Lecture 24. (Sketch.)

- Hwk3 is out. It is due December 18. After today's lecture, you have everything you need to solve all problems.
- Project presentations next week!

Rademacher recap (same slide as before!).

Concentration controlled one function at a time. To control many functions, out main tool is (unnormalized) Rademacher complexity:

$$\mathrm{URad}(V) := \mathbb{E} \sup_{u \in V} \langle \epsilon, u \rangle, \qquad \mathrm{Rad}(V) := \frac{1}{n}\mathrm{URad}(V).$$

Given data $S := (Z_1, \dots, Z_n)$ and functions $\mathcal{F}$, define vectors

$$\mathcal{F}_{|S} := \{(f(Z_1), \dots, f(Z_n)) : f \in \mathcal{F}\}.$$

Our main generalization tool involves $\mathrm{URad}(\mathcal{F}_{|S})$, and is a consequence of our two symmetrization lemmas and McDiarmid's inequality.

Theorem. Let $\mathcal{F}$ be given with $f(z) \in [a, b]$ a.s.. With probability $\geq 1 - \delta$,

$$\sup_{f \in \mathcal{F}} \mathbb{E} f - \widehat{\mathbb{E}}_n f \leq \frac{2}{n}\mathrm{URad}(\mathcal{F}_{|S}) + 3(b-a)\sqrt{\frac{\ln(2/\delta)}{2n}}.$$

Theorem (Massart finite lemma).
$\mathrm{URad}(V) \leq \sup_{u \in V} \|u\|_2 \sqrt{2 \ln |V|}$.

This lemma can be applied to infinite $\mathcal{F}$ by discretizing $\mathcal{F}_{|S}$.

1. Covering numbers; Pollard's bound.

We'll discretize via **covering numbers**.

Definition. Given a set U, scale ϵ, norm $\|\cdot\|$, $V \subseteq U$ is a **(proper) cover** when

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

Let $\mathcal{N}(U, \epsilon, \|\cdot\|)$ denote the **covering number**: the minimum cardinality (proper) cover.

Remarks.

- "Improper" covers drop the requirement $V \subseteq U$. (We'll come back to this.)
- Most treatments define special norms with normalization $1/n$ baked in; we'll use unnormalized Rademacher complexity and covering numbers.
- At the end of an early lecture we gave "primitive covers"; those used $\mathcal{F}$ not $\mathcal{F}_{|S}$ and $\|\cdot\|_{\mathrm{u}}$.

Theorem (Pollard bound). Given $U \subseteq \mathbb{R}^n$,

$$\mathrm{URad}(U) \leq \inf_{\alpha>0} \left(\alpha\sqrt{n} + \left(\sup_{a\in U} \|a\|_2\right) \sqrt{2\ln \mathcal{N}(U, \alpha, \|\cdot\|_2)} \right).$$

Remarks.

- $\|\cdot\|_2$ comes from applying Massart. It's unclear how to handle other norms without some technical slop.

Proof. Let $\alpha > 0$ be arbitrary, and suppose $\mathcal{N}(U, \alpha, \|\cdot\|_2) = \infty$ (otherwise bound holds trivially). Let V denote a minimal cover, and $V(a)$ its closest element to $a \in U$. Then

$$\begin{aligned}
\mathrm{URad}(U) &= \mathbb{E} \sup_{a\in U} \langle \epsilon, v\rangle \\
&= \mathbb{E} \sup_{a\in U} \langle \epsilon, v - V(a) + V(a)\rangle \\
&= \mathbb{E} \sup_{a\in U} \left(\langle \epsilon, V(a)\rangle + \langle \epsilon, v - V(a)\rangle\right) \\
&\leq \mathbb{E} \sup_{a\in U} \left(\langle \epsilon, V(a)\rangle + \|\epsilon\| \cdot \|v - V(a)\|\right) \\
&\leq \mathrm{URad}(V) + \alpha\sqrt{n} \\
&\leq \sup_{b\in V}(\|b\|_2)\sqrt{2\ln|V|} + \alpha\sqrt{n} \\
&\leq \sup_{a\in U}(\|a\|_2)\sqrt{2\ln|V|} + \alpha\sqrt{n},
\end{aligned}$$

and the bound follows since $\alpha > 0$ was arbitrary.

