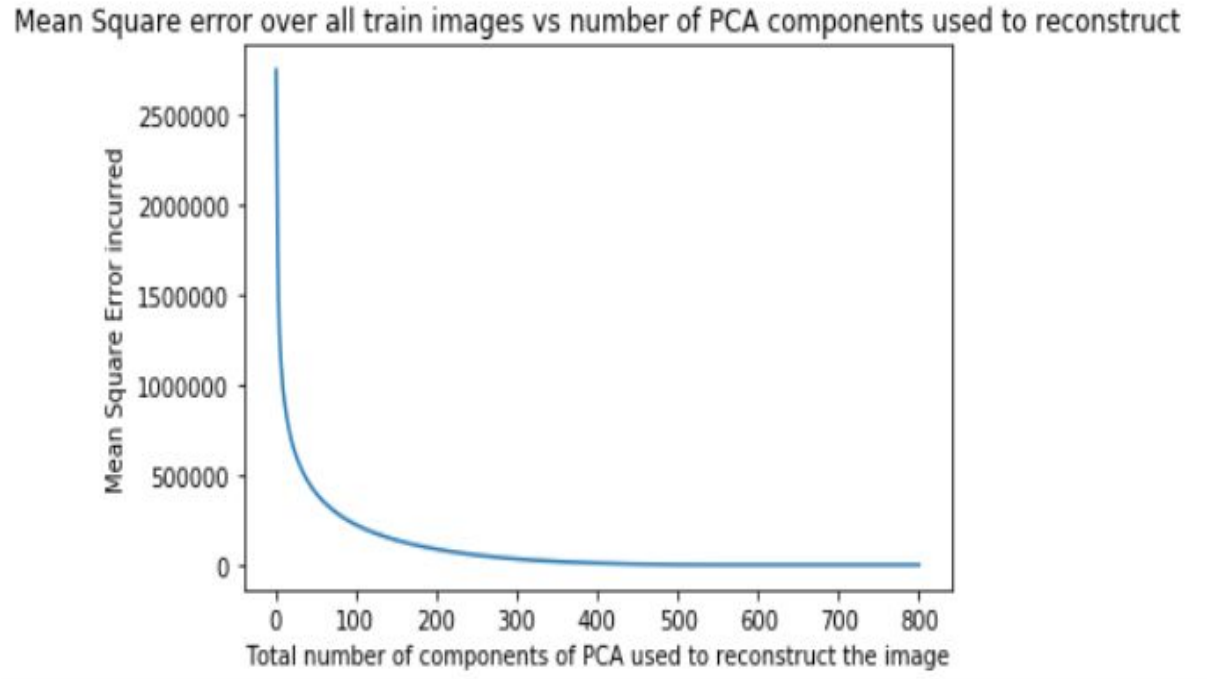
Principal Component Analysis
Report

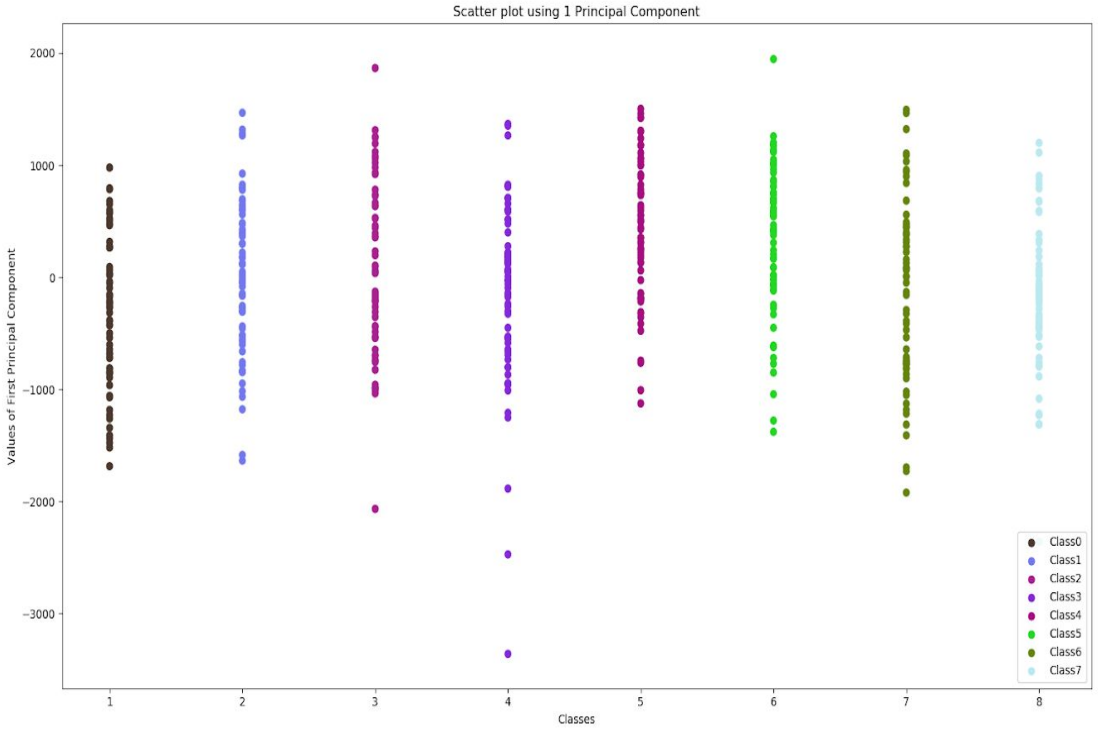
Saurabh Chand Ramola
20161106

All the training images are taken for the PCA and resized from (256,256) to (32,32) to aid in faster computation.The mean is subtracted from all the samples and the covariance matrix is found. To find the eigen vectors, the covariance matrix of the transpose of the images array is found and that is fed into the numpy function linalg.eig which finds the eigen values as well as the eigen vectors of the matrix.

