Alexandria University
Faculty of Computer and Data Science
AI Department

Pattern recognition
Assignment

Due to:1/6/2022

Face recognition

Problem Statement

We intend to perform face recognition. face recognition means that for a given image you can tell the subject id. Our database of subjects is very simple. It has 40 subjects. Below we will show the needed steps to achieve the goal of the assignment.

1. Download the dataset and understand the format (10 Points)

a. URL dataset is available at the following link.
https://www.kaggle.com/kasikrit/att-database-of-faces

b. The dataset has 10 images per 40 subjects. Every image is a grayscale image of size 92x112.

2. Generate the Data Matrix and the Label vector (20 Points)

a. Convert every image into a vector of 10304 (92x112) values corresponding to the Image Size.

b. Stack the 400 vectors into a single Data Matrix D and generate the label vector y. The labels are integers from 1:40 corresponding to the subject id.

3. Split the Dataset into Training and Test sets (20 Points)

(don't split randomly, as you might end with a whole class/subject in the test set that the model never saw in the training set)

a. From the Data Matrix D (400x10304) keep the odd rows for training and the even rows for testing. This will give you 5 instances per person for training and 5 instances per person for testing.
b. Split the label vector accordingly.

4. Classification using PCA.(40 points)
   a. Use the pseudo-code below for computing the projection matrix U. Define the alpha ={0.8,0.85,0.9,0.95}
   b. Project the training set and test sets separately using the same projection matrix.
   c. Use a simple classifier (first Nearest Neighbor to determine the class labels).
   d. Report Accuracy for every value of alpha separately.
   e. Can you find a relation between alpha and classification accuracy?

ALGORITHM 7.1. Principal Component Analysis

PCA $(\mathbf{D}, \alpha)$:

1 $\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ // compute mean
2 $\mathbf{Z} = \mathbf{D} - \mathbf{1}\cdot\boldsymbol{\mu}^T$ // center the data
3 $\boldsymbol{\Sigma} = \frac{1}{n}(\mathbf{Z}^T\mathbf{Z})$ // compute covariance matrix
4 $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\boldsymbol{\Sigma})$ // compute eigenvalues
5 $\mathbf{U} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_d\end{pmatrix} = \text{eigenvectors}(\boldsymbol{\Sigma})$ // compute eigenvectors
6 $f(r) = \frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{d}\lambda_i}$, for all $r = 1, 2, \ldots, d$ // fraction of total variance
7 Choose smallest r so that $f(r) \geq \alpha$ // choose dimensionality
8 $\mathbf{U}_r = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r\end{pmatrix}$ // reduced basis
9 $\mathbf{A} = \{\mathbf{a}_i \mid \mathbf{a}_i = \mathbf{U}_r^T\mathbf{x}_i, \text{for } i = 1, \ldots, n\}$ // reduced dimensionality data

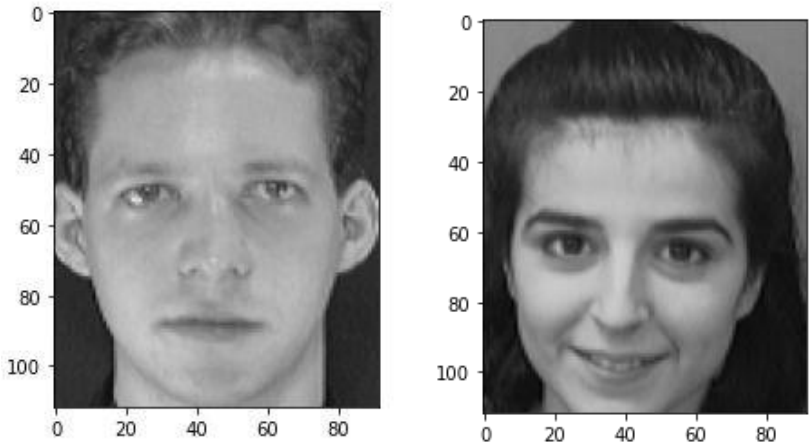
5. **Classifier Tuning using KNN (20 points)**
   a. Set the number of neighbors in the K-NN classifier to 1,3,5,7.
   b. Plot (or tabulate) the performance measure (accuracy) against the K value. This is to be done for PCA as well.

7. Bonus

a. [10 points] Use different Training and Test splits. Change the number of instances per subject to 7 and keep 3 instances per subject for testing. compare the results you have with the ones you got earlier with a 50% split.

b. Using a different classifier for tuning your algorithm.

8. Data sample



Alexandria University
Faculty of Computer and Data Science
AI Department

Pattern recognition
Assignment

Submission Instructions

1. Submit separate codes for the task and put them in a folder called “src”.
2. Comment your codes sufficiently.
3. Keep the datasets in a separate folder called “data”.
4. Submit a README file that will contain the instructions on how to execute your codes and all source codes, report, result, and the README file compressed file (.tar.gz or .zip).
5. The compressed file should be named as [ProjectNo_IDs_Team_NUMBER.zip].
6. Submit it in the same email.

Palagrizim will be severely punished!!!!