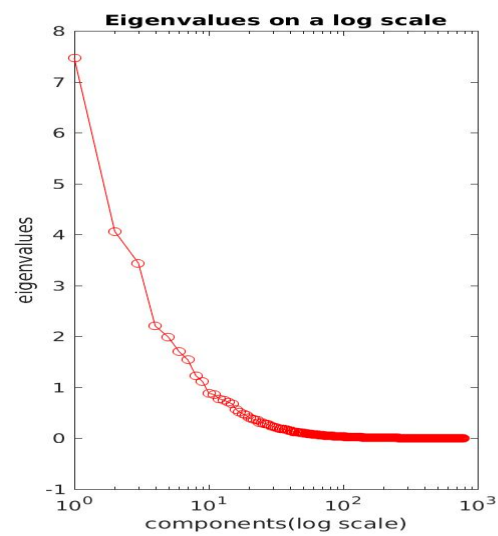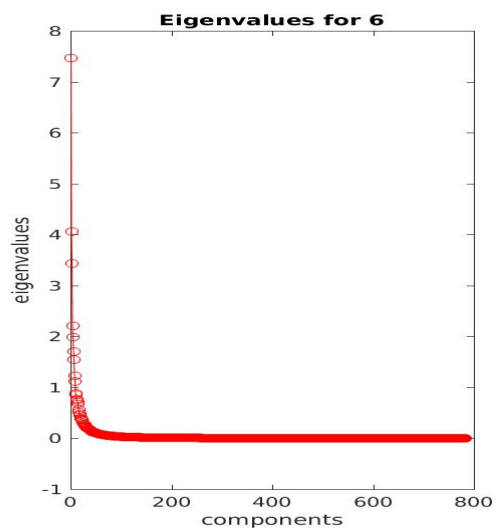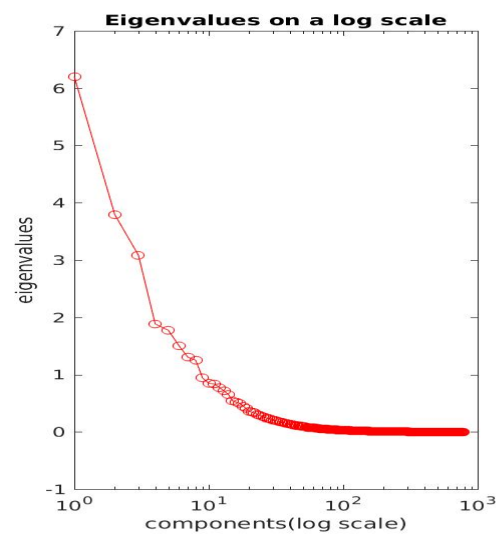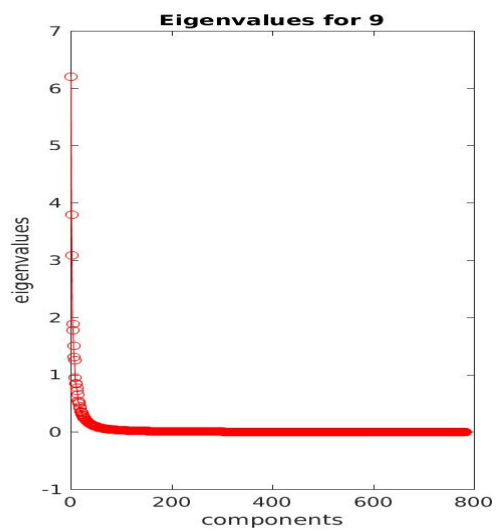