University of Tehran

School of Electrical and Computer Engineering

Pattern Recognition

Assignment 4

Due Date: 17th Azar

Corresponding TA:

Mahdie Soltaninejad – mahdie.soltaninejad@gmail.com

Mohammad Amin Khosrozadeh – m.a.khosrozadeh@gmail.com

TAs Office Hours:

Wednesday (8th Azar): 10-12

Thursday (9th & 16th Azar): 14-16

Saturday (11th Azar): 16-18

Room 402 (Cognitive Systems Lab)

Azar 96

IMPORTANT NOTES

1. For problem 1, download and UNZIP the "face" dataset. You should have 40 folders and each folder should have 10 .png images in it. Each folder contains 10 images of a person's face under different lighting and pose.

2. In problems 3 to 5, you have to use the small version of uploaded dataset, named "**TinyMNIST**". The dataset is available on your course page in "Dataset" section. Load the dataset with the attached file, named "TinyMNIST_loader.py".

3. You are **not** allowed to use pre-defined algorithms in this assignment.

PROBLEM 1

Use PCA method to project data onto the principle components.

a. Indicate the first and last 10 eigenvectors (called eigen-faces) in your report (ordered by eigenvalues) and discuss about what these images are representing. (The eigenvectors should be reshaped to 2-d matrices to form the eigen-faces).

b. Plot the singular values and explain how to find good point for truncating the Singular Value Decomposition.

c. Reconstruct an approximation of data and recover the data by projecting them back onto the original high dimensional space. Plot samples of reconstruct data and compare it with real data.

PROBLEM 2

In a C class classification problem, Within and Between Class Scatter Matrices (S_W and S_B) are defined as follows:

$$S_w = \sum_{k=1}^{C} \sum_{x^q \in w_k} (x^q - \mu_k)(x^q - \mu_k)^t$$

$$S_B = \sum_{k=1}^{C} N_k \, (\mu_k - \mu)(\mu_k - \mu)^t$$

a. Show that rank(S_B) ≤ C −1. Under what condition rank(S_B) = C −1?
b. What can you say about the maximum number of nonzero Eigen- (singular) values of the Separability Matrix ($S_w^{-1} S_B$) ?
c. Now, consider a two class problem C = 2. Derive expressions in terms of within class scatter matrix (S_w) for the eigenvalues and eigenvectors of the Separability Matrix ($S_w^{-1} S_B$).
d. What is the linear transformation that maximizes trace($S_w^{-1} S_B$), while reducing the number of features as much as possible.

PROBLEM 3

By using forward selection, select optimal number of features that best represent data using Bayes optimal classifier with Gaussian parametric estimate of pdf's. Draw CCR based on number of selected features.

PROBLEM 4

Without considering class labels, find linear dependency among features in train data. Find a linear transformation y=(x−b) that whitens the data, i.e. Σ y=I & μy=0.

a. Now consider class labels on whitened data, and compute within (S_W) and between scatter (S_B) matrices. Use Linear Discriminant Analysis (LDA) to form 196 combined features sorted in descending order of their effect on separability measure trace ($S_w^{-1} S_B$). Plot the eigen-values in descending order.

b. Choose different subspaces and project the data onto these subspaces. Plot separability measure vs. the number of components used for each subspace, and discuss about the effect of combining features (subspace dimension) on seperability percentage.

c. Compare results of problem 3 and ‘4-a’ and mention drawbacks of forward selection approach.

PROBLEM 5

a. After applying PCA method, design a Bayes optimal classifier using minimum number of best combined features.
b. Repeat part a without PCA and Compare the results.