CSCI567 Machine Learning (Fall 2008) Assignment #2

Instructor: Dr. Sofus A. Macskassy
TA: Cheol Han

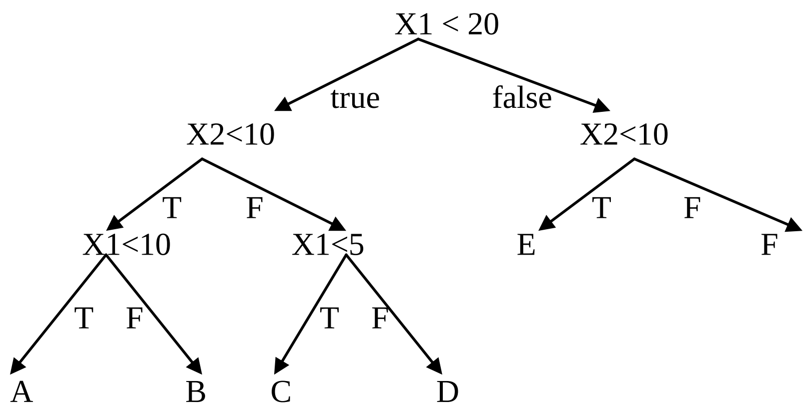
Due time: 5:00pm, Sep 30, 2008

Student Name: ______________________

Student ID: ______________________

Score: ______________________

1. (**Decision Tree**, 25 points) Consider the following decision tree:



(a) [10 points] Draw the decision boundaries defined by this tree. Each leaf of the tree is labeled with a letter, write this letter in the corresponding region of instance space.

(b) [15 points] Give another decision tree that is syntactically different but defines the same decision boundaries. This demonstrates that the space of decision trees is syntactically redundant. Is this redundancy a statistical problem (i.e., does it affect the accuracy of the learned trees)? Is it a computational problem (i.e., does it increase the computational complexity of finding an accurate tree)?

2. (**Decision Tree**, 20 points) Consider the following training examples:

X1	X2	X3	Y
0	1	0	0
1	1	0	0
0	0	1	1
0	0	0	0
1	0	1	0
0	1	1	1

What feature would be chosen for the split at the root of a decision tree using the mutual information criterion? Show your work (step by step).

3. (**Mutual Information**, 15 points) When constructing decision trees, split-points are determined based on mutual information gain. Prove that when computing mutual information for a continuous variable, split-points where the labels on each side of the threshold are the same will never be the optimal split. (Hint: an example is on slide #13 on the lecture 7 slides, e.g. using 1.5 will never be the optimal.)

4. (**Logistic Regression**, 15 points)
Consider the loss function:

decision	true label y	
	0	1
predict 0	0	2
predict 1	5	0

(a) [10 points] Derive the linear decision boundary for when logistic regression should predict 1. Hint: look at lecture 4 notes, slide 41 for an outline of how to do this. I am looking for a linear equation similar to the last equation on that slide. Show your work.

(b) [5 points] Generalize. What is the linear decision boundary for an arbitrary loss function where L(0,0)=L(1,1)=0 and where L(0,1) and L(1,0) are both positive values. Give your answer in terms of **w**, **x**, L(0,1) and L(1,0).

5. (**LDA,** 25 points) Based on Problem 4, page 103 in Alpayden book.

Derive

$$\log \frac{P(y=1\,|\,\mathbf{x})}{P(y=0\,|\,\mathbf{x})}$$

For the following discriminant functions (see Table 5.1 or slide 28 in the lecture 5 slides):

(a) [10 points] shared, hyperspheric

(b) [15 points] different, hyper-ellipsoidal