COS 511: Theoretical Machine Learning

Homework #4 — Due:

Rademacher & boosting — March 25, 2014

Problem 1

[15] Let $\mathcal{F}_1, \ldots, \mathcal{F}_n$ be families of real-valued functions on some space $\mathcal{Z}$, and let $a_1, \ldots, a_n$ be arbitrary (fixed) real numbers. Let $\mathcal{G}$ be the class of all functions g of the form

$$g(z) = \sum_{i=1}^{n} a_i f_i(z)$$

where $f_i \in \mathcal{F}_i$ for $i = 1, \ldots, n$. For any sample S, find $\hat{\mathcal{R}}_S(\mathcal{G})$ exactly in terms of $a_1, \ldots, a_n$, and $\hat{\mathcal{R}}_S(\mathcal{F}_1), \ldots, \hat{\mathcal{R}}_S(\mathcal{F}_n)$. Be sure to justify your answer.

Problem 2

[15] Suppose, in the usual boosting set-up, that the weak learning condition is guaranteed to hold so that $\epsilon_t \leq \frac{1}{2} - \gamma$ for some $\gamma > 0$ which is *known* before boosting begins. Describe a modified version of AdaBoost whose final classifier is a simple (unweighted) majority vote, and show that its training error is at most $(1 - 4\gamma^2)^{T/2}$.

Problem 3

Let $\mathcal{X}_n = \{0, 1\}^n$, and let $\mathcal{G}_n$ be any class of boolean functions $g : \mathcal{X}_n \to \{-1, +1\}$. In this problem, we will see, roughly speaking, that if a function f can be written as a majority vote of polynomially many functions in $\mathcal{G}_n$, then under any distribution, f can be weakly approximated by some function in $\mathcal{G}_n$. But if f cannot be so written as a majority vote, then there exists some "hard" distribution under which f cannot be approximated by *any* function in $\mathcal{G}_n$.

Let $\mathcal{M}_{n,k}$ be the class of all boolean functions that can be written as a simple majority vote of k (not necessarily distinct) functions in $\mathcal{G}_n$; that is, $\mathcal{M}_{n,k}$ consists of all functions f of the form

$$f(x) = \operatorname{sign}\left(\sum_{j=1}^{k} g_j(x)\right)$$

for some $g_1, \ldots, g_k \in \mathcal{G}_n$. Assume k is odd.

- [15] Show that if $f \in \mathcal{M}_{n,k}$ then for all distributions D on $\mathcal{X}_n$, there exists a function $g \in \mathcal{G}_n$ for which
$$\Pr_{x \sim D}\left[f(x) \neq g(x)\right] \leq \frac{1}{2} - \frac{1}{2k}.$$
- [15] Show that if $f \notin \mathcal{M}_{n,k}$ then there exists a distribution D on $\mathcal{X}_n$ such that for every $g \in \mathcal{G}_n$,
$$\Pr_{x \sim D}\left[f(x) \neq g(x)\right] > \frac{1}{2} - \sqrt{\frac{n \ln 2}{2k}}.$$