Lecture 24: November 26

Lecturer: Alessandro Rinaldo *Scribes: Boxiang Lyu*

24.1 Recap

Previously, we want to bound the random process

$$\|P_n - P\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \frac{1}{n}\left|\sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)])\right|, \quad X_1,\dots,X_n \overset{i.i.d.}{\sim} P$$

Our first result is

$$\mathbb{P}\left(\|P_n - P\|_{\mathcal{F}} \geq 2R_n(\mathcal{F}) + t\right) \leq \exp\left\{-\frac{nt}{2B^2}\right\}$$

where we assume

$$\|f\|_\infty = \sup_{x\in\mathcal{X}} |f(x)| \leq B, \quad \forall f \in \mathcal{F}$$

$$\frac{\mathbb{E}[\|P_n - P\|_{\mathcal{F}}]}{2} \leq R_n(\mathcal{F}) = \mathbb{E}_{\underset{\sim}{X},\underset{\sim}{\epsilon}}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\right|\right]$$

where $\mathcal{X} = (X_1,\dots,X_n)$, $\underset{\sim}{\epsilon} = (\epsilon_1,\dots,\epsilon_n) \overset{i.i.d.}{\sim}$ Rademacher, and $\underset{\sim}{X} \perp\!\!\!\perp \underset{\sim}{\epsilon}$ We can then focus on bounding $R_n\mathcal{F}$. We recall the definition that $\mathcal{F}$ has polynomial discrimination with parameter $\nu \geq 1$ when $|\mathcal{F}(X_1^n) \leq (n+1)^\nu$ for all n and $X_1^n = (X_1,\dots X_n) \subset \mathcal{X}$, where $\mathcal{F}(X_1^n)$ is defined as

$$\mathcal{F}(X_1^n) = \{(f(X_1),\dots,f(X_n)), f\in\mathcal{F}\} \subseteq \mathbb{R}^n$$

If $\mathcal{F}$ has polynomial discrimination, then $|R_n(\mathcal{F})| \leq 2B\sqrt{\nu\frac{\log(n+1)}{n}}$.

24.2 VC Theory

For all X_1^n, $|\mathcal{F}(X_1^n)| \leq 2^n$, where $\mathcal{F}$ is a class of functions taking binary values. $\mathcal{F}$ is a VC class when $|\mathcal{F}(X_1^n)|$ grows polynomially in n.

Definition 24.1 *Given a class $\mathcal{F}$ of $\{0,1\}$ valued functions we say that the n-tuple $X_1^n = (X_1,\dots,X_n) \subset \mathcal{X}$ is shattered by $\mathcal{F}$ if $|\mathcal{F}(X_1^n)| = 2^n$. VC dimension of $\mathcal{F}$ is the largest n such that some n-tuple X_1^n is shattered by $\mathcal{F}$. Write this $V(\mathcal{F})$ or V.*

If $n > V$ then no n-tuple X_1^n can be shattered by $\mathcal{F}$.

Remark 24.2 *Take $f \in \mathcal{F} \to \{0,1\}$ valued, then let $A = A(f) = \{x \in \mathcal{X}, f(x) = 1\}$ be a one to one correspondence between functions in $\mathcal{F}$ and the class $\mathcal{A}$ of subsets of $\mathcal{X}$ obtained this way.*

$$\mathcal{A} = \{A(f), f \in \mathcal{F}\}$$
$$\textit{VC-dim of } \mathcal{F} = \textit{VC-dim of } \mathcal{A}$$

In fact for any X_1^n,

$$\mathcal{F}(X_1^n) = \mathcal{A}(X_1^n) = \{A \cap X_1^n, A \in \mathcal{A}\}$$

Back to our example where $\mathcal{F} = \{\mathbb{1}_{(-\infty,z]}(\cdot), z \in \mathbb{R}\}$, $\mathcal{A} = \{(-\infty, z], z \in \mathbb{R}\}$, the VC-dimension is 1 because for all X_1^n

$$|\mathcal{F}(X_1^n)| = |\mathcal{A}(X_1^n)| \leq n + 1$$

. Consider when $\mathcal{A} = \{(a, b], -\infty < a < b\infty\}$, the VC dim is 2. In fact for all X_1^n, $\mathcal{A}(X_1^n) \leq (n + 1)^2$. If $n \geq V$ then $|\mathcal{A}(X_1^n)| < 2^n$ for all X_1^n but it could be close to being polynomial.

