Lecture 10: Resilience Beyond Mean Estimation

Logistics:

Pset 2: due next Tuesday
↳ covers through Lecture 9

revisit ↓ when $D = TV$ but L is arbitrary

Recap: $D = TV$

p^* $D(p^*, \tilde{p}) \le \varepsilon$ $\tilde{p}$

$X_1, \dots, X_n$

$\hat{\theta}(X_1, \dots, X_n)$ ..... $L(p^*, \hat{\theta})$

$L(p, \theta) = \|\theta - \mu(p)\|_2$ ← mean estimation

$n = \infty$: minimum distance functionals
→ modulus of continuity
→ bounded covariance, resilience

$n \to \infty$: finite-sample bounds:
- relaxing $TV \to \widetilde{TV}$
- expanding set $\mathcal{G} \to M$

} VC-dim bound, Lecture-Talagrand

algorithms: (non-convex program that implements MD functional ⇒ approx. optimum)

Examples of other losses:

① higher moment estimation

$L(p, \theta) = \|\theta - \mathbb{E}[XX^T]\|$ ← operator norm

$\theta^*(q) = \mathbb{E}_q[XX^T]$

$L(p, \theta) = \|I - \theta^{-1/2} \mathrm{Cov}_p[X] \theta^{-1/2}\|_F$

② regression

$L(p, \theta) = \mathbb{E}_{x,y \sim p}\left[(y - \langle \theta, x \rangle)^2\right]$

$L(p, \theta) = \mathbb{E}_{x,y \sim p}\left[\,|y - \langle \theta, x \rangle|\,\right]$

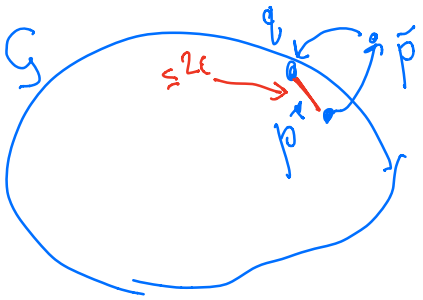
← best linear fit for q

③ classification

$$L(p,\theta) = \mathbb{P}_{x,y\sim p}\left[y \neq \text{sign}(\langle\theta, x\rangle)\right]$$ ← best classifier

---

MD functional: $MD(p, L) = \theta^*(q)$, where $q = \underset{q\in G}{\text{argmin}}\ TV(p,q)$ and $\theta^*(q) = \underset{\theta}{\text{argmin}}\ L(q,\theta)$



Loss of MD functional: at most

$$\star\ \varepsilon_n(G, L, 2\varepsilon)\ L(p, \theta^*(q))\ \star$$

$$= \sup_{p,q\in G,\ TV(p,q)\leq 2\varepsilon}$$

Generalizing resilience

Recall. p is (ρ,ε)-resilient if

$$\|\mu_p - \mu_r\| \leq \varepsilon$$

whenever $r \leq \frac{p}{1-\varepsilon}$.

Prop. If p, q (ρ,ε)-resilient and $TV(p,q) \leq \varepsilon$, then $\|\mu_p - \mu_q\|_2 \leq 2\rho$.

---

p is (ρ, ε)-resilient if

$$L(p, \theta^*(r)) \le \rho \quad \longleftarrow \text{not enough}$$

whenever $r \le \frac{p}{1-\varepsilon}$!

---

p, q : $TV(p, q) \le \varepsilon$

$\Rightarrow r \le \frac{p}{1-\varepsilon}, \quad r \le \frac{q}{1-\varepsilon}$

$\Rightarrow L(p, \theta^*(r)) \le \rho$

$L(q, \theta^*(r)) \le \rho$

!

WANT: $L(p, \theta^*(q)) \le 2p$

"bridge"

($\downarrow$) $L(r, \theta^*(p)) \le p$ if $r \le \frac{p}{1-\varepsilon}$.

$B(r,\theta)$

↑ we do well on all ε-deletions $B(r,\theta)$

($\uparrow$) If $r \le \frac{p}{1-\varepsilon}$ and $L(r, \theta) \le p_1$,

then $L(p, \theta) \le p_2$.

$G_\downarrow(p_1, \varepsilon)$

If we do well on any ε-deletion,

$G_\uparrow(p_1, p_2, \varepsilon)$ we do well on orig. dist^n.

Proposition. Let $G = G_\downarrow(p_1, \varepsilon) \cap G_\uparrow(p_1, p_2, \varepsilon)$

Then $m(G, L, \varepsilon) \le p_2$.

