Fall 2016
Pattern Recognition Theory 18-794

Homework 2

Due: Thursday, Oct. 20, 2016 by 12:30 pm in hard copy at CIC 1307 and (optional) soft copy at back up

To receive full credit, show the steps to your solution. Please be clear, neat, and eligible when you write your solutions. If the TAs cannot follow your work, you will not receive credit for that problem. Homework will be done individually: each student must hand in their own answers. It is acceptable for students to collaborate in figuring out answers and helping each other solve the problems. We will be assuming that, as participants in a graduate course, you will be taking the responsibility to make sure you personally understand the solution to any work arising from such collaboration. You must also indicate on each homework with whom you collaborated. **This homework has 120 points.**

Problem 1 (50 points) [Dipan]

PCA vs LDA. Given the following dataset for a 2-class problem with 2-D features:

$$c_1 = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\},\ c_2 = \left\{ \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 4 \end{bmatrix} \right\}$$

For this problem, although you are allowed to use MATLAB for aiding computations, you are required to show all intermediate steps.

1. Using global PCA, find the best direction onto which the data will be projected on.

   (a) (8 pts) Express the equation of the line along this direction passing through the samples mean in the following form: $\mathbf{w}^T\mathbf{x} + w_0 = 0$, where

   $$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

   and w_0 is a scalar known as the bias. Plot the $\mathbf{w}^T\mathbf{x} + w_0 = 0$ line along with all the sample points as well the line along $\mathbf{w}$ which recall starts from the mean.

(b) (6 pts) Project and reconstruct all the sample points. Plot the reconstructed points. edited

(c) (2 pts) Find the total mean square error MSE for all sample points (between the original points and the reconstructed points).

(d) (4 pts) Find the Fisher Ratio for this projection defined by

$$FR = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2},$$

where m_i is the mean of the projected samples of class i, and σ_i^2 is the equivalent variance. You can compute the FR on the projected 1-D points (rather than the reconstructed points which are 2-D vectors).

2. Using Fisher Linear Discriminant Analysis (LDA), determine the best one-dimensional space onto which the above data should be projected.

   (a) (8 pts) Express the equation of the line along this direction passing through the samples mean in the following form: $\mathbf{w}^T\mathbf{x} + w_0 = 0$. Plot the $\mathbf{w}^T\mathbf{x} + w_0 = 0$ line along with all the sample points.

   (b) (6 pts) Project and reconstruct all the sample points. Plot the reconstructed points. Make sure they lie on a line.

   (c) (2 pts) Find the total MSE for all sample points (between the original points and the reconstructed points). Compare that to the total MSE found in the first question.

   (d) (4 pts) Find the Fisher Ratio for this projection. You can compute the FR on the projected points (rather than the reconstructed points which are 2-D vectors). Compare that to the FR found in part 1. Interpret your result.

Problem 2 (20 points) [Dipan]

Another goal of PCA is to obtain the linear subspace which minimizes the projection error caused by dimension reduction. The two goals, maximum variance and minimum error, have the same formulation as PCA.

1. (10 pts) Assume all data samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ are centered. Formulate the optimization problem to solve for a vector $\mathbf{w}$ spanning the linear subspace which minimizes the sum of the squared reconstruction errors

$$\sum_{i=1}^{n} d_i^2,$$

where d_i, the reconstruction error of $\mathbf{x}_i$, is defined as the distance between $\mathbf{x}_i$ and its projection onto the linear subspace. Assume that $\mathbf{w}$ has a unit norm since we are interested only in the direction of the vector $\mathbf{w}$. The answer must be written in terms of $\mathbf{w}$ and $\mathbf{x}_i$ without d_i.

2. (10 pts) From your answer in (a), show that the optimization in part (a) is equivalent to the PCA optimization using the following equivalence:

$$\mathbf{\Sigma} = \sum_i \mathbf{x}_i \mathbf{x}_i^T.$$

Problem 3 (50 points) [Nancy]

Download the MNIST database at http://yann.lecun.com/exdb/mnist/ To help you read the images in the given format, there are a few MATLAB MNIST loaders available online. Alternatively you can download the dataset in MATLAB format from here http://cs.nyu.edu/ roweis/data.html

You are to implement PCA from scratch and not use the MATLAB built in functions. You are allowed to use SVD and/or eigenvalue decomposition. Please submit a print out of your code aw well.

If you run out of memory, you can use the first 1000 images from each class for all sections of this problem. In case you still run out of memory, use the first 100 images from each class instead.

1. Plot the mean image of digit 1 and plot the first 5 global PCA vectors corresponding to the dataset corresponding to two cases: using Gram Matrix Trick and without Gram Matrix trick. Do not forget to remove the mean before you run PCA. Measure the time taken for the PCA vector computation in each case.

2. Comment on the time taken in each case. Does it make sense to use the Gram trick for this particular dataset ?

3. Pick any random image from the dataset which can be any image from any class and project onto the eigen space (the global PCA eigen space that you obtained in part 1 of this problem) and reconstruct it using the first n eigen vectors (not just the n^{th} vector but all of them up until n) (for the two cases, Gram trick and without it) where $n = \{1, 2, 5, 10, 20\}$. Do not forget to remove the mean before projecting the image, and also to add it back once it has been reconstructed using only the first n eigen vectors.

For each of the 5 reconstructions, compute the mean square error (MSE) of the reconstructed image. Note that the MSE between two vectors a and b is given by

$$MSE(\mathbf{a}, \mathbf{b}) = ||\mathbf{a} - \mathbf{b}||_2^2$$

Display all reconstructed images and the original in one figure, with the corresponding MSE value as the title of each subfigure

Hint

1. Use the function *subplot* to plot several images in one figure
2. Using the functions *tic* and *toc* to compute the time processing
3. Plot an image in MATLAB using *imagesc*
4. Remember to convert the images from *uint8* data type to the *double* data type to avoid errors.
5. Be very careful while selecting features
6. Try out different distance metrics when building the similarity matrix.