Remarks.

- The same proof handles improper covers with minor adjustment: for every $b \in V$, there must be $U(b) \in U$ with $\|b - U(v)\| \leq \alpha$ (otherwise, b can be moved closer to U), thus

$$\sup_{b\in V} \|b\|_2 \leq \sup_{b\in V} \|b - U(b)\|_2 + \|U(b)\|_2 \leq \alpha + \sup_{a\in U} \|a\|_2.$$

- To handle other norms, superficially we need two adjustments: Cauchy-Schwarz can be replaced with Hölder, but it's unclear how to replace Massart without slop relating different norms.

2. The Dudley entropy integral.

- As made clear in the homework, the Pollard bound is *not tight*.
- We will present a different bound, the *Dudley entropy integral*, and in a remark at the end explain that it is tight with Rademacher complexity (and note the Pollard bound!).

- The Dudley entropy integral works at *multiple scales*.
  - Suppose we have covers $(V_N, V_{N-1}, \ldots)$ at scales $(\alpha_N, \alpha_N/2, \alpha_N/4, \ldots)$.
  - Given $a \in U$, choosing $V_i(a) := \arg\min_{b \in V_i} \|a - b\|$,
    $$a = (a - V_N(a)) + (V_N(a) - V_{N-1}(a)) + (V_{N-1}(a) - V_{N-2}(a)) + \cdots .$$
    We are thus rewriting a as a sequence of **increments** at different scales.
  - One way to think of it is as writing a number as its binary expansion
    $$x = (0.b_1 b_2 b_3 \ldots) = \sum_{i \geq 1} \frac{(b_i.b_{i+1} \ldots) - (0.b_{i+1} \ldots)}{2^i} = \sum_{i \geq 1} \frac{b_i}{2^i}.$$
    In the Dudley entropy integral, we are covering these **increments** b_i, rather than the number x directly.
  - One can cover increments via covering numbers for the base set, and that is why these basic covering numbers appear in the Dudley entropy integral. But internally, the argument really is about these increments.

Theorem (Dudley). Let $U \subseteq [-1, +1]^n$ be given with $0 \in U$.

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

Proof. We'll do the discrete sum first. The integral follows by relating an integral to its Riemann sum.

- Let $N \geq 1$ be arbitrary.
- For $i \in \{1, \ldots, N\}$, define scales $\alpha_i := \sqrt{n} 2^{1-i}$.
- Define cover $V_1 := \{0\}$; since $U \subseteq [-1, +1]^n$, this is a minimal cover at scale $\sqrt{n} = \alpha_1$.
- Let V_i for $i \in \{2, \ldots, N\}$ denote any minimal cover at scale α_i, meaning $|V_i| = \mathcal{N}(U, \alpha_i, \|\cdot\|_2)$.

Proof (continued).

Since $U \ni a = (a - V_N(a)) + \sum_{i=1}^{N-1} (V_{i+1}(a) - V_i(a)) + V_1(a)$,

$$\begin{aligned}&\mathrm{URad}(U) \\ &= \mathbb{E} \sup_{a \in U} \langle \epsilon, a \rangle \\ &= \mathbb{E} \sup_{a \in U} \left(\langle \epsilon, a - V_N(a) \rangle + \sum_{i=1}^{N-1} \langle \epsilon, V_{i+1}(a) - V_i(a) \rangle + \langle \epsilon, V_1(a) \rangle \right) \\ &\leq \mathbb{E} \sup_{a \in U} \langle \epsilon, a - V_N(a) \rangle \\ &\quad + \sum_{i=1}^{N-1} \mathbb{E} \sup_{a \in U} \langle \epsilon, V_{i+1} - V_i(a) \rangle \\ &\quad + \mathbb{E} \sup_{a \in U} \langle \epsilon, V_1(a) \rangle .\end{aligned}$$

Let's now control these terms separately.