Lemma 24.3 *<u>Sauer Lemma</u>: Let V be the VC dim of $\mathcal{A}$ then for each n-tuple $X_1^n = (X_1, \ldots, X_n)$, for all $n \geq 1$*

$$|\mathcal{A}(X_1^n)| = |\{X_1^n \cap A, A \in \mathcal{A}\}| \leq \sum_{i=1}^{V} \binom{n}{V} \leq (n + 1)^V$$

Let $S_\mathcal{A}(n) = \max_{X_1^n} |\mathcal{A}(X_1^n)|$ be the shatter coefficient of $\mathcal{A}$. If $\mathcal{A}$ has VC dimension V then $S_\mathcal{A}(n) \leq (n+1)^V$. We can then obtain the classical result

$$\mathbb{E}\left[\sup_{A \in \mathcal{A}} |P_n(A) - P(A)|\right] \leq \sqrt{2\frac{\log S_\mathcal{A}(2n)}{n}}$$

where $P(A) = \frac{\#\{X_i, X_i \in A\}}{n}$.

24.3 Controlling/Calculating the VC Dimension

Let $\mathcal{A}$ and $\mathcal{B}$ be collections of subsets of $\mathcal{X}$ with VC dimensions $V_\mathcal{A}$ and $V_\mathcal{B}$ then

1. the class $\mathcal{A}^C = \{A^C, A \in \mathcal{A}\}$ has VC dimension $V_\mathcal{A}$.
2. the class $\mathcal{A} \coprod \mathcal{B} = \{A \cup B, A \in \mathcal{A}, B \in B\}$ is such that $S_{\mathcal{A} \coprod \mathcal{B}}(n) \leq S_\mathcal{A}(n) S_\mathcal{B}(n)$
3. the class $\mathcal{A} \prod \mathcal{B} = \{A \cup B, A \in \mathcal{A}, B \in \mathcal{B}\}$ is such that $S_{\mathcal{A} \prod \mathcal{B}}(n) \leq S_\mathcal{A}(n) S_\mathcal{B}(n)$
4. the class $\mathcal{A} \times \mathcal{B} = \{A \times B, A \in \mathcal{A}, B \in \mathcal{B}\}$ is such that $S_{\mathcal{A} \times \mathcal{B}} \leq S_\mathcal{A}(n) S_\mathcal{B}(n)$
5. $S_\mathcal{A}(n + m) = S_\mathcal{A}(n) S_\mathcal{A}(m)$
6. If $\mathcal{C} = \mathcal{A} \cup \mathcal{B} = \{C : C \in \mathcal{A}$ or $C \in \mathcal{B}$ or both$\}$ then $S_\mathcal{C}(n) \leq S_\mathcal{A}(n) + S_\mathcal{B}(n)$

Examples

1. $\mathcal{A} = \{A_1, \ldots, A_m\}$, $V_\mathcal{A} \leq \log_2 m$, $S_\mathcal{A}(X_1^n) \leq |\mathcal{A}| = m$ for all n.
2. $\mathcal{A} = \{(-\infty, z_1] \times \cdots \times (-\infty, z_d], (x_1, \ldots, x_d) \in \mathbb{R}^d\}$, $V_\mathcal{A} = d$

3. $\mathcal{A}$ collection of rectangles in $\mathbb{R}^d$. $V_A = 2d$

Vector Space Structure: Let $\mathcal{G}$ be a vector space of dimension r of functions on $\mathbb{R}^d$. Let

$$\mathcal{A} = \{\{x \in \mathbb{R}^d; g(x) \geq 0\}, g \in \mathcal{G}\}$$

then VC dim of $\mathcal{A} \leq dim(\mathcal{G}) = r$.

