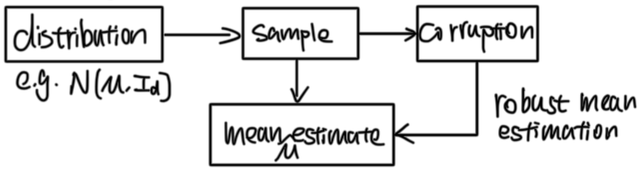
Lecture 12

Robust against adversial corruptions



unknown $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$, $0 \preceq \Sigma \preceq I_d$

sample $y_1^0, \cdots, y_n^0 \overset{iid}{\sim} N(\mu, \Sigma)$, $Y^0 = (y_1^0, \cdots, y_n^0)$

corruption $C_\varepsilon(Y) = \{ Z \mid Y - Z \text{ has } \le \varepsilon \cdot n \text{ non-zero columns} \}$

adversial error estimator $\hat\mu(Z)$

$$\sup \{ \|\hat\mu(Z) - \mu\|^2 \mid Z \in C_\varepsilon(Y^0) \}$$

Inefficient estimator given ε-corrupted sample $Z \in C_\varepsilon(Y^0)$

- find $Y \in C_\varepsilon(Z)$ such that $\mathrm{Cov}(\underbrace{\mathrm{Unif}(\{Y_1, \cdots, Y_n\})}) \preceq 2 \cdot I_d$
- output mean of

Lemma let y, y' random vectors. suppose $\Pr(y \neq y') \le \varepsilon$ and $\mathrm{Cov}(y), \mathrm{Cov}(y') \preceq I$.

Then, $\|\mathbb{E}[y - y']\| \lesssim \sqrt{\varepsilon}$. $*\, \mathrm{Cov}(y), \mathrm{Cov}(y') \preceq \sigma^2 \cdot I$, $\|\mathbb{E}[y] - \mathbb{E}[y']\| \le O(\sqrt{\varepsilon}) \cdot \sigma$.

Proof Let $u = y - y'$, $\bar u = \mathbb{E}[u] = \mathbb{E}[y] - \mathbb{E}[y']$

$$\|\bar u\|^2 = \langle \bar u, \bar u \rangle = \mathbb{E}[\langle \bar u, u \rangle] = \mathbb{E}[\mathbb{1}_{u \neq 0} \cdot \langle \bar u, u \rangle]$$

$$\le \underbrace{\sqrt{\mathbb{E}[\mathbb{1}^2_{u\neq 0}]}}_{\le \varepsilon} \cdot \underbrace{\sqrt{\mathbb{E}[\langle \bar u, u \rangle^2]}}_{\le \underbrace{\sqrt{\mathrm{Var}(\langle \bar u, y \rangle)}}_{\le \|\bar u\|} + \sqrt{\mathrm{Var}(\langle \bar u, y' \rangle)} + \|\bar u\|^2}$$

$$\le \sqrt{\varepsilon} \cdot (2\|\bar u\| + \|\bar u\|^2)$$

$$\longrightarrow \|\bar u\|^{\cancel{2}} \le \frac{2\varepsilon^{1/2}}{1 - \varepsilon^{1/2}} \cancel{\|\bar u\|} \longrightarrow \|\bar u\| \le \frac{2\varepsilon^{1/2}}{1 - \varepsilon^{1/2}} \le 4\sqrt{\varepsilon}$$

Theorem

$\exists$ estimator $\hat\mu$ such that w.h.p. over Y^0, ε-adversial error $\le O(\sqrt{\varepsilon})$, $n \ge O(d/\varepsilon)$.

Proof: By standard concentration bounds, whp.

