Robust and Heavy-Tailed Mean Estimation Made Simple, via Regret Minimization

Samuel B. Hopkins (UC Berkeley) Jerry Li (MSR) Fred Zhang (UC Berkeley)

NEURAL INFORMATION PROCESSING SYSTEMS

Our Result

- A much simpler analysis of a classic solution to the robust mean estimation problem, based on off-the-shelf regret bound of multiplicative update.
- A unified view on robust and heavy-tailed mean estimation, through convex duality.
- A simple and improved analysis of the gradient descent-based algorithm by Cheng, Diakonikolas, Ge and Soltanolkotabi (ICML '20).

Robust Mean Estimation

Problem. Given d-dimensional samples $X_1, \ldots, X_n \sim D$, but where an ϵn samples are arbitrarily corrupted, estimate the mean μ of D.

Why is it hard? The ϵn contaminated samples are corruptions introduced by a malicious adversary, which can replace the data in an arbitrary fashion.

Naïve idea fails Naïve estimators such as the empirical mean can suffer arbitrarily-high inaccuracy as a result of these malicious samples.

Heavy-Tailed Mean Estimation

Problem. Given d-dimensional samples $X_1, \ldots, X_n \sim D$, estimate μ by an estimator $\hat{\mu}$ such that $\|\mu - \hat{\mu}\|$ is small with high probability (or equivalently, estimate μ with optimal confidence intervals).

Why is it hard? Since our only assumption about D is that it has finite covariance, D may have heavy tails.

Naïve idea fails. Standard estimators such as the empirical mean can therefore be poorly concentrated.

Overview

Learning in the presence of outliers is a central task in modern statistics and machine learning.

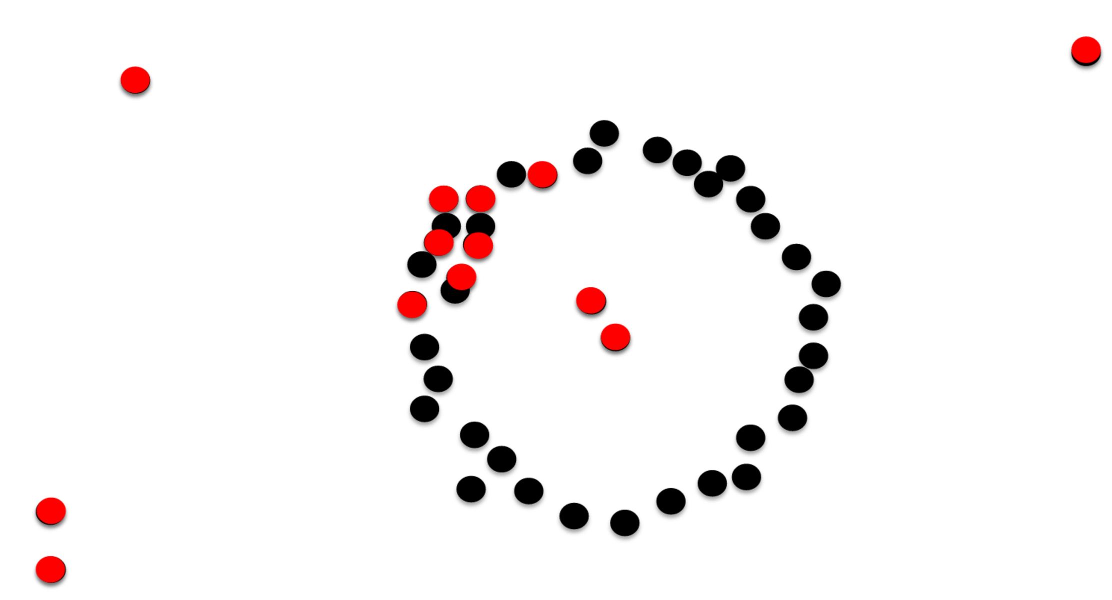

Figure 1. Can we find the mean under the outliers?

Outliers have many sources.

- Modern data sets can be exposed to random corruptions or even malicious tampering, as in data poison attacks.
- Data drawn from heavy-tailed distributions can naturally contain outlying samples—heavy-tailed data are found often in network science, biology, and beyond.

In this work, we revisit the most fundamental high-dimensional estimation problem, estimating the mean of a distribution from samples, in the two basic and widely-studied robust settings.

A Simple Analysis

Only based on the regret bound of multiplicative weights update, a classic result from statistical learning and convex optimization:

$$\frac{1}{T}\sum_{t=1}^{T}\left\langle w^{(t)}, \tau^{(t)}\right\rangle \le \frac{1}{T}(1+\eta)\sum_{t=1}^{T}\left\langle w, \tau^{(t)}\right\rangle + \frac{\rho \cdot \mathrm{KL}(w\|w^{(1)})}{T\eta}.$$

No-regret is all you need to defeat outliers.

A Unified View

We show that solving the following problem directly leads to near optimal solution to both problems:

Spectral Sample Reweighting

Given $\{x_i\}_{i=1}^n$ in $\mathbb{R}^d$, the spectral sample reweighing problem asks for a set of weights w and a spectral center $\nu \in \mathbb{R}^d$ such that for $\alpha = O(1)$

$$\left\|\sum_{i\le n} w_i(x_i-\nu)(x_i-\nu)^\top\right\| \le \alpha \cdot \min_{w,\nu}\left\|\sum_{i\le n} w_i(x_i-\nu)(x_i-\nu)^\top\right\|$$

We show that the classic Filter algorithm for robust mean estimation solves this problem. And there's a simple analysis!

Proof relies of SDP duality and a fancy *gaussian sampling* argument, more in the paper.

References

[1] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

[2] Yu Cheng, Ilias Diakonikolas, Rong Ge, and Mahdi Soltanolkotabi. High-dimensional robust mean estimation via gradient descent. In *International Conference on Machine Learning (ICML '20)*, 2020.

[3] Ilias Diakonikolas, Gautam Kamath, Daniel Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robust estimators in high-dimensions without the computational intractability. *SIAM Journal on Computing*, 48(2):742–864, 2019.

[4] Gábor Lugosi and Shahar Mendelson. Sub-gaussian estimators of the mean of a random vector. *Annals of Statistics*, 47(2):783–794, 2019.

Contact: `z0@berkeley.edu`