Homework 9
A Little Decision Trees

Due: 6:00pm on 11/25/09

Problem 9.1

For each of the following problems, say if you think decision trees would work well, decision trees with bagging would work, or if decision trees are the wrong thing to use entirely. Please give a reason for your answer. (You might note that any time decision trees are appropriate, DT with bagging is also appropriate. Choose the one that you feel is more applicable for the situation. You dont want to use bagging unless you think it will be a significant improvement over DT without bagging.)

In CIT 219, everyone has gathered for CS141 lecture. Today is a special day though because everyone is wearing single color shirts. Can you use decision trees to see if the class falls into the following categories:

a. Diverse: Everyone in the class is wearing a different color shirt.

b. Well Taught: As you know, the color sense of the Professor and the TAs often determines the quality of the class. If Meinolf, Matt and Yuri wear Blue, Green, or Purple shirts, the class is Well Taught.

c. Dominant: A majority of people in the class are wearing the same color shirt.

Problem 9.2

As usual, we want to know if mushrooms are edible.
Let A_{tall} and $A_{spotted}$ be two binary attributes.

Here is what we know about mushrooms:

$$P(A_{tall} = true) = .5 \qquad P(A_{spotted} = true) = .5$$

$$P(edible = true|A_{tall} = true) = .1 \qquad P(edible = true|A_{tall} = false) = .3$$

$$P(edible = true|A_{spotted} = true) = .125 \qquad P(edible = true|A_{spotted} = false) = .25$$

a. Compute the Gini index value for A_{tall} and $A_{spotted}$

b. Based on the Gini values, which attribute should you branch on?

c. Compute the entropy for A_{tall} and $A_{spotted}$

d. Based on the entropy values, which attribute should you branch on?

Problem 9.3

Take a look at Russel and Norvig page 659 ”Choosing attribute tests” and notice that their definition of the information gain only works for two-class classification (where their two classes are p and n... positive and negative). How could you extend the definition to work with multi-class classification? This way we aren't restricted to binary classifications. For example, we could classify different text documents into k different categories. (Also you will be using this formula for your upcoming decision tree project)

Note: The book's method of choosing the best attribute is a bit inefficient. $\arg\max_A Gain(A) = \arg\max_A[I(\frac{p}{p+n}, \frac{n}{p+n}) - Remainder(A)] = \arg\min Remainder(A)$ as $I(\frac{p}{p+n}, \frac{n}{p+n})$ is constant over all attributes (remember n is the number of negative examples and p is the number of positive examples in binary classification). we want you to define $Remainder(A)$ for multi-class classification.

Hint: Remember conceptually, $Remainder(A) = \sum_{choice \in A} q_{choice} I(q_{choice_p}, q_{choice_n})$, where q_{choice} is the probability of being in choice A and q_{choice_p} is the probability that given we are in choice of an example is classified as positive and q_{choice_n} is the same for negative instead.