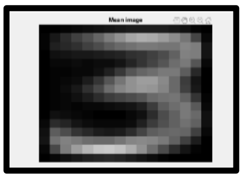
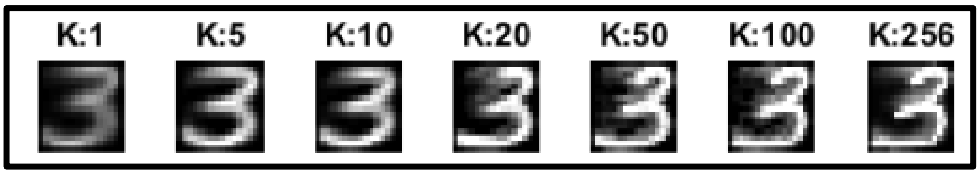
Principal Component Analysis on Handwritten Digits

Performing PCA on handwritten images of the digit 3 taken from the US Postal Service database. To access these images, load the Matlab data called *threes.mat* by typing load threes -ascii. This loads a 2-megabyte matrix called threes. Each line of this matrix is a single 16-by 16-image of a handwritten 3 that has been expanded out into a 256-long vector. You can look at the i-th image by typing the command *imagesc(reshape(threes(i, :),16,16),[0,1]).* To have a black-white picture use the command *colormap('gray')* first.

PCA is applied to a dataset of 500 images of 16x16 pixels reshaped to be as single 256-dimensional vector images of handwritten digit '3'. **Centering the data** on origin is the first step of PCA reduction that is done by subtracting mean of each image from all other images. The **covariance matrix** of the dimensions is then computed. From the eigenvectors and the eigenvalues are calculated. Eigenvectors with highest eigenvalues are selected to project the original image in a lower dimensionality with minimum loss as these selected eigenvectors contain most information of the original image.



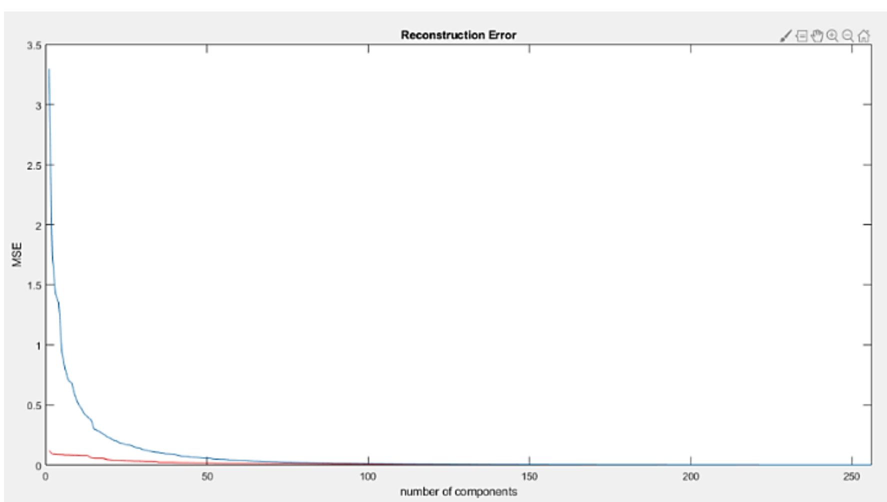


The mean of the dataset of images of digit 3

Some example reconstructions of 'K' eigenvectors with highest eigenvalues

The **reconstruction error** decreases as the number of eigenvectors increase. After k=50, although there are only a few eigenvectors, there is a sharp decline and the MSE goes < 0.02. This is followed by a low gradient slope that becomes almost a straight line by k=256. So, just with 20 highest eigenvectors we can almost perfectly reconstruct the handwritten images, thus reducing the dimensionality from 256 to just 20.

However, even with all eigenvectors it is still not possible to get a reconstruction error of zero because the mean image added to the reconstructed image makes the projection slightly different than the original image.



Reconstruction error (MAE) with increasing number of eigenvectors