Lecture 18: Metric Entropy Continued

Isabella Zhu

15 April 2025

§1 One-Step Discretization

As a reminder from last class, we look at variables indexed by f

$$W_f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i f(x_i), \quad f \in F$$

The conditions we need to calculate probabilities on tail bounds are

$$\mathbb{E}[W_f] = 0$$

and

$$\mathbb{E}[e^{\lambda(W_f - W_g)}] \leq e^{\lambda^2 \gamma^2(f,g)/2}$$

for some metric γ on F. One example of such a metric is

$$\gamma^2(f,g) = \frac{1}{n} \sum_{i=1}^{n} (f(x_i) - g(x_i))^2$$

Corollary 1.1

(*1-step discretization for empirical process.*) We have Rademacher complexity

$$\mathbb{E}\left[\sup_{f \in F} . \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(X_i)\right] \leq \delta + R\sqrt{2 \log N(\delta; F, ||.||_n)}$$

where $R = \sup_{f \in F} ||f||_n$ and $N(\delta; F; ||.||_n)$ is the δ-covering of F in $||.||_n$. This is valid for every $\delta > 0$.

Proof. Left as exercise.

§1.1 Examples

Example 1.2

(Parametric class.) Parametric classes typically have metric entropy scaling

$$\log N(\delta) \approx d \log(1/\delta)$$

Example 1.3

(Non-parametric entropy.) We might have a class where

$$\log N(\delta) \approx (1/\delta)^{\alpha}$$

for some $\alpha > 0$.

Remark 1.4. The key takeaway here is that bounds for nonparametric classes suck.

§2 Non-Parametric Classes

Example 2.1

(Class with metric entropy scaling as $1/\delta$.) Consider the function class

$$F = \{f : [0,1] \to \mathbb{R} \mid f(0) = 0, \;\; |f(x) - f(\tilde{x})| \le |x - \tilde{x}|\}$$

and let $||f||_{\infty} = \sup_{x \in [0,1]} |f(x)|$. Prove that

$$\log N(\delta; F, ||.||_{\infty}) = \frac{1}{\delta}$$

Proof. Chunk the unit interval into a bunch of points $0, \delta, 2\delta, \ldots T\delta$ such that $T\delta \le 1$. We construct our function as follows: at every tick, decide whether to go up or down.

Formally, let $\beta \in \{-1, 1\}^T$. Consider the function

$$f_{\beta}(x) = \sum_{t=1}^{T} \beta_t \delta \varphi \left(\frac{x - x_t}{\delta} \right)$$

where φ is the ramp function

$$\varphi(x) = \begin{cases} x & x \in [0,1] \\ 1 & x \ge 1 \\ 0 & x \le 0 \end{cases}$$

By construction, if $\beta \neq \beta'$, then $||f_{\beta} - f_{\beta'}||_{\infty} \ge 2\delta$. So we get a lower bound of $2^{1/\delta}$ functions. The proof for the upper bound is left as an exercise.

§3 Dudley Entropy Integral

We have that

$$\mathbb{E}\left[\sup_{\theta \in \Pi} X_{\theta}\right] \le \delta\text{-localized term} + \overbrace{\mathbb{E}\left[\max_{j} \left(X_{\theta^j} - X_{\tilde{\theta}}\right)\right]}^{\triangleq \mathbb{E}[Z]}$$

where $\{\theta^1, \ldots, \theta^N$ was a δ-cover. The naive approach is to apply union bound immediately

$$\gamma(\theta^j, \tilde{\theta}) \le D$$

Theorem 3.1

(Dudley) We have that

$$\mathbb{E}[Z] \leq c \int_{\delta}^{D} \sqrt{\log N(u; \Pi, \gamma)} du \leq cD \sqrt{\log N(\delta; \Pi, \gamma)}$$

The second inequality is trivial as N decreases in u.

Proof. Will cover next class.

Example 3.2

(Sharp result for parametric entropy.) Suppose we have $\log N(\delta) \approx \alpha \log(1/\delta)$. Then, for $D = 1$, we get

$$\text{Rademacher complexity} \leq \delta + \frac{c}{\sqrt{n}} \int_{\delta}^{1} \sqrt{\alpha \log(1/u)} du \lesssim c' \sqrt{\frac{d}{n}}$$

where we chose $\delta = 0$.