In The Name of God, The Compassionate, The Merciful

Name: .....

Student ID#: .....

Statistical Pattern Recognition (CE-725)
Department of Computer Engineering
Sharif University of Technology

Midterm Exam - Spring 2010
(120 minutes - 110 points)

1. Overall Concepts: True or False (10 points)

a. ..... The error surface followed by the gradient descent Backpropagation algorithm changes if we change the training data.

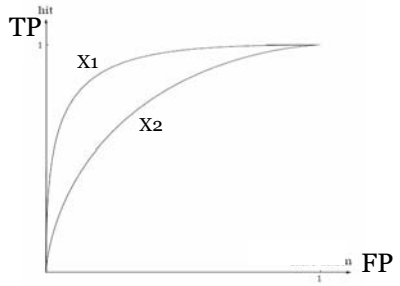
b. ..... In a two category case, if $p(x|w_1)$ and $p(x|w_2)$ are normal distributions with the same variance but different means, there exist a linear discriminant function which gives the same decision for all inputs as the minimum error rate classifier.

c. ..... Given that all $p(x|w_i)$ are normal distributions, the decision region for each w_i from the minimum error rate classifier is always a single connected region.

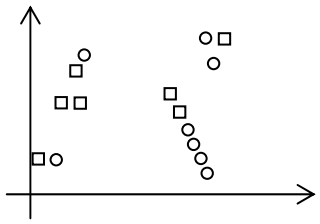
d. ..... Regardless of the true underlying model, the minimum error rate classifier using the maximum likelihood estimate gives always the smallest average classification error within the assumed model.

2. Features and classifiers performance (15 points)

a. For a detection problem, two different features x_1 and x_2 give two different ROC curves shown in the following figure. Which feature x_1 or x_2 gives a better performance? Why?



b. Consider the following data points. What is the accuracy of k-nearest neighbor on them using one-leave-out cross validation method?

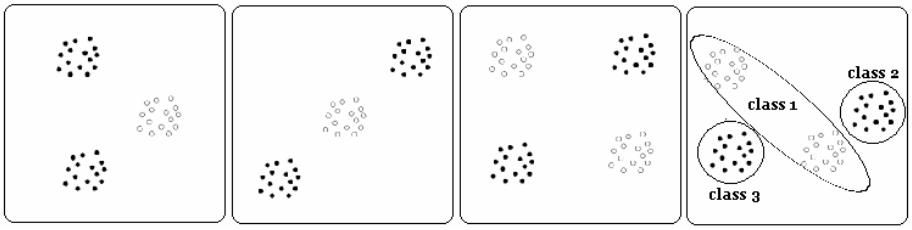


k	k-nearest neighbor accuracy
1	
2	
3	

3. Probabilistic methods (30 points)

a. In each of the following data sets (each data set in a figure) we would like to learn a classifier that achieves zero training error on this data. To do that we allow each classifier to divide the data into more than two classes, however, for each classifier there must be a subset of the classes that perfectly match class A and the

complementary set of classes must match cluster B. For example, in the last figure classes 2 and 3 contain all of one's points and class 1 contains all of other's points and so this classification is a legitimate solution to this problem.



For a Gaussian Bayes classifier and for each of above figures, state the minimum number of classes required to achieve the above goal. In all cases assume equal class priors.

	Min number of classes
Figure 1	
Figure 2	
Figure 3	
Figure 4	3

b. Consider a one dimensional two category classification problem. Suppose the class conditional densities are given by $p(x|c_1)=N(2,1)$ and

$$p(x|c_2)=\begin{cases}1/3 & 0\le x\le 3\\ 0 & \text{otherwise}\end{cases}$$

b1. Design the ML classifier.

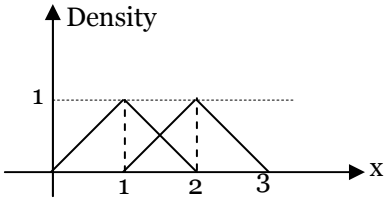
b2. Suppose priors are $P(c_1)=2/3$ and $P(c_2)=1/3$. Design the MAP classifier.

B3. Suppose priors are equal and the loss functions are as follows: loss for deciding c_1 when the true class is c_2 is 30, loss for deciding c_2 when the true class c_1 is 20. Loss for deciding the true class is 0. Design the Minimum Bayes risk classifier.

