# CS215 ASSIGNMENT 2

Utkarsh Varun

Question 4

## Contents

# 1 Introduction

The Code for this question is in Q4.m file which is in code folder

We have also included "mnist.mat" file in code folder so that everytime when Q4.m executes it can take the input from the mnist.mat file

There are total 20 images generated n Q4.m code is executed 10 of which are of Part b and others are of Part c and these have been submitted in results folder

# 2 Part a

Here for every digit from 0 to 9 we have calculated the mean $mu$ , Covariance matrix C and the principal mode of variation determined by the eigenvector $v1$ and the corresponding eigen-value $\lambda 1$ (where $\lambda 1$ is the largest of all eigenvalues) of the covariance matrix C

Out of 60,000 examples we have some number of 0's and different number of 1's,2's and so on

For every digit i.e from 0 to 9 we have calculated the mean $mu$ and stored it in **'allmean'** variable which is of size 784x10 2D array. Each column contains the mean vector of each digit starting from 0 to 9

For every digit i.e from 0 to 9 we have calculated the covariance matrix C using mean and stored it in **'allcovariance'** variable which is of size 784x784x10 3D array. Each covariance matrix is of size 784x784 and we have 10 different digits so we get 10 covariance matrix C which is represented by that 3rd dimension

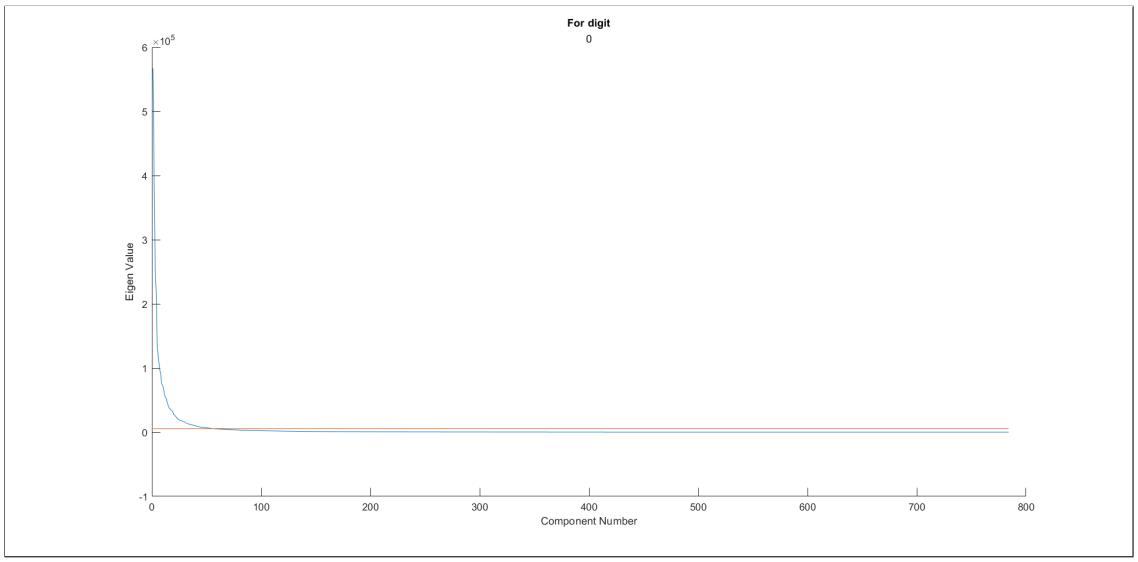
For every digit i.e from 0 to 9 we have calculated the largest eigen value of the covariance matrix C and stored it in **'largest_eigen_values'** variable which is of size 1x10 where columns contains the largest eigen value of a digit starting from 0 to 9. we have also stored the corresponding eigen vector of the largest eigen value also known as principal mode of variation in a variable **'pmov_eigen_vectors'** which is of size 784x10 where columns contains that before mentioned eigen vector starting from digit 0 to digit 9.(size of each digit's principle mode of variation or eigen vector of largest eigen value is 784x1)

**I have submitted all the sub parts asked in this part as .mat files in results folder as showing such big matrices is difficult**

# 3 Part b

Here for every digit from 0 to 9 we have sorted the 784 eigen values of the covariance matrix and plotting as a graph and identifying how many significant modes of variation ( or larger eigen values ) are present

