CSC 2416 Machine Learning Theory

ASSIGNMENT # 3

Given: November 22, Due: December 8

1. Consider a variant of Adaboost in which we set $\alpha_t = \alpha$ on every round of boosting, where $\alpha > 0$ is a fixed parameter that is set ahead of time. Call this algorithm $Boost(\alpha)$. Assume that on each round t of boosting, it is known ahead of time that ϵ_t will be at most $1/2 - \gamma$ for some $\gamma > 0$. Suppose that we set
$$\alpha = \frac{1}{2}\ln\left(\frac{1+2\gamma}{1-2\gamma}\right)$$

   a. Show how to modify the training error analysis of Adaboost to derive an upper bound on the training error of the final hypothesis H produced by Boost(α) after T rounds. Your bound should be in terms of γ and T only (and should not depend on α, ϵ_t, etc.)

   b. Use your result from part a. to show that the final hypothesis H will be consistent with m training examples after T rounds if $T > \frac{\ln m}{2\gamma^2}$.

   c. Assume that the weak learning algorithm generates hypotheses h_t which belong to a finite class $\mathcal{H}$. Use Occam's razor to show that if we choose T as in part b., then with probability $1-\delta$, the generalization error of H is at most
$$\frac{T\ln|\mathcal{H}| + \ln(1/\delta)}{m}$$

2. Let $X = \{-1,+1\}^n$ For $x \in X$, let x_i denote the ith component of x, so $x = x_1,\ldots,x_n$. Let $\mathcal{M}_k$ be the set of concepts $c : X \to \{-1,+1\}$ for which there exist indices $i_1,\ldots,i_k \in \{1,\ldots,n\}$ (not necessarily distinct) such that $c(x) = sign(\sum_{j=1}^{k} x_{i_j})$, for all $x \in X$. (We define $sign(0) = 0$.) Note that this implies that if $c \in \mathcal{M}_k$ then $\sum_j x_{i_j}$ cannot be equal to 0 for any x since $c(x) \in \{-1,+1\}$.) In this problem, you will show that a concept c is γ-weakly learnable by the features $x_1,\ldots,x_n$ for $\gamma = \Omega(1/poly(n))$ if and only if $c \in \mathcal{M}_k$ for $k = poly(n)$.

   a. For any concept $c : X \to \{-1,+1\}$ and distribution $\mathcal{D}$ on X, show that
$$E_{x\in\mathcal{D}}[c(x)\cdot x_i] = 1 - 2Pr_{x\in\mathcal{D}}[c(x) \neq x_i].$$

   b. Let c be as in the above equation. Argue that
$$\sum_{j=1}^{k} E_{x\in\mathcal{D}}[c(x)\cdot x_{i_j}] \geq 1$$
for every distribution $\mathcal{D}$ on X.

c. Let $c \in \mathcal{M}_k$. Use parts a. and b. to show that for every distribution $\mathcal{D}$ on X, there exists an index $i \in \{1, \ldots, n\}$ such that

$$Pr_{x \in \mathcal{D}}[x_i \neq c(x)] \leq (1/2) - (1/2k).$$

d. Consider a weak learning algorithm A that works as follows. Given a training set $(x^1, y^1), \ldots, (x^m, y^m)$ (where for each i, $x^i \in X$), and distribution $\mathcal{D}_t$ over the examples, A outputs the hypothesis $h_t(x) = x_{i^*}$ where i^* is the minimum error with respect to the examples and distribution. That is,

$$i^* = argmin_{1 \leq i \leq n} \sum_{j : x_i^j \neq y^j} \mathcal{D}_t(j).$$

Let $c : X \to \{-1, +1\}$ be any concept and let $\gamma > 0$. Suppose that, for every distribution $\mathcal{D}$ on X, there exists an index i such that

$$Pr_{x \in \mathcal{D}}[x_i \neq c(x)] \leq 1/2 - \gamma.$$

Use your analysis of Boost(α) and the weak learner described above to show that $c \in \mathcal{M}_k$ if $k > \frac{n \ln 2}{2\gamma^2}$.

3. The problem BBAYES (boolean bayes) is defined as follows. The input is a tuple $(G, T, A, B, \alpha, \beta)$ where: $G = (V, E)$ is a directed acyclic graph; T is a set of conditional probability tables t_v for each vertex $v \in V$; A, B are disjoint subsets of V; α is an assignment of values to the variables of A, and β is an assignment of values to the variables of B. (You can assume if you wish that the probabilities given in the tables t_v are values from $\{1/n, 2/n, \ldots, n/n\}$, where n is the number of vertices in G.) The output should be the value of $Pr[A = \alpha \mid B = \beta]$.

   a. Prove that BBAYES is in sharpP.

   b. Prove that BBAYES is sharpP hard by giving a polynomial-time reduction from sharpSAT to BAYES.

   c. Consider now the generalization of BBAYES, DBAYES (discrete bayes), where now for each variable $v \in V$, the range of v is some finite set $Range(v)$. Generalize your argument in a. to prove that DBAYES is in sharpP.

   NOTE: The complexity class sharpP is defined as follows. Let f be a function from the natural numbers to the natural numbers. Then f is in sharpP if there exists a nondeterministic, polynomial-time Turing machine, M with the property that for every input x (x is a natural number written in binary notation), the number of accepting computations of M on input x is equal to $f(x)$ (again $f(x)$ is the binary representation of some natural number.) (Note that if $|x| = n$, and M runs in time $poly(n)$, then the maximum number of accepting computations of M on input x is $2^{poly(n)}$.) For example, the function sharp3SAT takes as input a natural encoding of a 3CNF formula, f, and outputs the number of satisfying assignments to f. This problem is easily seen to be in sharpP, and can also be shown to be sharpP complete. That is, it can be shown that sharp3SAT can be solved in polynomial-time if and only if all functions in sharpP can be solved in polynomial time. The above question asks you to prove that BBAYES is also sharpP complete.