Accuracy First: Selecting a DP Level for Accurate ERM

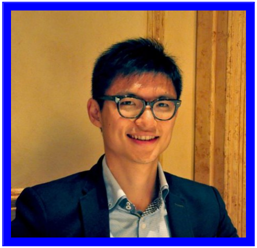
BIRS 2018, NIPS 2017, TPDP 2017

Seth V. Neel

May 3, 2018

Authors

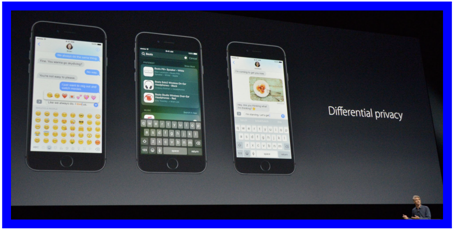





Seth Neel, Bo Waggoner, Katrina Ligett, Steven Wu, Aaron Roth

Motivation

- After over a decade of intense study, DP is beginning to see large scale deployments by companies like Apple and Google.



- ERM is the core task in machine learning
- Privacy is a priority, but absent regulation, accuracy is likely the first order concern
- **Natural question:** *Subject to a given accuracy level, what is the best privacy level one can obtain?*

From theory to practice

Theorem (Generic ML privacy theorem)

When run with privacy level ϵ, Alg achieves accuracy α.

From theory to practice

Theorem (Generic ML privacy theorem)

When run with privacy level ϵ, Alg achieves accuracy α.

e.g., $\alpha = \frac{10000}{\epsilon}$.

From theory to practice

Theorem (Generic ML privacy theorem)

When run with privacy level ϵ, Alg achieves accuracy α.

e.g., $\alpha = \frac{10000}{\epsilon}$.

e.g., $\alpha = O\left(\frac{1}{\epsilon}\right)$.

From theory to practice

Theorem (Generic ML privacy theorem)

When run with privacy level ϵ, Alg achieves accuracy α.

e.g., $\alpha = \frac{10000}{\epsilon}$.

e.g., $\alpha = O\left(\frac{1}{\epsilon}\right)$.

Engineer at Data Corp: How should I choose ϵ?

From theory to practice

Theorem (Generic ML privacy theorem)

When run with privacy level ϵ, Alg achieves accuracy α.

e.g., $\alpha = \frac{10000}{\epsilon}$.

e.g., $\alpha = O\left(\frac{1}{\epsilon}\right)$.

Engineer at Data Corp: How should I choose ϵ?

What if accuracy is **critical** to the system?

This work[1]

Question

Given an accuracy requirement, can we run a learning algorithm **as privately as possible**?

Setting: empirical risk minimization.

Given data and a loss function, find an "accurate" hypothesis.

[1] *Accuracy First: Selecting a Differential Privacy Level for Accuracy-Constrained ERM.* Joint with Katrina Ligett, Seth Neel, Aaron Roth, and Z. Steven Wu. *NIPS*, 2017.

Private Accurate ERM

- Empirical risk function:

$$L(\theta, D) = \tfrac{1}{n} \sum_{i=1}^{n} \ell(\theta, (X_i, y_i)) + \frac{\lambda}{2} ||\theta||_2^2$$

- Let $\theta^* = \text{argmin}_{\theta \in C} L(\theta, D)$

- Given accuracy tolerance α, find the most private θ_{priv} :

$$L(\theta_{priv}, D) \leq L(\theta^*, D) + \alpha$$

Private ERM

- Many algorithms: output/objective/covariance perturbation, exponential mechanism, SGD [Koufogiannis 2017, Smith 2017, Williams 2010, Chaudhuri 2008, Bassily 2014]

- Accuracy guarantees: ϵ privacy $\implies f(\epsilon)$ accuracy

- Given accuracy α solve for $\epsilon = f^{-1}(\alpha)$

How to go beyond worst-case analysis?

Naive Search: Doubling...

- For $t \in [T]$ generate ϵ_t-private hypothesis θ_t
- Check privately if $L(\theta_t, D) \leq L(\theta^*, D) + \alpha$
  - if **yes**: **stop**, output $(\theta_1, \ldots, \theta_t)$
  - if **no:** double ϵ_t
- Final ex-post privacy loss is:
  (cost publishing $\{\theta_i\}_{i=1}^t$) + (cost checking accuracy $\{\theta_i\}_{i=1}^t$)

How to formalize the privacy guarantee?