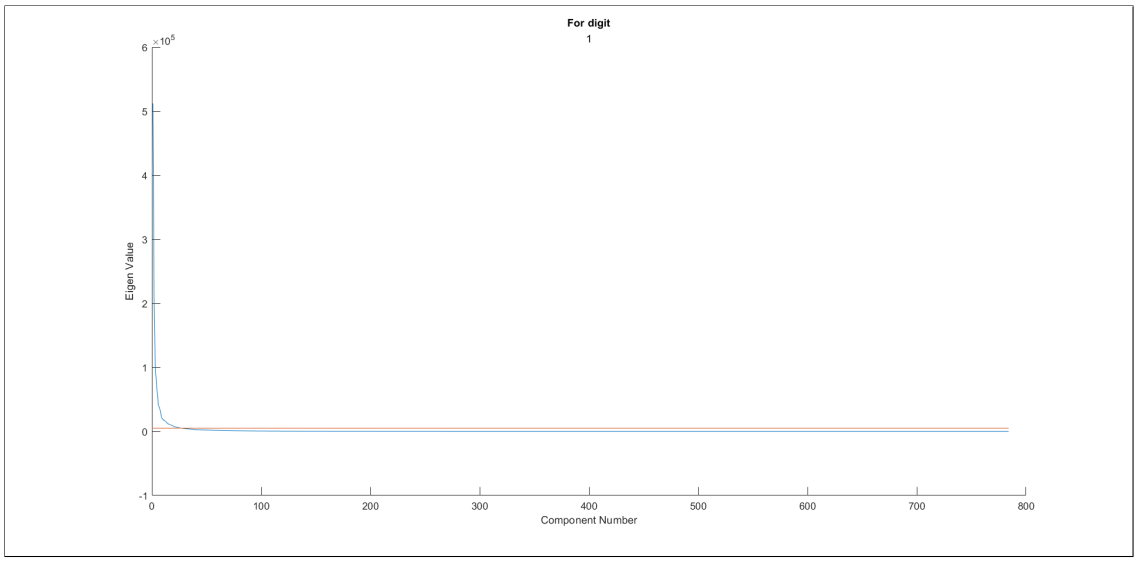
## 3.1 For digit 0




The largest eigen value in this case for digit 0 is 567257 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 0 we have 56 eigen values greater than the 1 percent of largest eigen value. Therefore there are **56** significant modes of variation for digit 0. The significant modes of variation is not equal to 784 and is far less than 784

