ASSIGNMENT-1
PRINCIPAL COMPONENT ANALYSIS

Use MATLAB or python for this assignment. If using python look into numpy, scipy, and scikit-learn. MATLAB has all the functions inbuilt.
Octave can also be used.

This task should be fairly straightforward if you have a clear understanding of pca and eigen-decomposition. If you haven't ever coded anything of this sort before, the implementation part may be a little daunting. But there are a lot of online tutorials available to learn things. MATLAB/OCTAVE may be easier to use.
MATLAB : IITB provides the licence for matlab. Connect to vpn to download and use it.

TASK1-PRINCIPAL MODE OF VARIATION

Download the dataset comprising images of handwritten digits in
http://yann.lecun.com/exdb/mnist/
Loading data: MATLAB/OCTAVE:imread() (don't forget to convert the values to double)
Python: see scipy.io or scikit-learn (or anything else you like :P)

In this task, we will try to visualize the principal mode of variation for the datasets.

For each of the digits:

1) Compute the mean image μ
2) Compute the covariance matrix C
3) The first mode of variation determined by the eigenvector v1 and the corresponding eigenvalue λ1 (where λ1 is the largest eigenvalue) of the covariance matrix C

NOTE: Before computing μ and C, convert each 28x28 pixel image matrix to 784x1 column vector by concatenating its columns (see reshape in numpy and MATLAB). For visualization, it would have to be reshaped back to 28x28.

REPORT THE FOLLOWING:

1) For each digit, sort the eigenvalues of the covariance matrix and plot them as a graph. How many “principal”/significant modes of variation do you see (i.e. large eigenvalues). Is this number comparable to 28*28?
2) For each digit, show three images, side by side (i) $\mu - \sqrt{\lambda 1} v1$ (ii) $\backslash mu$ (iii) $\mu + \sqrt{\lambda 1} v1$, to show the principal mode of variation of the digits around their mean. OBSERVATION: for a certain digit, what does the principal mode of variation tell you about how people write that digit?

   NOTE: you may use the eig inbuilt function in MATLAB and python libraries(eg numpy.linalg). I would not encourage to use pca function directly that is present in matlab and sklearn(i think), but if you may, go ahead.