Homework 3

Study Chapter 12 from the Classnotes and the slides on the class website.
Read Chapter 12.1 and 12.3 from Bishop's book.

1) Derive a linear transformation $y = Ax$ based on principal components which has the property that the covariance $\text{Cov}(y) = \text{Identity}$.

2) Imagine you have a collection of gray-level images of faces. Provide pseudo-code to compute the “eigenfaces” of this collection. (see below for 4 examples of eigenfaces).

3) Perform a PCA on the Iris data. Make sure you first center the data. Extract the first two eigenvalues and eigenvectors.

4) Project the data down to a two-dimensional subspace and produce a scatter plot of the data. Make sure you plot each of the three classes differently (using color or different markers).

5) Generate 200 data-points of dimension 100 using $X = \text{randn}(100,200)$ in matlab. What are the theoretical values for the spectrum of the covariance matrix $X^T X / N$? (the spectrum is the collection eigen-values). If you do a PCA for this data and you check the eigen-values, what do you find? Plot the eigen-values sorted by their decreasing value. Why is there discrepancy between the theoretical values and measured values (I would like to see a different answer than: “because the data were generated randomly”)? If you increase the number of data-cases, will the effect increase or decrease?