CS 479/679 Pattern Recognition
Spring 2016 – Prof. Bebis
Programming Assignment 4 - Due: 5/2/2016

In this assignment, you will experiment with two different classifiers for gender classification: SVMs and Bayesian classifier.

Data Set and experiments: The dataset to be used in your experiments contains 400 frontal images from 400 distinct people, representing different races, with different facial expressions, and under different lighting conditions. The 400 images have been equally divided between males and females. Histogram equalization has been applied to each normalized image to account for different lighting conditions. The data, which is available from the course's webpage, contains images of two different sizes: 16x20 and 48x60; you would need to experiment with each image size separately and compare your results. For each classifier, you need to report the **average** error rate using a three-fold cross-validation procedure. For this, we have randomly divided the dataset three times as follows:

>   Fold 1: Training (69M, 65F), Validation (73M, 60F), Test (58M, 75F)
>   Fold 2: Training (62M, 72F), Validation (58M, 75F), Test (80M, 53F)
>   Fold 3: Training (71M, 63F), Validation (67M, 66F), Test (62M, 71F)

Note that the validation set is typically used for parameter optimization. Since you will not need to optimize any parameters in this assignment, use both the validation set and test set for testing purposes by simply combining them into one set. Using each fold, compute the test error and then average all three errors to report the **average** error.

For each image, we have pre-computed its eigen-face representation; you should be training/testing each classifier using the first 30 eigen-features only (i.e., the ones corresponding to the top 30 eigenvectors). The file naming convention for each file is as follows: *trPCA_xx* for training, *valPCA_xx* for validation, and *tsPCA_xx* for testing; see "**descr**" file for more information. Note that the eigenvalues and eigenvectors have been provided in the files *EVs_xx* and *PCs_xx* for completeness, however, you will not need them in your experiments.

Experiment 1: Apply Support Vector Machines (SVMs) for gender classification. You will be using the *LibSVM* implementation. Experiment both with polynomial and RBF kernels as well as different C values. For consistency, try d=1, 2, and 3 for the polynomial kernel (note that LibSVM provides two extra parameters for the polynomial kernel; to be consistent with the lecture, set $\gamma=1$ and $c_0=0$). In the case of the RBF kernel, try $\sigma=1$, 10, and 100. For the C value, try C=1, 10, 100, and 1,000. Report your best results both for the 16x20 and 48x60 datasets. ***Warning:*** make sure that the data is provided in the format required by LibSVM; otherwise, it will not work correctly.

Experiment 2: For comparison purposes, apply the Bayes classifier for the same problem. Model the male and female classes using a Gaussian distribution and use ML estimation to estimate the parameters for each class. Use equal prior probabilities (e.g., $P(\omega_1)= P(\omega_2)$). Compare your results with those obtained using SVMs.

PROJECT REPORT SUBMISSION REQUIREMENTS

1. <u>Cover Page</u>. The cover page should contain Project title, Project number, Course number, Student's name, Date due, and Date handed in.

2. <u>Technical discussion</u>. This section should include the techniques used and the principal equations (if any) implemented.

3. <u>Discussion of results</u>. A discussion of results should include major findings in terms of the project objectives, and make clear reference to any figures generated.

4. <u>Division of work:</u> Include a statement that describes how the work was divided between the two group members.

5. <u>Program listings</u>. Includes listings of all programs written by the student. Standard routines and other material obtained from other sources should be acknowledged by name, but their listings should not be included.

A hard copy is required for items 1-4, submitted to the instructor in the beginning of the class on the due date. Item 5 should be emailed to the instructor, as a zip file, before class on the due date.