

20. Covering number bounds

Recap

- We have proved upper bounds on the RC, for:
 - linear models

$$\mathbb{E} \sup_{f \in \mathcal{F}} \langle \vec{\varepsilon}, f(Z^n) \rangle \leq C_{\text{norm}} \cdot \sqrt{\frac{1}{n}}$$

- two-layer networks

$$\mathbb{E} \sup_{f \in \mathcal{F}} \langle \vec{\varepsilon}, f(Z^n) \rangle \leq C_{\text{norm}} \cdot \sqrt{\frac{m}{n}}$$

- **Key.** **Peeling** technique
 - Somewhat specialized — depending on norm choices
- **Today.** A slightly more general technique — Covering number bounds

Overview

- Consider \mathcal{G} — a **finite approximation** of \mathcal{F}
 - Assume that, for any $f \in \mathcal{F}$, there exists $g \in \mathcal{G}$ such that

$$\|f(Z^n) - g(Z^n)\| \leq \gamma$$

- Then, we know that

$$\begin{aligned} \mathbb{E} \sup_{f \in \mathcal{F}} \langle \vec{\varepsilon}, f(Z^n) \rangle &= \mathbb{E} \sup_{f \in \mathcal{F}} \inf_{g \in \mathcal{G}} \left(\langle \vec{\varepsilon}, f(Z^n) - g(Z^n) \rangle + \langle \vec{\varepsilon}, g(Z^n) \rangle \right) \\ &\leq \mathbb{E} \sup_{f \in \mathcal{F}} \inf_{g \in \mathcal{G}} \left(\langle \vec{\varepsilon}, f(Z^n) - g(Z^n) \rangle \right) + \mathbb{E} \sup_{g \in \mathcal{G}} \langle \vec{\varepsilon}, g(Z^n) \rangle \\ &\leq \|\vec{\varepsilon}\|_2 \cdot \gamma + \mathbb{E} \sup_{g \in \mathcal{G}} \langle \vec{\varepsilon}, g(Z^n) \rangle \\ &\leq \gamma \sqrt{n} + \sqrt{n \log |\mathcal{G}|} \end{aligned}$$

Overview

- Then, we know that

$$\mathbb{E}\mathfrak{R}(\mathcal{F}) \leq \frac{1}{\sqrt{n}} \cdot \left(\gamma + \sqrt{\log |\mathcal{G}|} \right)$$

- **Key quantity.** The tradeoff of \mathcal{G} corresponding to γ
 - How to achieve a smooth tradeoff? **Chaining**

$$\begin{aligned} \langle \vec{\varepsilon}, f(Z^n) \rangle &= \langle \vec{\varepsilon}, f(Z^n) - g(Z^n) \rangle + \langle \vec{\varepsilon}, g(Z^n) \rangle \\ &= \langle \vec{\varepsilon}, f(Z^n) - g(Z^n) \rangle + \langle \vec{\varepsilon}, g(Z^n) - h(Z^n) \rangle + \langle \vec{\varepsilon}, h(Z^n) \rangle \\ &\quad \dots \end{aligned}$$

Formalisms

Definition (**cover**).

Given a set U , scale ϵ , and norm $\| \cdot \|$, a set $V \subseteq U$ is a (proper) cover when we have

$$\sup_{a \in U} \inf_{b \in V} \|a - b\| \leq \epsilon$$

Definition (**covering number**).

The covering number $\mathcal{N}(U, \epsilon, \| \cdot \|)$ is the minimum cardinality of a cover for the set U .

- Note. In fact, we often can remove the restriction that $V \subseteq U$
 - Called “net”
- Note. We have seen these, when we discussed Maurey’s empirical method

Warm up

- **Question.** Define an infinity-norm ball

$$U = \{u \in \mathbb{R}^d \mid \|u\|_\infty \leq 1\}$$

- What is the covering number in infinity-norm, i.e.,

$$\mathcal{N} \left(U, \|\cdot\|_\infty, \frac{1}{10} \right)$$

Covering vs. Packing

Definition (**Packing**).

A finite set $V \subseteq U$ is an ϵ -packing of a set U , whenever

$$\|v_i - v_j\| > \epsilon, \quad \forall v_i, v_j \in V, \quad i \neq j$$

- Let $\mathcal{M}(U, \|\cdot\|, \epsilon)$ denote the cardinality of the maximal packing — i.e., packing number

Proposition.

We have

$$\mathcal{M}(U, \|\cdot\|, 2\epsilon) \leq \mathcal{N}(U, \|\cdot\|, \epsilon) \leq \mathcal{M}(U, \|\cdot\|, \epsilon)$$

Proof sketch

$$\mathcal{M}(U, \|\cdot\|, 2\epsilon) \leq \mathcal{N}(U, \|\cdot\|, \epsilon) \leq \mathcal{M}(U, \|\cdot\|, \epsilon)$$

- **Right ineq.** Let V be a maximal packing.
 - No other point in $u \in U$ can be added.
 - Thus is a cover
 - \mathcal{N} is the “minimal” covering
- **Left ineq.** By contradiction
 - Suppose that there exists a cover V_c and packing V_p with $|V_p| \geq |V_c| + 1$
 - By pigeonhole, some elements of V_p should fall in the same covering ball
 - Can't do!

