Private Linear Regression via a Down-Sensitivity to Privacy Reduction

Ittai Rubinstein ittair@mit.edu
Chris Ge cge7@mit.edu
Samuel B. Hopkins samhop@mit.edu
Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract. We present a sample- and time-efficient (ε, δ)-differentially private (DP) algorithm for d dimensional linear regression with a sample complexity of

$$n_{\text{STAR}} = \widetilde{O}\left(\frac{d}{\alpha^2} + \frac{d\log(1/\delta)}{\alpha\varepsilon} + \frac{d\log(1/\delta)}{\varepsilon}\right) + o(d)\,.$$

This improves upon prior polynomial-time algorithms whose sample complexity either depends on the condition number of the design matrix κ (for DP-SGD with gradient clipping), scales quadratically with the dimension (for Sum-of-Squares algorithms) or with the privacy parameter (for outlier removal algorithms such as insufficient statistics perturbation or ISSP),

$$n_{\text{SoS}} = \widetilde{\Omega}\left(\frac{d^2}{\alpha^2}\right)\,, \quad n_{\text{DP-SGD}} = \widetilde{\Omega}\left(\frac{d\sqrt{\kappa}}{\varepsilon}\right)\,, \quad n_{\text{ISSP}} = \widetilde{\Omega}\left(\frac{d}{\varepsilon^2}\right)\,.$$

Our algorithm is based on a novel *subsample-test-aggregate* (STA) approach for ensuring privacy given only bounded *down-sensitivity* – robustness to removal, but not addition, of a small number of samples. The intuition that down-sensitivity should be related to privacy is not new, but STA formalizes this by providing an *efficient black-box reduction from down-sensitivity to privacy* which we expect to be applicable beyond the setting of linear regression.