(p_1, p_2, ε)-resilient

Pf. Suppose $p, q \in G$ and $TV(p, \eta) \le \varepsilon$.

Then take $r \le \frac{p}{1-\varepsilon}$, $r \le \frac{q}{1-\varepsilon}$.

$L(p, \theta^*(q)) \le$

$L(r, \theta^*(q)) \le \rho_1$ by $S_\downarrow$ on q

$\Rightarrow L(p, \theta^*(q)) \le \rho_2$ by S_p on p.

$\blacksquare$

So far: $S_\downarrow \cap S_\uparrow$ not "too big" $\Rightarrow$ modulus bounded

Next: $S_\downarrow \cap S_\uparrow$ not "too small"

Example 1. Second moment estimation

$$L(p, \theta) = \| \theta - \mathbb{E}_p[XX^T] \|$$

σ: moment bound

Claim. If p has bounded k^{th} moments for $k > 2$, then p is $(\rho, 2\rho, \varepsilon)$-resilient with $\rho = O(\sigma^2 \varepsilon^{1-2/k})$ ← mean estimation: $O(\sigma \varepsilon^{1-1/k})$

$\uparrow$ $\mathbb{E}_p[\langle v, X \rangle^k] \le \sigma^k \cdot \|v\|_2^k \quad \forall v \in \mathbb{R}^d$.

Goal: $L(r, \theta^*(p)) \le \rho_1$ if $r \ge \frac{p}{1-\varepsilon}$

$\|\mathbb{E}_r[xx^T] - \mathbb{E}_p[xx^T]\| \le \rho_1$ if $r \ge \frac{p}{1-\varepsilon}$.

Dual norm: nuclear norm, or sum of singular values

$$\|\mathbb{E}_r[xx^T] - \mathbb{E}_p[xx^T]\|$$

$$= \sup_{\|Z\|_* \le 1} \mathbb{E}_r[\langle xx^T, Z\rangle] - \mathbb{E}_p[\langle xx^T, Z\rangle]$$

If $Z = U\Lambda V^T$, then $\|Z\|_* = \sum_i \Lambda_{ii}$

$U\Lambda U^T$

$Z = \pm vv^T$, where $\|v\|_2 = 1$

$$= \sup_{\|v\|_2 \le 1} \left| \mathbb{E}_r[\langle v, x\rangle^2] - \mathbb{E}_p[\langle v, x\rangle^2] \right|$$

Suppose $(k/2)^{th}$ moment of $\langle v, x\rangle^2$ is at most σ^2

$\Rightarrow$ bound of $O(\sigma^2 \varepsilon^{1-1/k})$

$\mathbb{E}[|\langle v, x\rangle^2 - \mathbb{E}\langle v, x\rangle^2|^{k/2}]$

$$\leq \mathbb{E}\left(\langle v, X\rangle^2\right)^{k/2} = \mathbb{E}\left[\langle v, X\rangle^k\right]$$

Suppose: if $L(r, \Theta) \leq \rho$, then $L(p, \Theta) \leq 2\rho$

if $\|\mathbb{E}_r[XX^T] - \Theta\| \leq \rho$, then $\|\mathbb{E}_p[XX^T] - \Theta\| \leq 2\rho$

$$\|\mathbb{E}_p[XX^T] - \Theta\| \leq \|\mathbb{E}_r[XX^T] - \Theta\| + \|\mathbb{E}_r[XX^T] - \mathbb{E}_p[XX^T]\|$$

$\leq \rho$ by supposition

$\leq \rho$ by preceding GL argument

∎

$$\sup_{\|v\|_2 \leq 1} \left|\mathbb{E}_r[\langle v, X\rangle^2] - \mathbb{E}_p[\langle v, X\rangle^2]\right|$$

Fix v: $\left|\mathbb{E}_r[\langle v, X\rangle^2] - \mathbb{E}_p[\langle v, X\rangle^2]\right|$

$Y = \langle v, X\rangle^2$ ← 1-dim! dist?