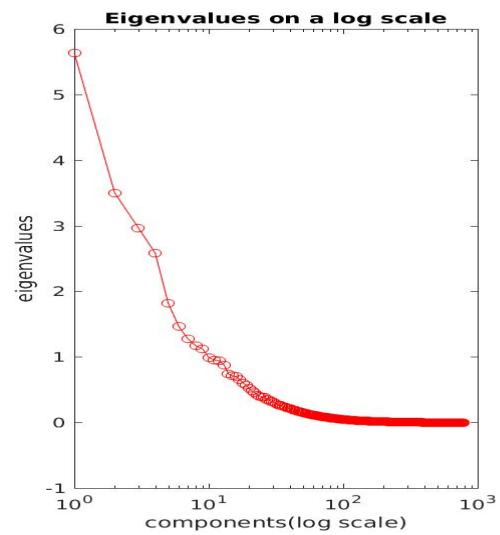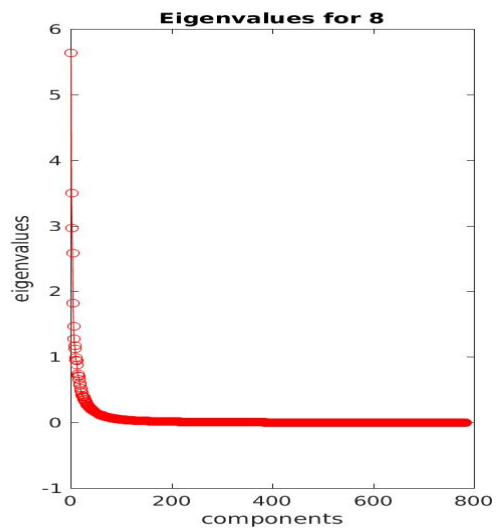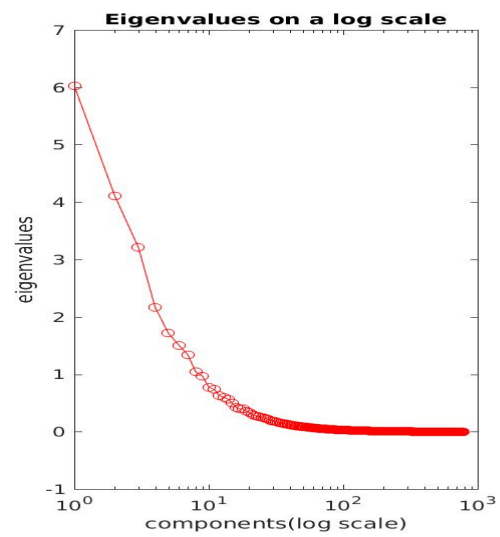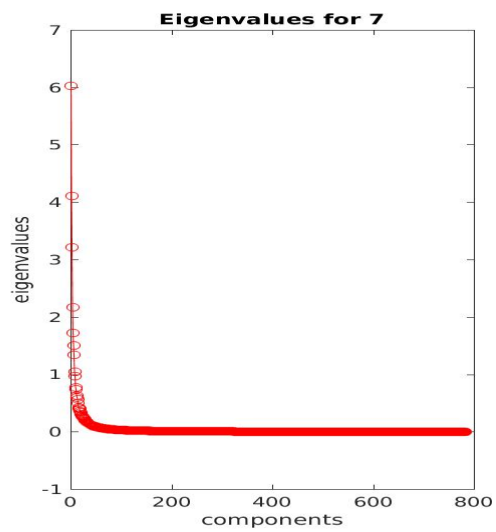
4.3 : Part C