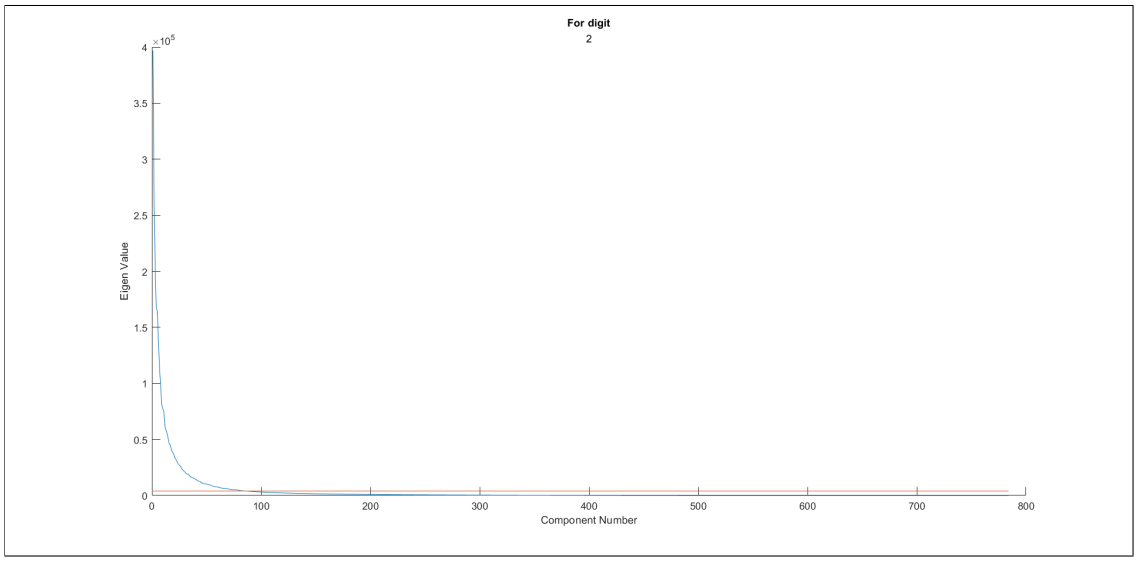
## 3.2 For digit 1




The largest eigen value in this case for digit 1 is 512065 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 1 we have 27 eigen values greater than the 1 percent of largest eigen value. Therefore there are **27** significant modes of variation for digit 1. The significant modes of variation is not equal to 784 and is far less than 784

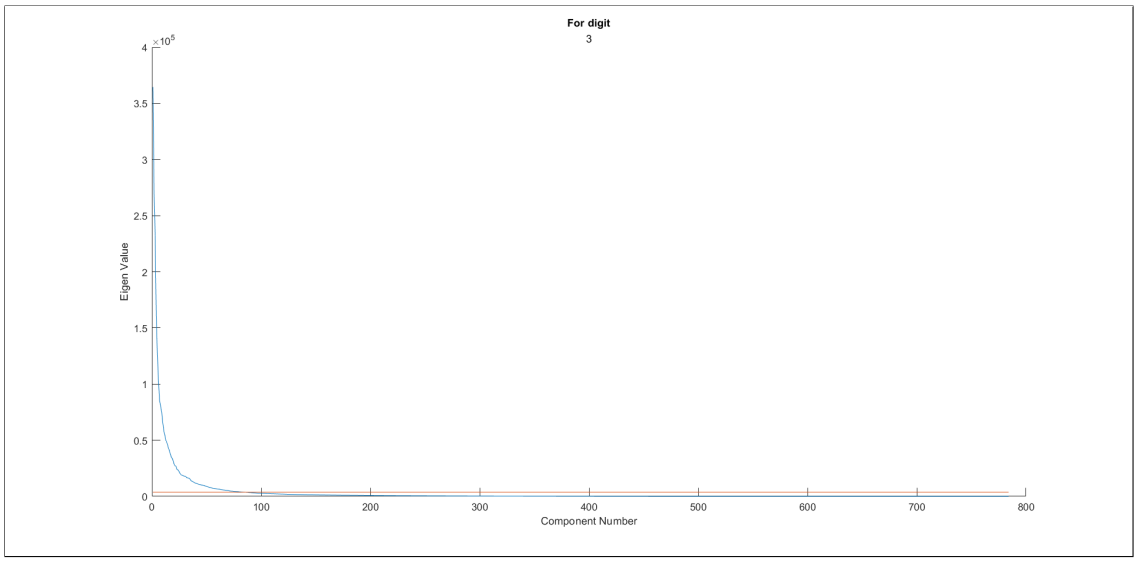
## 3.3 For digit 2




The largest eigen value in this case for digit 2 is 396935 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 2 we have 86 eigen values greater than the 1 percent of largest eigen value. Therefore there are **86** significant modes of variation for digit 2. The significant modes of variation is not equal to 784 and is far less than 784