In each case above, give the decision boundaries and decision regions.

4. Non parametric modeling (20 points)

Suppose that the class conditional densities are as bellow:



We Sample two sets each of size n, from each class. Assume that the method of Nearest Neighbor as the classification rule. We want to classify the point x=1.8. For parts a and b assume that h=0.2. For parts c and d, h is arbitrary.

a. What is the probability that the distance of the nearest neighbor point of the first class to x is greater than h?

b. What is the probability that the distance of the nearest neighbor point of the second class to x greater than h?

c. What is the probability that the nearest neighbor point to x is from the first class? (Hint: Use the results of parts a and b).

In The Name of God, The Compassionate, The Merciful

d. Repeat a, b and c for the case of k-nearest neighbor (that is find the probability of the distance of the k-nearest neighbor of x is greater than h).

5. Learning (20 points)

a. Is the classification problem of "assigning a binary string to w_1 if the number of non-zero components be odd and to w_2 otherwise" linearly separable? Why?

b. Given the following set of prototypes:

S1: (0, 0), (1, 1) S2: (0, 1), (1, 0)

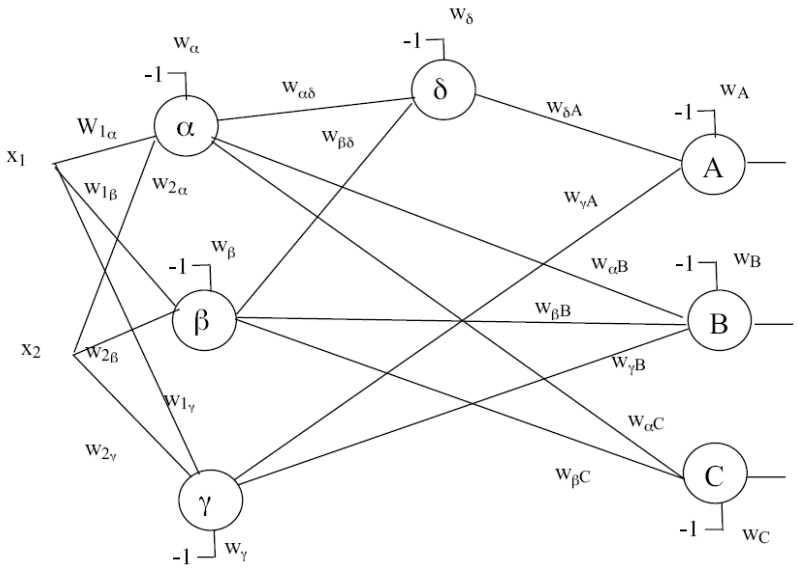
Work in augmented space. Apply pseudo inverse procedure to finding the w, considering $b = (1, 1, 1, 1)^T$, and note that:

$$\text{If } A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & 0 & -1 \end{bmatrix} \text{ then } (A^TA)^{-1}A^T = A^- = \frac{1}{2}\begin{bmatrix} -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 3/2 & -1/2 & -1/2 & -1/2 \end{bmatrix}$$

Does the resulting weight vector classify all prototypes correctly? Why?

6. Artificial Neural Networks (15 points)

a. Consider the following network with the specified weights:



α		β		γ		δ		A		B		C	
$w_{1\alpha}$	0	$w_{1\beta}$	1	$w_{1\gamma}$	0	$w_{\alpha\delta}$	1	$w_{\delta A}$	1	$w_{\alpha B}$	1	$w_{\alpha C}$	-1
$w_{2\alpha}$	1	$w_{2\beta}$	0	$w_{2\gamma}$	1	$w_{\beta\delta}$	-1	$w_{\gamma A}$	1	$w_{\beta B}$	1	$w_{\beta C}$	1
w_{α}	0	w_{β}	4	w_{γ}	3	w_{δ}	0.5	w_{A}	0.5	$w_{\gamma B}$	-1	w_{C}	0.5
										w_{B}	1.5		

How does this network partition the 2D space? (figure out the space and regions).

b. Prove that any multilayer neural net with linear neuron units (i.e., $f(p)=\alpha p$, where α is a constant) can be represented by a single layer net (also with linear neuron units).

Good Luck!