9/20/23

Lecture 5: SoS and Robust Regression

$(x_i, y_i)_{i=1}^N$: corrupted dataset $\qquad x_i \in \mathbb{R}^d$

$(x_i^*, y_i^*)_{i=1}^N$: uncorrupted dataset $\qquad y_i \in \mathbb{R}$

$y_i^* = \langle w^*, x_i \rangle + \xi_i$, $\quad \xi_i \sim N(0, \sigma^2)$

$$a_i^* = \begin{cases} 1 & \text{if } i \text{ is clean} \\ 0 & \text{o.w.} \end{cases}$$

so when $a_i^* = 1$, $\quad (x_i^*, y_i^*) = (x_i, y_i)$

SoS Program

Variables: w (d-dimensional)

$a_1, \ldots, a_N$ (1-dimensional)

Constraints: 1) $a_i^2 = a_i$ (Boolean indicators)

2) $\frac{1}{N} \sum a_i \geq 1 - \eta$ (η fraction corruptions)

Obj: $\min \tilde{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^N a_i (y_i - \langle w, x_i \rangle)^2\right]$

Clean MSE:

$$\frac{1}{N}\sum_{i=1}^N a_i^* (y_i^* - \langle w, x_i^* \rangle)^2 \leq \frac{1}{N}\sum_{i=1}^N (y_i^* - \langle w, x_i^* \rangle)^2 \qquad (*)$$

Note:

$$1 = \underbrace{a_i a_i^*}_{\text{points we correctly identified as clean}} + \underbrace{a_i(1-a_i^*)}_{\text{points we incorrectly identified as clean}} + \underbrace{(1-a_i)}_{\text{points we identified as corrupted}}$$

So

$$(6) = \frac{1}{N}\sum_i a_i a_i^0 \left(y_i^0 - \langle w, x_i^0 \rangle\right)^2 \qquad ①$$

$$+ \frac{1}{N}\sum_i a_i (1-a_i^0) \left(y_i^0 - \langle w, x_i^0 \rangle\right)^2 \qquad ②$$

$$+ \frac{1}{N}\sum_i (1-a_i) \left(y_i^0 - \langle w, x_i^0 \rangle\right)^2 \qquad ③$$

①: when $a_i^0 = 1$, $y_i = y_i^0$ and $x_i = x_i^0$

$$① = \frac{1}{N}\sum_i a_i a_i^0 \left(y_i - \langle w, x_i \rangle\right)^2$$

$$\overset{(a_i^0 \le 1)}{\le} \frac{1}{N}\sum_i a_i \left(y_i - \langle w, x_i \rangle\right)^2 \leftarrow \text{objective value}$$

$$\le \boxed{\text{OPT}}$$ (objective value achieved by w^0, i.e. clean MSE)

Note: $\{a_i^0\}$, w^0, $\{x_i^0\}$, $\{y_i^0\}$ is feasible solution to SoS program

$$②: \frac{1}{N}\sum_i \underbrace{a_i}\underbrace{(1-a_i^0)}\underbrace{\left(y_i^0 - \langle w, x_i^0 \rangle\right)^2}$$

(Cauchy-Schwartz)

$$\le \underbrace{\left(\frac{1}{N}\sum_i (1-a_i^0)^2\right)^{1/2}}_{\le \eta^{1/2}} \cdot \left(\frac{1}{N}\sum_i \underbrace{a_i^2}_{\le 1} \left(y_i^0 - \langle w, x_i^0 \rangle\right)^4\right)^{1/2}$$

$$\leq \eta^{1/2} \cdot \left(\frac{1}{N} \sum_i \left(y_i^* - \langle w, x_i^* \rangle \right)^4 \right)^{1/2}.$$

③: $\frac{1}{N} \sum_i \underbrace{(1-a_i)} \underbrace{\left(y_i^* - \langle w, x_i^* \rangle \right)^2}$

(Cauchy-Schwartz)

$$\leq \underbrace{\left(\frac{1}{N} \sum_i (1-a_i)^2 \right)^{1/2}} \cdot \left(\frac{1}{N} \sum_{i=1}^N \left(y_i^* - \langle w, x_i^* \rangle \right)^4 \right)^{1/2}$$

Note: $\frac{1}{N} \sum_i (1-a_i)^2 \overset{\text{(Booleanity)}}{=} \frac{1}{N} \sum_i (1-a_i) \overset{(\eta \text{ fraction corruptions})}{\leq} \eta$

$$\leq \eta^{1/2} \cdot \left(\frac{1}{N} \sum_{i=1}^N \left(y_i^* - \langle w, x_i^* \rangle \right)^4 \right)^{1/2}$$

How to bound $\frac{1}{N} \sum_{i=1}^N \left(y_i^* - \langle w, x_i^* \rangle \right)^4$?