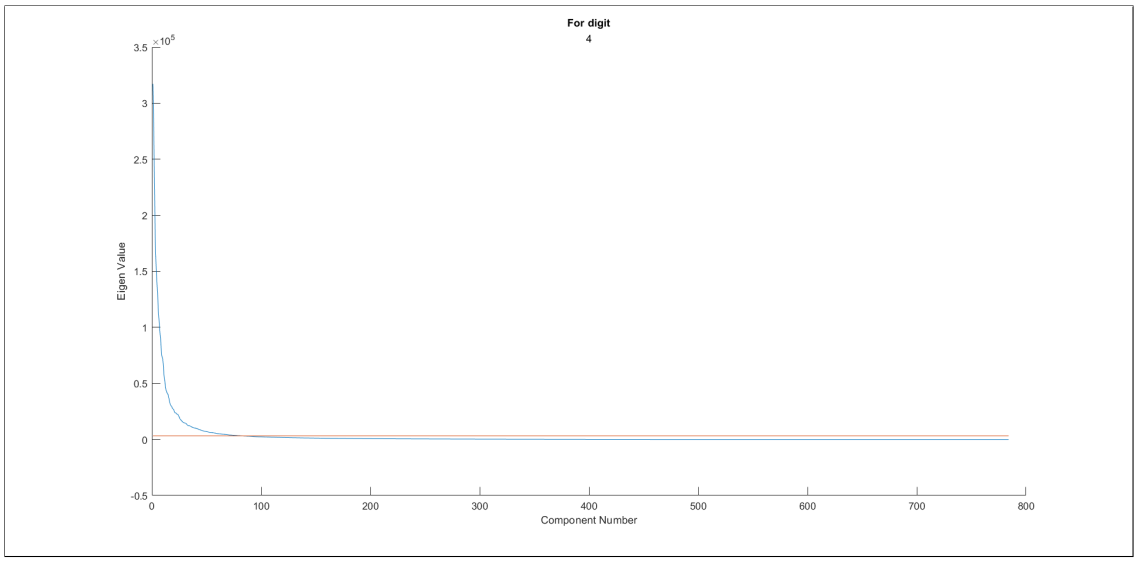
## 3.4 For digit 3




The largest eigen value in this case for digit 3 is 364484 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 3 we have 86 eigen values greater than the 1 percent of largest eigen value. Therefore there are **86** significant modes of variation for digit 3. The significant modes of variation is not equal to 784 and is far less than 784

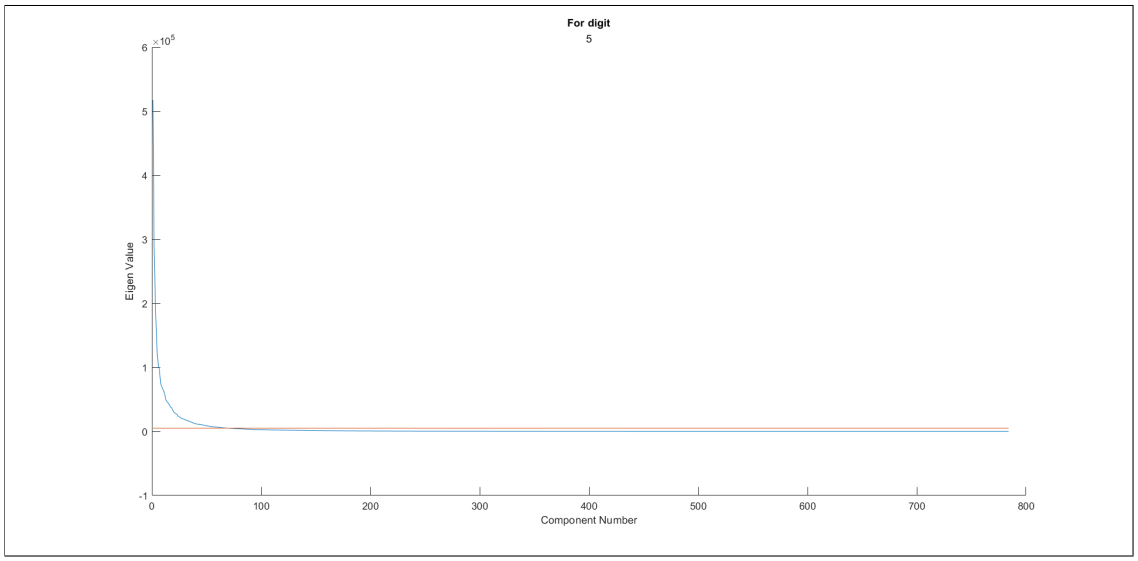
## 3.5 For digit 4




The largest eigen value in this case for digit 4 is 317197 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 4 we have 83 eigen values greater than the 1 percent of largest eigen value. Therefore there are **83** significant modes of variation for digit 4. The significant modes of variation is not equal to 784 and is far less than 784

