Indian Institute of Technology Kanpur
CS777 Topics in Learning Theory

Scribe: Bhaskar Mukhoty
Instructor: Purushottam Kar
Date: February 05, 2018

LECTURE
11

VC-Dimension

1 Introduction

In the last class we discussed Massart's finite class lemma, which gives upper bound on Radamacher Complexity $R_n(\mathcal{F})$ when the size of function class $\mathcal{F}$ is finite, i.e. $|\mathcal{F}| < \infty$.

Definition 11.1. (Restriction of $\mathcal{F}$ to S): Let $\mathcal{F} \subseteq \{-1,1\}^{\mathcal{X}}$ be a function class and $S = \{x_1, x_2, ...x_n\} \subset \mathcal{X}$. The restriction of $\mathcal{F}$ to S is the set of possible functions in $\mathcal{F}$, from S to $\{-1,1\}$. That is,

$$\mathcal{F}_S = \{(f(x_1), f(x_2), .., f(x_n)) : f \in \mathcal{F}\}$$

As all $a \in \mathcal{F}_S$ is a vector $\{-1,1\}^n$, we have $|\mathcal{F}_S| \leq 2^n$ and $\|a\|_2 = \sqrt{n}$

The MFCL lemma states that,

$$\underset{S}{R}(\mathcal{F}) = \mathbb{E}_{\hat{\epsilon}_i} \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(x_i) \right| = \mathbb{E}_{\hat{\epsilon}_i} \sup_{a\in\mathcal{F}_S} \left| \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i a_i \right| \leq \frac{c}{n}\sqrt{2\lg|\mathcal{F}_S|} \tag{1}$$

where, $c = \sup_{a\in\mathcal{F}_S} \|a\|_2$.

Since, each function in $f \in \mathcal{F}$, has only one evaluation for the set points S, $f : (x_1, x_2, .., x_n) \mapsto (f(x_1), f(x_2), ..., f(x_n))$, we have, $|\mathcal{F}_S| \leq |\mathcal{F}|$, and,

$$R_n(\mathcal{F}) = \mathbb{E}_S \underset{S}{R}(\mathcal{F}) \leq \mathbb{E}_S \frac{c}{n}\sqrt{2\lg|\mathcal{F}_S|} \leq \frac{c}{n}\sqrt{2\lg|\mathcal{F}|} \tag{2}$$

Also, for the binary classifier, we can have the trivial bound as,

$$R_n(\mathcal{F}) \leq \mathbb{E}_S \frac{c}{n}\sqrt{2\lg|\mathcal{F}_S|} = \frac{\sqrt{n}}{n}\sqrt{2\lg 2^n} = \sqrt{2\lg 2}$$

In the previous lecture note we have proved that, if $\mathcal{F}, \mathcal{C}_\epsilon$ be function classes defined over the set $\mathcal{X}$, where $\mathcal{C}_\epsilon$ is a ϵ-covered of $\mathcal{F}$. Then,

$$R_n(\mathcal{F}) \leq \epsilon + R_n(\mathcal{C}_\epsilon)$$

Definition 11.2. (Growth Function:) The growth of a function class $\mathcal{F}$, denoted by $\Pi_n(\mathcal{F}) : \mathbb{N} \to \mathbb{N}$, is defined as:

$$\Pi_n(\mathcal{F}) = \max_{\mathcal{S}\subset\mathcal{X}, |\mathcal{S}|=n} |\mathcal{F}_S| \tag{3}$$

An immediate corollary of the above is:

$$R_n(\mathcal{F}) \leq \mathop{\mathbb{E}}_S \frac{c}{n}\sqrt{2\lg|\mathcal{F}_S|} \leq \frac{c}{n}\sqrt{2\lg\Pi_n(\mathcal{F})}$$

Definition 11.3. (Shattering): A function class $\mathcal{F}$ shatters a finite set $S \subset \mathcal{X}$, if the restriction $\mathcal{F}_S$, has all the possible function from S to $\{-1,1\}$. That is, $|\mathcal{F}_S| = 2^{|S|}$

Definition 11.4. (VC-dimension): VC-dimension of a function class $\mathcal{F}$ is the size of the largest set that $\mathcal{F}$ can shatter.

$$VC(\mathcal{F}) = \max\{n : \exists\,|S| = n, |\mathcal{F}_S| = 2^n\} \tag{4}$$

Let us define a function class, $\mathcal{F}_{threshold} = \{x \mapsto \mathbb{I}\{x > a\} - \mathbb{I}\{x \ \leq a\}, a \in \mathbb{R}\}$.

In order to see the VC-dimension of $\mathcal{F}_{threshold}$, consider two points having features $S = \{x_1, x_2\} \subset \mathbb{R}$, with $x_1 < x_2$, we can never have $f(x_1) = 1$ and $f(x_2) = -1$ for any $f \in \mathcal{F}_{threshold}$. So $(\mathcal{F}_{threshold})_S = 3 < 2^2$. So that, $VC(\mathcal{F}_{threshold}) = 1$

Exercise 11.1. What is $VC(\mathcal{F}_{threshold} \cup -\mathcal{F}_{threshold})$?

Exercise 11.2. $\mathcal{F}_{lin} = \{sgn\langle w, x\rangle\}$ for $w \in \mathbb{R}^d$. Prove that $VC(\mathcal{F}_{lin}) \leq d$.

Exercise 11.3. Show that VC-dimension of convex polygon is infinity.

2 Sauer-Shelah-Parles Lemma

From the definition of VC-dimension it is direct that if $VC(\mathcal{F}) = d$, then for and $n \leq d$ we have $\Pi_n(\mathcal{F}) = 2^n$. The lemma states that $\Pi_n(\mathcal{F}) \leq \sum_{i=0}^{d} \binom{n}{i} \leq (\frac{en}{d})^d$, which implies that for $n > d$, the growth function is polynomial over n.Shalev-Shwartz and Ben-David (2014)

As a corollary, suppose $\mathcal{A} \subset \{-1,1\}^n$ and $|\mathcal{A}| > \sum_{i=0}^{d} \binom{n}{i}$ then, there exists a set S, s.t. $|S| = d+1$, such that $\mathcal{A}$ shatters S.

Another useful fact would be:

$$R_n(\mathcal{F}) \leq \frac{\sqrt{n}}{n}\sqrt{2\lg\Pi_n(\mathcal{F})} \leq \frac{1}{\sqrt{n}}\sqrt{2d\lg\frac{en}{d}} \leq \sqrt{\frac{2d\lg n}{n}}$$

for $d \geq e$.

Suppose $\mathcal{F} \subset \{-1,1\}^{\mathcal{X}}$, $VC(\mathcal{F}) \leq d$, then we can give the following bound:

$$\mathbb{P}(\sup_{f\in\mathcal{F}}\left|er_{\mathcal{D}}^{l^{0-1}}[f] - er_S^{l^{0-1}}[f]\right| > \frac{1}{2}\sqrt{\frac{2d\log n}{n}} + \epsilon) \leq 2\exp(\frac{-n\epsilon^2}{4}) \tag{5}$$

So that, with high probability,

$er_{\mathcal{D}}^{l^{0-1}}[\hat{f}_{ERM}] \leq er_{\mathcal{D}}^{l^{0-1}}[\hat{f}_{ERM}] + \frac{1}{2}\sqrt{\frac{2d\log n}{n}} + \epsilon$

References

Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.