Summary of Paper: Privacy Aware Learning

Presenter: Joseph Tobin

Department of Computer Science, University of Virginia

https://qdata.github.io/deep2Read/

Privacy Aware Learning

- Authors: John C. Duchi, Michael I. Jordan, Martin J. Wainwright

We study statistical risk minimization problems under a version of privacy in which the data is kept confidential even from the learner. In this local privacy framework, we establish sharp upper and lower bounds on the convergence rates of statistical estimation procedures. As a consequence, we exhibit a precise trade-off between the amount of privacy the data preserves and the utility, measured by convergence rate, of any statistical estimator.

Differential Privacy: Brief Overview

- Motivated by “Netflix problem”
  - Data was “anonymized” so researchers could compete to create a better recommendation algorithm

Definition 1. *A randomized function $\mathcal{K}$ gives ϵ-*differential privacy *if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq Range(\mathcal{K})$,*

$$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon) \times \Pr[\mathcal{K}(D_2) \in S] \quad (1)$$

- In other words, one’s privacy should at most marginally increase as a result of participating in a database
- Differentially private algorithms have many desirable properties

Introduction

- Approach problem from statistical decision theory
- Formulate goals of a learning system in terms of (standard) loss function
- Given data X, we want to find parameters *Theta* that minimize l(X, *Theta*)
- This paper focuses on *local privacy*, where each X_i is kept private from our algorithm *M*
- Instead of operating on X_1, X_2, … X_N, we operate on perturbed samples Z_1, Z_2, … Z_N

Introduction

- Standard measure of privacy is differential privacy
- Roughly states that parameters must not depend too much on *n* samples, and for any given vector *x*, it should be hard to tell whether *x* is in the set {X_1, … X_n}
- A method has alpha-differentiable privacy if

$$\sup_{S \in \sigma(\Theta)} \sup_{x_1,\ldots,x_n} \sup_{x'_1,\ldots,x'_n} \frac{Q(S \mid X_1 = x_1, \ldots, X_n = x_n)}{Q(S \mid X_1 = x'_1, \ldots, X_n = x'_n)} \leq \exp(\alpha).$$

Main Results

- Treat privacy preservation as a game between those protecting privacy and “nature”
- This method uses mutual information I(Z_i, X_i), where we say the “distribution Q generating Z from X is private only if I(X,Z) is small for all possible distributions P on X”
  - Cannot get very little information about X from Z no matter what the distribution X is sampled from
  - I(Z_i, X_i) is expected Kullback-Leibler (KL) divergence

Main Results

- For a given privacy level I*, sharp upper and lower bounds for convergence are developed to guarantee a level of privacy such that I(X_i, Z_i) <= I*
- For these bounds, they show we need problem-dependent constants
- Stochastic gradient is one of a potential class of procedures that achieves optimal convergence rates

Finding a good generating distribution (Q)

- Want to balance between accuracy and privacy
- Create a ball around point whose radius is dependent on the gradient dl(x, Theta) at that point.
- Want to find a saddle point (P*, Q*) such that

$$\sup_P I(P, Q^*) \leq I(P^*, Q^*) \leq \inf_Q I(P^*, Q),$$

Finding a good generating distribution (Q)

Definition 1. *The conditional distribution Q^* satisfies* optimal local privacy *for the sets $C \subset D \subset \mathbb{R}^d$ at level I^* if*

$$\sup_P I(P, Q^*) = \inf_Q \sup_P I(P, Q) = I^*,$$

where the supremum is taken over distributions $P \in \mathcal{P}(C)$ and the infimum is taken over regular conditional distributions $Q \in \mathcal{Q}(C, D)$.

- If Q* satisfies optimal local privacy, then it guarantees that even for the worst possible distribution P on X, the information communicated about X is limited

Definitions for main theorem

- l measures performance of Theta on X,

$$R(\theta) := \mathbb{E}[\ell(X, \theta)].$$

$$\epsilon_n(\mathcal{M}, \ell, \Theta, P) := R(\theta_n) - \inf_{\theta \in \Theta} R(\theta) = \mathbb{E}_P[\ell(X, \theta_n)] - \inf_{\theta \in \Theta} \mathbb{E}_P[\ell(X, \theta)].$$

$$\epsilon_n^*(\mathfrak{L}, \Theta) := \inf_{\mathcal{M}} \sup_{\ell \in \mathfrak{L}(P), P} \mathbb{E}_{P,Q}[\epsilon_n(\mathcal{M}, \ell, \Theta, P)],$$

Final Error Bounds

Corollary 1. *Let the conditions of Theorem 1(b) hold, and assume that* $M_\infty \geq 2L$. *Assume* Q^* *satisfies optimal local privacy at information level* I^*. *For universal constants* $c \leq C$,

$$c \cdot \frac{rL\sqrt{d \log d}}{\sqrt{nI^*}} \leq \epsilon_n^*(\mathfrak{L}, \Theta) \leq C \cdot \frac{rL\sqrt{d \log d}}{\sqrt{nI^*}}.$$

Corollary 2. *Let the conditions of Theorem 2 hold and assume that* $M_1 \geq 2L$. *Assume that* Q^* *satisfies optimal local privacy at information level* I^*. *For universal constants* $c \leq C$,

$$c \cdot \frac{rLd}{\sqrt{nI^*}} \leq \epsilon_n^*(\mathfrak{L}, \Theta) \leq C \cdot \frac{rLd}{\sqrt{nI^*}}.$$

Conclusion

- Obtained sharp tradeoffs between privacy protection and estimation rates
- Open question: are there restrictions on the class of loss functions where using $(Z_1, \ldots Z_n)$ is sufficient for inference?
- Future work: consideration of alternate restrictions

References

- http://link.springer.com/chapter/10.1007/978-3-540-79228-4_1
- http://dl.acm.org/citation.cfm?id=2180305
- http://people.csail.mit.edu/asmith/PS/sensitivity-tcc-final.pdf
- https://arxiv.org/pdf/1210.2085v2.pdf
- http://stanford.edu/~jduchi/projects/DuchiJoWa12_nips.pdf
- https://research.neustar.biz/2014/09/08/differential-privacy-the-basics/