$\|\frac{1}{n} \sum y_i^0 - \mu\| \le O(\sqrt{\varepsilon}) \implies \|\hat\mu^0 - \mu\| \le O(\sqrt{\varepsilon})$

$\mathrm{Cov}(\mathrm{Unif}(\{y_1^0, \cdots, y_n^0\})) \preceq 2 \cdot I$

estimator $\hat\mu(Z)$:

- find $Y = (y_1, \cdots, y_n) \in C_\varepsilon(Z)$ such that $\mathrm{Cov}(\mathrm{Unif}(\{y_0, \cdots, y_n\})) \preceq 2 \cdot I$
- output $\hat\mu(Z) = \frac{1}{n} \sum_{i=1}^{n} y_i$

analysis: let $(y, y') \sim \mathrm{Unif}(\{(y_1, y_1^0), \cdots, (y_n, y_n^0)\})$

$$\Pr(\cancel{y} \neq \cancel{y}') = \frac{1}{n} \cdot \#\text{ non-zero columns of } Y - \underline{Y}^\circ$$

$$\leq \frac{1}{n}\left(\#\text{ non-zero columns of } Y - Z + \#\text{ non-zero columns of } Z - \cancel{y}^\circ\right)$$

$$\leq 2\varepsilon$$

$$\xrightarrow{\text{lemma}} \| \mathbb{E}[\cancel{y}] - \mathbb{E}[\cancel{y}'] \| \leq O(\sqrt{\varepsilon})$$

$$\longrightarrow \| \hat{\underline{\mu}}^\circ - \hat{\mu}(Z) \| \leq O(\sqrt{\varepsilon})$$

$$\longrightarrow \| \mu - \hat{\mu}(Z) \| \leq O(\sqrt{\varepsilon})$$

Efficient estimator

polynomial system A with variables

- $y_1, \cdots, y_n \in \mathbb{R}^d$
- $\mu \in \mathbb{R}^d$
- $w_1, \cdots, w_n \in \mathbb{R}$
- $R \in \mathbb{R}^{d \times d}$

$$A = \begin{cases} w_i^2 = w_i, & i \in [n] \\ w_i (z_i - y_i) = 0, & i \in [n] \\ \sum_{i=1}^n w_i \geq (1-\varepsilon) n \\ \mu = \frac{1}{n}\sum_{i=1}^n y_i \\ \frac{1}{n}\sum_{i=1}^n (y_i - \mu)(y_i - \mu)^T + \boxed{R^T R} = 2 I_d \end{cases}$$

to enforce $\frac{1}{n}\sum_{i=1}^n (y_i - \mu)(y_i - \mu)^T \preceq 2 I_d$

$\xrightarrow{\text{lemma}}$ if $(Y^\circ, \mu^\circ, w^\circ, R^\circ) \models A$

$(Y, \mu, w, R) \models A$

then $\| \mu - \mu^\circ \| \lesssim \sqrt{\varepsilon}$

* Any two solutions to A are close
* In order to get poly-time estimator, we just need low-degree SOS proofs for the above.

Theorem

Let $(Y^\circ, \mu^\circ, w^\circ, R^\circ) \models A$, then $A \,\big|\!\frac{Y,\mu,w,R}{8}\, \| \mu^\circ - \mu \|^2 \leq 100\varepsilon$

or more general

Theorem

$A \vdash_4 \| \mu - \mu^\circ \|^2 \leq O(\varepsilon)$

Proof:

With the SOS meta theorem, there exists $\hat{\mu}(Z)$ such that

$\| \hat{\mu}(Z) - \mu^\circ \|^2 \leq O(\varepsilon) \Longrightarrow \| \hat{\mu}(Z) - \mu^\circ \| \leq O(\sqrt{\varepsilon})$

Some useful SOS inequalities

- $\vdash_2^{a,b} \langle a, b \rangle \leq \frac{1}{2}\|a\|^2 + \frac{1}{2}\|b\|^2$
- $\vdash_4^{a,b} \langle a, b \rangle^2 = \langle a \otimes b, b \otimes a \rangle \leq \frac{1}{2}\|a\|^2 \cdot \|b\|^2 + \frac{1}{2}\|a\|^2 \cdot \|b\|^2 = \|a\|^2 \cdot \|b\|^2$
- $C > 0$, $\{X^2 \leq CX\} \vdash_2 X < C$

$\{X^2 \leq CX\} \vdash_2 X = \left(\frac{1}{\sqrt{C}} X\right) \cdot \sqrt{C} \leq \frac{X^2}{2C} + \frac{C}{2} \leq \frac{X}{2} + \frac{C}{2}$

$$A \vdash_2 \frac{1}{n}\sum_{i=1}^n (1 - w_i w_i^*)^2 = \frac{1}{n}\sum_{i=1}^n (1 - w_i w_i^*)$$

