CS 410/510
Machine Learning
Winter, 2006

Homework 4:
Decision trees, neural networks, and evaluating hypotheses (Written problems)

Due Thursday, February 9.

Hand in solutions to the problems below, either neatly written by hand or formatted using a word processor or text editor. Show your work as clearly as possible.

1. Consider the problem of diagnosing avian influenza in a patient, given the following binary-valued (i.e., true or false) attributes:

Fever = patient has fever
Cough = patient has cough
Breathing = patient has difficulty breathing
HumanContact = patient had recent contact with human infected with avian influenza
AvianContact = patient had recent contact with bird infected with avian influenza

Consider the training set S given below. In each of the six training examples, true or false values for each of the five attributes are given, as well as the correct classification for each example.

Patient	**Fever**	**Cough**	**Breathing**	**HumanContact**	**AvianContact**	**AvianInfluenza**
P1	T	T	F	F	F	F
P2	T	T	T	T	F	T
P3	F	F	T	F	T	F
P4	F	T	T	F	T	T
P5	T	T	F	F	T	T
P6	T	F	F	T	T	F

(a) Give the information gain, Gain(S, A), for each attribute A with respect to the training set S. Show your work clearly.

(b) Find a decision tree that ID3 would return, by tracing the steps of ID3 by hand, using information gain as the splitting criterion. (If there is a tie for highest information gain, choose one of the highest-gain attributes at random.) Include your work that shows how you traced the steps of ID3. Verify that your resulting tree is consistent with the training data.

(c) Compute the gain ratio of the attributes with highest and second highest information gain (with respect to the original training set S). Again, break ties by choosing randomly among the tied attributes. Show your work. If ID 3 had used gain ratio instead of information gain in part (b) above, would the root of the tree have been a different attribute?

(d) Give an example of a new attribute for this problem that would have higher information gain than the highest one you found in part (a), but would not be useful for classifying unseen test data.

2. (a) Consider a perceptron with two input units, one bias unit, and one output unit, and suppose all input activations are binary: either -1 or 1. The bias unit gives a constant input activation of 1. (Consider the value -1 to represent "false" and the value 1 to represent "true".) Give values for the two weights, w_1 and w_2, and the bias θ, that will implement the two-bit Boolean function (x_1 AND x_2) OR (NOT x_2).

(b) What is the equation of the separation line defined by the perceptron in 2(a), given the values for weights and bias that you gave? Plot this line and show that it correctly separates the two classes for the function (x_1 AND x_2) OR (NOT x_2).

(c) Now take the same perceptron, weights, and bias that you gave in in 2(a) above, and suppose that it is given the following instance $\vec{x} = (-1, 1)$, with target output $t = 1$. What will the weights and bias be after an application of the perceptron learning rule? (Assume the learning rate is 1 and the momentum is 0.)

3. Suppose you are given hypothesis h_1 and sample S_1 of size 100, drawn from unknown distribution D. You find that h_1 misclassifies 20 of the examples in

S_1. (a) Give the approximate standard deviation of $error_S(h_1)$ over samples S of size 100. (Use the formula for approximation given in Section 5.3.4 of the textbook). (b) Use the results from part (1) to give the upper bound U of a one-sided confidence interval such that $error_D(h_1) \leq U$ with 99% confidence. (c) Now suppose you are given a second hypothesis, h_2, that is independent of h_1. You are also given a second sample S_2 of size 100, also drawn from distribution D. You find that h_2 misclassifies 12 of the examples in S_2. What is the probability that $error_D(h_1) > error_D(h_2)$?