6.824 Artificial Intelligence, Fall 1994
Problem Set 2

Due: Beginning of class Wednesday, Sept. 28

To facilitate grading we would like the solution for each problem to be turned in seperately — we will collect four seperate solution piles, one for each problem.

Problem 1. Suppose that every attribute is Boolean, i.e., takes on the discrete values **True** and **False**. We will consider disjunctive concepts over this instance space. These are the concepts can be represented by a Boolean formula of the form $P \vee Q \vee \neg R \vee \ldots \vee \neg W$, i.e., a conjunction of literals where each literal is either an attribute name or the negation of an attribute name. Assume that you have n Boolean attributes. Consider running Mitchell's version space algorithm over the hypothesis space consisting of all disjunctive concepts. Suppose that the first example classifies the instance in which every attribute is false as a negative example. Give a series of positive examples such that when these examples follow this universally false negative example the lower fringe computed by the algorithm grows exponentially as a function of n. Describe both the series of examples and the lower fringe that it generates. (Hint: The negative example is not needed but simplifies the analysis. There is a solution in which the size of the lower fringe is $2^{\frac{n}{2}}$.)

Problem 2. In this problem we again consider the instance space defined by n Boolean attributes. However we restrict concepts to those that can be expressed as Boolean formulas involving the operations $\vee$ and $\neg$ where the number of occurances of $\vee$ is at most k. For example, $x_1 \vee \neg(x_2 \vee \neg x_3)$ is a Boolean combination of literals containing two occurances of $\vee$. Note that every occurance of $\vee$ involves exactly two arguments. Also note that formulas of the form $\neg\neg\Phi$ can be replaced by Φ.

(a) Show that for fixed k the number of examples sufficient for PAC-learning over this concept space of k-binary operators is polynomial in $1/\epsilon, 1/\delta$, and n (the number of attributes). Don't worry about getting tight bounds on the sample complexity required for PAC learning — just give some polynomial that is sufficient. Show the derivation of your polynomial. (Hint: Consider the maximum length of a character string representation of the concept and the maximum number of such character strings.)

(b) Explain why the polynomial sample complexity you gave in part **a** does not immediately imply that this concept space is PAC learnable in polynomial time.

Problem 3. Consider instances with n linear attributes where each instance assigns each attribute x_i an integer value from 1 to 2^k (k is the number of bits used to represent the integer x_i). We consider conjunctive concepts defined by conditions of the form

$$x_1 \in [l_1,\ u_1] \wedge x_2 \in [l_2,\ u_2] \wedge \ldots \wedge x_n \in [l_n,\ u_n]$$

where each l_i and u_i is an integer in the range $[1,\ 2^k]$ and $l_i \leq u_i$.

From the cardinality of this concept space give a number of examples that suffices to ensure PAC learning, i.e., such that with probability $1 - \delta$ all concepts consistent with the examples have error rate no larger than ϵ. Your answer should be given in terms of n, k, δ and ϵ.

Problem 4. Again consider the hypothesis space defined in problem 3. Derive a number of examples sufficient for PAC learning using the VC dimension of this hypothesis space. Your answer should depend only on n, δ, and ϵ (and not on k). The answers to both problem 3 and this problem should be linear in n (holding the other parameters fixed). Give a relationship between k and ϵ which guarantees that the sample complexity from problem 3 is smaller than the sample complexity given by the VC dimension (Hint: you only need to look at the linear term containing n in your two sample complexity formulas.) Assume ϵ is $\frac{1}{32}$ (or 2^{-5}) and give an approximate value for k such that if we use no more than k bits per number then the bound from problem 3 no smaller than the bound given by the VC dimension.