Robust Linear Regression via Least Squares

Yuanhao Wang

February 9, 2021

1 Setting

Consider a an uncorrupted i.i.d. dataset $\{(\mathbf{x}_i, y_i^*)\}_{i\in[n]} \sim \mathcal{D}$ such that $y_i^* = \mathbf{x}_i^\top \theta^* + \xi_i$, where ξ_i is mean-zero and $1-$subgaussian. Assume that the adversary corrupts $m = \epsilon n$ of the labels y_i^* and the algorithm observes corrupted labels $\{y_i\}_{i\in[n]}$. In other words, there exists $\mathbf{b} \in \mathbb{R}^n$ such that $\|\mathbf{b}\|_0 \le m$, $\|\mathbf{b}\|_\infty \le 1$ such that

$$y_i = \mathbf{x}_i^\top \theta^* + \xi_i + b_i.$$

We make the following additional assumption on the distribution on $\mathbf{x}$:

Assumption 1 ($(C,4)$-hypercontractivity)**.** *$\exists C > 0$: $\forall v \in \mathbb{R}^d$,*

$$\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[(\mathbf{x}^\top v)^4] \le C \cdot \left(\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[(\mathbf{x}^\top v)^2]\right)^2.$$

Note that this property is invariant under arbitrary linear transformation and is satisfied by any Gaussian distribution [1]. For a $(C,4)$-hypercontractive distribution, we have the following facts.

Fact 1 (Fact 3.4 [1])**.** *Define $\epsilon_1 := \frac{Cd^2}{\sqrt{n\delta}}$ and $\Sigma := \mathbb{E}_{\mathbf{x}\sim\mathcal{D}}\mathbf{x}\mathbf{x}^\top$. With probability $1-\delta$,*

$$(1-\epsilon_1)\Sigma \preccurlyeq \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top \preccurlyeq (1+\epsilon_1)\Sigma.$$

Fact 2. *If the distribution of $\mathbf{x}$ is $(C,4)$-hypercontractive and isotropic (i.e. $\mathbb{E}\mathbf{x}\mathbf{x}^\top = \mathbf{I}$), then*

$$\Pr[\|\mathbf{x}\|_2 > t] \le \frac{C \cdot \mathrm{poly}(d)}{t^4}.$$

Proof. Consider a random $v \sim N(0, \mathbf{I})$.

$$\mathbb{E}_v(\mathbf{x}^\top v)^4 = \|\mathbf{x}\|^4 \cdot \mathbb{E}_{\xi\sim N(0,1)}\xi^4 = \Theta(1)\cdot\|\mathbf{x}\|^4.$$

Therefore,

$$\begin{aligned}\mathbb{E}_{\mathbf{x}}[\|\mathbf{x}\|^4] &\le \Theta(1)\cdot\mathbb{E}_{\mathbf{x},v}(\mathbf{x}^\top v)^4 \le C\Theta(1)\cdot\mathbb{E}_v\left(\mathbb{E}_{\mathbf{x}}(\mathbf{x}^\top v)^2\right)^2 \\ &\le C\cdot\Theta(1)\cdot\mathbb{E}_v\|v\|^4 = C\cdot\mathrm{poly}(d).\end{aligned}$$

The claim follows from Markov's inequality. □

Fact 3. *If $\mathcal{D}$ is $(C,4)$-hypercontractive and isotropic, $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are i.i.d. samples from $\mathcal{D}$. Denote $\sigma(\cdot)$ to be the decreasing order of $\|\mathbf{x}_i\|_2$. Then with probability $1-\delta$,*

$$\sum_{i=1}^{m} \|\mathbf{x}_{\sigma(i)}\|_2 \le \delta^{-1/4} n^{1/4} m^{3/4} \mathrm{poly}(C,d).$$

Proof. Fix $k \in [m]$. Set $t = \alpha \left(\frac{Cn}{k}\right)^{1/4} \mathrm{poly}(d)$. By Fact 2,

$$\begin{aligned}
\Pr[\|\mathbf{x}_{\sigma(k)}\|_2 > t] &\le \binom{n}{k} \Pr[\|\mathbf{x}\| > t]^k \\
&\le \binom{n}{k} \cdot \left(\frac{C \cdot \mathrm{poly}(d)}{t^4}\right)^k \\
&\le \frac{n^k}{k!} \cdot \frac{k^k}{\alpha^{4k} n^k} \le \left(\frac{e}{\alpha^4}\right)^k.
\end{aligned}$$

