ENEE633/CMSC828C/ENTS669E

Project 1 **Due 11/08/2018**

1 Introduction

The aim of this project is to implement different classifiers to achieve face recognition. You are given a set of faces and their corresponding labels. You need to split the data into training and testing sets and use the training data to train your classifiers. The whole pipeline is described in the following section.

2 Classifiers

In this project, you are requested to implement the following classifiers for face recognition. You can try different scenarios. For example, you can test the effect of expressions, illumination variations, etc. Besides, you can vary the ratio between training data and testing data. Feel free to play with the data!

Bayes' Classifier Assuming the underlying distribution is Gaussian, you need to implement the Maximum Likelihood estimation with Gaussian assumption followed by Bayes' classification.

k-NN Rule Implement the k-Nearest Neighbors (k-NN) rule to classify the test data. Vary k to see its effect and discuss it in the report.

Kernel SVM Implement the Kernel SVM classifier by solving the dual optimization problem (or using stochastic gradient descent on the cost function). Use the Radial Basis Function (RBF) kernel $\exp(-\|\mathbf{x}-\mathbf{y}\|^2/\sigma^2)$ and the polynomial kernel $(\mathbf{x}^\top\mathbf{y}+1)^r$. In order to choose the optimal values of σ^2 and r, you need to use cross-validation.

Boosted SVM Implement the AdaBoost algorithm for the class of *linear* SVMs. Report the improvement of the boosted classifiers with respect to the iterations of AdaBoost, and compare it with the performance of the Kernel SVMs.

PCA You have to implement Principal Component Analysis (PCA) and apply it to the data before feeding it into the classifiers above.

LDA Similar to PCA, you have to implement LDA followed by the training/application of the classifiers.

3 Datasets

You are provided with the following datasets (.mat files). Please run the experiments over at least ONE of the datasets. You are encouraged to experiment with as many datasets as possible! No matter which dataset is used, you should divide it into two disjoint sets, one for training and another for testing.

DATA Cropped images of 200 subjects, 3 images each, each image of size 24x21. The file data.mat has a variable face of size (24x21x600). The images corresponding to the person labeled n, n = 1, . . . , 200 can be indexed in Matlab as face(:,:,3*n-2), face(:,:,3*n-1) and face(:,:,3*n). The first image is a neutral face, the second image has a facial expression, and the third image has illumination variations.

POSE Cropped images of 68 subjects under 13 different poses. You can access the i_{th} pose of the j_{th} subject as pose(:, :, i, j).

ILLUMINATION Cropped images of 68 subjects under 21 different illuminations. You can access the i_{th} illumination of the j_{th} subject as illum(:, i, j). Note that each image has been already reshaped to a vector.

OTHERS You can also try other datasets like YALE dataset (which has expression variations) and PIE dataset.

The classification task is either identifying the subject label from a test image (for datasets DATA, POSE, and ILLUMINATION) or neutral vs. facial expression classification (for dataset DATA). You may apply the Kernel SVM and Boosted SVM to the second task (neutral vs. facial expression) only, but the Bayes' classifier and k-NN must be used for both tasks.

4 Submission Guidelines

In this project, you can use any programming language you want, such as Matlab (recommended), C++, Python, etc. You should not use any online toolbox that provides these classifiers. You need to submit your project in the following structure: A zip file with the name **YourDirectoryID_Proj1.zip** on to ELMS/Canvas. A main folder with the name **YourDirectoryID_P1** and the following sub-folders/files:

Code A folder containing all your codes.

Report Your report should be in PDF format and no more than 5-6 pages (excluding figures). You do not need to explain ML estimation, Bayes' rule, k-NN rule, Kernel SVM, AdaBoost, PCA and LDA in detail. Just concentrate on the experiments, the specific choices of parameters, and discuss the results you have obtained.

Readme A .txt file describing how to run your code.