Proof (continued). The first and last terms are easy:

$$\mathbb{E} \sup_{a \in U} \epsilon V_1(a) = \mathbb{E} \langle \epsilon, 0 \rangle = 0,$$

$$\mathbb{E} \sup_{a \in U} \langle \epsilon, a - V_N(a) \rangle \leq \mathbb{E} \sup_{a \in U} \|\epsilon\| \|a - V_N(a)\| \leq \sqrt{n} \alpha_N = n 2^{1-N}.$$

For the middle term, define **increment class** $W_i := \{V_{i+1}(a) - V_i(a) : a \in U\}$, whereby $|W_i| \leq |V_{i+1}| \cdot |V_i| \leq |V_{i+1}|^2$, and

$$\begin{aligned}\mathbb{E} \sup_{a \in U} \langle \epsilon, V_{i+1}(a) - V_i(a) \rangle &= \mathrm{URad}(W_i) \\ &\leq \left(\sup_{w \in W_i} \|w\|_2 \right) \sqrt{2 \ln |W_i|} \leq \left(\sup_{w \in W_i} \|w\|_2 \right) \sqrt{4 \ln |V_{i+1}|},\end{aligned}$$

$$\sup_{w \in W_i} \|w\| \leq \sup_{a \in U} \|V_{i+1}\| + \|a - V_i(a)\| \leq \alpha_{i+1} + \alpha_i = 3\alpha_{i+1}.$$

Combining these bounds,

$$\mathrm{URad}(U) \leq n 2^{1-N} + 0 + \sum_{i=1}^{N} 6\sqrt{n} 2^{-i} \sqrt{\ln \mathcal{N}(U, 2^{-i}\sqrt{n}, \|\cdot\|_2}.$$

$N \geq 1$ was arbitrary, so applying $\inf_{N \geq 1}$ gives the first bound.

Proof (integral form). Since $\ln \mathcal{N}(U, \beta, \|\cdot\|_2)$ is nonincreasing in β, the integral upper bounds the Riemann sum:

$$\begin{aligned} URad(U) &\leq n2^{1-N} + 6 \sum_{i=1}^{N-1} \alpha_{i+1} \sqrt{\ln \mathcal{N}(U, \alpha_{i+1}, \|\cdot\|)} \\ &= n2^{1-N} + 12 \sum_{i=1}^{N-1} (\alpha_{i+1} - \alpha_{i+2}) \sqrt{\ln \mathcal{N}(U, \alpha_{i+1}, \|\cdot\|)} \\ &\leq \sqrt{n}\alpha_N + 12 \int_{\alpha_{N+1}}^{\alpha_2} \sqrt{\ln \mathcal{N}(U, \alpha_{i+1}, \|\cdot\|)} \, d\beta. \end{aligned}$$

To finish, pick $\alpha > 0$ and N with

$$\alpha_{N+1} \geq \alpha > \alpha_{N+2} = \frac{\alpha_{N+1}}{2} = \frac{\alpha_{N+2}}{4},$$

whereby

$$\begin{aligned} \mathrm{URad}(U) &\leq \sqrt{n}\alpha_N + 12 \int_{\alpha_{N+1}}^{\alpha_2} \sqrt{\ln \mathcal{N}(U, \alpha_{i+1}, \|\cdot\|)} \, d\beta \\ &\leq 4\sqrt{n}\alpha + 12 \int_{\alpha}^{\sqrt{n}/2} \sqrt{\ln \mathcal{N}(U, \alpha_{i+1}, \|\cdot\|)} \, d\beta. \end{aligned}$$

Remarks.

- Tightness of Dudley: Sudakov's lower bound says there exists a universal C with

$$\mathrm{URad}(U) \geq \frac{c}{\ln(n)} \sup_{\alpha > 0} \alpha \sqrt{\ln \mathcal{N}(U, \alpha, \|\cdot\|)},$$

  which implies $\mathrm{URad}(U) = \widetilde{\Theta}$ (Dudley entropy integral).
- Taking the notion of increments to heart and generalizing the proof gives the concept of **chaining**. One key question there is tightening the relationship with Rademacher complexity (shrinking constants and log factors in the above bound).
- Another term for covering is "metric entropy".
- Recall once again that we drop the normalization $1/n$ from URad and the choice of norm when covering.