Choosing $\alpha = \Omega(\delta^{-1/4})$ gives $\Pr[\|\mathbf{x}_{\sigma(k)}\|_2 > t] \le \delta/k^2$. Thus, by a union bound, with probability $1 - (\pi^2/6)\delta$,

$$\sum_{i=1}^{m} \|\mathbf{x}_{\sigma(i)}\|_2 \le \sum_{k=1}^{m} \delta^{-1/4} \left(\frac{Cn}{k}\right)^{1/4} \mathrm{poly}(d) \le \delta^{-1/4} n^{1/4} m^{3/4} \mathrm{poly}(C,d).$$

□

2 Why Least Square Works

We now show that the ordinary least square estimator achieves robustness against adversarial corruption. Define

$$\hat{\theta} := \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top\right)^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i.$$

Define $\hat{\Sigma} := \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top$. Then

$$\begin{aligned}
\hat{\theta} &= \frac{1}{n} \hat{\Sigma}^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i \cdot (\mathbf{x}_i^\top \theta^*) + \mathbf{x}_i \cdot \xi_i + \xi_i \cdot b_i\right) \\
&= \theta^* + \frac{1}{n} \hat{\Sigma}^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i \cdot (\xi_i + b_i)\right).
\end{aligned}$$

Hence

$$\|\hat{\theta} - \theta^*\|_\Sigma \le \frac{1}{n} \left\| \hat{\Sigma}^{-1} \sum_{i \in [n]} \mathbf{x}_i \xi_i \right\|_\Sigma + \frac{1}{n} \left\| \hat{\Sigma}^{-1} \sum_{i \in [n]} \mathbf{x}_i b_i \right\|_\Sigma.$$

The first term is known to be $O\left(\frac{d^2}{\sqrt{n}}\right)$ with high probability. It remains to bound the second term. Define $\mathbf{z}_i := \Sigma^{-1/2}\mathbf{x}_i$ to be the whitened inputs. By Definition 1, the distribution of $\mathbf{z}_i$ is also $(C, 4)$-hypercontractive. Also, $\mathbb{E}\mathbf{z}_i\mathbf{z}_i^\top = \mathbf{I}$. Thus Fact 3 applies. By Fact 3, with probability $1 - \delta$

$$\sum_{i=1}^{n} \|\mathbf{z}_i\|_2 \cdot I[b_i \neq 0] \leq \delta^{-1/4} n^{1/4} m^{3/4} \mathrm{poly}(C, d).$$

It follows that with probability $1 - \delta$

$$\begin{aligned}
\left\| \hat{\Sigma}^{-1} \sum_{i=1}^{n} \mathbf{x}_i b_i \right\|_{\Sigma} &\leq \sum_{i=1}^{n} \|\Sigma^{1/2}\hat{\Sigma}^{-1}\mathbf{x}_i b_i\|_2 \\
&= \sum_{i=1}^{n} \|\Sigma^{1/2}\hat{\Sigma}^{-1}\Sigma^{1/2}\mathbf{z}_i b_i\|_2 \\
&\leq \|\Sigma^{1/2}\hat{\Sigma}^{-1}\Sigma^{1/2}\|_2 \cdot \sum_{i=1}^{n} \|\mathbf{z}_i\|_2 \cdot I[b_i \neq 0] \\
&\leq \left(1 + \frac{Cd^2}{\sqrt{n\delta}}\right) \cdot \delta^{-1/4} n^{1/4} m^{3/4} \mathrm{poly}(C, d) \\
&= n\epsilon^{0.75} \cdot \mathrm{poly}(C, d, 1/\delta).
\end{aligned}$$

In other words, with probability $1 - \delta$,

$$\|\hat{\theta} - \theta^*\|_{\Sigma} \leq O\left(\frac{d^2 \ln(1/\delta)}{\sqrt{n}}\right) + O\left(\epsilon^{0.75} \mathrm{poly}(C, d, 1/\delta)\right).$$

References

[1] Ainesh Bakshi and Adarsh Prasad. Robust linear regression: Optimal rates in polynomial time. *arXiv preprint arXiv:2007.01394*, 2020.