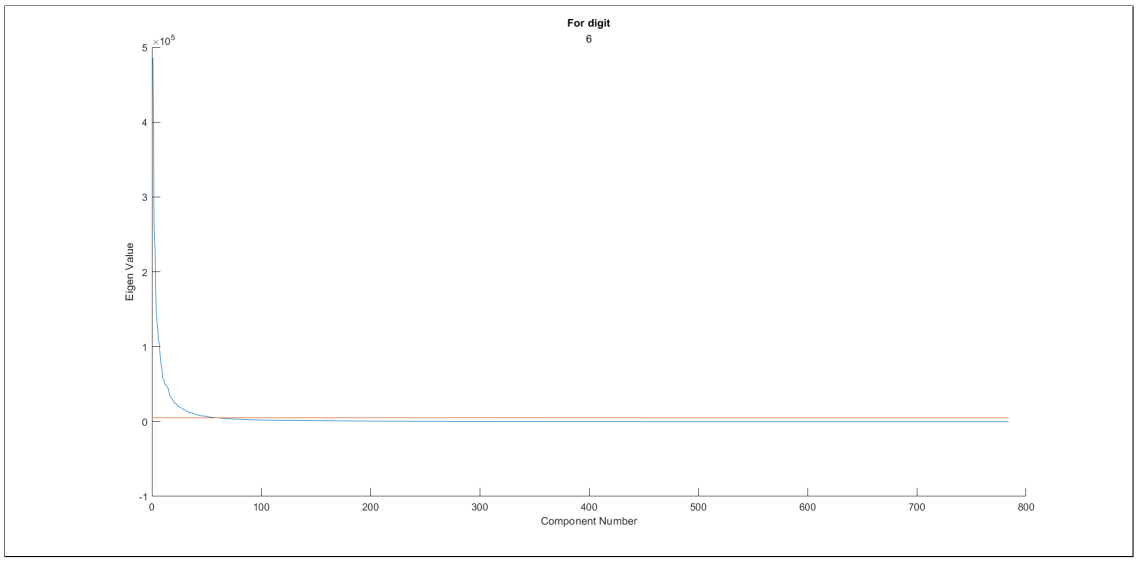
## 3.6 For digit 5




The largest eigen value in this case for digit 5 is 517505 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 5 we have 69 eigen values greater than the 1 percent of largest eigen value. Therefore there are **69** significant modes of variation for digit 5. The significant modes of variation is not equal to 784 and is far less than 784

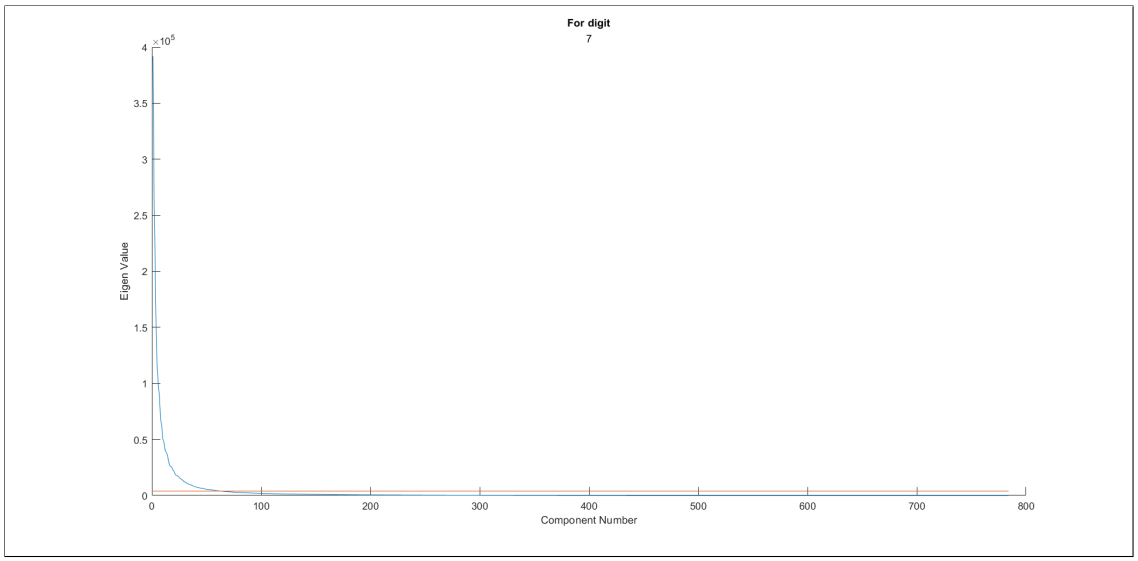
## 3.7 For digit 6




The largest eigen value in this case for digit 6 is 485468 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 5 we have 60 eigen values greater than the 1 percent of largest eigen value. Therefore there are **60** significant modes of variation for digit 6. The significant modes of variation is not equal to 784 and is far less than 784

