Is Interaction Necessary for Distributed Private Learning?

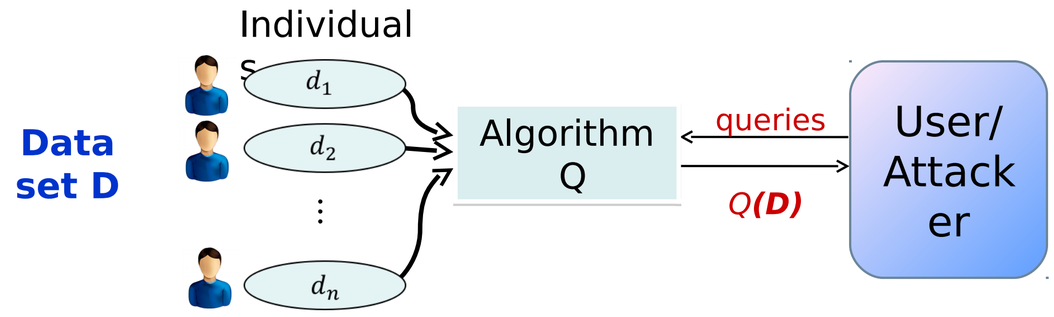
Adam Smith, Abhradeep Thakurta, Jalaj Upadhyay, Mahdy Zolghadr

Recent large-scale deployments of differentially private algorithms employ the local model for privacy, where data are randomized on individual's devices before being sent to an server that computes approximate, aggregate statistics. The server need not be trusted for privacy, leaving data control in users' hands.

For an important class of convex optimization problems (including logistic regression, support vector machines, and the Euclidean median), the best known locally differentially private algorithms, are highly interactive. With n users in the protocol, they use n rounds of back and forth communication. The server exchanges messages with each user only once, but must do so in sequence.

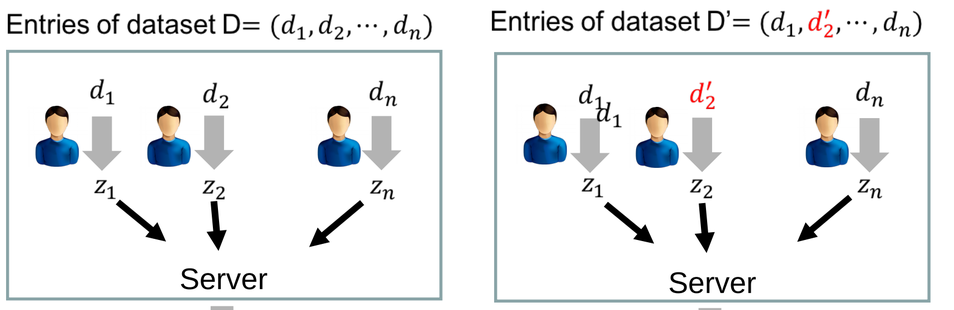
We ask: how much interaction is necessary to optimize convex functions in the local DP model?

What Can We Learn Privately?

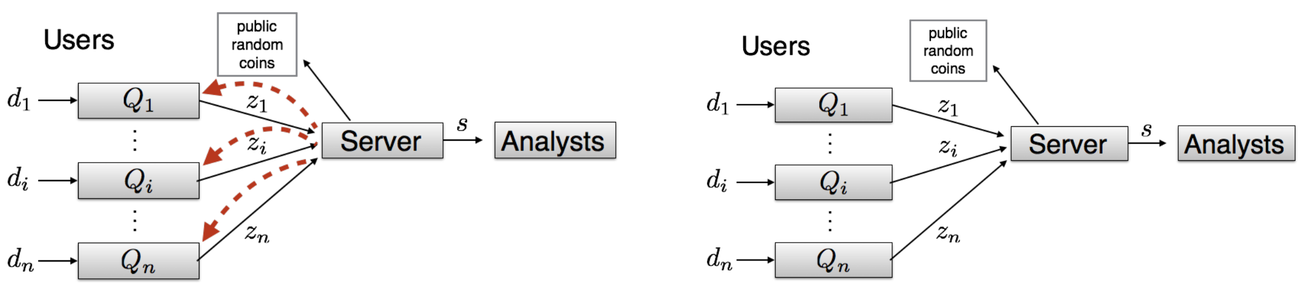


- Consider database of sensitive individual data (e.g., medical records, purchase history)
- Want: Run learning/statistical algorithms while preserving privacy

