Open Problem: Is Margin Sufficient for Non-Interactive Private Distributed Learning?

Amit Daniely *Hebrew University and Google*

Vitaly Feldman *Google Brain*

Editors: Alina Beygelzimer and Daniel Hsu

1. Overview

We consider learning in distributed systems where each client i (or user) holds a data point $z_i \in Z$ drawn i.i.d. from some unknown distribution P and the goal of the server is to solve some statistical learning problem using the data stored at the clients. In addition, the communication from the client to the server has to satisfy the guarantees of local differential privacy (LDP) (Warner, 1965; Evfimievski et al., 2003; Kasiviswanathan et al., 2011). In this model each user i applies a differentially-private algorithm to their point z_i and then sends the result to the server. The specific algorithm applied by each user is determined by the server. In the general version of the model the server can determine which LDP algorithm the user should apply on the basis of all the previous communications the server has received. Yet multi-round protocols are prohibitively slow in practice due to network latency and, as a result, currently deployed large-scale systems are limited to non-interactive protocols (Erlingsson et al., 2014; Apple's Differential Privacy Team, 2017; Ding et al., 2017). Despite the major practical importance of LDP and significant interest in the research community, very little is known about which learning problems can be solved by such non-interactive systems. In particular, it is unknown whether many standard learning algorithms have private and non-interactive counterparts.

The power of interaction in distributed private learning was first addressed by Kasiviswanathan et al. (2011) who demonstrated existence of an artificial class of Boolean functions C over $\{0,1\}^d$ with the following property. C can be PAC learned efficiently relative to the uniform distribution over $\{0,1\}^d$ by an interactive LDP protocol but requires $2^{\Omega(d)}$ samples to learn by any non-interactive learning algorithm. Their techniques does not extend to learning without distributional assumptions. Deriving a technique that applies to distribution independent learning was one of the main open problems in this area (Kasiviswanathan et al., 2011).

Recent work by the authors (Daniely and Feldman, 2018) demonstrates that sample complexity of non-interactive LDP learning of a class of functions C in the distribution independent PAC model is lower bounded by the margin complexity of C. In particular, classes such as decision lists and linear classifiers over $\{0,1\}^d$ require an exponential number of samples to learn non-interactively even though they are known to be learnable efficiently by interactive algorithms. Here we ask whether every class that has polynomial margin complexity can be learned efficiently by a non-interactive LDP algorithm. Learning of large-margin classifiers is a classical learning problem and various algorithms for the problem are widely used in practice. Hence answering our open problem

will elucidate the feasibility of solving this problem with differential privacy in practice. Partial progress towards answering this question is given by the authors (Daniely and Feldman, 2018) who show that there exists an efficient LDP learning algorithm for large-margin linear classifiers whose interactive queries only use unlabeled samples.

Other related work: Smith et al. (2017) address the question of the power of non-interactive LDP algorithms in the closely related setting of stochastic convex optimization. They derive new non-interactive LDP algorithms for the problem albeit requiring an exponential dependence in the dimension number of samples. They also give a strong lower bound for non-interactive algorithms that are further restricted to obtain only local information about the optimized function. Subsequently, upper and lower bounds on the number of queries to the gradient/second-order oracles for algorithms with few rounds of interaction have been studied by several groups. Acharya et al. (2018) implicitly give a separation between interactive and non-interactive protocols for the problem of identity testing for a discrete distribution over k elements, albeit a relatively weak one ($O(k)$ vs $\Omega(k^{3/2})$ samples). A very recent work of Joseph et al. (2019) explores a different aspect of interactivity in LDP. Specifically, they distinguish between two types of interactive protocols: fully-interactive and sequentially-interactive ones. Fully-interactive protocols place no restrictions on interaction whereas sequentially-interactive ones only allows asking one query per user. They give a separation showing that sequentially-interactive protocols may require more samples than fully interactive ones (although only by a polynomial factor). This separation is not relevant to the open problem as we are interested in the power of completely non-interactive protocols. There are also a number of lower bounds on the sample complexity of LDP algorithms demonstrating that LDP is less efficient than the central model of differential privacy (*e.g.* Duchi et al. (2013); Duchi and Rogers (2019)).

