18.1 Finite Class Lemma

Last time, we saw ways to use the concentration of Δ_n to control $\mathfrak{R}_n(\mathcal{F})$ (and thus $R(\widehat{f})$). Today we will discuss other ways to bound Rademacher complexity.

Lemma 18.1 (finite class (v1))**.** *If* $\mathcal{A} = \{a^{(1)}, ..., a^{(N)}\} \subset \mathbb{R}^n$ *is a finite set with* $||a^{(i)}|| \leq L \forall i$ *and* $N \geq 2$, *then*

$$\mathfrak{R}_n(\mathcal{A}) \leq \frac{4L\sqrt{\log N}}{n}$$

Proof. By some tricks for sub-Gaussian random variables as in last week's example,

$$\begin{aligned}
Y_j &= \frac{2}{n}\sum_{i=1}^{n} \sigma_i a_i^{(j)} && (18.1)\\
\mathbb{E}\big[e^{tY_j}\big] &\leq e^{2L^2t^2/n^2} && (18.2)\\
&\text{and} && (18.3)\\
\mathbb{E}\big[e^{-tY_j}\big] &\leq e^{2L^2t^2/n^2}, \text{ so} && (18.4)\\
\mathbb{E}\left[\max_j |Y_j|\right] &= \mathbb{E}[\max Y_1, -Y_1, ..., Y_N, -Y_N] && (18.5)\\
&\leq 2L\sqrt{log2N}/n(hereN \geq 2, 2N \leq N^2) && (18.6)\\
&\leq 4L\sqrt{\log N}/n && (18.7)
\end{aligned}$$

Thus, with probability $> 1-\delta$,

$$R_n(a^{(j)} \leq \widehat{R}_n(a^{(j)} + 8L\sqrt{\log N}/n + \sqrt{\frac{\log 1/\delta}{n}}$$

■

Lemma 18.2 (finite class (v2))**.** *This version of the lemma doesn't use Rademacher complexity.*

$$\begin{aligned}
P(\sup_j |\widehat{R}_n(a^{(j)}) - R_n(a^{(j)})| \geq \epsilon) &\leq \sum_{j=1}^{N} P(|\widehat{R}_n(a^{(j)}) - R_n(a^{(j)})| \geq \epsilon) && (18.8)\\
&\leq 2Ne^{-n\epsilon^2/(2L)^2} && (18.9)\\
\textit{so w.p. } &> 1-\delta, && (18.10)\\
R_n(a^{(j)}) &\leq \widehat{R}_n(a^{(j)}) + 2L\sqrt{\frac{2\log N + \log 1/\delta}{n}} && (18.11)
\end{aligned}$$

How can we extend this to infinite classes? Recall that $\mathcal{F}$is projected onto the data z^n : $\mathcal{F}(z^n) = \{f(z_1), ..., f(z_n) : f \in \mathcal{F}\}$. This is the set whose size we want to find.

Suppose $f \mapsto \{0,1\}\,\forall f \in \mathcal{F}$. Then $\mathcal{F}(z^n) \subseteq \{0,1\}^n$ and thus, $\forall f \in \mathcal{F}$, $\sqrt{\sum_{i=1}^n |f(z_i)|^2} \leq \sqrt{n}$.

So $\widehat{\mathfrak{R}}_n(\mathcal{F}(z^n)) \leq 4\sqrt{\log|\mathcal{F}(z^n)|/n}$

How can we get a tighter bound?

18.2 VC-dimension

Definition 18.3 (Shattering)**.** *Let $\mathcal{C}$be a class of subsets of Z. We say that $S = \{z_1, ..., z_n\} \subset \mathbb{Z}$ (finite) is* ***shattered*** *by $\mathcal{C}$if $\forall S' \subseteq S : \exists C \in \mathcal{C}$ s. t. $S' = S \cap C$. In this case we say C "picks out" S'.*

Equivalently, we can say S is shattered by $\mathcal{C}$if:

$\forall b \in \{0,1\}^n : \exists C \in \mathcal{C}$ s. t. $(I(z_1 \in C), ..., I(z_n \in C)) = b \iff \{(I(z_1 \in C), ..., I(z_n \in C)) : C \in \mathcal{C}\} = \{0,1\}^n$