Applications:

1. $\mathcal{A} = \{\{x \in \mathbb{R}^d, x^T a \geq b\}, a \in \mathbb{R}^d, b \in \mathbb{R}\}$, class of half spaces in $\mathbb{R}^d$, $V(\mathcal{A}) \leq d+1$
2. $\mathcal{A} = \{\mathcal{B}(a,r), a \in \mathbb{R}^d, r > 0\}, \mathcal{B}(a,r) = \{x \in \mathbb{R}^d : \|x-a\|^2 \leq r^2\}$ then $V(A) \geq d+2$

Proof: Write

$$\sum_{i=1}^d (x_i - a_i)^2 - r = \sum_{i=1}^d x_i^2 + \sum_{i=1}^d a_i^2 - 2\sum_{i=1}^d x_i a_i - r$$

Let $g_1, g_2, \ldots, g_{d+2}$ be functions on $\mathbb{R}^d$ of the form

$$g_1(\mathcal{X}) = \sum_{i=1}^d x_i^2$$
$$g_2(\mathcal{X}) = x_1$$
$$\vdots$$
$$g_{d+1}(\mathcal{X}) = x_d$$
$$g_{d+2}(\mathcal{X}) = 1$$

where $\mathcal{X} = (x_1, \ldots, x_d)$. ■

Traditional Approach to VC Theory: We want to bound

$$\mathbb{P}\left(\sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \geq \lambda\right), \lambda > 0$$

where $P(A) = \frac{\#\{Y_i, Y_i \in A\}}{n}$, $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} P \perp\!\!\!\perp (X_1, \ldots, X_n)$.

Proof: <u>Part 1</u>: Symmetrization if $\lambda^2 n \geq 2$

$$\begin{aligned}\mathbb{P}\left(\sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \geq \lambda\right) &\leq 2\mathbb{P}\left(\sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \geq \lambda/2\right) \\ &= 2\mathbb{P}_{\underset{\sim}{X}, \underset{\sim}{Y}, \underset{\sim}{\epsilon}}\left(\sup_{A \in \mathcal{A}} \frac{1}{n}\left|\sum_{i=1}^n \epsilon_i(\mathbb{1}\{X_i \in A\} - \mathbb{1}\{Y_i \in A\})\right| \geq \lambda/2\right) \\ &= 2\mathbb{E}_{\underset{\sim}{X}, \underset{\sim}{Y}}\left[\mathbb{P}_{\underset{\sim}{\epsilon} \perp\!\!\!\perp \underset{\sim}{X}, \underset{\sim}{Y}}\left(\sup_{A \in \mathcal{A}} W_A | X, Y\right)\right]\end{aligned}$$

where $W_A = \frac{1}{n}\left|\sum_{i=1}^n \epsilon_i(\mathbb{1}\{X_i \in A\} - \mathbb{1}\{Y_i \in A\})\right|$ conditionally on $\underset{\sim}{X}, tY$. W_A is an average of iid RV's taking values in $\{-1, 1\}$. For fixed A,

$$P(W_A > \lambda/2 | \underset{\sim}{X}, \underset{\sim}{Y}) \leq 2\exp\left\{-\frac{n\lambda^2}{8}\right\}$$

by Hoeffding. Let $A^*(\underset{\sim}{X},\underset{\sim}{Y}) \subset \mathcal{A}$ be such that

$$\{A \cap (\underset{\sim}{X},\underset{\sim}{Y}), A \in \mathcal{A}^*(\underset{\sim}{X},\underset{\sim}{Y}) = \{A \cap \{\underset{\sim}{X},\underset{\sim}{Y}\}, A \in \mathcal{A}\}$$

then $|\mathcal{A}^*(\underset{\sim}{X},\underset{\sim}{Y}) \leq S_{2\mathcal{A}}(2n)|$. We then have that

$$\begin{aligned}\mathbb{P}_{\underset{\sim}{\epsilon}|\underset{\sim}{X},\underset{\sim}{Y}}\left(\sup_{A\in\mathcal{A}} W_A \geq \lambda/2|\underset{\sim}{X},\underset{\sim}{Y}\right) &= \mathbb{P}_{\underset{\sim}{\epsilon}|\underset{\sim}{X},\underset{\sim}{Y}}\left(\max_{A\in\mathcal{A}^*(\underset{\sim}{X},\underset{\sim}{Y})} W_A \geq \lambda/2|\underset{\sim}{X},\underset{\sim}{Y}\right)\\ &\leq S_{\mathcal{A}}(2n)\cdot 2\exp\{-\frac{n\lambda^2}{8}\}\end{aligned}$$

■