3.13 Covering number (metric entropy)

Covering number : counts the number of ε-balls needed to cover the hypothesis class.

Goal : try to have shattering coefficients for real-valued functions.

[Definition 12] metric space

- A metric space $(\mathcal{X}, \rho)$: set $\mathcal{X} \supseteq F$ and ρ is a metric.
- $\rho : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, non-negative, symmetric, satisfy the triangle inequality and evaluate to 0 iff its arguments are equal.
- If $\rho(f, f') = 0$ is possible, the we say ρ is pseudometric.

[Definition 13] ball

Let $(\mathcal{X}, \rho)$ be a metric space, Define ε-ball as

$$B_\varepsilon(f) := \{ f' \in \mathcal{X} : \rho(f, f') \le \varepsilon \}$$

[Definition 14] covering number

(i) An ε-cover of a set $F \subseteq \mathcal{X}$ w.r.t. ρ is a finite subset $C = \{f_1, \cdots, f_m\} \subseteq \mathcal{X}$ s.t. $F \subseteq \bigcup_{j=1}^{m} B_\varepsilon(f_j)$.

(ii) Define the ε-covering number of F w.r.t. ρ to be :

$$N(\varepsilon, F, \rho) := \min \{ m : \exists \{f_1, \cdots, f_m\} \subseteq \mathcal{X},\ F \subseteq \bigcup_{j=1}^{m} B_\varepsilon(f_j) \}$$

(iii) The metric entropy of F is $\log N(\varepsilon, F, \rho)$

As $\varepsilon \downarrow$, $N(\varepsilon, F, \rho) \uparrow$. What is the tradeoff ?

[Example 7] all functions

- Let $F = \mathcal{X}$ be all functions from $\mathbb{R}$ to $[0,1]$

$g \sim L_2(P_n)$, on $z_1, \cdots, z_n$.

- In order to cover F. fix any $f \in F$.

For each z_i, For a segmentation of $[0,1]$: $Y = \{2\varepsilon, 4\varepsilon, \cdots, 1\}$. For $f(z_i) \in [0,1]$, we can pick $g(z_i) \in Y$ s.t. $|f(z_i) - g(z_i)| \le \varepsilon$. $g(z)$ for $z \ne z_i$ can be chosen arbitrarily. Averaging over all z_i, we get $\rho(f,g) \le \varepsilon$. We just need to calculate the possible permutation of Y. Furthermore. $|Y| = \frac{1}{2\varepsilon}$, so

$$N(\varepsilon, F, L_2(P_n)) \le \left(\frac{1}{2\varepsilon}\right)^n$$

the metric entropy is $O(n \log(1/\varepsilon))$, which is too large. To see this, by Massart's finite lemma, $\hat{R}_n(F) \sim O(\sqrt{\frac{n\log(1/\varepsilon)}{n}}) = O(1/\varepsilon)$, not going to zero.

[Example 8] non-decreasing function

- Let $F = \{f : \mathbb{R} \to [0,1],\ f \text{ is non-decreasing}\}$
- Let $z_1, \cdots, z_n$ be n fixed points (in an increasing order)
- $Y = \{\varepsilon, 2\varepsilon, \cdots, 1\}$. Fix any function $f \in F$. For each $y \in Y$, consider z_i for which $f(z_i) \in [y - \varepsilon, y]$. Set $g(z_i) = y$ for these points. Note: g is non-decreasing across $z_1, \cdots, z_n$ and g satisfies $\rho(f,g) \le \varepsilon$.
- Count the number of possible g. Key observation: each g is non-decreasing, we can associate each level $y \in Y$ with leftmost point z_i for $g(z_i) = y$; the choice of leftmost points for each level unique defines g. Thus:

$$N(\varepsilon, F, L_2(P_n)) = O(n^{1/\varepsilon})$$

and the metric entropy is $O(\frac{1}{\varepsilon}\log n)$, better than example 7.