Instructor: Rocco Servedio
Class Manager: Andrew Wan

Computer Science 4252: Introduction to Computational Learning Theory
Problem Set #1 Spring 2006

Due 5:00pm Monday, Feb 13, 2005

See the course Web page `http://www.columbia.edu/~atw12/learning` for instructions on how to submit homework electronically.

Problem 1 In the first two problems you'll explore relationships between some of the concept classes we've discussed.

(i) Show that any decision list on variables $x_1, \ldots, x_n$ can be expressed as an n-term DNF formula.

(ii) Show that any decision list on variables $x_1, \ldots, x_n$ can be expressed as a linear threshold function.

Problem 2
(i) Show that the class of 2-term DNFs is not contained in the class of linear threshold functions over $\{0,1\}^n$.

(ii) Show that the class of linear threshold functions over $\{0,1\}^n$ is not contained in the class of $p(n)$-term DNFs for any polynomial $p(n)$.

Problem 3
(i) We mentioned in class that the elimination algorithm can be used to learn arbitrary disjunctions (not necessarily monotone) over n variables. What mistake bound can you give for using the elimination algorithm to learn an arbitrary disjunction?

(ii) What is the best mistake bound you can give for an online algorithm which learns the class of arbitrary (not necessarily monotone) disjunctions of size k over n variables? You should describe an algorithm, state a mistake bound, and explain why the algorithm achieves the mistake bound.

Problem 4 Recall that in the Winnow1 algorithm a *promotion* step is performed on examples x such that $h(x) = 0$ and $c(x) = 1$. In a promotion step the weight w_i is multiplied by 2 for all i such that $x_i = 1$. Consider the following modified version of Winnow1: instead of multiplying these weights w_i by 2, multiply each of these weights by a value α such that after the promotion step the sum of these weights will be exactly $2n$. Thus the value of α may be different on different promotion steps.

Suppose that the target concept c is a monotone disjunction of k variables. Give the best bound you can on the number of mistakes which this modified algorithm will make.

Problem 5 The Infinite Attribute mistake bound model is a Boolean on-line learning model where the number of Boolean attributes $x_1, x_2, \ldots$ may be infinite. It is assumed, however, that in each example only a finite number of attributes x_i take value 1. An example is presented to the learner as a list of its positive attributes. The learning scenario is the same as in the (standard) mistake-bounded model.

Describe an on-line learning algorithm that makes at most $O(k \log n)$ mistakes when learning any monotone disjunction of k literals, if every example presented to the learner has at most n positive attributes.

Problem 6

(i) Compare the performance of the Perceptron and Winnow algorithms for learning a monotone disjunction of size k over $\{0,1\}^n$. Which algorithm has the better mistake bound?

(ii) Now compare Perceptron and Winnow for learning the majority function $x_1 + \cdots + x_n \geq n/2$ over $\{0,1\}^n$. Which algorithm has the better mistake bound?