For every digit, the graph of $\mu - \sqrt{\lambda_1}.v_1$, μ and $\mu + \sqrt{\lambda_1}.v_1$ was plotted and the following inferences were made:

- The plots plot of μ shows an **average spread** of the way people usually write the digit.

- As discussed earlier, the eigenvector corresponding to the maximum eigenvalue is the **principle mode of variation**. This is the drection in which the **variance 'varies' the most**.

- We can say that this **eigenvector** is a **linear combination** of the 784 variables of our image in a way such that it gives the most significant distinction in data. This obtained feature can be called the **most optimum** feature.

- Hence, $\mu - \sqrt{\lambda_1}v_1$ and $\mu + \sqrt{\lambda_1}.v_1$ will represent a (sort of) **range** in which **people usually** will **write** the respective digit, and if any input data lies within this range, then it will be highly probable that the input data matches this digit.

- For intance, in the case of the digit one, we observe two oblique lines in the two plots. This implies that the digit one, which ideally is a standing line, can be given some margin of error, and allow the user to write a 'one' with a slight slanting line as well.
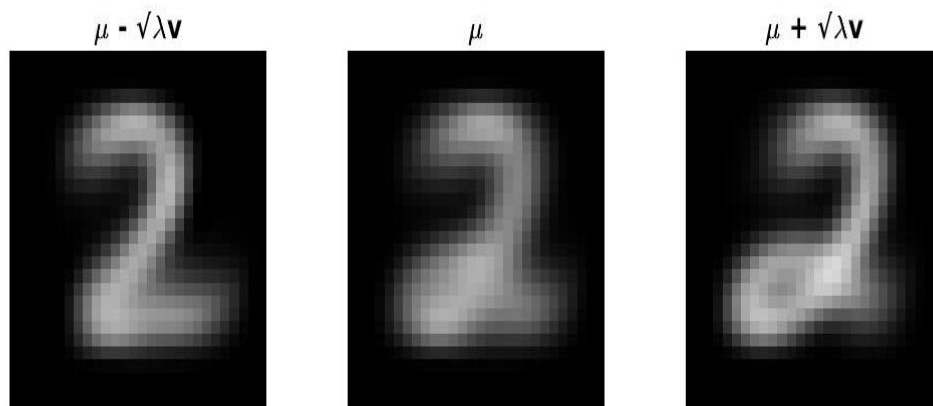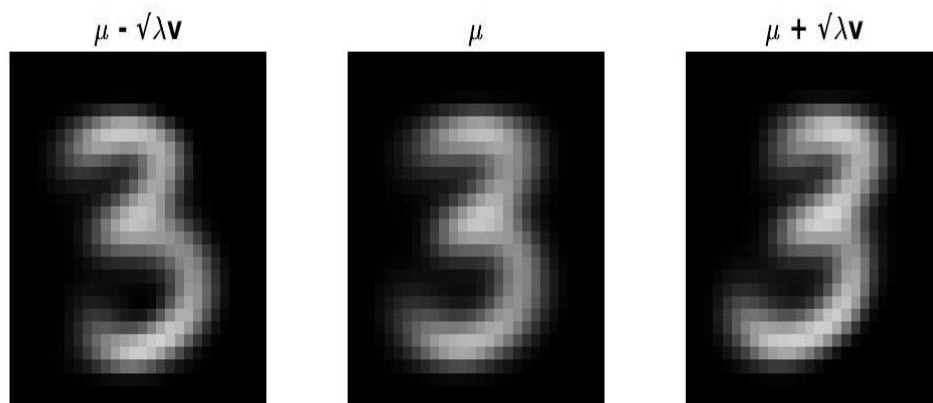
The plots of the comparison follow:
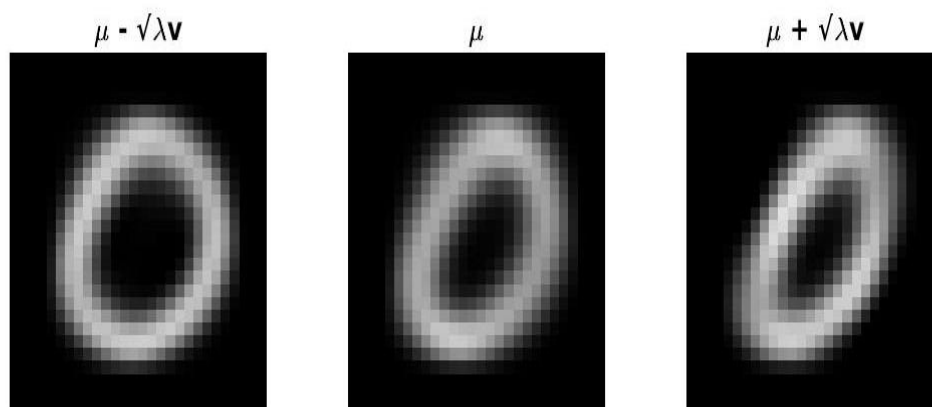
Analysis for the digit 0

| $\mu - \sqrt{\lambda}\mathbf{v}$ | μ | $\mu + \sqrt{\lambda}\mathbf{v}$ |
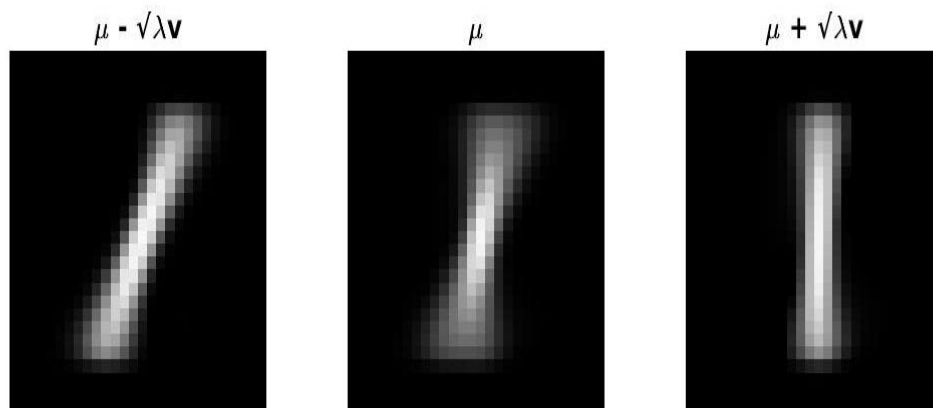
Analysis for the digit 1