How much can $\mathbb{E}[Y]$ change under deletions? ← mean estimation resilient

ℓ^{th} moment bounded by τ

$\Rightarrow O(\tau \varepsilon^{1-1/\ell})$

Can take $\tau = \sigma^2$, $\ell = k/2$

↓ $k=2$: proved via Chebyshev

General fact:
follows from Orlicz norm bounds

Example 2: Linear regression.

$$L(p,\theta) = \mathbb{E}\left[(y-\langle\theta,x\rangle)^2\right] - \mathbb{E}\left[(y-\langle\theta^*,x\rangle)^2\right]$$

θ^*: optimal regressor

$\Rightarrow$ so that $L(p,\theta^*(p)) = 0$

$z = y - \langle\theta^*, x\rangle \Rightarrow$ "residual error"

Two conditions:

(1) $\mathbb{E}_p[XZ^2X^T] \preceq \sigma^2 \cdot \mathbb{E}_p[XX^T]$

(2) $\mathbb{E}_p[\langle v,x\rangle^4] \le K^4 \cdot \mathbb{E}_p[\langle v,x\rangle^2]^2$

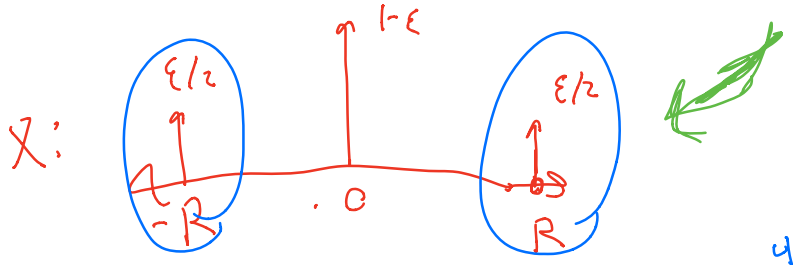
(1) Suppose X, Z independent

"features" "noise"

$$\mathbb{E}_p[XZ^2X^T] = \mathbb{E}_p[Z^2]\cdot\mathbb{E}_p[XX^T] \preceq \sigma^2\cdot\mathbb{E}_p[XX^T]$$

$$\Leftrightarrow \mathbb{E}_p[Z^2] \le \sigma^2$$

(2) Bounding 4th moment in terms of second



$Y = \langle \theta^*, X \rangle$ $\quad \mathbb{E}_p[X^4] = \varepsilon \cdot R^4 \quad R = 1 \Rightarrow$ 4th moment ε

$$\frac{\mathbb{E}_p[X^4]}{\mathbb{E}_p[X^2]^2} = \frac{\varepsilon \cdot R^4}{(\varepsilon \cdot R^2)^2} = \frac{1}{\varepsilon}$$

$$L(p,\theta) = \underbrace{\mathbb{E}[(y - \langle \theta, X \rangle)^2]}_{\text{quadratic}} - \underbrace{\mathbb{E}[(y - \langle \theta^*, X \rangle)^2]}_{\text{quadratic}}$$

θ^*: optimal regressor

Also have.

$$L(p,\theta) = (\theta - \theta^*)^T S_p (\theta - \theta^*),$$ ↖ "almost" a norm

where $S_p = \mathbb{E}_p[XX^T]$.

Goal: If $L(r,\theta) \le \rho_1$, then $L(p,\theta) \le \rho_2$

If $(\theta - \theta^*(r))^T S_r (\theta - \theta^*(r)) \le \rho_1$,

then $(\theta - \theta^*(p))^T S_p (\theta - \theta^*(p)) \le \rho_2$

Will show "$S_r \approx S_p$"

$\frac{1}{2} S_r \le S_p \le 2 S_r$

2nd vs. 4th moment condition

"hypercontractivity"

S_1: $L(r, \theta^*(p)) \le \rho_1$

$(\theta^*(r) - \theta^*(p))^T S_r (\theta^*(r) - \theta^*(p)) \le \rho_1$

"$\theta^*(r)$ and $\theta^*(p)$ are close"

$\mathbb{E}_r[XX^T]$ vs. $\mathbb{E}_p[XX^T]$

$S_r \stackrel{?}{\le} \frac{1}{2} S_p$

$v^T S_r v \stackrel{?}{\le} \frac{1}{2} v^T S_p v \quad \forall\, v$

$\mathbb{E}_r[\langle v, X\rangle^2] \stackrel{?}{\le} \frac{1}{2} \mathbb{E}_p[\langle v, X\rangle^2] \quad \forall\, v$

use resilience

$\left|\mathbb{E}_r[\langle v, X\rangle^2] - \mathbb{E}_p[\langle v, X\rangle^2]\right| \le C(\varepsilon) \cdot \mathbb{E}_p[\langle v, X\rangle^4]^{1/2}$

$\le \frac{1}{2} \mathbb{E}_p[\langle v, X\rangle^2]$

pairs (X, Y)

X in $\mathbb{R}^d$, Y in $\mathbb{R}$