Assignment 3

This assignment is due back in class on Thursday April 3.

Problem 1

We modify the on-line learning model seen in class extending its instance space to an infinite set of attributes \( A = \{a_1, a_2, \ldots \} \). An example in this new model is a finite subset of \( A \). An example \( x \subseteq A \) is interpreted as having all attributes in \( x \) set to 1 and the rest set to 0. Consider the class \( MonD_\infty \) of monotone disjunctions over the infinite attribute space \( A \). This class includes concepts such as \( a_1 \lor a_3 \lor a_{1001} \). Negative literals are not allowed.

Find an algorithm that learns \( MonD_\infty \) in this new learning model and derive its mistake bound.

*Hint:* The bound can depend on the size (number of literals) of the target disjunction. Try to make use of a variant of the Elimination algorithm which uses both positive and negative examples.

Problem 2

We consider a modification of the Perceptron Algorithm, the Perceptron Algorithm with Margins (PAM). This algorithm also updates the current weight vector \( w \) when its prediction margin is very close to 0 (by prediction margin we mean the value of \( l(\bar{x})(\bar{w} \cdot \bar{x}) \)). Hence, the hypothesis is updated whenever a positive example (negative example, resp.) is within distance \( \tau \) from the hyperplane given by \( \bar{w}, \theta \).

**Algorithm PAM**

1. Set \( w_i = 0 \) for all \( i = 1, \ldots, n \) and \( \theta = 0 \)
2. On example \( \bar{x} \), do:
3. \[ \text{if } l(\bar{x})(\bar{w} \cdot \bar{x}) \leq \tau \]
4. \[ \text{then } \bar{w} = \bar{w} + l(\bar{x})\bar{x} \]

where \( \tau \) is a small positive number (the desired margin) and \( l(\bar{x}) \) denotes the true label of example \( \bar{x} \) (+1 if positive, -1 if negative).

Show that for any \( \bar{v} \) and \( \delta \) and for any sequence of examples with separation \( \delta \) via \( \bar{v} \), and such that for each example \( \bar{x} \) in the sequence \( |\bar{x}| \leq R \), the number of mistakes made by the algorithm is at most

\[
\frac{R^2 + 2\tau}{\delta^2}.
\]

*Note:* The mistake bound proved here is worse than the one we proved in class, but the resulting hypothesis is guaranteed to have a higher margin when classifying. Intuitively, it should be more
stable to slight variations in the distribution. In practice this seems to improve the classification accuracy of the resulting classifiers.

**Problem 3**

For $x \in \{0, 1\}^n$ let $t(x) = \land_{i=1}^n x_i$. For example for $x = 101$, $t(x) = x_1 x_3$. For two assignments $x, y \in \{0, 1\}^n$, $x \land y$ denotes their bitwise AND. For example for $x = 101$ and $y = 110$, $x \land y = 100$. Show that the following algorithm efficiently learns the class of monotone DNF expressions and give bounds for the numbers of equivalence and membership queries.

1. Initialize $S = \emptyset$ and $h = 0$.
2. While $EQ(h)$ provides a (positive) counterexample $x$
   (a) For each element $s$ in $S$ do:
      - if $MQ(s \land x)$ returns “positive” then replace $s$ with $s \land x$ and quit for loop.
   (b) if no element was replaced then add $x$ to $S$.
   (c) Let $h = \lor_{s \in S} t(s)$
3. Output $h$

**Hint:** Argue that two elements of $S$ cannot satisfy the same term of the target. Then try to quantify how you make “progress” after every counter example.

**Problem 4 (extra credit)**

1. Define a notion of reducibility appropriate for learning from Equivalence and Membership Queries. Using this notion show that if $C^1$ reduces to $C^2$ and $C^2$ is efficiently learnable from Equivalence and Membership Queries then $C^1$ is efficiently learnable from Equivalence and Membership Queries.
   Note that you would want this notion of reducibility to be useful in part (2).
2. A read-3 DNF formula over $\{0, 1\}^n$ is a DNF formula using literals in $x_1, \ldots, x_n, \overline{x_1}, \ldots, \overline{x_n}$ in which every variable appears at most 3 times (whether negated or unnegated). For example $x_1 x_2 \overline{x_3} x_5 \lor \overline{x_2} x_4 x_5 \lor x_3 x_4 x_5 \lor \overline{x_1} \overline{x_2} \overline{x_4}$ is read-3 but $x_1 x_2 \overline{x_3} x_5 \lor \overline{x_2} x_4 x_5 \lor x_3 x_4 x_5 \lor \overline{x_1} \overline{x_2} \overline{x_4}$ is not (since $x_5$ appears 4 times). The class of read-3 DNF is composed of the union of such formulae for $n \geq 1$.
   Let $p()$ be any polynomial and let $C$ be a sub-class of DNF formulae. We say that $C$ is $p()$-bounded if every formulae $c \in C$ over $\{0, 1\}^n$ has at most $p(n)$ terms. Show that if $C$ is $p()$-bounded then it reduces to read-3 DNF (using reducibility of part (1)).