University of Tehran

School of Electrical and Computer Engineering

Pattern Recognition

Assignment 4

Due Date: 21st of Ordibehesht

Corresponding TAs:

Ida Mirzadeh – ida_mrd@yahoo.com

Maryam Ebrahimi – Maryam.ebrahimi25@yahoo.com

Farvardin – Ordibehesht 98

PROBLEM 1

Consider two normal distributions with arbitrary mean but equal covariance.

$$p\left(y|\tilde{\theta}_i\right) = \frac{1}{\sqrt{2\pi}\tilde{s}} \exp[-(y-\tilde{\mu}_i)^2/\left(2\tilde{s}^2\right)]$$

Where $\tilde{\theta}_i = \begin{pmatrix}\tilde{\mu}_i \\ \tilde{s}\end{pmatrix}$ for i = 1, 2 . Denote the samples after projection as $\widetilde{D}_i$.

Prove that the Fisher linear discriminant can be derived from the negative of the log-likelihood ratio.

Hint: log-likelihood ratio is defined as $r = \frac{\ln p(\tilde{D}|\tilde{\theta}_1)}{\ln p(\tilde{D}|\tilde{\theta}_2)}$

PROBLEM 2

Let $p_x(x|w_i)$ be arbitrary densities with means $\boldsymbol{\mu_i}$ and covariance matrices $\Sigma_{\boldsymbol{i}}$ (not necessarily normal) for i = 1,2. Let $y = \boldsymbol{w^T x}$ 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 maximized by

$$\boldsymbol{w} = (\Sigma_{\mathbf{1}} + \Sigma_{\mathbf{2}})^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$$

PROBLEM 3

Assume we have 2 classes Y_1 and Y_2, with sizes n_1 and n_2 respectively. The expression

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

measures the total within group scatter. Show that this within group scatter can be written as (m_i and s_i are mean and variance corresponding to class i):

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

For problems 4, 5, 6 and 7 use FASHION MNIST dataset, you can use mnist-loader.py.

PROBLEM 4

By implementing forward selection algorithm, select the optimal number of features for best performance in classification by using Naive Bayes optimal classifier[1]. (Consider Gaussian parametric estimate of pdf's). Plot correct classification rate as a function of the number of selected features to find the optimal number of features.

PROBLEM 5

Implement PCA method from **scratch**[2]

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. Next, apply Naive Bayes optimal classifier[3], with Gaussian parametric estimate of pdf's and report the CCR.
c. Repeat part 'b' without PCA and Compare the results (CCR).

PROBLEM 6

Without considering class labels, Implement Linear Discriminant Analysis from scratch (At first whiten the data and then compute between scatter matrix (S_B) and within scatter matrix (S_W)).

a. 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[4], 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).

---

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

[2] You're not allowed to use available functions in sklearn or other similar libraries for PCA

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

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

PROBLEM 7

Use PCA[5] and LDA[6] functions from sklearn python package with different number of principle components. After feature reduction, use Gaussian classifier and calculate CCR of classification result.

Plot CCR as a function of number of principle components for both PCA and LDA methods in one diagram. Compare the effect of number of principle components on CCR in each case. Between PCA and LDA which method gives the better result on this dataset? Explain your reasoning.

PROBLEM 8 (Bonus)

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.[۷]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. 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 train dataset as your predicted label for the test image.

a) Plot Recognition rate[8] as a function of the number of principle components by changing the input parameter of PCA function. Explain the results.
b) 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.

---

[۵]https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

[۶]https://scikit-learn.org/0.16/modules/generated/sklearn.lda.LDA.html

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

[۸]the percentage of human faces correctly recognized

Notes

1. Please make sure you reach the deadline because there would be no extra time available.
2. Late policy would be as bellow:
   - 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 cause lost in your scores.
3. 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 CamScanner or google drive application.
4. Simulation problems need report as well as source codes and results. This report must be prepared as a standard scientific report.
5. You have to prepare your final report including the analytical problems answer sheets and your simulation report in a single pdf file.
6. 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.zip

7. Plagiarisms would be strictly penalized.
8. You may ask your questions from corresponding TAs.