Lecture 22: November 16, 2016

Lecturer: Alessandro Rinaldo *Scribes: Neil Spencer*

Consider the problem of constrained least squares (possible nonparametric). Here, we wish to bound $||\hat{f} - f^*||_n^2$ with high probability. Typically, the driving term for such a bound is the critical inequality

$$\frac{G_n(\delta, \mathcal{F})}{\delta} \leq \frac{\delta}{2\sigma}$$

It is often sharp for $||\hat{f} - f^*||_n^2$.

22.1 Example 1

For example, consider $\mathcal{F} = \{f_\theta, \theta \in B_q(R_q)\}$ where $f_\theta =< x, \theta >$ with $x, \theta \in \mathbb{R}^d$ and

$$B_q(R_q) = \left\{ \theta \in \mathbb{R}^d : \sum_{i=1}^{d} |\theta_i|^q \leq Rq \right\}$$

with $q \in (0,1)$. Here, $\theta \in B_q(R_q)$ implies that the coordinates of θ decay to 0 quickly (largest ones much greater than the others).

Consider

$$\hat{f} \in \text{argmin}_{f \in \mathcal{F}} \frac{1}{2n} \sum_{i=1}^{n} (y_i - f(x_i))^2 \text{ for } x_1, \ldots, x_n \text{ fixed.}$$

Then

$$\begin{aligned} ||\hat{f} - f^*||_n^2 &= \frac{||x(\hat{\theta} - \theta^*)||^2}{n} \\ &\leq R_q \left(\frac{\sigma^2 \log(d)}{n} \right)^{1-q/2} \end{aligned}$$

where X is $n \times d$ with x_i as its ith row. This rate is minimax optimal. Here, the function class $\mathcal{F}$ is not convex and is not star-shaped. We instead work with the class $d\mathcal{F} = \mathcal{F} - \mathcal{F} = \{f - g, f, g \in \mathcal{F}\}$ which is contained in $\mathcal{F}(2R_q) = \{f_\theta, \sum |\theta_i|^q \leq 2R_q\}$ and use the fact that

$$\log(N(\mu, \mathcal{F}, ||\cdot||_{\text{euclid}})) \leq C_q R_q^{2/q} \left(\frac{1}{\mu}\right)^{2q/(2-q)} \log(d)$$

The same bound applies to the log covering number of $B_n(\delta, \mathcal{F}(2R_q))$ in $||\cdot||_n$ with $f \in \mathcal{F}(2R_q) : ||f||_n \leq \delta$. So we need to evaluate

$$\frac{1}{\sqrt{n}} \int_0^\delta \sqrt{\log N(\mu, B_n(\delta, \mathcal{F}(2R_q)))} d\mu \leq R_q^{1/(2-q)} \sqrt{\frac{\log(d)}{n}} \delta^{1-q/(2-q)}$$

This is an upper bound on $G_n(\delta, \mathcal{F})$ (the local Gaussian complexity). Simply set it equal to δ^2 and solve for it. Sometimes upper bound the local entropy with that of the entire space.

22.2 Example 2

Consider nonparametric regression.

$$\mathcal{F} = \{f : [0,1] \to \mathbb{R}, |f(x) - f(y)| \leq L|x - y| \text{ for all } x, y[0,1]\}$$

and

$$y_i = f^*(x_i) + \delta w_i \text{ with } w_1, \ldots, w_n \sim N(0,1).$$

For this problem, consider the enlarged class $\mathcal{F}(2L) \supset \mathcal{F}(L) - \mathcal{F}(L)$. Use the fact that the metric entropy at scale μ is proportional to L/μ with respect to the L_∞ norm. So,

$$\frac{1}{\sqrt{n}} \int_0^\delta \sqrt{\log(N(\mu, B_n(\delta, \mathcal{F}(2L))))} d\mu$$

$$\propto \int_0^\delta \left(\frac{L}{\mu}\right)^{1/2} d\mu = C' \sqrt{\frac{L\delta}{n}}$$

From the equation $\sqrt{L\delta_n/n} = \delta^2/\sigma$ which implies $\delta_n^2 = \left(L\sigma^2/n\right)^{2/3}$. Thus,

$$||\hat{f} - f^*||_n \propto\leq \left(\frac{L\sigma^2}{n}\right)^{2/3} \text{ with high probability}$$

22.3 Oracle Inequality for Constrained Least Squares

So far we have assumed that $f^* \in \mathcal{F}$. If $f^* \notin \mathcal{F}$, then our benchmark becomes

$$\inf_{f \in \mathcal{F}} ||f - f^*||_n^2$$

This the error that an oracle who knows f^* would make.

