COLLEGE OF ENGINEERING & TECHNOLOGY

Department : Computer Engineering
Lecturer : Dr. Manal Helal
Course Name : Pattern Recognition
Course Code : CC716 **Total Marks:**
Date : Jan. 19-2016 **Start time : 3 pm Time allowed: 1.5 Days**

Final Examination Paper

Answer the following questions:

Exam Instructions:

- This is an open Book(s) and slides exam.
- You can use matlab, R, python or any package installed in your machine if required.
- No Internet access is allowed.
- The exam will be curve graded on best effort, and worth 20% of your final mark.

Statistical Classification **[7 points]**

1) When learning a logistic regression classifier, you run gradient ascent for 50 iterations with the learning rate, η=0.3, and compute J(θ) after each iteration. Compute the conditional log-likelihood J(θ) after each iteration (where θ denotes the weight vectors). You find that the value of J(θ) increases quickly then levels off. Based on this, which of the following conclusions seems most plausible?

A) Rather than use the current value of η, it'd be more promising to try a larger value for the learning rate (say η =1.0).

B) η =0.3 is an effective choice of learning rate.

C) Rather than use the current value of η, it'd be more promising to try a smaller value (say η =0.1).

2) Consider a binary classification problem with variable $X_1 \in \{0, 1\}$ and label $Y \in \{0, 1\}$.
The true generative distribution $P(X_1, Y) = P(Y)P(X_1|Y)$ is shown as Table 1 and Table 2.

Table 1: P(Y)

Y = 0	Y = 1
0.8	0.2

Table 2: $P(X_1|Y)$

	$X_1 = 0$	$X_1 = 1$
Y = 0	0.7	0.3
Y = 1	0.3	0.7

Now suppose we have trained a Naive Bayes classifier, using infinite training data generated according to Table 1 and Table 2. In Table 3, please write down the predictions from the trained Naive Bayes for different configurations of X_1. Note that $\hat{Y}(X_1)$ in the table is the decision about the value of Y given X_1. For decision terms in the table, write down either $\hat{Y} = 0$ or $\hat{Y} = 1$; for probability terms in the table, write down the actual values (and the calculation process if you prefer, e.g., 0.8 * 0.7 = 0.56).

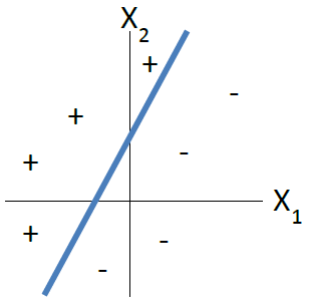
Table 3: Predictions from the trained Naive Bayes

	$P(X_1, Y = 0)$	$\hat{P}(X_1, Y = 1)$	$\hat{Y}(X_1)$
$X_1 = 0$			
$X_1 = 1$			

What is the expected error rate of this Naive Bayes classifier on testing examples that are generated according to Table 1 and Table 2? In other words, P ($\hat{Y}$ (X_1) ≠ Y) when (X_1, Y) is generated according to the two tables. Hint: P ($\hat{Y}$ (X_1) ≠ Y) = P ($\hat{Y}$ (X_1) ≠ Y, X_1 = 0) + P ($\hat{Y}$ (X_1) ≠ Y, X_1 = 1).

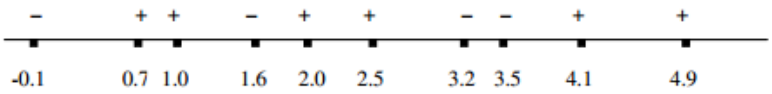
Linear Classification **[3 points]**

3) What are the weights w_0, w_1, and w_2 for the perceptron whose decision surface is illustrated below? You should assume that the decision surface crosses the X_1 axis at -5 and crosses the X_2 axis at 8.



Non- Linear Classification **[15 points]**

4) Consider the following dataset with one real-valued input and one binary output (+ or -). The following questions assume that we are using k-nearest-neighbor learning with unweighted Euclidean distance to predict y for an input x. **[2 points]**



A) What is the leave-one-out cross-validation error of 1-NN on this dataset. Give your answer as the number of misclassifications and circle them in the diagram above.

B) What is the leave-one-out cross-validation error of 3-NN on this dataset. Give your answer as the number of misclassifications and circle them in the diagram below.

5) Consider the following set of training examples: **[5 points]**

Example	Is Heavy?	Is Smelly?	Is Spotted?	Is Smooth	Is Poisonous?
A	0	0	0	0	0
B	0	0	1	0	0
C	1	1	0	1	0
D	1	0	0	1	1
E	0	1	1	0	1
F	0	0	1	1	1
G	0	0	0	1	1
H	1	1	0	0	1
U	1	1	1	1	?
V	0	1	0	1	?
W	1	1	0	0	?

You know whether or not mushrooms A to H are poisonous, but you do not know about U to W. Consider only mushrooms A to H in questions A and B.

A) What is the entropy of "Is Poisonous"?

B) What attribute should you choose as the root of a decision tree? Hint: You can decide without computing the information gain for all four attributes.

C) What is the information gain of the attribute you chose in the previous question?

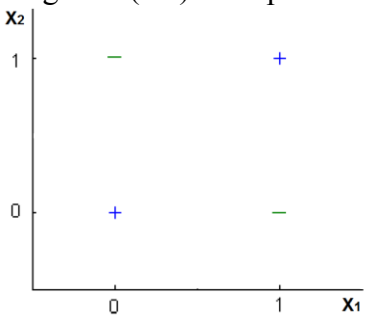
D) Build the decision tree to classify mushrooms as poisonous or not.

E) Decide whether U to W mushrooms are poisonous or not.

COLLEGE OF ENGINEERING & TECHNOLOGY

Department : Computer Engineering
Lecturer : Dr. Manal Helal
Course Name : Pattern Recognition
Course Code : CC716 **Total Marks:**
Date : Jan. 19-2016 **Start time : 3 pm Time allowed: 1.5 Days**

6) Consider a classification problem with two Boolean variables $X_1, X_2 \in \{0, 1\}$ and label $Y \in \{0, 1\}$. In the Figure, two positive ("+") and two negative ("-") examples are shown. **[4 points]**



A) Draw (or just simply describe) a decision tree that can perfectly classify the four examples in the Figure

B) What will happen if you try to train a Gaussian Based Bayed Classifier on such a dataset? Assume that the classifier is able to learn arbitrarily covariance matrix.

C) Suppose we learn a Naive Bayes classifier from the examples in the Figure, using MLE (maximum likelihood estimation) as the training rule. Write down all the parameters and their estimated values (note: both P(Y) and P (X_i|Y) should be Bernoulli distributions). Also, does this learned Naive Bayes perfectly classify the four examples?

D) Is there any logistic regression classifier using X_1 and X_2 that can perfectly classify the examples in the Figure? Why?

7) Describe how multiclass datasets are handled in Bayesian, perceptron, neural networks, and SVM classification algorithms? **[4 points]**