2. Formal definitions and problem statement

Local Differential Privacy: In the local differential privacy (LDP) model (Warner, 1965; Evfimievski et al., 2003; Kasiviswanathan et al., 2011) it is assumed that each data sample obtained by the server is randomized in a differentially private way. This is modeled by assuming that the server running the learning algorithm accesses the dataset via an oracle defined below.

Definition 2.1 *An ϵ-local randomizer $R : Z \to W$ is a randomized algorithm that satisfies $\forall z_1, z_2 \in Z$ and $w \in W$, $\mathbf{Pr}[R(z_1) = w] \leq e^{\epsilon} \mathbf{Pr}[R(z_2) = w]$. For a dataset $S \in Z^n$, an LR_S oracle takes as an input an index i and a local randomizer R and outputs a random value w obtained by applying $R(z_i)$. An algorithm is ϵ-LDP if it accesses S only via the LR_S oracle with the following restriction: for all $i \in [n]$, if $\mathrm{LR}_S(i, R_1), \ldots, \mathrm{LR}_S(i, R_k)$ are the algorithm's invocations of LR_S on index i where each R_j is an ϵ_j-randomizer then $\sum_{j \in [k]} \epsilon_j \leq \epsilon$.*

For a non-interactive LDP algorithm one can assume without loss of generality that each sample is queried only once since the application of k fixed local randomizers with $\sum_{j \in [k]} \epsilon_j \leq \epsilon$ can be seen as an execution of a single ϵ-randomizer. This model can be contrasted with the standard, or central, model of differential privacy where the entire dataset is held by the learning algorithm whose output needs to satisfy differential privacy (Dwork et al., 2006). This is a stronger model and an ϵ-LPD algorithm also satisfies ϵ-differential privacy. For an algorithm in this model we say that the algorithm is *non-interactive* (or *non-adaptive*) if all its queries are determined before observing any of the oracle's responses.

PAC Learning and margin complexity: Our results are for the standard realizable PAC model of learning (Valiant, 1984).

Definition 2.2 *Let X be a domain and C be a class of Boolean functions over X. An algorithm $\mathcal{A}$ is said to PAC learn C with error α if for every distribution D over X and $f \in C$, given access (via oracle or samples) to the input distribution over examples $(x, f(x))$ for $x \sim D$, the algorithm outputs a function h such that $\mathbf{Pr}_D[f(x) \neq h(x)] \leq \alpha$ with probability at least $2/3$.*

We say that the learning algorithm is efficient if its running time is polynomial in $\log |X|$, $\log |C|$ and $1/\epsilon$.

We say that a class of Boolean ($\{-1, 1\}$-valued) functions C is closed under negation if for every $f \in C$, $-f \in C$. For dimension d, we denote by $\mathcal{B}^d(1)$ the unit ball in ℓ_2 norm in $\mathbb{R}^d$.

Definition 2.3 *Let X be a domain and C be a class of Boolean functions over X. The* margin complexity *of C, denoted $\mathsf{MC}(C)$, is the minimal number $M \geq 0$ such that for some d, there is an embedding $\Psi : X \to \mathcal{B}^d(1)$ for which the following holds: for every $f \in C$ there is $w \in \mathcal{B}^d(1)$ such that*

$$\min_{x \in X}\{f(x) \cdot \langle w, \Psi(x)\rangle\} \geq \frac{1}{M}.$$

The authors prove the following lower bound for non-interactive LDP learning of any class of Boolean functions closed under negation (Daniely and Feldman, 2018).

Theorem 2.4 *Let C be a class of Boolean functions closed under negation. Assume that there exists a non-interactive 1-LDP algorithm $\mathcal{A}$ that, with success probability at least $2/3$, PAC learns C distribution-independently with error less than $1/2$ using at most n examples. Then $n = \Omega(\mathsf{MC}(C)^{2/3})$.*

Our open problem asks whether the converse of Thm. 2.4 is true (up to a polynomial).

