EEL6825 – Fall 99

Exam 2 November 17, 1999

NAME: ______________________________

This exam is open-book and calculator. You may use any books or papers that you like. There are four problems and one extra credit in this exam, you have three full class periods. State your assumptions and reasoning for each problem. Justify your steps and clearly indicate your final answers.

1	**/25**
2	**/25**
3	**/25**
4	**/25**
*****	**/5**
TOTAL	

1. (25 points)

   Class ω_1 points are given by:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

   Class ω_2 points are given by:

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

   Answer the following questions:

   (a) (5 points) Classify the point $\mathbf{x} = [1 \ 1 \ 0]^T$ using the 3-NN voting classification technique.

(b) (5 points) Compute $\hat{p}(\mathbf{x})$ at the location $\mathbf{x} = [1\ 1\ 0]^T$ for class ω_1. Use a Parzen Windows pdf estimate with a uniform window of $h = 3$.

(c) (5 points) Compute $\hat{p}(\mathbf{x})$ at the location $\mathbf{x} = [1\ 1\ 0]^T$ for class ω_2. Use a Parzen Windows pdf estimate with a uniform window of $h = 3$.

(d) (5 points) Compute $\hat{p}(\mathbf{x})$ at the location $\mathbf{x} = [1\ 1\ 0]^T$ for class ω_1. Use a k nearest neighbor (volumetric) pdf estimate with $k = 1$.

(e) (5 points) Compute $\hat{p}(\mathbf{x})$ at the location $\mathbf{x} = [1\ 1\ 0]^T$ for class ω_2. Use a k nearest neighbor (volumetric) pdf estimate with $k = 1$.

2. (25 points)

Class ω_1 points are given by:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Class ω_2 points are given by:

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Answer the following questions regarding the use of a neural network to solve this two-class classification problem. Assume the sigmoid activation function of the neural network to be:

$$f(a) = \begin{cases} 1 & \text{if } a > 0 \\ -1 & \text{else} \end{cases}$$

(a) (5 points) Draw the simplest neural network configuration that can correctly classify all of the data points. (Note: you don't have to include actual weight values here since you will compute them in the subsequent parts of this problem.)

(b) (10 points) The final output of your neural network should be +1 for class 1 and -1 for class 2. Do you need a hidden layer for this problem? Explain. If you require a hidden layer, provide all the weight values for the hidden layer. Explain your reasoning.

(c) (10 points) Compute all the remaining weight values (e.g. output layer). The final output of your neural network should be +1 for class 1 and -1 for class 2. Explain your reasoning.

3. (25 points)

Class ω_1 points are given by:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Class ω_2 points are given by:

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Answer the following questions regarding the KL dimensionality reduction of this problem.

(a) (5 points) Estimate the combined covariance matrix ($\hat{\Sigma}$) of the data. Assume equal apriori probabilities if you need to. Make sure to use an unbiased estimator.

(b) (5 points) If you correctly computed part (a) then the eigenvectors of $\hat{\Sigma}$ are

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

Note: These eigenvectors are listed in no particular order and are not normalized. Compute the normalized eigenvectors and their respective eigenvalues. If you realize that you did part (a) wrong, clearly state this or, better yet, find your mistake.

(c) (5 points) If you had to drop a single feature, which feature would it be? What is the error that you would then pay in terms of representation and in terms of classifier performance?

(d) (10 points) Map the four sample points to the new two-dimensional space. What are the numeric values of the four new locations.

4. (25 points) Short Answer.

   (a) (5 points) We all know that the resubstitution error of a 1-NN classifier is always 0%. How can this be since this outperforms the "optimal" Bayes classifier on most problems? Assume the Bayes classifier is designed using the correct distributions and parameters and that we have an extremely large number of data points.

   (b) (5 points) In order to separate his data into training and test sets, a student computes the magnitude of each point. He then places the smallest 50% of the points into the training set and the remaining points in the test set. Discuss the merits of this technique.

(c) (5 points) We showed in class that increasing k in k-NN classification leads to improved performance. Why then is 1-NN the most popular classifier in practical implementations? (Ignore computational considerations).

(d) (5 points) Does this program properly implement the d-squared plot from homework 3 for the **one dimensional** class data given. Why or why not? Do not consider efficiency, generality, elegance or cleverness (Jeremy developed these programs).

```
% Program to generate d^2 plot
%  assumes 4 total data points and 2 from each class
close all      % close all figure windows
clear all      % clear the matlab workspace
a = [1,2];     % class one points
b = [-1,-2];   % class two points
c = [a,b];     % vector of all data
for i=1:4      % loop through all of the data
        for j=1:2
                dist_to_a(j) = ((c(i)-a(j))^2)^.5 ;  % class 1 distances
                dist_to_b(j) = ((c(i)-b(j))^2)^.5 ;  % class 2 distances
        end
        dist_to_nearest_a(i) = min(dist_to_a); % minimum class 1 distance
        dist_to_nearest_b(i) = min(dist_to_b); % minimum class 2 distance
end
figure
plot(dist_to_nearest_a,dist_to_nearest_b,'b*'); % d-squared plot
```

(e) (5 points) Which of these three Matlab program segments has an error and what is the error? Hint: one of the program will cause an error in Matlab and the other two will not. Again, do not worry about efficiency, etc.

i.
```
% Program segment to process NBA player data
nba = [2 6 6 220; 1 6 3 190 ; 3 6 10 240 ]; % 3 NBA players
[n,d]=size(nba);    % n is total number of players
n1=0; % number of guards
n2=0; % number of forwards
n3=0; % number of centers
for i=1:n                     % loop through all of the players
   if nba(i,1)==1             % check if it is a guard
      n1=n1+1;                % if so, increment number of guards
      height1(n1)=nba(i,2)+nba(i,3)/12;  % copy height
      weight1(n1)=nba(i,4);              % copy weight
   elseif nba(i,1)==2         % do same procedure for forwards
      n2=n2+1;
      height2(n2)=nba(i,2)+nba(i,3)/12;
      weight2(n2)=nba(i,4);
   else                       % remaining case must be centers
      n3=n3+1;
      height3(n3)=nba(i,2)+nba(i,3)/12;
      weight3(n3)=nba(i,4);
   end
end
```

ii.
```
% Program segment to plot a sine wave
figure                 % open a new figure
N=1000;                % number of points in sine wave
omega=.05;             % frequency
phase=.1;              % phase
x=0:(N-1);             % initialize input
y=zeros(1,N);          % initialize output to zero

for i=0:(N-1)          % loop through N times
  y(i)=sin(omega*x(i)+phase);
end

plot(x,y,'b*');        % finally plot the sine wave
```

iii.
```
% Program to plot normal PDF approx using sum of N uniform variables
figure               % open a new figure
N=10;                % number of uniform PDFs
n=10000;             % number of samples in each uniform PDF
m=40;                % number of bins to plot
%set up the PDF's
Z=zeros(1,n);        % initialize uniform PDF to all zeros
for i=0:(N-1)        % loop through N times
  Z = Z + rand(1,n); % Add up N independent samples of uniform PDF
end
%histogram of Z using m bins
hist(Z,m);           % finally plot histogram with m bins
```

5. (5 points) EXTRA CREDIT Consider the following classifier: Find the distance to the closest point in ω_1 and call this ℓ_1. Find the distance to the closest point in ω_2 and call this ℓ_2. Classify as class ω_1 if $\ell_1 \leq \alpha\ell_2$ and as class ω_2 if $\ell_1 > \alpha\ell_2$ where α is a carefully chosen positive constant. Assume N_1 data points in ω_1 and N_2 data points in ω_2. How would you best choose the value of α?