Mehryar Mohri
Foundations of Machine Learning 2018
Courant Institute of Mathematical Sciences
Homework assignment 1
Sep 18, 2018
Due: Oct 01, 2018

A. Probability review

Let $h\colon X \to \{0,1\}$ be a hypothesis and let S denote an i.i.d. sample of size m. For any $\epsilon > 0$, the following two-sided inequality holds:

$$\Pr_S(|\widehat{R}_S(h) - R(h)| > \epsilon) \leq 2e^{-2m\epsilon^2}.$$

Show that the variance of $\widehat{R}_S(h)$ satisfies $\mathrm{var}[\widehat{R}_S(h)] \leq \frac{\log(2e)}{2m}$. (*Hint:* use the identity $\mathrm{E}[X^2] = \int_0^\infty \Pr[X^2 > t]dt$.)

B. PAC learning

1. Consider the concept class C formed by threshold functions on the real line, $C = \{[c,\infty)\colon \forall c \in \mathbb{R}\} \cup \{(-\infty, c]\colon \forall c \in \mathbb{R}\}$. Give a PAC-learning algorithm for C. The analysis is similar to that of the axis-aligned rectangles given in class, but you should carefully present and justify your proof.

2. Give a PAC-learning algorithm for the concept class C_2 on $\mathbb{R}^2$ that is formed by intersections of axis-aligned half-spaces: C_2 consists of concepts of the following forms:

$$\{(x,y)\colon x \geq c_x, y \geq c_y\},$$
$$\{(x,y)\colon x \geq c_x, y \leq c_y\},$$
$$\{(x,y)\colon x \leq c_x, y \geq c_y\},$$
$$\{(x,y)\colon x \leq c_x, y \leq c_y\},$$

   where $c_x, c_y \in \mathbb{R}$. You should carefully justify all steps of your proof.