$A \vdash_2 1 - w_i w_i^* \leq 2 - w_i - w_i^* \rightarrow$ $\leq \frac{1}{n}\sum_{i=1}^n (1 - w_i) + \frac{1}{n}\sum_{i=1}^n (1 - w_i^*)$

$\leq 2\varepsilon$

* $\sum_{i=1}^n w_i \geq (1-\varepsilon) n$

$\Rightarrow \frac{1}{n}\sum_{i=1}^n w_i \geq 1 - \varepsilon$

$\Rightarrow \frac{1}{n}\sum_{i=1}^n (1 - w_i) \leq \varepsilon$

$$A \vdash_4 \|\mu-\mu^*\|^2 = \langle \mu-\mu^*, \mu-\mu^* \rangle$$

$$= \langle \mu-\mu^*, \tfrac{1}{n}\sum_{i=1}^n y_i - \tfrac{1}{n}\sum_{i=1}^n y_i^* \rangle$$

$$= \frac{1}{n}\sum_{i=1}^n \langle \mu-\mu^*, y_i - y_i^* \rangle$$

$$= \frac{1}{n}\sum_{i=1}^n (1-w_i w_i^*) \langle \mu-\mu^*, y_i - y_i^* \rangle$$

$$= \frac{1}{n}\sum_{i=1}^n (1-w_i w_i^*) \langle \mu-\mu^*, y_i - \mu + \mu^* - y_i^* + \mu - \mu^* \rangle$$

$$\leq \frac{1}{n}\sum_{i=1}^n (1-w_i w_i^*) \langle \mu-\mu^*, y_i - \mu \rangle$$

$$+ \frac{1}{n}\sum_{i=1}^n (1-w_i w_i^*) \langle \mu-\mu^*, \mu^* - y_i^* \rangle + 2\varepsilon \|\mu+\mu^*\|^2$$

$$A \vdash_8 \|\mu-\mu^*\|^4 \leq 4\Big(\frac{1}{n}\sum_{i=1}^n (1-w_i w_i^*) \langle \mu-\mu^*, y_i - \mu \rangle + \frac{1}{n}\sum_{i=1}^n (1-w_i w_i^*) \langle \mu-\mu^*, \mu^* - y_i^* \rangle \Big)^2$$

$$\leq 8 \Big(\underbrace{\frac{1}{n}\sum_{i=1}^n (1-w_i w_i^*)}_{\leq 2\varepsilon} \langle \mu-\mu^*, y_i - \mu \rangle \Big)^2 + 8 \Big(\underbrace{\frac{1}{n}\sum_{i=1}^n (1-w_i w_i^*)}_{\leq 2\varepsilon} \langle \mu-\mu^*, \mu^* - y_i^* \rangle \Big)^2$$

$$\leq 16\varepsilon \Big(\frac{1}{n}\sum_{i=1}^n \langle \mu-\mu^*, y_i - \mu \rangle \Big)^2 + 16\varepsilon \Big(\frac{1}{n}\sum_{i=1}^n \langle \mu-\mu^*, \mu^* - y_i^* \rangle \Big)^2$$

Recall $\frac{1}{n}\sum_{i=1}^n (y_i^* - \mu^*)(y_i^* - \mu^*)^T \preceq 2I_d$, which implies

$$A \vdash_2 \frac{1}{n}\sum_{i=1}^n \langle \mu-\mu^*, \mu^* - y_i^* \rangle^2 = (\mu-\mu^*)^T \Big[\frac{1}{n}\sum_{i=1}^n (y_i - \mu)(y_i - \mu)^T \Big] (\mu-\mu^*)$$

$$= (\mu-\mu^*)^T \big[2I_d - R^T R \big] (\mu-\mu^*)$$

$$\leq 2\|\mu-\mu^*\|^2$$

<u>together</u>

$$A \vdash_8 \|\mu-\mu^*\|^4 \leq 32\varepsilon \cdot \|\mu-\mu^*\|^2$$

$$\Rightarrow A \vdash_8 \|\mu-\mu^*\|^2 \leq 32\varepsilon = O(\varepsilon)$$