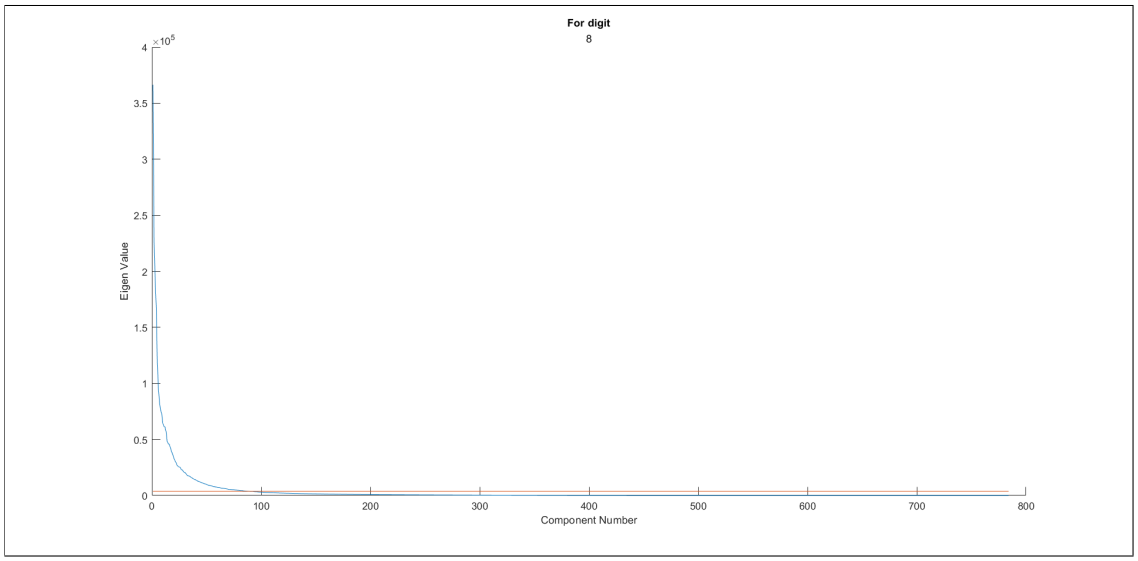
## 3.8 For digit 7




The largest eigen value in this case for digit 7 is 391915 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 7 we have 63 eigen values greater than the 1 percent of largest eigen value. Therefore there are **63** significant modes of variation for digit 7. The significant modes of variation is not equal to 784 and is far less than 784

