

DA311 Machine Learning Lab

Assignment 4

Date: August 29th, 2023

Please download the dataset from the following link:

https://github.com/tsharma12/IITG-DA311-Machine-Learning-Lab/tree/main/Week4_Dataset

Task 1

The gallery folder “**Gallery.zip**” contains images from 40 individuals, each of them providing 5 images. The pixel intensities of the 200 face images will be used for computing the KL Transform. By employing the method of efficient computation of the basis vectors for high dimensional data (discussed in class),

- (i) Display the Eigenface images corresponding to the top 5 Eigenvalues of the covariance matrix Σ .
- (ii) Plot a graph depicting the percentage of the total variance of the original data retained in the reduced space versus the number of dimensions. From this graph, find the number of dimensions required for projecting the face vectors so that:
 - (a) At least 85% of the total variance of the original data is accounted for in the reduced space.
 - (b) At least 95% of the total variance of the original data is accounted for in the reduced space.
- (iii) Reconstruct the image ‘face_input_1.pgm’ using the:
 - (a) Eigenface corresponding to the largest eigenvalue.
 - (b) Top 4 Eigenfaces.
 - (c) Top 15 Eigenfaces.
 - (d) Top 150 Eigenfaces.
 - (e) All the Eigenfaces.

Display the reconstructed image and the mean squared error in each case.

- (iv) Depict graphically the mean squared error obtained for different numbers of Eigenfaces. Note that the Eigenfaces will vary from 1 to 199 (total number of training samples-1).
- (v) Repeat parts (iii) and (iv) for the image face_input_2.pgm. Comment on your results.

Task 2

The test image folder “**Probe.zip**” contains 5 images of each of the 40 individuals.

- (a) Classify the test samples in this folder using a 1-nearest neighbor classifier (with Euclidean distance) in a reduced 25-dimensional subspace. Compute the classification accuracy.
- (b) Depict graphically the recognition accuracies obtained for different numbers of dimensions. For this part, you have to vary the dimensions from 1 to 199 (total number of training samples).