

**Winter 2002, CS 790**  
**Introduction to Pattern Recognition**  
**Homework #2**  
**Due date: 2/11**

*In recognition of the Wright State University policies of academic integrity, I certify that I have neither given nor received dishonest aid in this homework assignment.*

Name: \_\_\_\_\_ Signature: \_\_\_\_\_

**NOTES:**

1. Start each problem on a separate page
2. Provide hardcopies of all plots, MATLAB code and derivations.
3. Download the compressed file 'hw2.zip' from the course web page
4. Sign and return this page with your finished assignment.

**Problem 1 (10%)**

In a two-class, three dimensional classification problem, the feature vectors in each class are normally distributed with covariance matrix

$$\Sigma = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}$$

The respective mean vectors are  $[0 \ 0 \ 0]^T$  and  $[1 \ 1 \ 1]^T$ . Assume that  $p(\omega_2) = 2 \cdot p(\omega_1)$ . Derive the corresponding linear discriminant functions and the equation describing the decision surface. Use this decision surface to classify the example  $[0.1 \ 0.7 \ 0.8]^T$ .

**Problem 2 (10%)**

Generate 200 examples for each of the three densities:

$$\begin{array}{lll} \mu_1 = [1 \ 3 \ 3]^T & \mu_2 = [4 \ 0 \ 3]^T & \mu_3 = [0 \ -4 \ 0]^T \\ \Sigma_1 = \begin{bmatrix} 3 & 2 & -1.9 \\ 2 & 3 & 0 \\ -1.9 & 0 & 3 \end{bmatrix} & \Sigma_2 = \begin{bmatrix} 6 & -2.5 & -1.9 \\ -2.5 & 6 & -0.5 \\ -1.9 & -0.5 & 4 \end{bmatrix} & \Sigma_3 = \begin{bmatrix} 5 & -1.5 & -0.9 \\ -1.5 & 1.0 & -0.1 \\ -0.9 & -0.1 & 1 \end{bmatrix} \end{array}$$

To simulate noisy channels, augment the feature vector with 17 additional dimensions, each one from a Gaussian distribution  $N(\mu=0; \sigma=10)$  regardless of class. As a result, you will have 600 (20-dimensional) examples.

- (i) Compute the first two principal components of the data, and generate a scatter plot of the 600 examples. Use the 'text' function to plot examples by their class label. Use the 'axis' function to adjust the limits of the plot
- (ii) Compute the linear discriminant projections of the data, and generate a scatter plot of the 600 examples. Use the 'cs790\_lda' implementation included in 'hw2.zip'
- (iii) Discuss your results

### Problem 3 (20%)

Implement a Quadratic classifier for Problem 2

- (i) Split the data into training and test sets by randomly selecting 20% of the examples from each class for the test set.
- (ii) Classify the test data in the original 20-dimensional space. What is the average classification rate?
- (iii) Compute the PCA projections using ONLY the training data
- (iv) Project both training and test data using the first two PCA eigenvectors
- (v) Classify the test data in the 2-dimensional PCA subspace. What is the average classification rate?
- (vi) Compute the LDA projections using ONLY the training data
- (vii) Project both training and test data using the LDA eigenvalues
- (viii) Classify the test data in the 2-dimensional LDA subspace. What is the average classification rate?
- (ix) Discuss your results

### Problem 4 (10%)

Generate 150 examples from an N-dimensional distribution (N=10) with zero mean and unit covariance matrix. HINT: use the MATLAB command 'mvnrnd'

- (i) Arbitrarily assign the examples to one of three classes
- (ii) Compute the linear discriminant projections of the data, and generate a scatter plot of the examples. Use the 'text' function to plot examples by their class label
- (iii) Repeat steps (i) and (ii) using N=50, 75 and 100 dimensions
- (iv) Discuss your results

### Problem 5 (20%)

Implement a k Nearest Neighbor classifier for Problem 4

- (i) Split the data into training and test sets by randomly selecting 20% of the examples from each class for the test set
- (ii) Compute the LDA projections using ONLY the training data

- (iii) Project both training and test data using the LDA eigenvalues and generate a scatter plot. HINT: mark each test examples with an asterisk (e.g., a test example from class 1 would be shown as 1\*) so you can distinguish test data from training data
- (iv) Classify the test data and provide an average correct classification rate
- (v) Repeat steps (i) through (iv) for N=10,50,75 and 100
- (vi) Discuss your results

### Problem 6 (30%)

In the previous problems you had the opportunity to develop and evaluate two classifiers using synthetic data generated on your own. Things get a little more interesting now, because we are going to use a blind test to evaluate the performance of your implementations.

Extract the dataset 'hw2p6\_data.mat' from 'hw2.zip', which contains the following matrices:

- x1: training set (row vectors)
- clab1: training set labels
- x2: test set (row vectors)

You have to build a classification system using the training set (for which you have the correct labels) and generate class labels for the test set, (for which only I have the correct labels). You can assume I generated both sets using the same distribution.

- (i) Perform dimensionality reduction to visualize the structure of the data. Which technique does a better job at unfolding the structure of the data? Why?
- (ii) Split your training set as in Problems 3 and 5 to determine which classification approach will work best. Return the class labels for the test data using the approach that yields highest classification rate. In addition to providing hardcopies, send me the predicted labels in the BODY of an email message using the format below. Please do not use HTML formatting, just plain ASCII.

```

uclass2 = [
    3
    1
    2
    2
    3
    ...
];

```