STA 3000 Lecture 6.

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)] \right|.$$

$$\left\{ \begin{array}{l} \text{ULLN} \\ \text{Rate of convergence} \\ \text{UCLT.} \end{array} \right\} \longrightarrow \boxed{\begin{array}{l} \text{Consistency} \\ \text{rate} \\ \text{asymptotics} \end{array}}$$

- $\mathcal{F}$ is finite, ULLN
- $\mathcal{F}$ infinite, discretization.

Thm (Wald). Suppose $\theta \in K \subseteq \mathbb{R}^d$, K compact.

(i) $\mathbb{E}\left[\sup_{\theta \in K} |f(\theta; X)|\right] < +\infty$.

(ii) $f(\cdot, x_i)$ is continuous.

then

$$\sup_{\theta \in K} \left| \frac{1}{n}\sum_{i=1}^{n} f(\theta; X_i) - \mathbb{E}[f(\theta; X)] \right| \xrightarrow{P} 0$$

$$F(\theta) := \mathbb{E}[f(\theta; X)]$$

Proof. F is continuous (DCT), so uniformly cts.

Fix $\varepsilon > 0$, $\exists \delta$,

$$\text{s.t.} \sup_{\|\theta - \theta'\|_2 \le \delta} |F(\theta) - F(\theta')| \le \varepsilon.$$

We let $\{\theta_j\}_{j=1}^N$ be a δ-cover of K.

$$\left(F_n(\theta) := \frac{1}{n} \sum_1^n f(\theta; X_i) \right)$$

$$\sup_{\theta \in K} |F_n(\theta) - \cancel{F(\theta)}|$$

$$\le \max_{i \in [N]} \sup_{\|\theta - \theta_i\| \le \delta} \left\{ |F_n(\theta) - F_n(\theta_i)| + \underbrace{|F(\theta) - F(\theta_i)|}_{\le \varepsilon} + |F_n(\theta_i) - F(\theta_i)| \right\}$$

(2) (1)

$$\max_{i \in [N]} |F_n(\theta_i) - F(\theta_i)| \xrightarrow{P} 0$$

$$\text{First term} = \max_{i \in [N]} \sup_{\|\theta - \theta_i\|_2 \le \delta} \left| \frac{1}{n} \sum_{j=1}^n \left(f(\theta; X_j) - f(\theta_i; X_j) \right) \right|$$

$$\le \max_{i \in [N]} \frac{1}{n} \sum_{j=1}^n \sup_{\|\theta - \theta_i\| \le \delta} |f(\theta; X_j) - f(\theta_i; X_j)|.$$

(weak) LLN: $\xrightarrow{P} \max_{i \in [N]} \mathbb{E}\left[\sup_{\|\theta - \theta_i\| \le \delta} |f(\theta; X) - f(\theta_i; X)| \right]$.

Suffices to show.

$$(*) \quad \sup_{\theta \in K} \mathbb{E}\left[\sup_{\|\theta' - \theta\| \le \delta} |f(\theta; X) - f(\theta'; X)| \right] \longrightarrow 0 \quad (\text{as } \delta \to 0).$$

Proof of $(*)$:

$$M_\varepsilon(\theta) = \sup_{\|\theta' - \theta\| \le \varepsilon} |f(\theta'; X) - f(\theta; X)|.$$

Pointwise : $\underline{\lim_{\varepsilon \to 0^+} M_\varepsilon(\theta)} = 0.$

$$|M_\varepsilon(\theta)| \le \underbrace{2 \cdot \sup_{\theta' \in K} |f(\theta'; X)|}_{\in L^1}$$

DCT. $\mathbb{E}[M_\varepsilon(\theta)] \to 0 \quad (\varepsilon \to 0^+).$

Convergence is uniform in $\theta \in K$: Dini's thm

Bounds on ①

Naive approach

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n f(\theta; X_i) - F(\theta) \right| \ge \varepsilon \right) \le \cdots R_n(\varepsilon)$$

$$\mathbb{P}\left(\sup_{i \in [N]} |\cdots\cdots| \ge \varepsilon \right) \le N \cdot R_n(\varepsilon).$$

Suppose $R_n(\varepsilon) = \exp(-n\varepsilon^2)$

we have $\mathbb{P}\left(\sup_{i\in[N]} \left| - \quad \right| > \varepsilon\right) \leq N \cdot \exp(-n\varepsilon^2)$

Need $\varepsilon \asymp \sqrt{\frac{\log N}{n}}$.

