Math 547: Mathematical Foundations of Statistical Learning Theory Fall 2017

Lecture 33 — November 6

Instructor: Stas Minsker *Scribe: Mose Wintner*

33.1 Continued

Let $T_\varepsilon = \{u \in T : ||\mathbb{X}u||_1 / n \le \varepsilon\}$. Then

$$\mathbb{E} \sup_{u \in T_\varepsilon} ||u||_2 \le \sqrt{\frac{8\pi}{n}} \mathbb{E} \sup_{u \in T} |\langle g, u \rangle| + \sqrt{\frac{\pi}{2}} \varepsilon.$$

Take $u_0 \in T$.

$$\mathbb{E} \sup_{u \in T} |\langle g, u \rangle| = \mathbb{E} \sup_{u \in T} |\langle g, u - u_0 \rangle| + \mathbb{E}|\langle g, u_0 \rangle| \le \mathbb{E} \sup_{u, u_0 \in T} \langle g, u - u_0 \rangle + \mathbb{E}^{1/2} \langle g, u_0 \rangle^2$$

to drop the absolute value bars.

The RHS is

$$\mathbb{E}\left(\sup_u \langle g, u \rangle + \sup_{u_0} -\langle g, u_0 \rangle \right) + ||u_0||^2 = 2\mathbb{E} \sup_{u \in T} \langle g, u \rangle + ||u_0||^2.$$

33.2 Estimation from noisy observations

$Y = \mathbb{X}\lambda + v$ s.t. $||v||_1 / n \le \varepsilon$. Note that

$$\frac{1}{n} ||v||_2 \le \frac{1}{n} \sqrt{\sum |v_j|^2} \le \varepsilon.$$

Let $\hat{\lambda} \in \mathbb{R}^p$ satisfy $\hat{\lambda} \in K$, $\frac{1}{n} \left|\left| \mathbb{X}\hat{\lambda} - Y \right|\right|_1 \le \varepsilon$.

Theorem 1. For any such $\hat{\lambda}$,

$$\mathbb{E} \sup_{\lambda \in K} \left|\left| \hat{\lambda} - \lambda \right|\right|_2 \le \sqrt{8\pi} \left(\frac{w(K)}{\sqrt{n}} + \varepsilon' \right).$$

Proof: Let $T = K - K$, $\varepsilon \mapsto 2\varepsilon'$. Then we have

$$\mathbb{E} \sup_{u \in T_{2\varepsilon'}} ||u||_2 \le \sqrt{8\pi} \left(\frac{1}{\sqrt{n}} \mathbb{E} \sup_{u \in K - K} |\langle g, u \rangle| \right) + \sqrt{\frac{\pi}{2}} 2\varepsilon'.$$

So $\mathbb{E} \sup_{u \in K-K} |\langle g, u \rangle| = 2\mathbb{E} \sup_{u \in K} \langle g, u \rangle$ since $K - K$ is symmetric. Claim: $\hat{\lambda} - \lambda' \in T_{2\varepsilon'}$ for any $\lambda' \in K$. We have

$$\begin{aligned} \frac{1}{n} \left|\left| \mathbb{X}(\hat{\lambda} - \lambda') \right|\right|_1 &= \frac{1}{n} \left|\left| \mathbb{X}\hat{\lambda} - Y + v \right|\right| \\ &\le \frac{1}{n} \left|\left| X\hat{\lambda} - Y \right|\right|_1 + \frac{1}{n} ||v||_1 \le 2\varepsilon' \end{aligned}$$

33.3 Computational Considerations

Assume the set K is star-shaped, meaning that $tK \subset K$ for $t \in [0,1]$.

Definition 1. The gauge (Minkowski functional) of the set K is

$$||x||_K = \inf\{t > 0 : x/t \in K\}.$$

Uniformly bounded by 1.

Show that if K is convex and symmetric, this defines a norm with K as its unit ball.

As before, assume that $Y = \mathbb{X}\lambda + v$. Let $\hat{\lambda}$ be the minimizer of $||\lambda||_K$ subject to $\frac{1}{n}||\mathbb{X}\lambda - Y||_1 \le \varepsilon$.

Then $\hat{\lambda}$ satisfies the bound

$$\mathbb{E}\sup_{\lambda\in K}\left|\left|\hat{\lambda}-\lambda\right|\right|_2 \le \sqrt{8\pi}\left(\frac{w(K)}{\sqrt{n}}+\varepsilon\right).$$

By the previous theorem, it is enough to show $\hat{\lambda} \in K$, but $\left|\left|\hat{\lambda}\right|\right|_K \le ||\lambda||_K \le 1$ since $\lambda \in K$, so $\hat{\lambda}_K \in K$.

We can make this problem convex. We need a convex objective function to do this. The smallest convex set that contains K is its convex hull. The modified problem is then to minimize $||\lambda'||_{CH(K)}$ subject to $||\mathbb{X}\lambda' - Y||_1/n \le \varepsilon$. To do this take any $\hat{\lambda} \in CH(K)$ such that $\left|\left|\mathbb{X}\hat{\lambda} - Y\right|\right|_1/n \le \varepsilon$. It follows from the previous results that

$$\mathbb{E}\sup_{\lambda\in K}\left|\left|\hat{\lambda}-\lambda\right|\right|_2 \le \mathbb{E}\sup_{\lambda\in CH(K)}\left|\left|\hat{\lambda}-\lambda\right|\right|_2 \le \sqrt{8\pi}\left(\frac{w(CH(K))}{\sqrt{n}}+\varepsilon\right).$$

But $w(CH(K)) = w(K)$.

Assume that K is the set of all s-sparse vectors, meanign that all but s coordinates are 0. Then $w^2(K) \le Cs\log(p/s)$ for some constant C.