Recall $y_i^* = \langle w^*, x_i^* \rangle + \xi_i$, $\xi_i \leftarrow N(0, \sigma^2)$ so

$$\frac{1}{N} \sum_{i=1}^N \left(y_i^* - \langle w, x_i^* \rangle \right)^4 = \frac{1}{N} \sum_{i=1}^N \left(\langle w^* - w, x_i^* \rangle + \xi_i \right)^4$$

Note elementary inequality $(a+b)^4 \le 8(a^4+b^4)$, so the above

$$\le \boxed{\frac{8}{N}\sum_{i=1}^{N}\langle w^\circ - w, x^\circ_i\rangle^4} + \underbrace{\frac{8}{N}\sum_{i=1}^{N}\xi_i^4}_{\lesssim\, 8\, \mathbb{E}_{\xi\sim N(0,\sigma^2)}[\xi^4]}$$

$$= 24\sigma^4 = O(\sigma^4)$$

To summarize:

$$\text{clean MSE} \le \frac{1}{N}\sum_i (y^\circ_i - \langle w, x^\circ_i\rangle)^2$$

$$= ① + ② + ③$$

$$= OPT + 2\eta^{1/2}\cdot\left(\frac{1}{N}\sum_{i=1}^{N}(y^\circ_i - \langle w, x^\circ_i\rangle)^4\right)^{1/2}$$

$$\le OPT + 2\eta^{1/2}\cdot\left(\frac{8}{N}\sum_i\langle w^\circ - w, x^\circ_i\rangle^4 + O(\sigma^4)\right)^{1/2}$$

$$\le^{*} OPT + O(\eta^{1/2})\cdot\left[\left(\frac{1}{N}\sum_i\langle w^\circ - w, x^\circ_i\rangle^4\right)^{1/2} + \sigma^2\right]$$

---

* Technically not legit in SoS b/c of fractional power (1/2), but this step can be made rigorous by writing as $(\text{clean MSE} - OPT)^2 \le O(\eta)\cdot\left[\frac{1}{N}\sum_i\langle w^\circ - w, x^\circ_i\rangle^4 + \sigma^4\right]$

To bound $\frac{1}{N}\sum_i \langle w^* - w, x_i^* \rangle^4$, need assumption on distribution:

<u>Def</u>: q is <u>-hypercontractive</u> if

$$\mathop{\mathbb{E}}_{x\sim q}\left[\langle v, x\rangle^4\right] \leq \left(C \cdot \mathop{\mathbb{E}}_{x\sim q}\left[\langle v, x\rangle^2\right]\right)^2 \quad (**)$$

for all $v \in \mathbb{R}^d$, for some $C = O(1)$.

q is <u>certifiably 4-hypercontractive</u> if $(**)$ has an SoS proof.

<u>Example</u>: Any rotation of a product distribution (e.g. $N(\mu, \Sigma)$) is certifiably 4-hypercontractive.

So for $v = w^* - w$, we get

$$\left(\frac{1}{N}\sum_{i=1}^{N} \langle w^* - w, x_i^* \rangle^4\right)^{1/2} \leq \frac{C}{N}\sum_{i=1}^{N} \langle w^* - w, x_i^* \rangle^2$$

$$= \frac{C}{N} \sum_{i=1}^{N} \left(y_i^\circ - \langle w, x_i^\circ \rangle - \xi_i \right)^2$$

$(a+b)^2 \le 2a^2 + 2b^2$

$$\le \frac{2C}{N} \sum_i \left(y_i^\circ - \langle w, x_i^\circ \rangle \right)^2 + \frac{2C}{N} \sum_i \xi_i^2$$

$$\le \frac{2C}{N} \sum_i \left(y_i^\circ - \langle w, x_i^\circ \rangle \right)^2 + O(C\sigma^2)$$

So

$$\text{clean MSE} \le \frac{1}{N} \sum_i \left(y_i^\circ - \langle w, x_i^\circ \rangle \right)^2$$

$$\le \mathrm{OPT} + O(\eta^{1/2}) \left[\frac{2C}{N} \sum_i \left(y_i^\circ - \langle w, x_i^\circ \rangle \right)^2 + O(C\sigma^2) \right]$$

rearranging, we get

$$\left(1 - O(C\eta^{1/2})\right) \frac{1}{N} \sum_i \left(y_i^\circ - \langle w, x_i^\circ \rangle \right)^2$$

$$\le \mathrm{OPT} + O(C\eta^{1/2}\sigma^2),$$

so if $C\eta^{1/2}$ sufficiently small,

$$\text{clean MSE} \le \left(1 + O(C\eta^{1/2})\right)\left(\mathrm{OPT} + O(C\eta^{1/2}\sigma^2)\right)$$