Name: _____

# 15-681 Midterm Exam

80 minutes
Open book, open notes
Each question 20 points

1. Consider the following training examples. Assume the hypothesis space is the set of conjunctions of these features, as in the Chapter 2 reading.

| Example# | Sky | Air-Temp | Humidity | Wind | Water | Forecast | Enjoy-Sport? |
|---|---|---|---|---|---|---|---|
| 1 | Sunny | Warm | High | Strong | Cool | Change | Yes |
| 2 | Sunny | Warm | Normal | Strong | Warm | Same | No |
| 3 | Sunny | Warm | High | Weak | Warm | Same | No |
| 4 | Rainy | Cold | High | Strong | Warm | Change | Yes |

Give the version space resulting from this set of training data. Note, you need only show the final answer. You are not required to show the trace of the algorithm.

2. Suppose you saw 20 examples from a person’s calendar database and found that in 19 of the 20, the feature "thirty-minute-meeting?" was YES. Then, given a new example you might bet the answer for the new meeting should be yes as well. On the other hand, suppose you pick a coin from your pocket and flip it 20 times and in 19 it comes up heads. Then, you would probably still think that in a subsequent flip, heads and tails are equally likely.... Give a Bayesian interpretation of this phenomenon.

3. Consider H = {all decision trees of depth 2 or less, based on n boolean attributes}. Give a reasonable upper bound on the number of training examples that must be observed to be 95% confident that a tree found to be consistent with all of them will have less than 10% error. Assume the target concept is in H. (you may leave your answer in the form of an expression).

4. Consider the new floceptron unit with inputs $x_1$ and $x_2$, which produces a real-valued output defined as follows:

output = $w_0 + w_1 w_0 x_1 + w_2 w_0 x_2$

Describe an algorithm for learning the weights $w_0$, $w_1$ and $w_2$ of a single floceptron, based on gradient descent. Your training rule should attempt to minimize the floceptron error defined as follows:

Sum over all examples of [desired-output - floceptron-output]$^2$

5. Consider the following combination of the weighted majority algorithm and version spaces, for a hypothesis space of conjunctions over n boolean features (allowing negations). You begin with all conjunctions having weight 1. Every time you see a new example, you predict based on a weighted majority vote over all the conjunctions. Then, instead of throwing out the inconsistent conjunctions, you cut their weight by a factor of 2.

How many mistakes will this procedure make at most, as a function of n and the number of mistakes of the best conjunction? (You may use O() notation)