Open Problem 2.5 *Let C be a class of Boolean functions over some domain X. Does there exists a non-interactive 1-LDP algorithm that PAC learns C distribution-independently with error α and success-probability at least $2/3$ using $n = \mathrm{poly}(\mathsf{MC}(C)/\alpha)$ examples.*

Equivalence to statistical queries: For completeness we mention that the key tool for understanding the power of LDP protocols is the statistical query model of Kearns (1998) that is defined by having access to $\mathrm{STAT}_P(\tau)$ oracle, where P is the unknown data distribution.

Definition 2.6 *Let P be a distribution over a domain Z and $\tau > 0$. A statistical query oracle $STAT_P(\tau)$ is an oracle that given as input any function $\phi \colon Z \to [-1, 1]$, returns some value v such that $|v - \mathbf{E}_{z \sim P}[\phi(z)]| \leq \tau$.*

To solve a learning problem in this model an algorithm needs to succeed for any valid (that is satisfying the guarantees on the tolerance) oracle's responses. In other words, the guarantees of the algorithm should hold in the worst case over the responses of the oracle.

Kasiviswanathan et al. (2011) show that one can simulate $\mathrm{STAT}_P(\tau)$ oracle with success probability $1 - \delta$ by an ϵ-LDP algorithm using LR_S oracle for S containing $n = O(\log(1/\delta)/(\epsilon\tau)^2)$ i.i.d. samples from P. They also prove the converse of this simulation. Specifically, any query of an ϵ-LPD algorithm can be simulated (within total variation distance δ) using $O(1)$ queries to

$\mathrm{STAT}_P(\tau)$ for $\tau = \Theta(\delta/e^{2\epsilon})$. These simulations are interactive if and only if the original algorithm was interactive. Hence one can equivalently formulate our open problem as a problem about the power of non-adaptive SQ algorithms. This formulations also allows to derive corollaries for other models that are known to be equivalent to the SQ model (see (Daniely and Feldman, 2018) for additional details).

References

Jayadev Acharya, Clément L Canonne, and Himanshu Tyagi. Inference under information constraints i: Lower bounds from chi-square contraction. *arXiv preprint arXiv:1812.11476*, 2018.

Apple's Differential Privacy Team. Learning with privacy at scale. *Apple Machine Learning Journal*, 1(9), December 2017.

Amit Daniely and Vitaly Feldman. Learning without interaction requires separation. *arXiv preprint arXiv:1809.09165*, 2018.

Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *31st Conference on Neural Information Processing Systems (NIPS)*, pages 3574–3583, 2017.

John Duchi and Ryan Rogers. Lower bounds for locally private estimation via communication complexity. *arXiv preprint arXiv:1902.00582*, 2019.

John C. Duchi, Martin J. Wainwright, and Michael I. Jordan. Local privacy and minimax bounds: Sharp rates for probability estimation. In *NIPS*, pages 1529–1537, 2013.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: randomized aggregatable privacy-preserving ordinal response. In *ACM SIGSAC Conference on Computer and Communications Security*, pages 1054–1067, 2014.

Alexandre V. Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant. Limiting privacy breaches in privacy preserving data mining. In *PODS*, pages 211–222, 2003.

Matthew Joseph, Jieming Mao, Seth Neel, and Aaron Roth. The role of interactivity in local differential privacy. *CoRR*, abs/1904.03564, 2019. URL http://arxiv.org/abs/1904.03564.

Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM J. Comput.*, 40(3):793–826, June 2011.

M. Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM*, 45(6): 983–1006, 1998.

Adam D. Smith, Abhradeep Thakurta, and Jalaj Upadhyay. Is interaction necessary for distributed private learning? In *2017 IEEE Symposium on Security and Privacy, SP 2017*, pages 58–77, 2017.

L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *J. of the American Statistical Association*, 60(309):63–69, 1965.