Road Map

- Formalizes a notion of *ex-post* privacy: privacy loss is data-dependent

- Gives an ex-post analysis of the AboveThreshold algorithm with private queries

- Application to two private ERM algorithms

- Use of *gradual release* technique [Koufogiannis 2017] improves upon doubling method

Ex-post privacy loss

All outputs are private but some outputs of an algorithm may be more *private* than others. In Math:

Definition (ex-post privacy loss)

$$\mathsf{Loss}(o) = \max_{D,D':D\sim D'} \log \frac{P[\mathcal{A}(D)=o]}{P[\mathcal{A}(D')=o]}.$$

Ex-post DP

Definition (Ex-Post Differential Privacy)

We say that $\mathcal{A}$ satisfies $\mathcal{E}(o)$-*ex-post* differential privacy if for all $o \in \mathcal{O}$, $\text{Loss}(o) \leq \mathcal{E}(o)$.

- Related to the notion of privacy odometers [Rogers, Roth, Ullman, Vadhan 2016]
- Ex-post differential privacy has the same semantics as differential privacy, once the output of the mechanism is known: it bounds the log-likelihood ratio of the dataset being D vs. D', which controls how an adversary with an arbitrary prior on the two cases can update her posterior.

Our Approach

$$\overbrace{\underbrace{\{\theta_i\}_{i=1}^{t}}_{\text{publishing hypothesis}} + \underbrace{\{\theta_i\}_{i=1}^{t}}_{\text{checking accuracy}}}^{\text{privacy cost of search}}$$

1. To privately evaluate the error of each θ^t use AboveThreshold (Trick: Ex-post AboveThreshold)
2. Generate $\{\theta_i\}_{i=1}^{t}$ such that publishing any prefix $(\theta^1, \ldots \theta^k)$ released incurs only privacy loss ϵ_k (Trick: Noise Reduction)

Our framework: example

Our framework: example

1. Compute “true” output non-privately

True (non-private) θ

$$\begin{pmatrix} 1.0 \\ 4.0 \\ \dots \\ 2.0 \end{pmatrix}$$

Our framework: example

1. Compute "true" output non-privately
2. Use random walks to add noise to each coordinate

True (non-private) θ

$$\begin{pmatrix} 1.0 \\ \\ 4.0 \\ \\ \dots \\ \\ 2.0 \end{pmatrix}$$

Our framework: example

1. Compute “true” output non-privately
2. Use random walks to add noise to each coordinate

Our framework: example

1. Compute "true" output non-privately
2. Use random walks to add noise to each coordinate

Our framework: example

1. Compute "true" output non-privately
2. Use random walks to add noise to each coordinate
3. If not accurate enough, "rewind" the walks!

   use InteractiveAboveThreshold *to check accuracy*

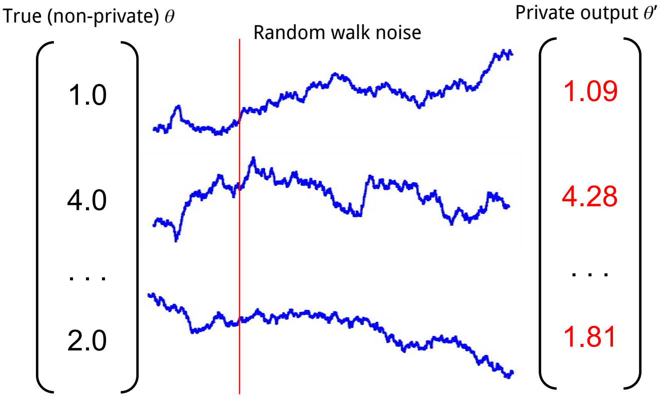
Our framework: example

1. Compute “true” output non-privately
2. Use random walks to add noise to each coordinate
3. If not accurate enough, “rewind” the walks!

   use InteractiveAboveThreshold *to check accuracy*

Our framework: example

1. Compute "true" output non-privately
2. Use random walks to add noise to each coordinate
3. If not accurate enough, "rewind" the walks!
   use InteractiveAboveThreshold *to check accuracy*



Ex-post Above Threshold I

- We want to publish the most private query $\theta_t \in \{\theta_i\}_{i=1}^{T}$ whose accuracy is above the threshold α
- Standard priv analysis: publish all the private queries and run AboveThreshold
- Intuitively, we want to generate and publish queries one at a time until the algorithm halts
- Pay only for the queries we publish: requires an *ex-post* analysis

---

Algorithm 2 InteractiveAboveThreshold: IAT$(D, \varepsilon, W, \Delta, M)$

Input: Dataset D, privacy loss ε, threshold W, ℓ_1 sensitivity Δ, algorithm M
Let $\hat{W} = W + \text{Lap}\left(\frac{2\Delta}{\varepsilon}\right)$
for each query $t = 1, \ldots, T$ **do**
  Query $f_t \leftarrow M(D)_t$
  if $f_t(D) + \text{Lap}\left(\frac{4\Delta}{\varepsilon}\right) \geq \hat{W}$: **then** Output (t, f_t); **Halt.**
Output $(T, \perp)$.

