Foundations of Data Science & Machine Learning

Summary — Week 05
Devansh Singh Rathore
111701011

B.Tech. in Computer Science & Engineering
Indian Institute of Technology Palakkad

April 14, 2021

Abstract

In this week's lectures, we continue discussing PAC. Further, we define growth function and VC dimension of Hypothesis classes.

1 PAC (Cont'd.)

→ **Examples for H:**

- **Ex1.** H = lines in $\mathbb{R}^2$ passing through origin.
- **Ex2.** Any finite H.
- **Ex3.** Any finite X.
- **Ex4.** H = lines in $\mathbb{R}^2$.
- **Ex5.** H = hyperplanes in $\mathbb{R}^n$, where n is finite

→ **Ex3. Special Case:** The Hypothesis class of lines in $\mathbb{R}^2$ separating a finite input space X is PAC-learnable with $n \geq (1/\epsilon)(2ln|x| + ln(1/\delta))$ training examples. This is called **"Smart Move"**.

→ **Ex3. - Smartest Move:**

- If $E_out(h)$ is bad ($> \epsilon$), then it's very likely (with prob. $> 1/2$), that for another $S' \sim D^n$, $E_{S'}(h)$ is also bad ($> \epsilon/2$).
- Hence instead of worrying about all X, we need to worry only about $S \cup S'$.

Let $S \sim D^n$
Take $h \in H : E_S(h) = 0 \wedge E_D(h) > \epsilon$
Then for some $S' \sim D^n$, $E_{S'}(h) > \epsilon/2$ (with high prob. $\approx\geq 1/2$)

Proof:

$$E_D(h) > \epsilon \Rightarrow P_D(h \triangle f) = p > \epsilon$$

$$\Rightarrow P_{x\sim D}(h(x) \neq f(x)) = p$$

$$\Rightarrow P_{S'\sim D^n}(|S' \cap (h \triangle f)| = k) \sim Binom_p(n, k)$$

$$\Rightarrow E_{S'\sim D^n}|S' \cap (h \triangle f)| = pn$$

$$\Rightarrow P_{S'\sim D^n}(|S' \cap (h \triangle f)| = pn/2) < 1/2 \text{ (Using Chernoff bound, though very loose bound)}$$

$$\Rightarrow P_{S'\sim D^n}[E_{S'}(h) < p/2] < 1/2$$

$$\Rightarrow P_{S'\sim D^n}[E_{S'}(h) \geq p/2] \geq 1/2$$

$$\Rightarrow P_{S'\sim D^n}[E_{S'}(h) \geq \epsilon/2] \geq 1/2$$

$\rightarrow$ Since in the last random experiment: $S \sim D^n$, $S' \sim D^n$
Event A:: $\exists h \in H : E_S(h) = 0 \wedge E_D(h) > \epsilon$
Event B:: $\exists h \in H : E_S(h) = 0 \wedge E_{S'}(h) > \epsilon/2$
We proved: P(B/A) $\geq 1/2$

$$1/2 \leq P(B/A) = P(B \cap A)/P(A) \leq P(B)/P(A)$$

$$P(A) \leq 2P(B)$$

$\rightarrow$ **Now we can forget X and just worry about** $S \cup S'$
Model: $T = S \cup S' \sim D^{2n}$
S and S' are obtained by uniform random equipartition.
Lemma: Let $A \subseteq S \cup S'$, then what is $P(A \subseteq S)$ in a random equipartition.

$$P(A \subseteq S) = \binom{2n-k}{n-k} / \binom{2n}{n} \quad (n = |S| = |S'|,\ k = |A|)$$

$$= ((n)(n-1)...(n-k+1))/((2n)(2n-1)...(2n-k+1))$$

$$< (1/2).(1/2)...(1/2) = (1/2)^k = (1/2)^{|A|}$$

So, for any $h \in H$,

$$P_{S,S'\sim D^n}[(E_S(h,f) = 0) \wedge (E_{S'}(h,f) > \epsilon/2)] \leq (1/2)^{|A|} \leq (1/2)^{\epsilon n/2}$$

$$\text{(Proof: } A = \{x \in S \cup S' : h(x) \neq f(x)\},\ |A| > \epsilon n/2 = k)$$

$\rightarrow$ Now considering $B_h := [(E_S(h,f) = 0) \wedge (E_{S'}(h,f) > \epsilon/2)]$, so

$$P_{S,S'\sim D^n}[\exists h \in H : B_h] \leq (\text{"}\#distinct\, h\text{"})(1/2)^{\epsilon n/2}$$

$$\text{(considering } h_1 = h_2 \text{ (i.e. } h_1 \approx h_2 \sim f) \text{ if } h_1|_{S\cup S'} = h_2|_{S\cup S'})$$

$$\leq |\{[h] : h \in H\}|(1/2)^{\epsilon n/2}$$

$$\leq (2n)^2(1/2)^{\epsilon n/2} \leq \delta/2\ (required)$$

2 Growth function & VC Dimension of Hypothesis Classes

$\rightarrow$ Extending idea from **Ex4.** case to **Ex5.** case.

Definition: The **Growth Function** of a hypothesis class H on a domain X is defined as:

$g_H(n)$ = Maximum no. of distinct function that can be obtained by restricting the function in H to a set of n points in X.

For $T \subseteq X$ let $H|_T = \{h|_T : h \in H\}$, where $h : X \rightarrow \{+1, -1\}$ and $h|_T : T \rightarrow \{+1, -1\}$,then:

$$g_H(T) = |H|_T| << |H|$$

$$g_N(n) = max\{g_H(T) : T \in \binom{X}{n}\}$$

$\rightarrow$ If H = linear separators in $\mathbb{R}^2$, then $g_N(n) = \binom{n+1}{2} + 1$

$\rightarrow$ Let H be the hypothesis class with growth function $g_H()$. For any $\epsilon, \delta \in (0, 1)$, any distribution D on $\mathbb{R}^2$ and any true labelling $f : \mathbb{R}^2 \rightarrow \{+1, -1\}$, $P_{S \sim D^n}[\exists h \in H : (E_S(h, f) = 0) \wedge (E_D(h, f) > \epsilon)] \leq \delta$ whenever n is large enough so that $g_H(2n)(1/2)^{\epsilon n/2} \leq \delta/2$ (will work if g_H is polynomial).

$\rightarrow$ When is $g_H() \in O(n^d)$ i.e polynomial bounded?

ANS. When the VC dimension of N is finite. ($g_H() \in O(n^d) <=> H$ has finite VC dimension)

Definition: A hypothesis class H over a domain X is said to **shatter** a set $T \subseteq X$ if every binary function on T can be obtained as the restriction $h|_T$ of some $h \in H$. The **VC Dimension** of H is the size of a largest set that is shattered by H.

$\rightarrow$ Examples:

H	**VC-d(H)**
Intervals in $\mathbb{R}$	2
Lines in $\mathbb{R}^2$	3
Axis aligned rectangles in $\mathbb{R}^2$	4
Polygons in $\mathbb{R}^2$	∞
Circles in $\mathbb{R}^2$	3
Half spaces in $\mathbb{R}^d$	d+1
Spheres in $\mathbb{R}^d$	d+1