10-601 Machine Learning: Assignment 4

- The assignment is due at 3:00pm (beginning of class) on **Monday, February 25, 2008**.
- Since you only have five days to work on the assignment, it will be worth **30 points**.
- Write your name at the top right-hand corner of each page submitted.
- Each student must hand in a hard-copy writeup. See the course webpage for collaboration policies.

Q1: Probably approximately correct (PAC) learning [15 pts]

Consider a variant of a decision tree learning algorithm that considers only examples described by boolean features $\langle X_1, \ldots X_n \rangle$, learns only boolean-valued functions ($Y \in \{+, -\}$), and outputs only 'regular, depth-2 decision trees.' A 'regular, depth-2 decision tree' is a depth two decision tree (a tree with four leaves) in which the left and right child of the root are *required to test the same attribute*. For instance, the following tree is a 'regular, depth-2 decision tree':

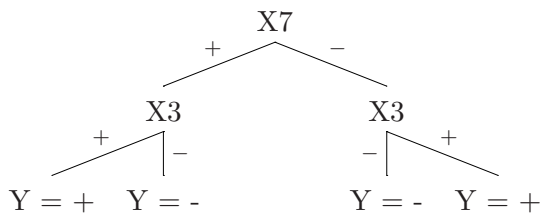

Figure 1: Example of regular boolean depth-2 decision tree.

1. Suppose you have noise-free training data for target concept c which you know can be described by a regular, depth-2 decision tree. How many training examples must you provide the learning algorithm in order to assure that with probability .99 the learner will output a tree whose true accuracy is at least .97? Assume you have data with 20 attributes in total (though of course you believe only two of these twenty will be needed to describe the correct tree).
2. [**Entirely optional, extra credit**] Now suppose you modify the algorithm to allow for instances to contain real-valued attributes instead of boolean attributes, and you allow each decision tree node to perform a boolean threshold test of the form $X_i > a$ where a is allowed to take on any real value. The tree is further constrained such that the second level nodes, which share the same attribute, must also use the same threshold. In this case, re-answer the above question: How many training examples must you provide the learning algorithm in order to assure that with probability .99 the learner will output a tree whose true accuracy is at least .97? In this case, assume each example has only two attributes, so the tree will end up using both.

Q2: Bayes nets [15 pts]

For the following questions, assume M_{indp} and M_{dep} are as defined in **Question 3** of **Homework 2**, but with $\mathbf{n = 4}$ instead of 10:

1. Draw a Bayes net representing the conditionally *independent* naïve Bayes model M_{indp}. How many total rows are in all of the fully specified conditional probability tables (cpt's) represented by this graph?

2. Draw a Bayes net representing the conditionally *dependent* model M_{dep}. How many total rows are in all of the fully specified cpt's represented by this graph?

3. Compare the number of cpt rows you found in each graph you drew above to the number of parameters you previously estimated for M_{indp} and M_{dep} in **Questions 3.1 and 3.5** of **Homework 2**. Did the number of cpt rows in your graph grow at the same rate as the number of parameters in your models? Why might this be? What does this imply about a Bayesian network's ability to succinctly represent a model's conditional probability distributions?