Statistical Foundations of Learning - Assignment 3

CIT4230004 (Summer Semester 2024)

Exercise 3.1: PAC Expected Risk of the Learner

(a) Show that H is agnostic PAC learnable.

Given:

$$\mathbb{E}_{S\sim D^m}\left[L_D(A(S))\right] \leq \inf_{h\in H} L_D(h) + \epsilon \text{ for } m \geq \tilde{m}_H(\epsilon)$$

To show H is agnostic PAC learnable, we need to prove that for every $\epsilon > 0$ and $\delta > 0$, there exists a sample size m such that:

$$\mathbb{P}\left(L_D(A(S)) \leq \inf_{h\in H} L_D(h) + \epsilon\right) \geq 1 - \delta$$

Given:

$$\mathbb{E}_{S\sim D^m}\left[L_D(A(S))\right] \leq \inf_{h\in H} L_D(h) + \epsilon$$

Using Markov's inequality:

$$\mathbb{P}\left(L_D(A(S)) > \inf_{h\in H} L_D(h) + \epsilon\right) \leq \frac{\mathbb{E}_{S\sim D^m}\left[L_D(A(S))\right] - \inf_{h\in H} L_D(h)}{\epsilon}$$

Since:

$$\mathbb{E}_{S\sim D^m}\left[L_D(A(S))\right] \leq \inf_{h\in H} L_D(h) + \epsilon$$

We get:

$$\mathbb{P}\left(L_D(A(S)) > \inf_{h\in H} L_D(h) + \epsilon\right) \leq \frac{\epsilon}{\epsilon} = 1$$

Therefore:

$$\mathbb{P}\left(L_D(A(S)) \leq \inf_{h\in H} L_D(h) + \epsilon\right) \geq 1 - \delta$$

Thus, H is agnostic PAC learnable.

(b) Show that H is not agnostic PAC learnable.

Given:

$$\mathbb{E}_{S\sim D^m}\left[L_D(A(S))\right] > \inf_{h\in H} L_D(h) + \epsilon \text{ for every } m$$

To show H is not agnostic PAC learnable, we need to prove that for every m, there exists a distribution D such that:

$$\mathbb{E}_{S\sim D^m}\left[L_D(A(S))\right] > \inf_{h\in H} L_D(h) + \epsilon$$

This implies that no matter how large the sample size m is, the expected risk of the learner is always greater than the infimum of the risks over all hypotheses $h \in H$ by at least ϵ.

Thus, H is not agnostic PAC learnable.

Exercise 3.2: Complement of a Learnable Class

Given:

$$H \subseteq \{\pm 1\}^X \text{ is PAC learnable}$$

$$H' = \{\pm 1\}^X \setminus H$$

To prove/disprove: H' is PAC learnable.

Proof:

If H is PAC learnable, there exists a learner A such that for every $\epsilon > 0$ and $\delta > 0$, there exists a sample size m such that:

$$\mathbb{P}\left(L_D(A(S)) \le \inf_{h\in H} L_D(h) + \epsilon\right) \ge 1 - \delta$$

The complement class H' contains all functions that are not in H. If H is learnable, it implies H has a finite VC dimension. However, H' might have an infinite VC dimension, making it not PAC learnable. Thus, H' is not PAC learnable.

Exercise 3.3: VC Dimensions

1. VC Dimension of H:

Given:

$$H = \left\{\sum_{i=0}^{k} 1\{t_{2i} \le x < t_{2i+1}\}, 0 \le t_0 < \ldots < t_{2k+1} \le 1\right\}$$

The function in H is a piecewise constant function that changes value at $t_0, t_1, \ldots, t_{2k+1}$. Each interval $[t_{2i}, t_{2i+1})$ can independently take the value 0 or 1.

To find the VC dimension, consider the set of points $\{x_1, x_2, \ldots, x_n\}$. We want to find the largest n such that these points can be shattered, meaning that every possible binary labeling of these points can be realized by some function in H.

For n points to be shattered: - Each point must be in some interval $[t_{2i}, t_{2i+1})$. - We need to form intervals such that each of the 2^n binary labelings can be represented.

Since there are $2k+2$ thresholds (endpoints of intervals) that can be chosen, we can create $2k+1$ intervals. Each interval can be independently labeled 0 or 1, allowing for 2^{2k+1} different combinations of labelings.

Therefore, H can shatter $2k+1$ points but not $2k+2$ points.

Thus, the VC dimension of H is $2k+1$.

2. VC Dimension of $H = \{\mathbf{sign}(\sin(ax)), a \in \mathbb{R}\}$:

Consider the set of functions:

$$H = \{\text{sign}(\sin(ax)) \mid a \in \mathbb{R}\}$$

The sign function changes from +1 to -1 at the zeros of the sine function, which occur at intervals of $\frac{\pi}{a}$.

To find the VC dimension, we need to find the largest number of points that can be shattered by these functions.

Consider the set of points $x_i = 10^{-i}$ for $i = 1, 2, \ldots, n$. We need to show that for any binary labels $y_1, y_2, \ldots, y_n$, there exists an a such that:

$$\text{sign}(\sin(ax_i)) = y_i$$

Choose:

$$a = \pi \left(1 + \sum_{i=1}^{n} (1 - y_i) 10^i \right)$$

This choice of a ensures that: - For $y_i = 1$, the argument $\pi(1 + \sum_{i=1}^{n}(1 - y_i)10^i)x_i$ is such that $\sin(ax_i) \geq 0$. - For $y_i = -1$, the argument is such that $\sin(ax_i) < 0$.

Since a can be chosen to achieve any combination of labels for any number of points n, the VC dimension of H is infinite.

Thus, the VC dimension of $H = \{\text{sign}(\sin(ax)), a \in \mathbb{R}\}$ is infinite.