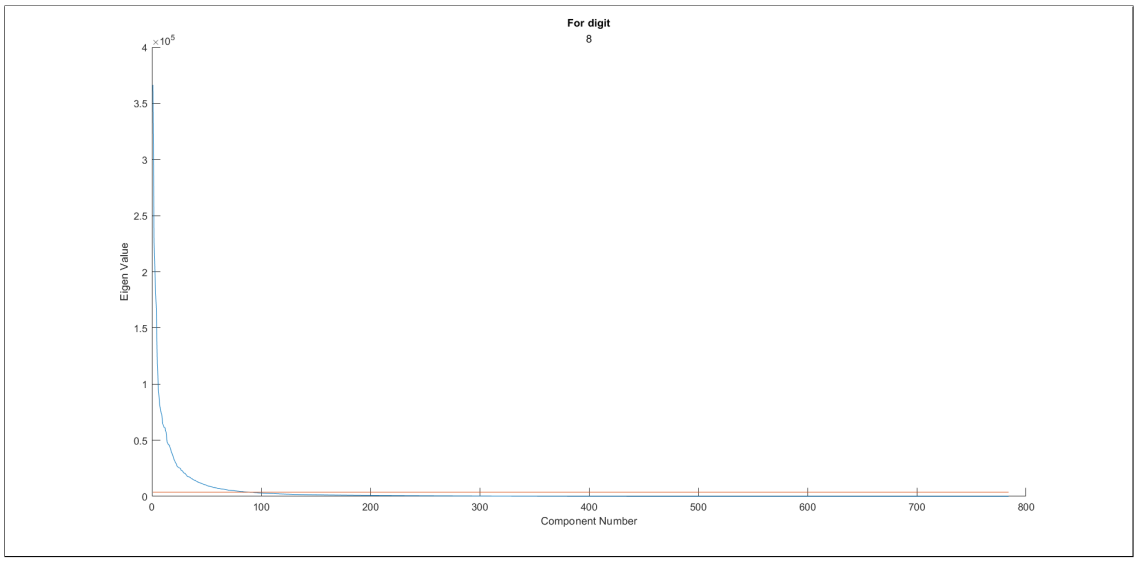
## 3.9 For digit 8




The largest eigen value in this case for digit 8 is 366391 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 8 we have 89 eigen values greater than the 1 percent of largest eigen value. Therefore there are **89** significant modes of variation for digit 8. The significant modes of variation is not equal to 784 and is far less than 784

## 3.10 For digit 9




The largest eigen value in this case for digit 9 is 403360 and there are 784 eigen values so by observing the graph i have also plotted a horizontal line at 1 percent of the largest eigen value and considering that eigen values above this line as significant modes of variation.

So For digit 9 we have 63 eigen values greater than the 1 percent of largest eigen value. Therefore there are **63** significant modes of variation for digit 9. The significant modes of variation is not equal to 784 and is far less than 784

# 4 Part c

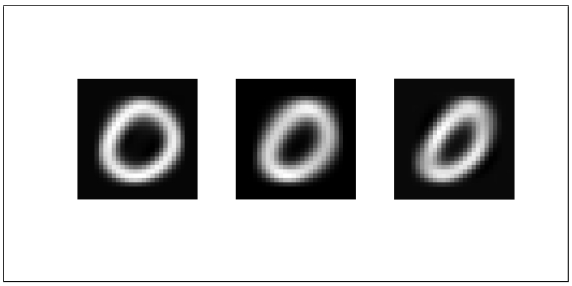
Here for each digit from 0 to 9 we have to show the images of

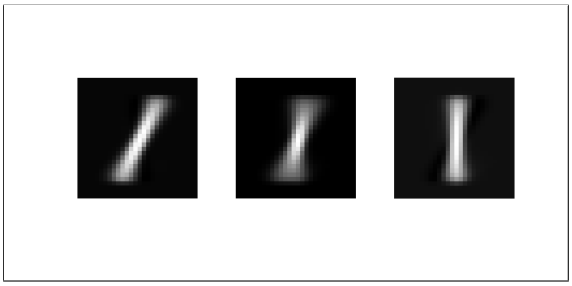
$(i)\mu - \sqrt{\lambda}1v1$ $(ii)\mu$ $(iii)\mu + \sqrt{\lambda}1v1$

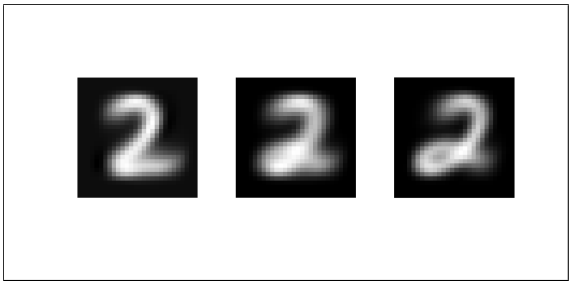
where $\mu$ is the mean and $\lambda 1$ is the largest eigen value and $v1$ is the principal mode of variation ( or the eigen vector corresponding to largest eigen value )

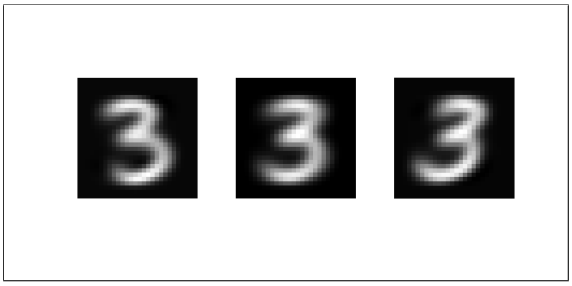
In each figure the left image is of $(i)$ part and middle one is of $(ii)$ part and the last one is of $(iii)$ part