Suppose only have second moment,

$$R_n(\varepsilon) = \frac{1}{n\varepsilon^2}.$$

Union bound $\Longrightarrow$ $\varepsilon \asymp \sqrt{\frac{N}{n}}$.

Bound on ②.

One step discretization is loose.

For ①, we use "~~symmetrization~~"

For ②, we use "chaining".

Notation: $P_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$

$$Pf = \mathbb{E}[f(X)].$$

Want to bound $\sup_{f\in\mathcal{F}} |P_n f - Pf|$

Thm (symmetrization).

$$R_n(\mathcal{F}) := \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(X_i)\right|\right]$$

where $\varepsilon_i \overset{iid}{\sim}$ Rademacher $\begin{pmatrix} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{pmatrix}$.

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |P_n f - Pf|\right] \le 2\cdot R_n(\mathcal{F}).$$

"Rademacher complexity".

Strategy: conditioning on $(X_i)_{i=1}^{n}$

bound $\mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{1}^{n} \varepsilon_i f(X_i)\right| \,\middle|\, (X_i)_{i=1}^{n}\right]$

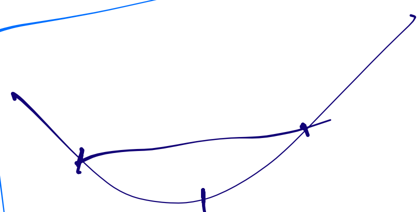
using union bound and discretization

Proof: $X_1' \cdots X_n' \overset{iid}{\sim} \mathbb{P}$

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |P_n f - Pf|\right]$$

$$= \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i}^{n} f(X_i) - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[f(X_i')]\right|\right].$$

(Jensen) $$\le \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \frac{1}{n}\sum_{1}^{n} f(X_i')\right|\right].$$



$$f(X_i) - f(X_i') \overset{d}{=} \varepsilon_i\left(f(X_i) - f(X_i')\right)$$

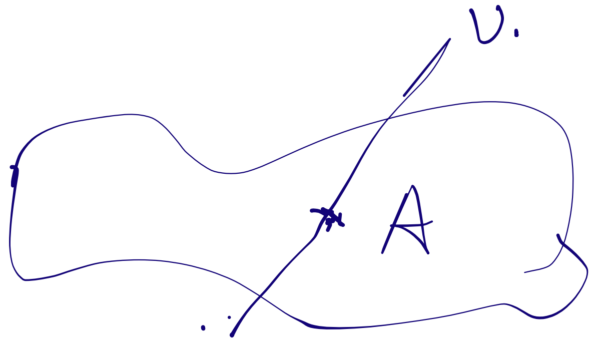
$$= \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i\left(f(X_i) - f(X_i')\right)\right|\right]$$

$$\le 2\cdot \mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(X_i)\right].$$

$A \subseteq \mathbb{R}^n.$

Interested in

$$\cancel{R_n(A)} := \mathbb{E}\left[\sup_{a\in A} \left|\frac{1}{n}\cancel{\sum_{i=1}^{n} \varepsilon_i a_i}\right|\right]$$



$$\sup_{a\in A} |a^T v|.$$

width along random direction.

$$G_n(A) = \mathbb{E}\left[\sup_{a\in A} \frac{1}{n}\sum_{i=1}^{n} g_i a_i\right]$$

where $g_i \overset{iid}{\sim} N(0,1)$.

(Qualitively equivalent).

Bounding Rademacher complexity.

- Finite A.

