Lecture 19: Metric Entropy Continued

Isabella Zhu

17 April 2025

The motivation is as follows: if N is a δ-cover and we have problem

$$\mathbb{E}[\max X_{\theta^j} - X_{\tilde{\theta}}]$$

we don't want to do the naive thing and apply union bound. Instead, we can integrate

$$c\int_{\delta}^{D}\sqrt{\log N(u;\Pi,\gamma_x)}du$$

§1 Dudley

Theorem 1.1

We have

$$\mathbb{E}\left[\max_{j=1,\ldots N} X_{\theta^j} - X_{\tilde{\theta}}\right] = c\int_{\delta}^{D}\sqrt{\log N(u;\Pi,\gamma_x)}du$$

Proof. We take a tree structure thing. The base layer is a δ-cover $\epsilon_L = \frac{D}{2^L}$. Going up the layers, we get our coarsest cover, which is $U^1 = \epsilon_1 = \frac{D}{2}$ cover. For any point, you're finding a sequence of points, one per layer, that goes up the chain while keeping as close as possible?

Formally, mapping $\gamma^L = \theta^j$, $\gamma^{L-1} = \Pi^{L-1}(\gamma_L)$ so finding the closest one to γ_L. We keep going up in this manner.

By telescoping argument, we get

$$X_{\theta^j} - X_{\gamma_1} = \sum_{t=2}^{L} X_{\gamma_t} - X_{\gamma_{t-1}}$$

We get that

$$\begin{aligned}
\mathbb{E}\left[\max_{j=1,\ldots N} X_{\theta^j} - X_{\tilde{\theta}}\right] &= \sum_{t=2}^{L}\mathbb{E}\left[\max(X_{\gamma^t} - X_{\gamma^{t-1}})\right] \\
&= \sum_{t=2}^{L}\mathbb{E}\left[\max(X_{\gamma^t} - X_{\Pi^{t-1}(\gamma^t)})\right] \\
&\leq \sum_{t=2}^{L}\mathbb{E}[2\epsilon]
\end{aligned}$$

$$\leq \sum_{t=2}^{L} 2\epsilon_t \sqrt{2 \log N(\epsilon_t)}$$

which is basically a Riemann approximation of the integral.

§2 Gaussian Complexity and Comparison

Let $(X_1, \ldots X_N)$ and $(Y_1, \ldots Y_N)$ be zero-mean Gaussian random variable vectors. We care about when we can make the statement

$$\mathbb{E}[\max X_j] \leq \mathbb{E}[\max Y_j]$$

We are not assuming the structure of either process.

Theorem 2.1

(Sudakov-Fernique) If we have

$$\mathrm{var}(X_j - X_k) \leq \mathrm{var}(Y_j - Y_k)$$

for every j and k, then this implies

$$\mathbb{E}[\max X_j] \leq \mathbb{E}[\max Y_j]$$

§2.1 Gaussian Contractions

The setup is we have $\mathbb{T} \subset \mathbb{R}^n$ and on each coordinate we have some function

$$|\phi_i(t_i) - \phi_i(\tilde{t_i})| \leq L|t_i - \tilde{t_i}|$$

Notation-wise, we have

$$\phi(\mathbb{T}) = \{\phi_1(t_1), \ldots, \phi_n(t_n) | t \in \mathbb{T}\}$$

Claim 2.2 — Similar to Rademacher, we have

$$\mathbb{E}\left[\max_{t \in \mathbb{T}} \sum_{i=1}^{n} w_i \phi_i(t_i)\right] \leq L \cdot \mathbb{E}\left[\max_{t \in \mathbb{T}} \sum_{i=1}^{n} w_i t_i\right]$$

for Gaussian complexity. We can rewrite this as

$$G_n'(\phi(\mathbb{T})) \leq L \cdot G_n'(\mathbb{T})$$

(Proof) We do the comparison stuff with Sudakov-Fernique. We have

$$\begin{aligned}
\mathrm{var}(X_t - X_s) &= \sum_{i=1}^{n} (\phi_i(t_i) - \phi_i(s_i))^2 \\
&\leq L^2 \sum_{i=1}^{n} (t_i - s_i)^2 \text{ (Lipschitz)} \\
&= L^2 \cdot \mathrm{var}(Y_t - Y_s)
\end{aligned}$$

Proposition 2.3

Let $\mathbb{T}$ be any set and X_θ zero-mean Gaussians variables. We have

$$\mathbb{E}\left[\max_{\theta\in\mathbb{T}} X_\theta\right] \geq c\delta\sqrt{\log M(\delta;\mathbb{T},\gamma_x)}$$

for any $\delta > 0$.

Proof. Let $\theta^1, \ldots \theta^M$ be a δ-packing. We have that

$$\mathbb{E}\left[\max_{\theta\in\mathbb{T}} X_\theta\right] \geq \mathbb{E}\left[\max X_{\theta^j}\right]$$

We know that

$$\mathrm{var}(X_{\theta^j} - X_{\theta^k}) \geq \delta^2 = \mathrm{var}(Z_j - Z_k)$$

which we can compare to $Z_j \sim N\left(0, \frac{\delta^2}{2}\right)$, we can say

$$\mathrm{var}(X_{\theta^j} - X_{\theta^k}) \geq \delta^2 = \mathrm{var}(Z_j - Z_k) \geq \frac{c\delta}{\sqrt{2}}\sqrt{2\log M}$$