**Winter 2001, CS 790**
**Introduction to Pattern Recognition**
**Homework #1**
**Due date: 1/28**

*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.*

*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 'hw1.zip' from the course web page*
4. *Sign and return this page with your finished assignment.*

**Problem 1 (25%)**

Given the set of patters for the digits {1,2,3,4} shown in Figure 1, design a set of features (not more than 10) that provides good separability between all classes. Discuss your rationale. Generate Matlab code to extract these features for each pattern. Generate two-dimensional scatter plots for each pair of features (e.g., f1 vs. f2, f3 vs. f4, etc) and select the two features that provide the best separability. Are the classes linearly separable?

NOTES:

- An electronic copy of these 16 patterns can be loaded from the file "hw1p1_data.mat" ("load hw1p1_data"). Each example is a 1x64 row vector. You may use the command "reshape" to convert each example into an 8x8 image.
- Use the command "text" to generate scatter plots so the examples are labeled.

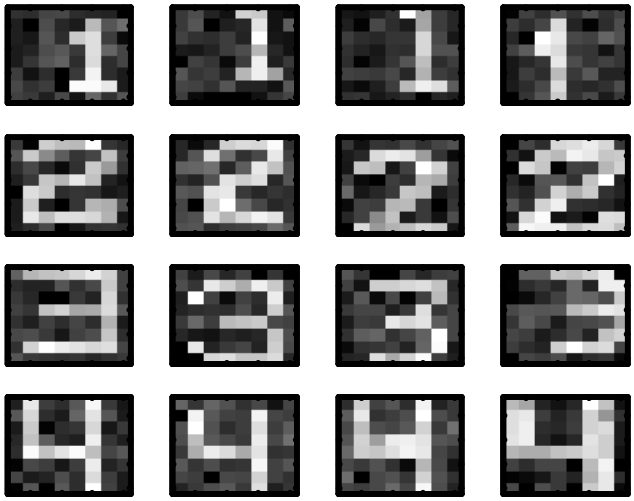

**Figure 1. Dataset of patterns for the first four digits**

**Problem 2 (20%)**

Load the binary file 'hw2p2_data.mat' into MATLAB. This will create a matrix "x", which consists of 100-dimensional feature vectors arranged by ROWS:

(i) Generate a scatter plot of the first three dimensions. Rotate the reference frame (command 'rotate3d') and try to find structure in the data by changing the viewpoint. Plot the best orientation that you can find. Can you identify any structure?
(ii) Compute the Principal Components of the data and generate a plot of the eigenvalues, sorted in decreasing order. How many eigenvalues are responsible for most of the variance in the data?
(iii) Generate a scatter plot of the first three PCA projections (the ones with largest eigenvalues). Rotate the reference frame and try to find structure in the data by changing the viewpoint. Plot the best such orientation. Can you identify any structure?
(iv) Discuss your findings.

**Problem 3 (25%)**

Consider a medical diagnosis problem where a fast biochemical test is used for screening patients. The test returns a result close to ZERO for healthy patients and close to ONE for infected patients, according to the following likelihood functions:

$$P(x \mid \omega_1) = N(\mu = 0,\ \sigma = 0.3)$$
$$P(x \mid \omega_2) = N(\mu = 1,\ \sigma = 0.1)$$

Assuming that, on average, 1 out of 10,000 patients is infected, and the following misdiagnosis costs:

- Diagnosing a healthy patient as "infected": expected $20,000 in medical bills for a comprehensive in-patient procedure.
- Diagnosing an infected patient as "healthy": expected $1M settlement for medical malpractice.

Analytically determine a decision rule for each of these criteria:

(i) Maximum Likelihood
(ii) Maximum A Posteriori
(iii) Minimum Bayes Risk

Discuss your results.

*DISCLAIMER: This problem is not intended to reflect any realistic medical scenario.*

**Problem 4 (10%)**

Generate N=100 random numbers from a Gaussian density N($\mu$=10,$\sigma$=3). Plot a histogram of the random numbers using 50 bins. Plot the theoretical density on the same figure. Do the theoretical and experimental distributions match? Repeat the same experiment for N=1000, 10000, 100000 and 1000000. Comment on the results.

HINT: Remember that the mass (area) of the histogram has to be equal to ONE.

**Problem 5 (20%)**

Consider a two-class uni-variate classification problem with equal priors and the following likelihood densities:

$$P(x \mid \omega_1) = N(0,1)$$
$$P(x \mid \omega_2) = N(3,1)$$

(i) Determine the optimal decision rule for minimizing the probability of error.
(ii) Determine the theoretical probability of error (HINT: use numerical integration with 'quad' or the error function 'erf').
(iii) Generate a dataset of test examples (100 examples per class). Label each example according to the p.d.f. from which it was drawn (its true label).
    a. Classify each example according to the decision rule derived in (i).
    b. Compare the predicted class label against the true class label of each example and estimate the correct classification rate of the decision rule.
    c. How does this classification rate compare with the probability of error in (ii)?
(iv) Repeat part (iii) for 1,000 test examples
(v) Repeat part (iii) for 1,000,000 test examples
(vi) Discuss your results.