Local Differential Privacy

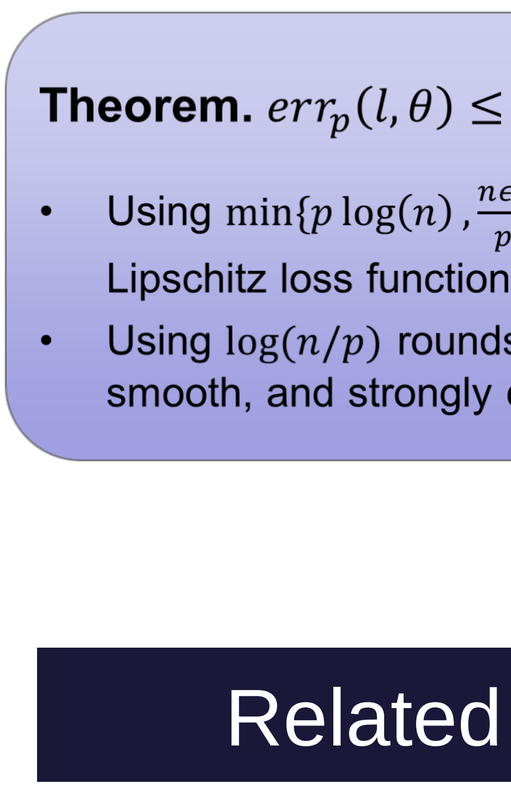


Definition [DMNS06]
Randomized mechanism $\mathfrak{M}$ is (ε, δ)-differentially private if, for all "neighboring" datasets, $\boldsymbol{A}$ and $\boldsymbol{A}'$, and for all events S, $\Pr[\mathfrak{M}(\boldsymbol{A}) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathfrak{M}(\boldsymbol{A}') \in S] + \delta$



Lower Bound on Interaction

Theorem. To achieve an empirical risk of α, any first (or higher order method), needs to make at least $O(\log(1/\alpha))$ adaptive interactions



One adaptive query needed to move to the next inner layer

Non-interactive Private Local Learning

Theorem. Let $\mathcal{C} \subseteq \mathbb{R}^p$, and let $\ell: \mathcal{C} \times \mathcal{D} \to \mathbb{R}$ be a 1-Lipschitz loss function. For every distribution $\mathcal{P}$ on $\mathcal{D}$, with probability $99/100$, one can output a $\theta \in \mathcal{C}$, such that

$$err_{\mathcal{P}}(\ell; \theta) \leq O\left(\left(\frac{\sqrt{p}\log^2(\epsilon^2 n)}{\epsilon^2 n} \right)^{\frac{1}{p+1}} \right)$$

Adaptive Private Local Learning

Theorem. $err_p(l, \theta) \leq \Theta\left(\sqrt{\frac{p}{n\,\epsilon^2}}\right)$

- Using $\min\{p\log(n), \frac{n\epsilon^2}{p}\}$ rounds of interaction for Lipschitz loss function
- Using $\log(n/p)$ rounds of interaction for Lipschitz, smooth, and strongly convex function

Related Publications

[KLNRS13] S. Kasiviswanathan, H. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What Can We Learn Privately? SIAM J. Computing, 2013.

[STU16] A. Smith, A. Thakurta, and J. Upadhyay. Is Interaction Necessary for Distributed Private Learning. To Appear in IEEE S&P, 2017.

NSF Grant IIS-1447700