A New Perspective On Robust Mean Regression

Abstract

Big data are often contaminated by outliers and heavy-tailed errors. To address this challenge, we propose the adaptive Huber regression for robust estimation and inference. The key observation is that the robustification parameter should adapt to sample size, dimension and moments for optimal tradeoff between biases and robustness. Our framework is able to handle heavy-tailed data with bounded $(1 + \delta)$-th moment for any $\delta > 0$. We establish a sharp phase transition for robust estimation of regression parameters in both finite dimensional and high dimensional settings: when $\delta \geq 1$, the estimator achieves sub-Gaussian rate of convergence without sub-Gaussian assumptions, while only a slower rate is available in the regime $0 < \delta < 1$ and the transition is smooth and optimal. As a consequence, the nonasymptotic Bahadur representation for finite-sample inference can only be derived when the second moment exists. Numerical experiments lend further support to our obtained theories.