9/23/24

Lecture 6: SoS and Robust Regression

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

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

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

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

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

SoS Program

Variables: w (d-dimensional)

$a_1, \dots, a_N$ (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: $\min \tilde{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^N a_i (y_i - \langle w, x_i \rangle)^2\right]$

Theorem [Klivans–Kothari–Meka '18]:

Suppose x_i^*'s are drawn from a mean-0 distribution over $\mathbb{R}^d$ which is certifiably 4-hypercontractive (see below for defn.). Then for any degree-k pseudodistribution $\tilde{\mathbb{E}}$ over solutions to the program minimizing the objective above,

$$\text{clean MSE}(\tilde{\mathbb{E}}[w]) \leq (1 + O(\eta^{1/2}))(OPT + O(\eta^{1/2}\sigma^2))$$

Claim: Suffices to give a deg-k SoS proof that $\frac{1}{N}\sum_{i=1}^{N}(y_i^\circ - \langle w, x_i^\circ\rangle)^2 \leq$ RHS

Pf: Note $\frac{1}{N}\sum_{i=1}^{N} a_i^\circ (y_i - \langle w, x_i\rangle)^2 \leq \frac{1}{N}\sum_{i=1}^{N}(y_i^\circ - \langle w, x_i^\circ\rangle)^2$ because $x_i = x_i^\circ$ and $y_i = y_i^\circ$ when $a_i^\circ = 1$. Claim then follows by following fact about pseudodistributions.

Fact (Pseudo Cauchy-Schwarz): For any polynomials p, q of deg $\leq k/2$ in $\vec{x}$ and any deg-k pseudodistribution $\hat{\mathbb{E}}$ over x,

$$\hat{\mathbb{E}}\left[p(x)q(x)\right] \leq \hat{\mathbb{E}}\left[p(x)^2\right]^{1/2}\hat{\mathbb{E}}\left[q(x)^2\right]^{1/2}$$

Pf: Suppose $\hat{\mathbb{E}}[p(x)^2], \hat{\mathbb{E}}[q(x)^2] > 0$. By linearity, can assume WLOG that $\hat{\mathbb{E}}[p(x)^2] = \hat{\mathbb{E}}[q(x)^2] = 1$. Note $p \cdot q \leq \frac{1}{2}(p^2 + q^2)$ in deg-k SoS, so $\hat{\mathbb{E}}[pq] \leq 1$ as desired.

Suppose next that $\hat{\mathbb{E}}[p(x)^2] = 0$. Then $\forall \alpha > 0$ $2\alpha p \cdot q \leq p^2 + \alpha^2 q^2$ in deg-k SoS, so $\hat{\mathbb{E}}[pq] \leq \alpha^2 \hat{\mathbb{E}}[q^2]$. Send $\alpha \to 0$ to conclude. $\square$

Let's try to decompose the polynomial

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

Note:

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

So

$$(\star) = \frac{1}{N}\sum_i a_i a_i^\circ (y_i^\circ - \langle w, x_i^\circ\rangle)^2 \qquad ①$$

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

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

①: when $a_i^\circ = 1$, $y_i = y_i^\circ$ and $x_i = x_i^\circ$

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

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

$$\overset{\text{(after taking pseudoexpectation)}}{\le} \boxed{OPT} \quad (\text{objective value achieved by } w^\circ, \text{ i.e. clean MSE})$$

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

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

(Cauchy-Schwartz)

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

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

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

(Cauchy-Schwartz)

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

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

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

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

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

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

To summarize:

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

$$= ① + ② + ③$$

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

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

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

---

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

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

<u>Def</u>: q is <u>4-hypercontractive</u> if

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

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

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

proof at the end

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

<u>Lemma</u>: Let $N \geq \Omega(d^4 \lg^2(d/\delta))$. w.p. at least $1-\delta$, uniform distribution over N i.i.d. draws from a certifiably 4-hypercontractive dist. is certifiably 4-hypercontractive.

So for $v = w^\circ - w$, we get

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

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

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

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

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

So

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

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

rearranging, we get

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

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

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

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

Proof IOU:

Claim: Any product distribution $q = q_1 \otimes \cdots \otimes q_d$ s.t. $\mathbb{E}_{z \sim q_i}[z] = 0$, $\mathbb{E}_{z \sim q_i}[z^2] = 1$, and $\mathbb{E}_{z \sim q_i}[z^4] \le C^2 \ \forall i$ is certifiably 4-hypercontractive.

Pf:
$$\mathbb{E}_{x \sim q} \langle v, x \rangle^4 = \sum_i v_i^4 \mathbb{E}(x_i^4) + 6 \sum_{i<j} v_i^2 v_j^2 \mathbb{E}[x_i^2 x_j^2]$$

$$\le \sum_i C^2 v_i^4 + 6 \sum_{i<j} v_i^2 v_j^2$$

$$= \max(3, C^2) \left(\sum v_i^2\right)^2 - \sum_i |C^2 - 3| \cdot v_i^4$$

The only SoS step
$$\le \max(3, C^2) \cdot \left(\sum v_i^2\right)^2$$

$$\mathbb{E}_{x \sim q} \langle v, x \rangle^2 = \sum_i v_i^2 \, \mathbb{E}(x_i^2) = \sum v_i^2 . \qquad \square$$

(note that definition of certifiable hypercontractivity is invariant under linear transformations of q)