Problem 1

[15] Let $F_1, \ldots, F_n$ be families of real-valued functions on some space $Z$, and let $a_1, \ldots, a_n$ be arbitrary (fixed) real numbers. Let $G$ be the class of all functions $g$ of the form

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

where $f_i \in F_i$ for $i = 1, \ldots, n$. For any sample $S$, find $\hat{R}_S(G)$ exactly in terms of $a_1, \ldots, a_n$, and $\hat{R}_S(F_1), \ldots, \hat{R}_S(F_n)$. 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 $X_n = \{0, 1\}^n$, and let $G_n$ be any class of boolean functions $g : X_n \rightarrow \{-1, +1\}$. 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 $G_n$, then under any distribution, $f$ can be weakly approximated by some function in $G_n$. 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 $G_n$.

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

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

for some $g_1, \ldots, g_k \in G_n$. Assume $k$ is odd.

- [15] Show that if $f \in M_{n,k}$ then for all distributions $D$ on $X_n$, there exists a function $g \in G_n$ for which

$$\Pr_{x \sim D} |f(x) \neq g(x)| \leq \frac{1}{2} - \frac{1}{2k}.$$

- [15] Show that if $f \notin M_{n,k}$ then there exists a distribution $D$ on $X_n$ such that for every $g \in G_n$,

$$\Pr_{x \sim D} |f(x) \neq g(x)| > \frac{1}{2} - \sqrt{\frac{n \ln 2}{2k}}.$$