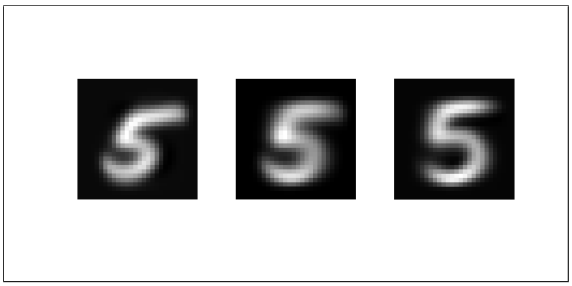
## 4.1 Images generated for each digit

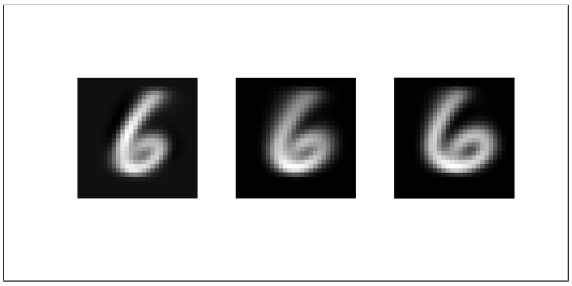


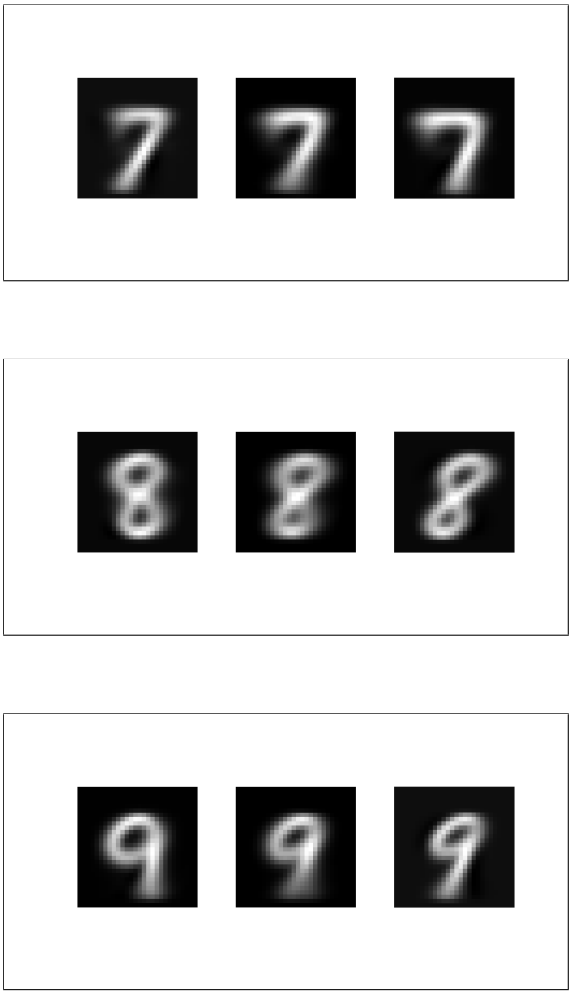










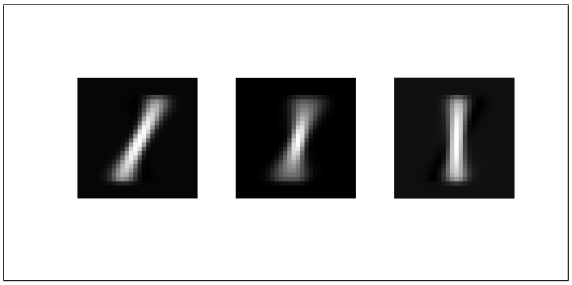


## 4.2 Observations on above generated images

$\mu - \sqrt{\lambda}1v1$ , $\mu + \sqrt{\lambda}1v1$ roughly tells us the region in which most of the users wrote a number because all the variations are covered by that range.

We can also justify this because it is almost similar to gaussian distribution because as most of the matter in gaussian distribution lies in the range of Mean−Variance to Mean + Variance here also it is almost similar to that.

### 4.3 Comments on digit 1



The region around mean image ( middle image ) is where most people write and with help of principle mode of variation we can say that $\mu - \sqrt{\lambda}1v1$ components have where people wrote 1 as slant and $\mu + \sqrt{\lambda}1v1$ components have where people wrote 1 as a straight/vertical .