Robust Linear Regression by Subquantile Minimization

Arvind Rathnashyam Fatih Orhan Josh Myers Jake Herman

Rensselaer Polytechnic Institute

(rathna, orhanf, myersj5, hermaj2)@rpi.edu

ML and Optimization Group U5

April 17, 2023

Presentation Overview

Huber Contamination Model

Problem

The *Huber Contamination Model* *is the following:*

$$\hat{P} = (1-\varepsilon)P + \varepsilon Q \textit{ where } \varepsilon \in (0, 0.5)$$

where P and Q represent the general linear models

$$\boldsymbol{y}_P = \boldsymbol{\beta}_P^\top \mathbf{P} + \epsilon_P$$

$$\boldsymbol{y}_Q = \boldsymbol{\beta}_Q^\top \mathbf{Q} + \epsilon_Q$$

$\boldsymbol{\beta}_P$ and $\boldsymbol{\beta}_Q$ are oracle regressors and ϵ_P and ϵ_Q represent 0-centered gaussian noise.

Our goal is to learn a model that learns a good distribution of P from $\hat{P}$

Motivation

Definition

Oblivious Noise is noise which is independent of the input data
Adaptive Noise is noise which is dependent upon the input data

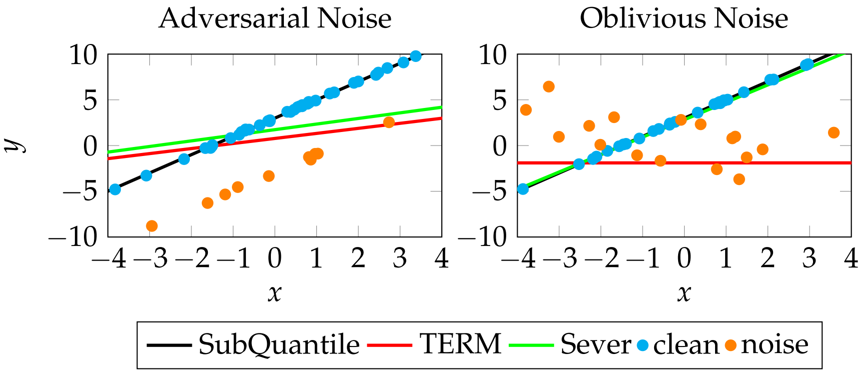

Figure: Sub-Quantile Performance on Adapative Outliers

Theorem

The expected optimal parameters of the corrupted model $\hat{P}$

$$\mathbb{E}\left[\boldsymbol{X}^{\dagger}\boldsymbol{y}\right] = (1-\varepsilon)\boldsymbol{\beta}_P + \varepsilon\boldsymbol{\beta}_Q$$

where $\boldsymbol{X}^{\dagger} \triangleq \left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}$, i.e. the Moore-Penrose Inverse.

This theorem motivates our reasoning for optimizing over the Subquantile. We want a method to reduce ε.

Statistical Preliminaries of the Subquantile

1. $$\mathcal{Q}_p = \inf \{x \in \mathbb{R} : p \leq F(x)\}$$

2. Let ℓ be the loss functions. We can now define risk as:

$$\mathcal{U} = \mathbb{E}\left[\ell(f(\boldsymbol{x};\boldsymbol{\theta}), y)\right]$$

3. The p-Quantile of the Empirical Risk is given by:

$$\mathbb{L}_p = \frac{1}{p}\int_0^p \mathcal{Q}_q(\mathcal{U})dq = \mathbb{E}\left[\mathcal{U}|\mathcal{U} \leq \mathcal{Q}_p(\mathcal{U})\right] = \max_{t\in\mathbb{R}}\left\{t - \frac{1}{p}\mathbb{E}\left[(t-\mathcal{U})^+\right]\right\}$$

4. For the least squares regression case:

$$\mathbb{L}_p = \max_{t\in\mathbb{R}}\left\{t - \frac{1}{np}\sum_{i=1}^{n}\left(t - \left(\boldsymbol{\theta}^\top\boldsymbol{x}_i - y_i\right)\right)^+\right\}$$

Multiple Columns

Subtitle

Heading

1. Statement
2. Explanation
3. Example

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer lectus nisl, ultricies in feugiat rutrum, porttitor sit amet augue. Aliquam ut tortor mauris. Sed volutpat ante purus, quis accumsan dolor.

