Please show your work. Partial credit will be given for reasonable attempts — put down what you can. Start early and if you get stuck, please go to sections or office hours for hints/help.

Please check the course web page for information and announcements:
You can work in teams of up to three but everybody needs to hand in his own solution. The exercise parts marked with an asterisk are optional (for those of you who can handle an extra challenge).

Late assignments will lose 10% for the first week late and 50% after that

To set up matlab for this assignment, type
addpath ../public
To load the data for this assignment, type
load hw4data

The matrix x contains a number of 2-dimensional samples. You will use these data for Exercises A,B, and C below.

This week’s reading: Chapter 5, Chapter 6(optional read ahead).

(A) 1. Make (and hand in ) a scatterplot of the samples
   2. Find the sample mean (and write it down in what you hand in)
   3. Find the sample covariance matrix (and write it down in what you hand in)
(B) 1. Find the eigenvectors and eigenvalues of the sample covariance matrix.
   2. What is the eigenvector corresponding to the largest eigenvalue?
   3. What is the eigenvector corresponding to the smallest eigenvalue?
   4. Use the ’eigsort’ function (in ../public) to sort the eigenvectors and eigenvalues in order of largest eigenvalue to smallest eigenvalue. If the matrices \( V_{old} \) and \( D_{old} \) are, respectively, the eigenvectors and eigenvalues returned by the ’eig’ function, the syntax for ’eigsort’ is:
\[
[V,D] = \text{eigsort}(V_{old},D_{old})
\]
   How is the matrix of sorted eigenvectors related to the matrix of unsorted eigenvectors?
(C) The PCA transformation of a point consists of subtracting the mean and then multiplying by the transpose of the matrix of eigenvectors.
   1. Consider the point
\[
\begin{pmatrix}
-35 \\
40
\end{pmatrix}
\]
in the original coordinate system. What are its coordinates in the new coordinate system found by PCA (i.e., what are the principle components of the point)?

2. Transform all 100 sample points with the PCA transformation, and make (and hand in) a scatterplot of the corresponding points in the new coordinate system.

3. Look at the scatterplots of the samples in the original and transformed coordinates. PCA involves a translation plus a rotation, and can also involve flips (mirror images). How are the two scatterplots related: Is it by
(a) translation + rotation
or
(b) translation + rotation + flip?

The matrix 'faces' contains 48 face images. These are images of 16 people in 3 different lighting conditions. Each column of 'faces' is a 60x60 face image. These are the same faces that were used in Turk and Pentland’s original paper on eigenfaces. You will use them for exercises D–H below.

(D) For this exercise, you only need to tell us the matlab commands that you used (nothing else to turn in).
  1. Use the 'viewcolumn' command to view the 7th face image, which is stored in the 7th column of the 'faces' matrix. (You don’t need to print this face image out or turn it in to us—it’s just for you.)
  2. Compute the mean face, and use 'viewcolumn' to see it.
  3. Subtract the mean from all of the data, and call the matrix of mean-subtracted data 'A'.
  4. Use the 'eig' command to compute the eigenvectors and eigenvalues of $A^T A$.
  5. Use the 'eigsort' command to sort the eigenvectors and eigenvalues in order of largest to smallest eigenvalue. Call the matrix of sorted eigenvectors 'V'.
  6. V contains the sorted eigenvectors of $A^T A$. Use A and V to calculate U, the matrix of eigenfaces.
  7. Use the 'normc' command to normalize the columns of U so they all have length 1.

(E) For this exercise, just tell us the matlab commands that you used.
Find the principal components of the 7th face image in the data set. Call the vector of principal components 'c'.

(F) For this exercise, turn in printouts of the pictures.
  1. Display the 3rd eigenface (use the viewolumn command).
  2. Reconstruct the 7th face using all 48 principle components. (For all reconstructions, remember to add the mean face!)
  3. Reconstruct the 7th face using only the first 10 principle components.

(G) Let’s explore what happens when a non-face image is projected into face space.
  1. The vector 'dog' contains a 60x60 picture of a dog. Display the picture of the dog. (Just tell us the matlab command that you used.)
  2. Using the eigenfaces you found, find the principal components of the dog image. (Just tell us the matlab commands you used.)
  3. Reconstruct the dog picture using all 48 principle components. (Print out the picture.)
  4. Does the reconstructed dog look like the original picture? Explain why the reconstruction looks the way it does.
Extra credit. Columns 1–3 of ‘faces’ contain the three images of person 1, columns 4–6 contain the three images of person 2, and so on, all the way up to columns 46–48, which contain the three images of person 16. You will explore how well the face space generalizes to new people who weren’t in the data set. (At the end of each part, we tell you what to turn in for that part.)

1. Choose one of the 16 people (3 of the 48 images) to exclude, and calculate the eigenfaces using only the remaining 15 people (45 images). (Just tell which three images you excluded.)

2. Choose one of the images of the excluded 16th person. (Plot the image.)

3. Use the eigenfaces you found in part (1) to calculate the principle components of the image you chose in part (2). Reconstruct the image using all 45 of its principle components. (Plot the reconstructed image.)

4. Based on parts (1)–(3), how well do you think the face space represents faces that were not in the original data set?