CUHK STAT6050: Statistical Learning Theory Spring 2026

Lecture 6: Rademacher complexity III

Pollard's bounds and Dudley's entropy integral

Lecturer: Ben Dai

"*There is Nothing More Practical Than A Good Theory.*" — Kurt Lewin

1 Entropy bounds

Now, we provide the connection between covering numbers and Rademacher complexity.

Theorem 1.1 (Pollard's bounds)**.**

$$\mathbb{E}_\rho \|\mathbf{Rad}_n(h)\|_{\mathscr{H}} \leq \inf_{\varepsilon>0} \Big(U \sqrt{\frac{2\log\big(N(\mathscr{H}, L_2(\mathbb{P}_n), \varepsilon)\big)}{n}} + \varepsilon \Big),$$

where $U = \sup_{h\in\mathscr{H}} \|h\|_{L_2}$ *and* $\|h\|_{L_2(\mathbb{P}_n)} = \big(n^{-1}\sum_{i=1}^n h^2(\mathbf{Z}_i)\big)^{1/2}$.

Theorem 1.2 (Dudley's Theorem)**.** *Let* $\sigma_n^2 = \sup_{h\in\mathscr{H}} \|h\|^2_{L_2(\mathbb{P}_n)}$. *Then*

$$\mathbb{E}_\rho \|\mathbf{Rad}_n(h)\|_{\mathscr{H}} \leq 12 \int_0^{\sigma_n} \sqrt{\frac{\log N(\mathscr{H}, L_2(\mathbb{P}_n), \varepsilon)}{n}}\, d\varepsilon.$$

Remark 1.3. Dudley's bound (Theorem 1.2) and Pollard's bound (Theorem 1.1) are not directly comparable in general. Dudley's bound integrates over all scales ε and is tighter when the entropy integral converges (e.g., VC-type classes). However, when the entropy grows too fast (e.g., $\log N \sim 1/\varepsilon^2$), Dudley's integral may diverge while Pollard's bound, by optimizing at a single scale, can still yield a finite bound.

1.1 Some examples for uniformly bounded classes

In this section, we give some examples of covering numbers and Rademacher complexity on some uniformly bounded classes, and compare Pollard's and Dudley's bounds.

Before that, we introduce the following facts which are useful to compute Dudley's bounds.

$$\int_0^1 \sqrt{\log(1/t)}dt < 1, \quad \int_0^1 \sqrt{1/t}dt = 2.$$

Note that

$$\|h - h'\|_{L_2(\mathbb{P}_n)} \leq \sup_{\mathbf{x}\in\mathscr{X}} |h(\mathbf{x}) - h'(\mathbf{x})| =: \|h - h'\|_{L_\infty}.$$

Then, according to Lemma 2.4 in Lecture 5, we have

$$N(\mathscr{H}, L_2(\mathbb{P}_n), \varepsilon) \leq N(\mathscr{H}, L_\infty, \varepsilon).$$

Lemma 1.4 (VC-type classes)**.** *Suppose $\mathscr{H}$ is uniformly bounded by U, and*

$$N(\mathscr{H}, L_2(\mathbb{P}_n), \varepsilon) \leq \left(\frac{cU}{\varepsilon}\right)^d, \quad \varepsilon > 0.$$

Then,

$$\mathbb{E}\|\mathbf{Rad}_n(h)\|_{\mathscr{H}} \leq c\sqrt{\frac{d}{n}}.$$

Example 1.5 (linear function class)**.** *Let $\mathscr{X} = [-1,1]^d$ and $\mathscr{H} = \{h(\mathbf{x}) = \boldsymbol{\theta}^\intercal \mathbf{x} : \|\boldsymbol{\theta}\|_1 \leq U\}$. Then,*

$$\mathbb{E}\|\mathbf{Rad}_n(h)\|_{\mathscr{H}} \leq c\sqrt{\frac{d}{n}}.$$

Example 1.6 (Sparse function class)**.** *Let $\mathscr{X} = [-1,1]^d$ and $\mathscr{H} = \{h(\mathbf{x}) = \boldsymbol{\theta}^\intercal \mathbf{x} : \|\boldsymbol{\theta}\|_2 \leq U, \|\boldsymbol{\theta}\|_0 \leq K\}$. Then,*

$$N(\mathscr{H}, L_2(\mathbb{P}_n), \varepsilon) \leq \binom{d}{K}\left(\frac{cU}{\varepsilon}\right)^K \leq \left(\frac{ed}{K}\right)^K \left(\frac{cU}{\varepsilon}\right)^K,$$

and

$$\mathbb{E}\|\mathbf{Rad}_n(h)\|_{\mathscr{H}} \leq c\sqrt{\frac{K\log(d)}{n}}.$$

Example 1.7 (Lipschitz functions)**.** *Suppose $\mathscr{H}$ is an L-Lipschitz function class from $\mathscr{X} = [0,1]^d$ to $[0,1]$, then*

$$N(\mathscr{H}, L_2(\mathbb{P}_n), \varepsilon) \leq c(1/\varepsilon)3^{L/\varepsilon},$$

and

$$\mathbb{E}\|\mathbf{Rad}_n(h)\|_{\mathscr{H}} \leq c(L/n)^{1/2}.$$

Example 1.8 (Non-decreasing function class)**.** *Suppose $\mathscr{H}$ is a non-decreasing function class from $\mathbb{R}$ to $[0,1]$. Then,*

$$N(\mathscr{H}, L_2(\mathbb{P}_n), \varepsilon) \leq n^{1/\varepsilon},$$

and

$$\mathbb{E}\|\mathbf{Rad}_n(h)\|_{\mathscr{H}} \leq c\left(\frac{\log(n)}{n}\right)^{1/2}.$$

Please refer to [Bartlett and Mendelson, 2002] for more examples, including decision trees, neural networks, and kernel methods.

References

[Bartlett and Mendelson, 2002] Bartlett, P. L. and Mendelson, S. (2002). Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482.