$(\forall \lambda > 0)$.

$$R_n(A) \le \frac{1}{\lambda}\log \mathbb{E}\left[\exp\left(\max_{a\in A} \frac{\lambda}{n}\sum_{1}^{n} \varepsilon_i a_i\right)\right].$$

$$\le \frac{1}{\lambda}\log\left(\sum_{a\in A}\prod_{i=1}^{n} \cdot\, \mathbb{E}\left[\exp\left(\frac{\lambda}{n}\varepsilon_i a_i\right)\right]\right).$$

$$\le \frac{1}{\lambda}\log\left(\sum_{a\in A}\exp\left(\frac{\lambda^2}{2n^2}\|a\|_2^2\right)\right).$$

$$\leq \frac{1}{\lambda} \log \left[|A| \cdot \exp\left(\frac{\lambda^2}{2n^2} \max_{a \in A} \|a\|_2^2 \right) \right].$$

$$\leq \frac{\log |A|}{\lambda} + \frac{\lambda}{2n^2} \max_{a \in A} \|a\|_2^2 .$$

(Choose ~~optimal~~ λ)

$$= \max_{a \in A} \frac{\|a\|_2}{\sqrt{n}} \cdot \sqrt{\frac{\log |A|}{2n}} .$$

One-step discretisation

$a^{(1)}, a^{(2)}, \dots, a^{(M)}$ δ-cover of A

under $\|\cdot\|_n$.

$$\boxed{\|a\|_n := \sqrt{\sum_{i=1}^{n} a_i^2 / n}}$$

$$\pi(a) = \arg\min_{a^{(j)}} \|a - a^{(j)}\|_n .$$

$$R_n(A) \leq \mathbb{E}\left[\sup_{a \in A} \left| \frac{1}{n} \varepsilon^T (a - \pi(a)) \right| \right] + \mathbb{E}\left[\max_{j \in [M]} \frac{1}{n} \varepsilon^T a^{(j)} \right].$$

$$\leq \max_{a \in A} \|a\|_n \sqrt{\frac{\log M(\delta)}{n}}$$

First term $\leq \mathbb{E}\left[\sup_{a \in A} \frac{1}{n} \cdot \|\varepsilon\|_2 \cdot \|a - \pi(a)\|_2\right]$

$$= \sup_{a \in A} \frac{1}{\sqrt{n}} \cdot \|a - \pi(a)\|_2 \leq \delta.$$

$$R_n(A) \leq \delta + \max_{a \in A} \|a\|_n \cdot \sqrt{\frac{\log M(\delta)}{2n}}.$$

e.g. for parametric models, $M(\delta) \asymp \left(\frac{1}{\delta}\right)^d$

$$R_n(A) \lesssim \sqrt{\frac{d \cdot \log n}{n}}.$$

eg (we'll see later) nonparametric

large gap.

Thm (chaining). $\exists$ universal constant $C > 0$

$$R_n(A) \leq \frac{C}{\sqrt{n}} \int_0^{+\infty} \sqrt{\log N(\delta; A, \|\cdot\|_n)}\, d\delta$$

Proof - $D = \max_{a \in A} \|a\|_n$.

A_m: minimal $D/2^m$ - covering of A.

$$|A_m| = N\left(\frac{D}{2^m}; A, \|\cdot\|_n\right).$$

$A_0 = \{0\}$

$\pi_m(a)$ is best approximation to a in A_m.

$$\|a - \pi_m(a)\|_n \leq \frac{D}{2^m}.$$

$$\frac{1}{n}\varepsilon^T a = \sum_{m=0}^{+\infty} \frac{1}{n}\varepsilon^T\left(\pi_{m+1}(a) - \pi_m(a)\right).$$

$$\mathbb{E}\left[\sup_{a\in A} \frac{1}{n}\varepsilon^T a\right] \leq \sum_{m=0}^{+\infty} \mathbb{E}\left[\sup_{a\in A} \frac{1}{n}\varepsilon^T\left(\pi_{m+1}(a) - \pi_m(a)\right)\right]$$

Bounding each term:

$$\mathbb{E}\left[\sup_{a\in A} \frac{1}{n}\varepsilon^T\left(\pi_{m+1}(a) - \pi_m(a)\right)\right] \leq \frac{3D}{2^{m+1}} \cdot \sqrt{\frac{\log(|A_m|\cdot|A_{m+1}|)}{2n}}.$$