Theorem 22.1 *Assume that $d\mathcal{F} = \mathcal{F} - \mathcal{F}$ is star-shaped. Let δ_n be any positive solution to*

$$\frac{G_n(\delta, d\mathcal{F})}{\delta} \leq \frac{\delta}{2\sigma}.$$

Then there exists $c_0, c_1, c_2 > 0$ such that for all $t \geq \delta_n$,

$$||\hat{f} - f^*||_n^2 \leq \inf_{\gamma \in (0,1)} \left\{ \frac{1+\gamma}{1-\gamma} ||f - f^*||_n^2 + \frac{c_0 t \delta_n}{\gamma} \right\}$$

with probability greater than or equal to $1 - c_1 \exp\left(-c_2 n t \delta_n / \sigma^2\right)$.

Remarks:

1. We can rephrase it as

$$||\hat{f} - f^*||_n^2 \leq \frac{1+\gamma}{1-\gamma} \inf_{f \in \mathcal{F}} ||f - f^*||_n^2 + \frac{c_0}{\gamma} + \delta_n$$

2. If $f^* \in \mathcal{F}$ then $||\hat{f} - f^*||_n^2 \leq\propto \delta_n^2$.
3. Setting $t = \delta_n$ with $f^* \notin \mathcal{F}$, we get

$$||\hat{f} - f^*||_n^2 \leq\propto \inf_{f\in\mathcal{F}} ||f - f^*||_n^2 + \delta_n^2$$

The third point is like bias-variance tradeoff.

Proof: Let f be an arbitrary function in in $\mathcal{F}$. Then, from

$$\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2 \leq \frac{1}{2n}\sum_{i=1}^{n}(y_i - f(x_i))^2$$

we get

$$\frac{1}{2}||\hat{\Delta}||_n^2 \leq \frac{1}{2}||f - f^*||_n^2 + \sigma/n|\sum_{i=1}^{N} w_i\Delta_i|$$

where $\hat{\Delta} = \hat{f} - f^*$ and $\Delta = f - f^*$. We focus on the second term. Consider two cases:

1. $||\Delta||_n \leq \sqrt{t\delta_n}$. Then,

$$\begin{aligned}
||\Delta||_n^2 &= ||\hat{f} - f + f - f^*||^2 \\
&\leq (||\Delta||_n + ||f - f^*||_n)^2 \\
&\leq \left(\sqrt{t\delta_n} + ||f - f^*||_n\right)^2 \\
&= t\delta_n + ||f - f^*||_n^2 + 2\sqrt{t\delta_n}||f - f^*||_n \\
&\text{Using the Young Fenchel inequality } xy \leq x^2/2\alpha + \alpha y^2/2 \; x, y \in \mathbb{R}.\alpha > 0 \\
&\leq t\delta_n + ||f - f^*||_n^2 + t\delta_n/\alpha + ||f - f^*||_n^2\alpha \\
&\leq t\delta_n(1 + \alpha^{-1}) + ||f - f^*||_n^2(1 + 2\alpha) \\
&\text{Set } \gamma = \alpha/(\alpha + 1) \\
&= \frac{t\delta_n}{\gamma} + ||f - f^*||_n^2\frac{1+\gamma}{1-\gamma}
\end{aligned}$$

2. Otherwise, we can assume $||\Delta||_2 > \sqrt{t\delta_n}$. Here, $\Delta \in \mathcal{F}$, so applying Lemma 13.2 with $u = \sqrt{t\delta_n}$ yields

$$\mathbb{P}(2|\frac{\sigma}{n}\sum_{i=1}^{n} w_i\Delta(x_i)| \geq (1 + 4\beta)\,||f - f^*||_n^2 + \frac{4}{\beta}t\delta_n$$

due to the Young-Fenchel inequality with any $\beta > 0$. Setting $\beta = \gamma/(2(1-\gamma))$ followed by some algebra completes the proof.

■