| $\mu - \sqrt{\lambda}\mathbf{v}$ | μ | $\mu + \sqrt{\lambda}\mathbf{v}$ |

Analysis for the digit 2

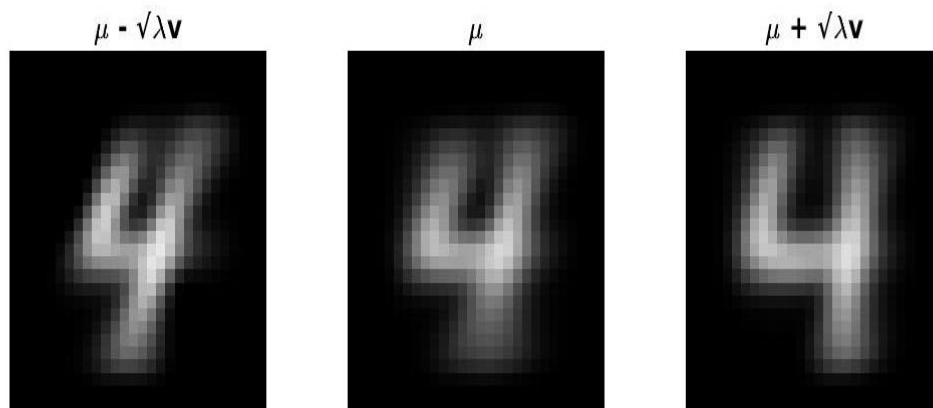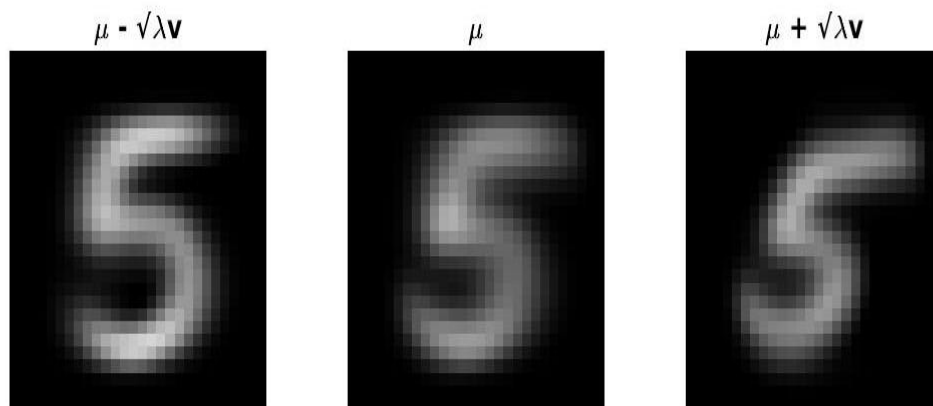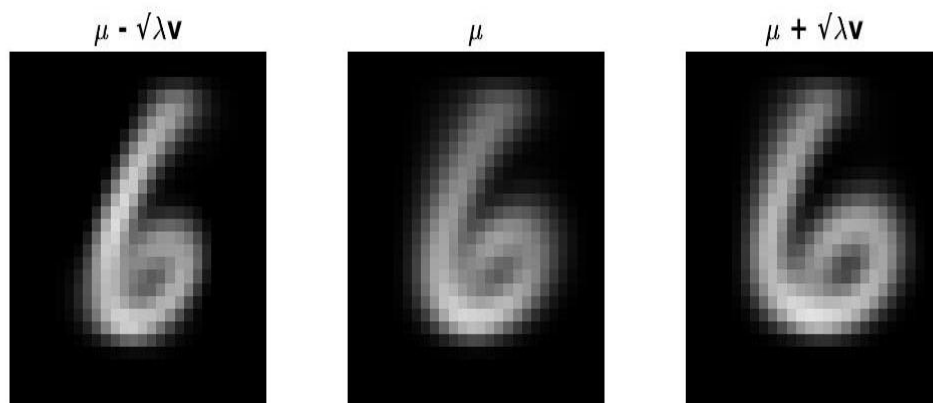| $\mu - \sqrt{\lambda}\mathbf{v}$ | μ | $\mu + \sqrt{\lambda}\mathbf{v}$ |

