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

Solution: For any $u > 0$, we have

$$\begin{aligned}
\text{var}[\widehat{R}(h)] &= \text{E}[(\widehat{R}(h) - R(h))^2] \\
&= \int_0^u \Pr[(\widehat{R}(h) - R(h))^2 > t]dt + \int_u^\infty \Pr[(\widehat{R}(h) - R(h))^2 > t]dt \\
&\leq u + \int_u^\infty 2e^{-2mt}dt \\
&= u + \frac{e^{-2mu}}{m} := f(u).
\end{aligned}$$

The function $f(u)$ is a convex function of u, with its minimum value attained when

$$f'(u_0) = 0 \iff 1 - 2e^{-2mu_0} = 0 \iff u_0 = \frac{\log 2}{2m}.$$

Plug in u_0 to get $f(u_0) = \frac{\log(2e)}{2m}$. □

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.

Solution: Let $S = \{(x_1, y_1), \cdots, (x_m, y_m)\}$ denote the labeled sample of size m. Without loss of generality, assume that the true concept is $[c, \infty)$ for some unknown $c \in \mathbb{R}$. Define

$$\widehat{l} = \max\{x_i : (x_i, y_i) \in S, y_i = -1\},$$
$$\widehat{r} = \min\{x_i : (x_i, y_i) \in S, y_i = 1\}.$$

By definition, $\widehat{l} \leq c \leq \widehat{r}$.

The algorithm returns the concept $R_S = [\widehat{c}, \infty)$ with $\widehat{c} = (\widehat{l} + \widehat{r})/2$. The error region of R_S is the interval $[\widehat{c}, c)$ when $\widehat{c} < c$, and $[c, \widehat{c})$ otherwise. In both cases, the error region is a subset of $(\widehat{l}, \widehat{r})$. Therefore,

$$\begin{aligned}\Pr[R(R_S) > \epsilon] &\leq \Pr[R((\widehat{l}, \widehat{r})) > \epsilon] \\ &\leq (1 - \epsilon)^m \leq e^{-m\epsilon}.\end{aligned}$$

Setting δ to be greater than or equal to the right-hand side leads to $m \geq \frac{1}{\epsilon} \log(\frac{1}{\delta})$. □

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.

Solution: The algorithm returns the tightest concept in C_2 containing points labeled with 1. The proof is similar to the proof of learning axis-aligned rectangles.