Drug Discovery

Objectives	Test RMSE (`Drug Discovery`)			
	$\epsilon = 0.1$	$\epsilon = 0.2$	$\epsilon = 0.3$	$\epsilon = 0.4$
ERM	$1.303_{(0.0665)}$	$1.790_{(0.0849)}$	$2.198_{(0.0645)}$	$2.623_{(0.1010)}$
CRR [**?**]	$\mathbf{1.079}_{(\mathbf{0.0899})}$	$1.125_{(0.0832)}$	$1.385_{(0.1372)}$	$1.725_{(0.1136)}$
STIR [**?**]	$1.087_{(0.1256)}$	$1.167_{(0.0750)}$	$1.403_{(0.0987)}$	$1.668_{(0.1142)}$
Robust Risk [**?**]	$1.176_{(0.1110)}$	$1.336_{(0.1882)}$	$1.437_{(0.1723)}$	$1.800_{(0.0820)}$
SMART [**?**]	$1.094_{(0.1065)}$	$1.323_{(0.0758)}$	$1.578_{(0.0799)}$	$1.984_{(0.2020)}$
TERM [**?**]	$1.326_{(0.0757)}$	$1.357_{(0.0990)}$	$1.310_{(0.0670)}$	$1.302_{(0.0851)}$
SEVER [**?**]	$1.111_{(0.0924)}$	$\mathbf{1.067}_{(\mathbf{0.0457})}$	$\mathbf{1.071}_{(\mathbf{0.0807})}$	$\mathbf{1.138}_{(\mathbf{0.1162})}$
Huber [**?**]	$1.412_{(0.0474)}$	$1.501_{(0.2918)}$	$2.231_{(0.9054)}$	$2.247_{(1.0399)}$
RANSAC [**?**]	$1.238_{(0.0529)}$	$1.643_{(0.1331)}$	$2.092_{(0.1935)}$	$2.679_{(0.1365)}$
SubQuantile($p = 1 - \epsilon$)	$\mathbf{0.887}_{(\mathbf{0.1046})}$	$\mathbf{0.936}_{(\mathbf{0.1051})}$	$\mathbf{0.927}_{(\mathbf{0.0729})}$	$\mathbf{1.015}_{(\mathbf{0.0978})}$
Genie ERM	$0.986_{(0.1039)}$	$0.955_{(0.0698)}$	$1.038_{(0.0886)}$	$1.030_{(0.0578)}$

Table: `Drug Discovery` Dataset. Empirical Risk over P with oblivious noise

Figure

Rensselaer

Definitions & Examples

Definition

A prime number is a number that has exactly two divisors.

Example

- 2 is prime (two divisors: 1 and 2).
- 3 is prime (two divisors: 1 and 3).
- 4 is not prime (three divisors: 1, 2, and 4).

You can also use the `theorem`, `lemma`, `proof` and `corollary` environments.

Theorem, Corollary & Proof

Theorem (Mass–energy equivalence)

$E = mc^2$

Corollary

$x + y = y + x$

Proof.

$\omega + \phi = \epsilon$ $\square$

Equation

$$\cos^3\theta = \frac{1}{4}\cos\theta + \frac{3}{4}\cos 3\theta \tag{1}$$

Verbatim

Example (Theorem Slide Code)

```
\begin{frame}
\frametitle{Theorem}
\begin{theorem}[Mass--energy equivalence]
$E = mc^2$
\end{theorem}
\end{frame}
```

Slide without title.

Citing References

An example of the `\cite` command to cite within the presentation:

This statement requires citation [Smith, 2022, Kennedy, 2023].

References

John Smith (2022)
Publication title
Journal Name 12(3), 45 – 678.

Annabelle Kennedy (2023)
Publication title
Journal Name 12(3), 45 – 678.

Acknowledgements

Smith Lab

- Alice Smith
- Devon Brown

Cook Lab

- Margaret
- Jennifer
- Yuan

Funding

- British Royal Navy
- Norwegian Government

The End

Questions? Comments?