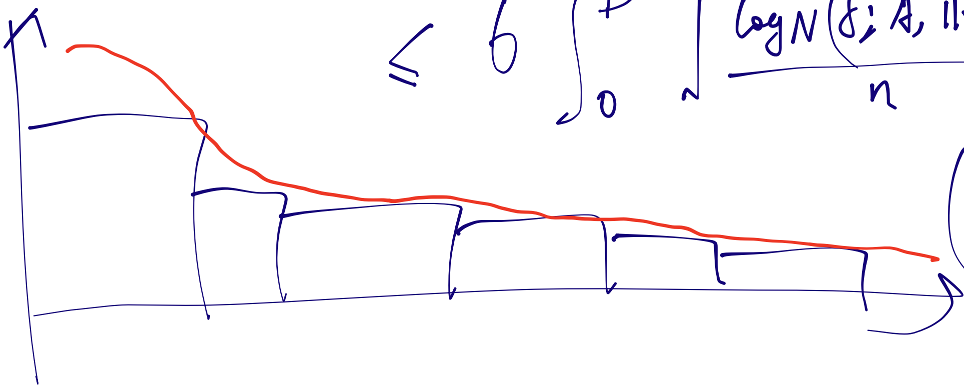
$$\leq \frac{3D}{2^{m+1}} \cdot \sqrt{\frac{\log|A_{m+1}|}{n}}.$$

Cardinality $\leq |A_m| \cdot |A_{m+1}|$

Diameter $\leq \frac{3}{2^{m+1}} \cdot D$.

$$\mathbb{E}\left[\sup_{a\in A} \frac{1}{n}\varepsilon^T a\right] \leq 3 \cdot \sum_{m=1}^{+\infty} \frac{D}{2^m} \sqrt{\frac{\log|A_m|}{n}}.$$

$$\leq 6 \cdot \sum_{m=0}^{+\infty} \int_{\frac{D}{2^{m+1}}}^{\frac{D}{2^m}} \sqrt{\frac{\log N(\delta; A, \|\cdot\|_n)}{n}}\, d\delta.$$



$$\leq 6 \int_0^D \sqrt{\frac{\log N(\delta; A, \|\cdot\|_n)}{n}}\, d\delta.$$

(Dudley).

Remarks.

1 – May diverge.

Fix: discretization to some level (will get back later)

2. "Generic chaining"

Thm. Assume $|f(x)| \leq G(x)$ (for any $f \in \mathcal{F}$)

Then.

$$\mathbb{E}\left[\sup |P_n f - Pf|\right]$$

$$\leq C \cdot \sqrt{\frac{\mathbb{E}[G(X)^2]}{n}} \int_0^1 \sqrt{\log \sup_Q N(\delta \cdot \|G\|_{L^2(Q)}; \mathcal{F}, L^2(Q))}\, d\delta.$$

Proof. $\mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E} f(X) \,\middle|\, (X_i)_{i=1}^n\right]$

$$\leq 2 \cdot \mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i) \,\middle|\, (X_i)_{i=1}^n\right].$$

$$\leq \frac{12}{\sqrt{n}}\mathbb{E}\left[\int_0^{\sqrt{P_n G^2}} \sqrt{\log N(\delta; A, \|\cdot\|_n)}\, d\delta\right].$$

$$A = \{ (f(x_1), f(x_2) \cdots, f(x_n)) : f \in \mathcal{F} \} \subseteq \mathbb{R}^n.$$

$$a \in A, \quad \|a\|_n = \sqrt{\frac{1}{n}\sum_1^n f(x_i)^2} \leq \sqrt{\frac{1}{n}\sum_1^n G(x_i)^2}.$$

$$\leq \frac{12}{\sqrt{n}} \cdot \mathbb{E}\left[\sqrt{P_n G^2} \cdot \int_0^1 \sqrt{\log N(\delta \cdot \sqrt{P_n G^2}; \mathcal{F}, L^2(P_n))}\, d\delta\right].$$

$$\leq \frac{12}{\sqrt{n}} \cdot \underbrace{\mathbb{E}\left[\sqrt{P_n G^2}\right]}_{\leq \sqrt{\mathbb{E} G^2}} \cdot \int_0^1 \log \sup_Q N(\delta \cdot \|G\|_{L^2(Q)}; \mathcal{F}, L^2(Q))\, d\delta.$$

Remark: If replacing covering/packing w/ bracketing. can get rid of $\sup_Q$, just $L^2(P)$.

