Midterm Exam

This is an *open book* exam—feel free to consult your class notes and other literature. Clarity of writing is important—not just "having the right answer."

The exam ends at 12:30. It contains four problems. You have 90 minutes to earn 90 points. Answer each question in the space provided. If you need more room, write on the back sides of the pages, and indicate that this is what you have done.

Good luck!

Name: _____

Problem 1. Short answer questions (24 points)

(a) A colleague shows you his latest invention: "nerdal nets." Each nerdal net specifies a division of R^n into positive and negative examples. Your colleague also has an efficient algorithm which, given as input a finite set S of labelled examples, will *always* return a nerdal net that correctly labels at least 2/3 of the examples in S. Does this imply that the class of nerdal nets is PAC-learnable? Explain briefly.

(b) We proved in class that if VC-dimension$(C) < \infty$, then $\Pi_C(m)$ grows at most polynomially in m. Argue that the converse is also true.

(c) Suppose we have a concept class C such that there exists a learning algorithm L that satisfies the definition of PAC-learnability of C by C except that L only works when $\delta = 1 - \epsilon$. Argue that C is nonetheless PAC-learnable by C.

Problem 2. (20 points)

Suppose a certain coin is either fair or double-headed. You believe that the *a priori* odds of it being fair are F to 1 (i.e., the *a priori* probability of the coin being fair is $\frac{F}{F+1}$), but you start watching the outcomes of it being flipped in order to learn more. As a function of F, how many heads in a row do you have to see before you become convinced that there is better than an even chance that the coin is double-headed?

Problem 3. (22 points)

Consider concept classes C_1 and C_2 defined on instance space X. Suppose that C_1 and C_2 have finite VC-dimension. Consider the concept class

$$C = C_1 + C_2 = \{c_1 + c_2 : c_1 \in C_1, c_2 \in C_2\}$$

where $c_1 + c_2 = \{x : c_1(x) \vee c_2(x)\}$.

(a) Argue that $C = C_1 + C_2$ is PAC-learnable. (Hint: Find a simple upper bound on $\Pi_C(m)$.)

(b) Argue that if C_1 and C_2 are efficiently PAC-learnable, then it doesn't necessarily follow that $C = C_1 + C_2$ is efficiently PAC-learnable.

Problem 4. (24 points)

The concept class $C_{n,l}$ of 3-term majority formulae is the set of all $\text{MAJ}(T_1, T_2, T_3)$, where each T_i is a conjunction of at most l (a constant) literals over the boolean variables $x_1, x_2, \ldots, x_n$. For any concept $c = \text{MAJ}(T_1, T_2, T_3)$, $c(x) = 1$ if x satisfies at least 2 of the terms T_1, T_2, and T_3, and $c(x) = 0$ otherwise. For example, a concept in $C_{20,4}$ might have 3 terms $x_1\overline{x}_5, x_2\overline{x}_7x_8x_{15}$, and $x_{10}x_{11}$, and any example $x \in \{0,1\}^{20}$ which makes at least 2 of these terms true is a positive example for this concept.

(a) Give an efficient algorithm for learning $C_{n,l}$ in the on-line mistake bounded model. Your algorithm should run in time polynomial in n. Give the mistake bound of your algorithm.

(b) Let C_n denote the concept class of 3-term majority, where the terms have unrestricted size. Show that C_n is PAC-learnable using 3-decision lists. (Hint: show that any 3-term majority function can be expressed as a 3-decision list, and then argue that this implies C_n is PAC-learnable using 3-decision lists.)