Volumetric argument

Proposition.

Let $\|\cdot\|$ be some norm on \mathbb{R}^d , and let $U = \{u \in \mathbb{R}^d \mid \|u\| \leq 1\}$. Then,

$$\frac{1}{\epsilon^d} \leq \mathcal{N}(U, \|\cdot\|, \epsilon) \leq \left(1 + \frac{2}{\epsilon}\right)^d$$

- **Idea.** Via **volumetric** argument

- Left. Union of N ϵ -balls is a superset of one unit ball — thus should have larger volume
- Right. Covering is smaller than or equal to packing
 - For packing, balls with radius $\epsilon/2$ are disjoint
 - The union of balls are contained in a $(1 + \epsilon/2)$ -ball

Rademacher vs. Cover

Theorem 15.1.

Given $U \subseteq \mathbb{R}^n$, we have

$$n \cdot \mathfrak{R}(U) \leq \inf_{\alpha > 0} \left(\alpha \sqrt{n} + \left(\sup_{u \in U} \|u\|_2 \right) \sqrt{2 \log \mathcal{N}(U, \alpha, \|\cdot\|_2)} \right)$$

- Note. This is simply a re-statement of what we saw in the beginning of this class.

Chaining

- Suppose that we have covers V_1, V_2, V_3, \dots at scales $1, 2^{-1}, 2^{-2}, \dots$
- Given some $a \in U$, choose $V_i(a) = \arg \min_{b \in V_i} \|a - b\|$

- Then, we have

$$a = (a - V_N(a)) + (V_N(a) - V_{N-1}(a)) + (V_{N-1}(a) - V_{N-2}(a)) + \dots$$

Theorem 15.2 (**Chaining**).

Let $U \subseteq [-1, +1]^n$ be given with $\mathbf{0} \in U$. Then,

$$\begin{aligned} n \cdot \mathfrak{R}(U) &\leq \inf_{N \in \mathbb{N}} \left(n \cdot 2^{-N+1} + 6\sqrt{n} \sum_{i=1}^{N-1} 2^{-i} \sqrt{\log \mathcal{N} \left(U, \|\cdot\|_2, \sqrt{n} 2^{-i} \right)} \right) \\ &\leq \inf_{\alpha > 0} \left(4\alpha\sqrt{n} + 12 \int_{\alpha}^{\sqrt{n}/2} \sqrt{\log \mathcal{N}(U, \|\cdot\|_2, \beta)} \, d\beta \right) \end{aligned}$$

Proof sketch

$$n \cdot \mathfrak{R}(U) \leq \inf_{N \in \mathbb{N}} \left(n \cdot 2^{-N+1} + 6\sqrt{n} \sum_{i=1}^{N-1} 2^{-i} \sqrt{\log \mathcal{N} \left(U, \|\cdot\|_2, \sqrt{n} 2^{-i} \right)} \right)$$

- Let V_1, V_2, \dots be covers with scale $\sqrt{n}, \sqrt{n}2^{-1}, \sqrt{n}2^{-2}, \dots$

$$\begin{aligned} \mathbb{E} \sup_{u \in U} \langle \varepsilon, u \rangle &= \mathbb{E} \sup_{u \in U} \left(\langle \varepsilon, u - V_N(u) \rangle + \sum_{i=1}^{N-1} \langle \varepsilon, V_{i+1}(u) - V_i(u) \rangle + \langle \varepsilon, V_1(u) \rangle \right) \\ &\leq \mathbb{E} \sup_{u \in U} \langle \varepsilon, u - V_N(u) \rangle + \mathbb{E} \sup_{u \in U} \left(\sum_{i=1}^{N-1} \langle \varepsilon, V_{i+1}(u) - V_i(u) \rangle \right) + \mathbb{E} \sup_{u \in U} \langle \varepsilon, V_1(u) \rangle \\ &\leq \|\varepsilon\| \cdot \|u - V_N(u)\| &= 0 \\ &\leq n \cdot 2^{1-N} \end{aligned}$$

Proof sketch

$$n \cdot \mathfrak{R}(U) \leq \inf_{N \in \mathbb{N}} \left(n \cdot 2^{-N+1} + 6\sqrt{n} \sum_{i=1}^{N-1} 2^{-i} \sqrt{\log \mathcal{N} \left(U, \|\cdot\|_2, \sqrt{n} 2^{-i} \right)} \right)$$

$$\mathbb{E} \sup_{u \in U} \langle \epsilon, u \rangle \leq n \cdot 2^{1-N} + \mathbb{E} \sup_{u \in U} \left(\sum_{i=1}^{N-1} \langle \epsilon, V_{i+1}(u) - V_i(u) \rangle \right)$$

- To bound the second term, apply the finite class lemma: $W_i = V_{i+1}(u) - V_i(u)$

- Cardinality. $|V_{i+1}| \cdot |V_i| \leq |V_{i+1}|^2$
- Max norm. $\leq \|V_{i+1}(u) - u + u - V_i(u)\|$
 $\leq \|V_{i+1}(u) - u\| + \|u - V_i(u)\|$
 $\leq \sqrt{n} 2^{-i} + \sqrt{n} 2^{1-i}$
 $= 3 \cdot \sqrt{n} 2^{-i}$

Next up

- Covering number bounds for neural nets
- VC-dimension