A bracket $[l_i, u_i]$ s.t. $l_i(x) \leq u_i(x)$ $\forall x$

$f \in [l_i, u_i]$ if $l_i \leq f \leq u_i$

$\{[l_i, u_i] : i \in [N]\}$ cover $\mathcal{F}$ if $\forall f \in \mathcal{F}$ $\exists i, f \in [l_i, u_i]$.

ε-bracket $\|u_i - l_i\|_{L^2(P)} \leq \varepsilon$.

Minimal bracket cover $N_{[\,]}(\varepsilon; \mathcal{F}, \|\cdot\|_{L^2(P)})$.

We can show

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf|\right] \leq \frac{C}{\sqrt{n}} \cdot \sqrt{P G^2} \cdot \int_0^1 \sqrt{\log N_{[\,]}(\delta \cdot \|G\|_{L^2(P)}; \mathcal{F}, \|\cdot\|_{L^2(P)})}\, d\delta.$$

(see van der Vaart & Wellner book).

---

e.g. $\Theta \subseteq \mathbb{B}(0, R)$. $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$.

$$|f_{\theta_1}(x) - f_{\theta_2}(x)| \leq M(x) \cdot \|\theta_1 - \theta_2\|_2.$$

Take $G(x) = |f_{\theta_0}(x)| + M(x) \cdot R$.

$$N(\delta \cdot \|G\|_{L^2(Q)}; \mathcal{F}, \|\cdot\|_{L^2(Q)}) \leq N(\delta \cdot R \cdot \|M\|_{L^2(Q)}; \mathcal{F}, \|\cdot\|_{L^2(Q)}).$$

Construction: take $\theta_1, \theta_2, \dots, \theta_N$ as min-ε-covering of Θ.

$\forall \theta \in \Theta$, $\exists j$ s.t. $\|\theta - \theta_j\|_2 \leq \varepsilon$.

$$\|f_\theta - f_{\theta_j}\|^2_{L^2(Q)}$$

$$= \int \left|f_\theta(x) - f_{\theta_j}(x)\right|^2 dQ(x)$$

$$\leq \int M(x)^2 \cdot \|\theta - \theta_j\|_2^2 \, dQ(x) = \|M\|^2_{L^2(Q)} \cdot \|\theta_j - \theta\|_2^2 .$$

$$\|f_\theta - f_{\theta_j}\|_{L^2(Q)} \leq \varepsilon \cdot \|M\|_{L^2(Q)} \leq \delta \cdot R \cdot \|M\|_{L^2(Q)}.$$

Need to choose $\varepsilon = \delta R$

So $N(\delta \cdot R \cdot \|M\|_{L^2(Q)}; \mathcal{F}, \|\cdot\|_{L^2(Q)})$

$$\leq N(\delta \cdot R; \mathbb{B}(0, R), \|\cdot\|_2).$$

$$\leq \left(1 + \frac{2}{\delta}\right)^d .$$

$$\int_0^1 \sqrt{\log \sup_Q N(\delta \cdots)} \, d\delta$$

$$\leq \int_0^1 \sqrt{d \log\left(1 + \frac{2}{\delta}\right)} \, d\delta$$

$$\leq c \cdot \sqrt{d}$$

$$\mathbb{E} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} f_\theta(X_i) - \mathbb{E}\left[f_\theta(X)\right] \right|$$

$$\leq C \cdot \sqrt{\frac{\mathbb{E}\left[f_{\theta_0}^2(X)\right] + \mathbb{E}\left[M(X)^2\right] \cdot R^2}{n}} \cdot \sqrt{d}.$$