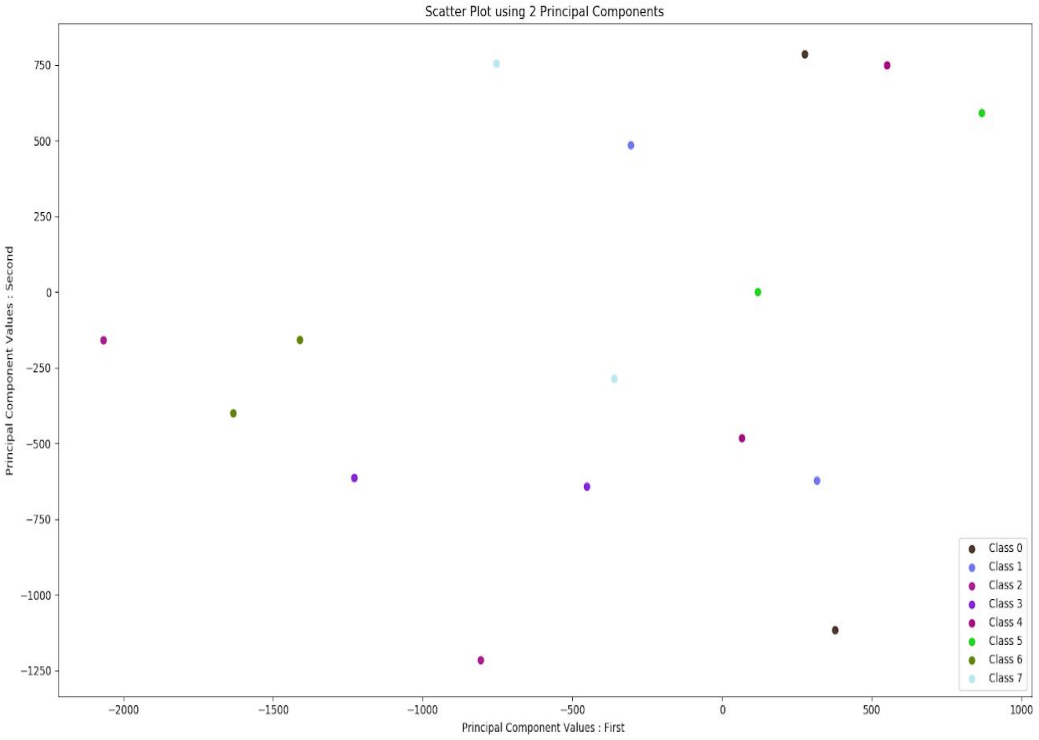
Further MSE is calculated for each image separately and then divided by original image to find the ratio. The average Mean Square Error is computed for all the 520 images for different number of eigenvectors and the following graph is observed:



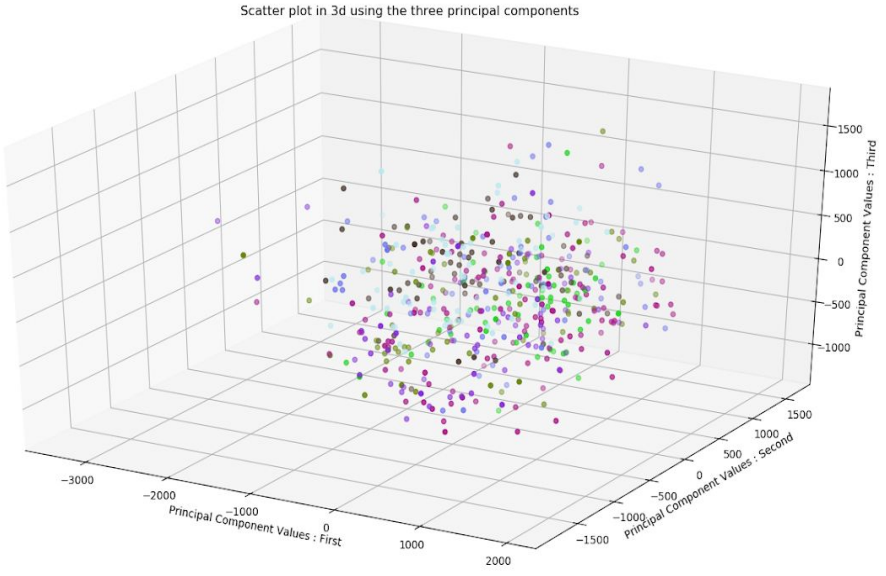
- Clustering in 1D:
    - We take the first component of the eigen-vector, i.e., the first coefficients of the eigen-vector.



- Clustering in 2D:
    - We take the first two component of the eigen-vector, i.e., the first two coefficients of the eigen-vector.



- Clustering in 3D:
    - We take the first three component of the eigen-vector, i.e., the first three coefficients of the eigen-vector.



Images showing how their resolution decreases as we take lesser eigen-vectors: