THE COST OF PRIVACY: OPTIMAL RATES OF CONVERGENCE FOR PARAMETER ESTIMATION WITH DIFFERENTIAL PRIVACY

Linjun Zhang

Rutgers University
E-mail: zlj11112222@gmail.com

Abstract: Privacy-preserving data analysis is a rising challenge in contem- porary statistics, as the privacy guarantees of statistical methods are often achieved at the expense of accuracy. In this paper, we investigate the tradeoff between statistical accuracy and privacy in mean estimation and linear regression, under both the classical low- dimensional and modern high-dimensional settings. A primary focus is to establish minimax optimality for statistical estimation with the (ε , δ)-differential privacy constraint. To this end, we find that classical lower bound arguments fail to yield sharp results, and new technical tools are called for. Inspired by the theoretical computer science literature on “trac- ing adversaries”, we formulate a general lower bound argument for minimax risks with differential privacy constraints, and apply this argument to high-dimensional mean estimation and linear regression problems. We also design computationally efficient algorithms that attain the minimax lower bounds up to a logarithmic factor. In par- ticular, for the high-dimensional linear regression, a novel private iter- ative hard thresholding pursuit algorithm is proposed, based on a pri- vately truncated version of stochastic gradient descent. The numerical performance of these algorithms is demonstrated by simulation stud- ies and applications to real data containing sensitive information, for which privacy-preserving statistical methods are necessary.