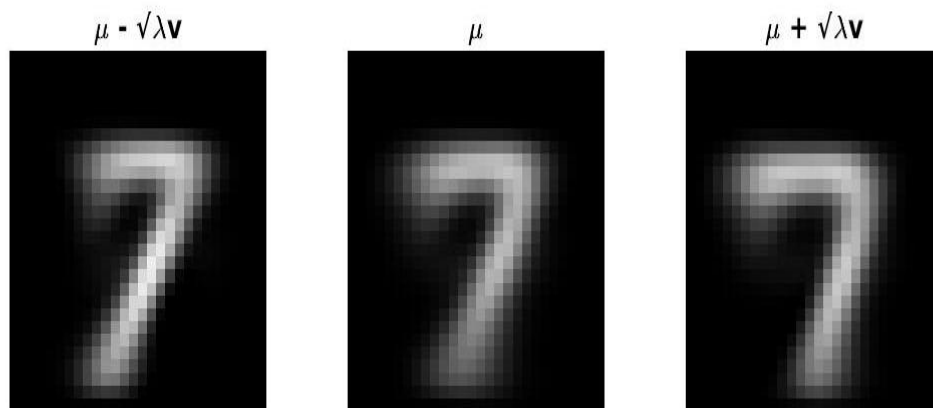Analysis for the digit 3

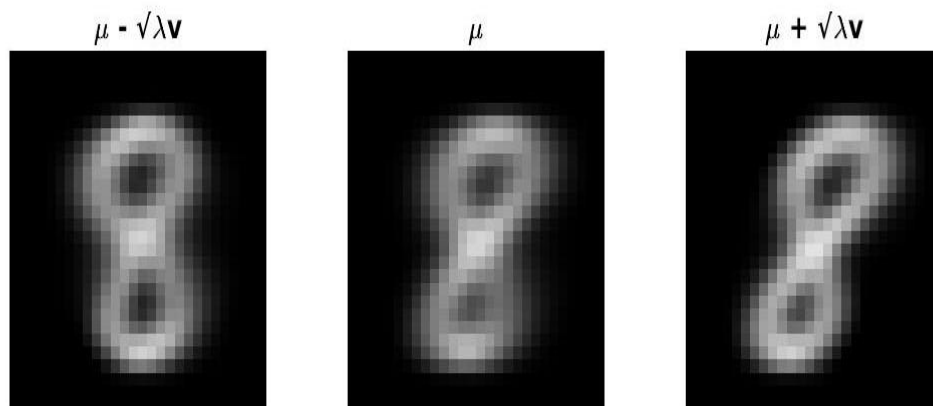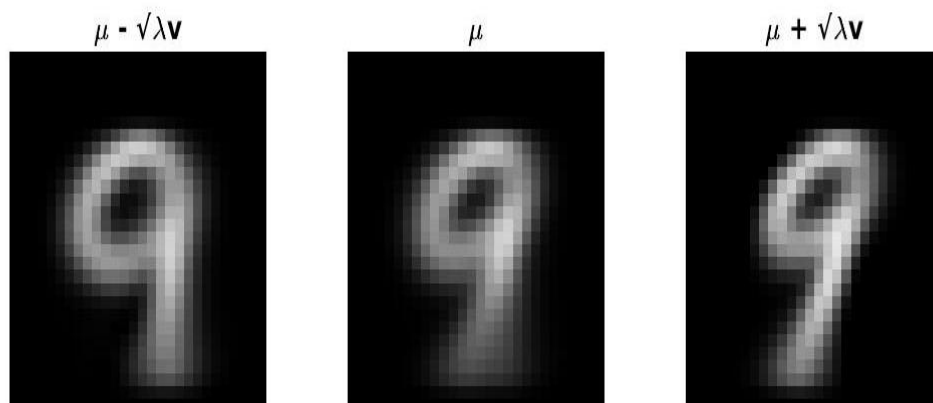Analysis for the digit 4

Analysis for the digit 5

Analysis for the digit 6

Analysis for the digit 7

Analysis for the digit 8

Analysis for the digit 9

4.4 : Usage of Code

The following are the instructions for the usage of the code:

- Load the code present in 'submission/code/q4/q4.m ' .

- In the same directory are functions implemented like myMean, myCov which return the mean and covariance of appropriate matrices.

- Simply run the code in 'q4.m ' and this wil automatically create the required plots.

- Lines 44, 58 (commented by default) hava a code to save jpg files of the respective plots. Comment/Uncomment these lines appropriately according to need.

- The results.mat present in 'results/mat' contains all the required results for the partA stored:

  Mean Vectors: Stored as a cell by the name MeanVectors, where the ith cell has the mean vector for the digit 'i'

  Covariance Matrix: Stored as a cell by the name CovarianceMatrices, where the ith cell has the Covariance Matrix for the digit 'i'

  First Modes of Variation: Stored as a cell by the name FirstModesOfVariation, where the ith cell has the First Mode of Variation for the digit 'i'

  Maximum Eigenvalues: Stored as a cell by the name MaxEigenValues, where the ith cell has the Maximum Eigenvalue for the digit 'i'