Lecture 21: Least Squares Continued

Isabella Zhu

24 April 2025

Reading: Chap 13, Wai 19.

§1 Recap

We write our regression model of the form

$$Y_i = f^*(x_i) + w_i$$

for $i = 1, \dots n$ and $\mathbb{E}[w_i|x_i] = 0$. Although the w_i are independent, they are not i.i.d.

Today we focus on the well-specified setting where

$$\hat{f} \in \operatorname*{argmin}_{f \in F} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2$$

and we analyze the empirical error

$$||\hat{f} - f^*||_n^2 = \frac{1}{n} \sum_{i=1}^{n} (\hat{f}(x_i) - f^*(x_i))^2$$

Example 1.1

Say $F = \{\text{L-Lipschitz functions on } [0,1]\}$. We are solving the minimization problem

$$\min_{\tilde{y}_i} \frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_i - y_i)^2$$

satisfying the constraint $|\tilde{y}_i - \tilde{y}_j| \le L|x_i - x_j|$. However, note that there are $\binom{n}{2}$ constraints. The more practical way to do this is using a 2-layer neural net

$$f(x) = \sum_{t=1}^{T} w_t \varphi(a_t x + b_t)$$

Example 1.2

(Shape constraints). Examples are if we are constraining our function to have some sort of property, such as monotonicity, convexity, etc.

For convexity, for example,

$$\min_{\tilde{y}_i} \frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2$$

we know that tangent line is a lower bound. The constraints here are

$$f(x_2) \geq f(x_1) + g_1(x_2 - x_1)$$

In general, g_1 is called a subgradient. We are optimizing such that

$$\tilde{y}_i \geq \tilde{y}_j + \tilde{g}_j(x_i - x_j)$$

Note that we are optimizing over $\tilde{y}_i$ and $\tilde{g}_j$.

§2 General Bound with Localization

We will state a result for when $w_i \sim N(0, \sigma^2)$. Consider

$$F^* = \{f - f^* | f \in F\}$$

The **localized Gaussian complexity** is

$$G_n(\delta) = \mathbb{E}\left[\sup_{||f-f^*||_n \leq \delta, f \in F} \frac{1}{n}\left|\sum_{i=1}^{n} w_i(f(x_i) - f^*(x_i))\right|\right]$$

Fact 2.1. Under some regularity conditions, we have that $\delta \to \frac{G_n(\delta)}{\delta}$ is non-increasing on $(0, \infty)$. The **critical radius** δ_n solves

$$\frac{G_n(\delta)}{\delta} = \frac{\delta}{2\sigma}$$

Theorem 2.2

Under given conditions, we have

$$\mathbb{P}(||\hat{f} - f^*||_n \geq c_0 t\delta_n) \leq e^{-nt^2/2}, \ \ \forall t \geq 1$$

In other words, we have

$$\mathbb{E}||\hat{f} - f^*||_n \leq c_1\left\{\delta_n^2 + \frac{\sigma^2}{n}\right\}$$

§2.1 Intuition

So what is the intuition for this critical radius situation? Recall that non-parametric least-squares satisfies a basic inequality

$$\frac{1}{2}||\hat{f} - f^*||_n^2 \leq \frac{\sigma}{n}\sum_{i=1}^{n} w_i(\hat{f}(x_i) - f^*(x_i))$$

Let's call $\hat{\delta}^2 \triangleq ||\hat{f} - f^*||_n^2$. By definition, we can say

$$\begin{aligned}\frac{1}{2}||\hat{f} - f^*||_n^2 &\leq \frac{\sigma}{n}\sum_{i=1}^{n} w_i(\hat{f}(x_i) - f^*(x_i)) \\ &\leq \sigma\left\{\frac{1}{n}\sup_{||f-f^*||_n \leq \hat{\delta}, f\in F}\sum_{i=1}^{n} w_i(f(x_i) - f^*(x_i))\right\}\end{aligned}$$

Now we do some handwaving, which is that on expectation stuff concentrates near expectation, which is true if δ is fixed but note $\hat{\delta}$ is not fixed. So, we get approximately

$$\frac{\hat{\delta}^2}{2} \lesssim \sigma G_n(\hat{\delta})$$

Corollary 2.3

We can obtain upper bounds on δ_n via Dudley.

$$\frac{c}{\sqrt{n}}\int_{\delta^2/4\sigma}^{\delta}\sqrt{\log N(t; F\cap B_\delta(f^*), ||.||_n)}dt \leq \frac{\delta}{2\sigma}$$

§3 Examples and Consequences

Example 3.1

For Lipschitz functions, we have $\log N(t; ||.||_n) \leq \log N(t; ||.||_\infty) \approx 1/\delta$. Taking square root and integrating, we get

$$\frac{1}{\sqrt{n}}\int_0^{\delta}\frac{1}{\sqrt{t}}dt \approx \sqrt{\delta/n}$$

which is the LHS while the RHS is $\frac{\delta}{2\sigma}$. Setting them equal to each other, we get

$$\frac{\sqrt{\delta}}{\sqrt{n}} = G_(\delta) \approx \frac{\delta^2}{\sigma}$$

Thus we get

$$\delta_n^2 = \left(\frac{\sigma^2}{n}\right)^{2/3}$$

Example 3.2

For convex functions, we have $\log N(t; F_{convex}, ||.||_n) \approx \frac{1}{\sqrt{\delta}}$. So,

$$\frac{1}{\sqrt{n}} \int_0^\delta \frac{1}{t^{1/4}} dt \approx \frac{\delta^{3/4}}{\sqrt{n}}$$

Then we solve

$$G_n(\delta) = \frac{\delta^{3/4}}{\sqrt{n}} \approx \frac{\delta^2}{\sigma} \implies \delta_n^2 = \left(\frac{\sigma^2}{n}\right)^{4/5}$$