Abstract: Robust Estimators in High Dimensions without the Computational Intractability

April 3, 2017

Abstract

Starting from the seminal works of Tukey (1960) and Huber (1964), the field of robust statistics asks: Are there estimators that provably work in the presence of noise? The trouble is that all known provably robust estimators are either hard to compute in high-dimensions, or lose large dimension-dependent factors in the error.

Here, we study a basic problem in robust statistics, posed in various forms in the above works. Given samples from a Gaussian or sub-gaussian distribution, can we estimate the mean and covariance of the underlying distribution? We give algorithms for this problem which are both optimal in the error (up to logarithmic factors) and which are efficiently computable. We also show several more applications to our techniques to product distributions, various mixture models, and weaker moment-based assumptions. Moreover, we demonstrate the practical applicability of our algorithm, on various synthetic and real data sets.

Based on joint work with Ilias Diakonikolas, Gautam Kamath, Daniel Kane, Ankur Moitra, and Alistair Stewart.