University of Tehran

School of Electrical and Computer Engineering

Pattern Recognition

Assignment 4

Due Date: 26 Azar 1398

Corresponding TA:

Ida Mirzadeh – a.mirzadeh@ut.ac.ir

Maryam Ebrahimi – Maryam_ebrahimi@ut.ac.ir

Ali sorouramini - ali.sorouramini@ut.ac.ir

Parmiss Tavassoli – p.tavassoli@ut.ac.ir

Azar 98

PROBLEM 1

Let $p_x(x|w_i)$ be arbitrary densities with means μ_i and covariance matrices Σ_i (not necessarily normal) for i = 1,2. Let $y = \ w^t x$ be a projection, and let the induced one-dimension densities $p(y|w_i)$ have means μ_i and variances σ_i^2.

Show that the criterion function

$$J_1(w) = \frac{(\mu_1 - \ \mu_2)^2}{\sigma_1^2 + \ \sigma_2^2}$$

Is maximize by

$$w = (\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2)$$

PROBLEM 2

The expression

$$J \ = \frac{1}{n_1 n_2} \sum_{y_i \in Y_1} \sum_{y_j \in Y_2} \left(y_i - y_j\right)^2$$

measures the total within group scatter. Show that this within group scatter can be written as:

$$J = (m_1 - m_2)^2 + \frac{1}{n_1} s_1^2 + \ \frac{1}{n_2} s_2^2$$

PROBLEM 3

Show that the total (or mixture) scatter matrix is the sum of the within-class and between-class scatter matrices.

PROBLEM 4

Consider an *N*-dimensional random vector *x*, which is approximated by

$$\hat{x} = \sum_{i=0}^{m-1} y_i e_i + \sum_{i=m}^{N-1} c_i e_i$$

Where c_i are non-random constants and e_i, $i = 0.1.2. \ldots. N-1$ constitute an orthonormal basis. Show that the minimum mean square error $E\|x - \hat{x}\|^2$ is achieved if:

a. $c_i = E[y_i]$, $i = m. \cdots. N-1$
b. The orthonormal basis consists of the Eigen vectors of Σ_x?
c. e_i, $i = m. \ldots. N-1$, correspond to the N-m smallest Eigen values.

PROBLEM 5

Divergence can be used as a separability measure for the classes $w_1.w_2$, with respect to the adopted feature vector x. For a multiclass problem, the divergence between classes w_i and w_j is defined as follows:

$$d_{ij} = \int_{-\infty}^{+\infty} \big(p(x|w_i) - p(x|w_j)\big) ln \frac{p(x|w_i)}{p(x|w_j)} dx$$

a. Show that if $d_{ij}(x_1.x_2.\dots.x_l)$ is the class divergence based on l features, adding a new one x_{l+1} cannot decrease the divergence, that is

$$d_{ij}(x_1.x_2.\dots.x_l) \le d_{ij}(x_1.x_2.\dots.x_l.x_{l+1})$$

b. Show that if the features are statistically independent, then the divergence is given by:

$$d_{ij}(x_1.x_2.\dots.x_l) = \sum_{m=1}^{l} d_{ij}(x_m)$$

PROBLEM 6 (computer assignment)

Dimension reduction with principal component analysis (PCA) is a common technique for image compression. The number of Principle components (PCs) that used in this task affects compression rate and image quality. The main idea of this problem is to compress images by PCA before face recognition process. As the result, images are transformed into a smaller set of characteristic feature images, which is then used for face recognition task.

You have to use “FACES” dataset that contains 28 images as the Train data and 14 images as Test data.[1] This data and “image_loader.py” file are attached to this assignment.

By applying PCA using sklearn library functions compress the images and explore effects of extracted PCs on face recognition rate by trying out different values for explained variability parameter. For face recognition, you can compute the Euclidean distance of the compressed test image to every compressed train images and choose the label of the closest image from the train dataset as your predicted label for the test image.

a) Select an image from the test dataset, compress the image by applying PCA then reconstruct the image and put the corresponding image**s** and compression rate for this sample into your report.
b) Plot Recognition rate[2] as a function of explained variability by changing the input parameter of PCA function. Explain the results.

---

[1] Every person is represented two times in Train dataset and one time in Test dataset

[2] the percentage of human faces correctly recognized

c) Reconstruct the compressed images after applying PCA with different number of components. Plot mean square error (MSE) between the compressed images and original images as a function of number of PCs. Explain the result.

<u>For problems 10, 11, 12 and 13 use FASHION MNIST dataset, you can use mnist-loader.py.</u>

PROBLEM 7 (computer assignment)

By putting together train and test dataset for all the classes and removing the class labels, examine whether there is any linear dependency among features. Find a linear transformation y = A(x - b) that whitens the data ($\Sigma y = I$[3] & $\mu y = 0$).

PROBLEM 8 (computer assignment)

Implement Forward selection algorithm from scratch. You can use Naïve Bayes optimal classifier[4] in forward selection as the classifier.

a. Plot CCR as a function of the number of selected feature.
b. Select the optimal number of features for best performance in classification. (You do not need to run the code completely. Run the code to the extent that you can show the optimal number of features)

PROBLEM 9 (computer assignment)

By calculating the Eigen-vector and Eigen-value of Covariance matrix, implement PCA method

a. Plot the Eigen-values of Covariance matrix in descending order.
b. Choose appropriate number of features (based on the result of part a). After projecting the data into new subspaces, apply Naive Bayes optimal classifier[5], with Gaussian parametric estimate of pdf's and report the CCR.
c. Repeat part 'b' without PCA and Compare the results (CCR).

For this question, you are not allowed to reduce the sample size nor the feature size before applying PCA.

---

[3] Identity Matrix

[4] https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

[5] https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

PROBLEM 10 (computer assignment)

Now consider class labels, and by calculating between scatter (S_B) and within scatter (S_W) matrices, implement Linear Discriminant Analysis (LDA) method.

a. Calculate the separability matrix and plot the Eigen-values of separability matrix in descending order.
b. Plot the separability measure vs. number of components.
c. Choose appropriate number of features (based on the result of part a and b). After projecting the data into new subspaces. Next, apply Naive Bayes optimal classifier[6], with Gaussian parametric estimate of pdf's and report the CCR. Explain the result.
d. Repeat part 'c' without LDA and Compare the results (CCR).
e. Between PCA and LDA methods, which approach was more helpful in the final classification task (in terms of correct classification rate versus component)? Explain your reasoning.

For this question, you are not allowed to reduce the sample size nor the feature size before applying LDA.

[6] https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

Notes

1. Please try your best to meet the deadlines.
2. Late submission is possible but would be penalized by a discount factor in your score.
3. Every student has a budget for late submission during the semester. This budget is two weeks for all the assignments. Late submission more than two weeks may be penalized.
4. Analytical problems can be solved on papers and there is no need to type the answers. The only thing matters is quality of your pictures. Scanning your answer sheets is recommended. If you are using your smartphones you may use scanner apps such as CamSanner or google drive application.
5. Simulation problems need report as well as source codes and results. This report must be prepared as a standard scientific report.
6. You have to prepare your final report including the analytical problems answer sheets and your simulation report in a single pdf file.
7. Finalized report and your source codes must be uploaded to the course page as a ".zip" file (not ".rar") with the file name format as bellow:

PR_Assignment #[Assignment Number]_Surname_Name_StudentID.pdf

8. Plagiarisms would be strictly penalized.
9. You may ask your questions from corresponding TA.