Some example classes $\mathcal{C}$:

$$\mathcal{C} = \{(-\infty, t] : t \in \mathbb{R}\}, \text{ half-open intervals} \tag{18.12}$$
$$\mathcal{C} = \{(a,b) : a \leq b\}, \text{ open intervals} \tag{18.13}$$
$$\mathcal{C} = \{(a,b) \cup (c,d) : a \leq b \leq c \leq d\}, \text{ unions of open intervals} \tag{18.14}$$
$$\mathcal{C} = \text{ all discs in } \mathbb{R}^d \tag{18.15}$$
$$\mathcal{C} = \text{all axis parallel rectangles in } \mathbb{R}^d \tag{18.16}$$
$$\mathcal{C} = \{x : \beta^T x \geq 0\}, \text{ all half spaces} \tag{18.17}$$
$$\mathcal{C} = \text{all convex sets} \tag{18.18}$$

Example 18.4. *Let $\mathcal{C} = \{(a,b) : a \leq b\}, S = \{1,2,3\}$. Which $S' \subset S$ can $\mathcal{C}$pick out? $\mathcal{C}$picks out all $S' \in \wp(S)$ EXCEPT $\{1,3\}$. So S is not shattered by $\mathcal{C}$.*

More generally, $\mathcal{C} = \{(a,b) : a \leq b\}$ can shatter any set of size two ($S = \{s,t\}$ where $s \neq t$), but only trivial sets of size 3.

Definition 18.5 (Vapnik-Chervonenkis (VC) dimension of $\mathcal{C}$)**.** *The* ***VC-dimension*** *is*

$$V(\mathcal{C}) := \max\{n \in \mathbb{N} : \exists S \subset \mathbb{Z} \text{ s. t. } |S| = n \text{ and } S \text{ is shattered by } \mathcal{C}\}.$$

If $V(\mathcal{C}) < \infty$ we say $\mathcal{C}$is a ***VC-class****.*

VC is well-defined: $\mathcal{S}_n < 2^n \Rightarrow \mathcal{S}_m < 2^m\,\forall m > n$. The hard part is always proving that no set of size $> n$ can be shattered.

Definition 18.6 (the n^{th} shatter coefficient)**.** $\mathcal{S}_n\mathcal{C} := \sup_{S \subset \mathbb{Z}, |S|=n} |\{S \cap C : C \in \mathcal{C}\}|$

The shattering coefficient gives us another way to define VC-dimension: $V(\mathcal{C}) := \max\{n \in \mathbb{N} : \mathcal{S}_n(\mathcal{C}) = 2^n\}$.

Example 18.7. *For $\mathcal{C} = \{(a,b) : a \leq b\}$, $\mathcal{S}_2(\mathcal{C}) = 4$, $\mathcal{S}_3(\mathcal{C}) = 7$.*

Example 18.8. *VC-dim for binary functions: $\forall f : \mathbb{Z} \to \{0,1\}, let C_f = \{z : f(z) = 1\}$.*

$\forall C \subseteq \mathbb{Z}$, let $f_C(x) = I(x \in C)$.

Definition 18.9. *Let $\mathcal{F}$be a class of functions $f : \mathbb{Z} \to \{0,1\}$. We say S is shattered by $\mathcal{F}$if S is shattered by $C_{\mathcal{F}} = \{I(f = 1) : f \in \mathcal{F}\}$, where $I(f = 1)$ is the indicator of C_f. In addition, let*

$$\mathcal{S}_n(\mathcal{F}) := \mathcal{S}_n(\mathcal{C}_{\mathcal{F}}) \tag{18.19}$$
$$V(\mathcal{F}) := V(\mathcal{C}_{\mathcal{F}}) \tag{18.20}$$

Example 18.10 (semi-infinite intervals)**.** *For $\mathcal{C} = \{(-\infty, t) : t \in \mathbb{R}\}$, $V(\mathcal{C}) = 1$. Clearly, $\mathcal{C}$ can shatter some 1-point set: take $S = \{0\}$. Then $(-\infty, -1] \cap S = \emptyset$ and $(-\infty, -1] \cap S = S$. So we know $V(\mathcal{C}) \geq 1$.*

To show there is no 2-point set, let $S = \{a, b\}, a < b$. Obviously, $\nexists t$ s. t. $(-\infty, t] = \{b\}$.

Example 18.11 (open intervals)**.** *Let $\mathcal{C} = \{(a, b) : a \leq b\}$.*

1. *$S = \{s, t\}$ $s < t$. choose $a_1 < a_2 < s < a_3 < t < a_4$. Then*

$$(a_1, a_2) \cap S = \emptyset \tag{18.21}$$
$$(a_2, a_3) \cap S = \{s\} \tag{18.22}$$
$$(a_1, a_4) \cap S = \{t\} \tag{18.23}$$
$$\tag{18.24}$$

2. *For any $S = \{s, t, u\}(s < t < u)$ and $a_1 < a_2$,*
   $\{s, u\} \subset (a_1, a_2) \;\Rightarrow\; \{t\} \subset (a_1, a_2)$