Exam 2: Take Home Portion

Instructions: You may use your notes from class, anything on the class website, and your homework solutions to assist you. You may not discuss your solutions with others in the class, and you are expected to do your own work.

The goal: Given a collection of photographs of people, construct a classifier based on gender. Here are some specifics:

1. Separate the data into a "Training" set and a "Validation" set. For the training set, use 7 (out of 13) randomly chosen photos for each gender for a total of 14 in the training set.

2. Method 1: Subspace classifier. For this classifier, each gender will have its own subspace- The subspace will be defined by the mean (average) as the origin, and it will be three dimensional (use the "best" three basis vectors from the SVD).

   The gender of a new face is defined to be the subspace that best matches the given face- That is, the subspace that is closer to the given face (see the second note in the discussion on the next page).

   You should run your algorithm several times using different choices for training, and then settle on the best set. The "error" is the number of faces from the testing set that were misclassified.

   Comment on your script file what your best error was, then send it to me.

3. Method 2: RBF Classifier. In this case, we will input a face, and output the gender.

   Write the RBF script yourself using two centers and Gaussian transfer function (use our function `rbf1.m`, then solve the RBF equation,

   $$W\Phi + \mathbf{b} = \mathbf{y}$$

   You may run this several times to see if you can find a best width for the two Gaussians, and also with different training sets to see if this value changes- perhaps the distance between centers is an interesting number?

   Comment on your best error at the end of your script file, then send it to me.

Be sure to read the section at the end of this document to learn more about the data we will use in this part.

Discussion of the Exam Problems

1. The Data: Consists of 26 photos, 13 of men, 13 of women. Each photo was originally 162×149 pixels, or a vector in $\mathbb{R}^{24,138}$. The data is in the file `MenAndWomen2.dat` on the class website, and is conveniently stored as a matrix Y that is 24138×26. The first 13 columns are women, the next 13 columns are men.

   If you would like to see what a photo is (for example, the first column), you can type:

   ```
   imagesc(reshape(Y(:,1),162,149))
   colormap(gray)
   axis equal; axis off
   ```

   If you have a matrix M consisting of your 7 men for classification (the matrix M would be $24,138 \times 7$), then the mean male face would be the following- (this is the "origin" of which we spoke for the subspace classifier). In constructing the basis vectors, we ought to subtract the mean from the other columns:

   ```
   MeanMale=mean(M,2);                                    %This is 24138 x 1
   MeanSubtracted=M-repmat(MeanMale,1,7);
   ```

   Now the data is ready for the SVD.

2. In the first problem, we define a "subspace" by an origin, $\mathbf{o}$, and a set of o.n. basis vectors stored as columns in some matrix, W. Then if $\mathbf{x}$ is a vector, we'll define the distance to this subspace as:

   $$\|(\mathbf{x} - \mathbf{o}) - WW^T(\mathbf{x} - \mathbf{o})\|$$

   To do the classification, compute this value for each of the two subspaces. If you're doing them in order, the first one is for women, the second is for men, and it will be easy to keep track of the misclassifications (see the template script file).

3. For the RBF, we're training and simulating by hand using `rbf1.m`. Important: Keep track of the sizes of your matrices as you work- especially at the end, when we incorporate the bias vector $\mathbf{b}$ into the matrix W (in our class notes, we constructed $\hat{W}\hat{\Phi}$.

   Furthermore, for the RBF, we need to create targets- In this case, we make the vector $[1, 0]^T$ the target for "women", and $[0, 1]^T$ the target for "men" (see the template script file).

   For the pseudo-inverse, use the `pinv` command rather than the slash command or the SVD itself.