Ex-post Above Threshold II

Suppose that the prefix $\{f_1, \dots f_t\}$ is ϵ_t-differentially private. Then ex-post AT is $(\epsilon + \epsilon_t)$-ex-post differentially private.

Proof.

$$\frac{\Pr[\mathrm{IAT}(D) = t, f_1, \dots f_t]}{\Pr[\mathrm{IAT}(D') = t, f_1, \dots f_t]} = \frac{\Pr[\mathrm{IAT}(D) = t \mid f_1, \dots f_t]}{\Pr[\mathrm{IAT}(D') = t | f_1, \dots, f_t]} \frac{\Pr[M(D) = f_1, \dots f_t]}{\Pr[M(D') = f_1, \dots f_t]}$$
$$\leq e^{\varepsilon_A} \cdot e^{\varepsilon_t} = e^{\varepsilon_A + \varepsilon_t},$$

- $\epsilon_0 \approx O(\frac{\log(T/\gamma)}{\alpha n})$; ϵ_t data-dependent - can be much smaller!

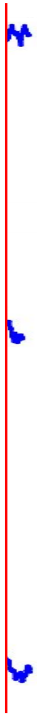
Intuition for privacy improvement

The **noisier** estimates reveal no private information conditioned on the **least noisy** one!

True (non-private) θ

$$\begin{pmatrix} ? \\ ? \\ \cdots \\ ? \end{pmatrix}$$

Random walk noise

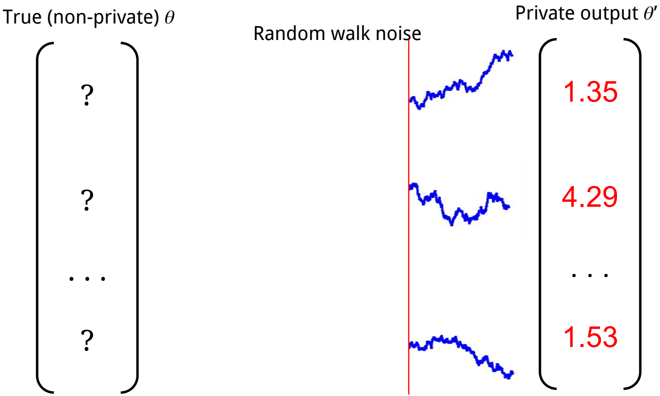


Private output θ'

$$\begin{pmatrix} 1.73 \\ 4.26 \\ \cdots \\ 1.41 \end{pmatrix}$$

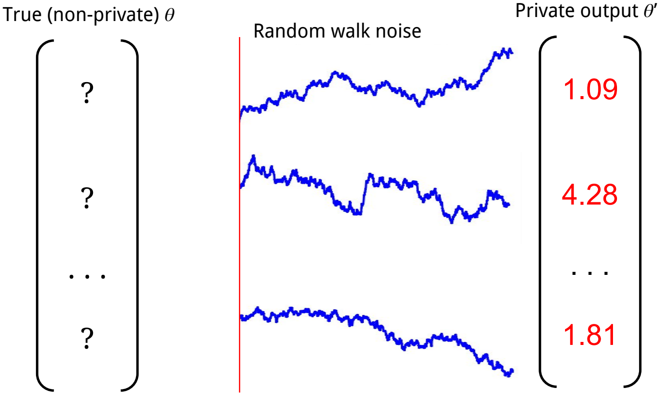
Intuition for privacy improvement

The **noisier** estimates reveal no private information conditioned on the **least noisy** one!



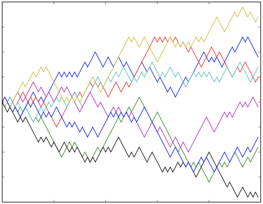
Intuition for privacy improvement

The **noisier** estimates reveal no private information conditioned on the **least noisy** one!



Noise Reduction [Koufogiannis 2017]

- Instead of generating private hypothesis $\{\theta_t\}$ independently via the Laplace Mechanism, use correlated noise technique
- Each θ_t is a post-processing of every θ_s, $s < t$
- Publishing the prefix $\{\theta_1, \dots \theta_t\}$ incurs only loss ϵ_t instead of $\sum_{s=1}^{t} \epsilon_s$, by post-processing



Gradual Private Release via Random Walk with Laplace Marginals

High-level paradigm

known algorithms for differentially-private learning

example above: output perturbation

⬇

InteractiveAboveThreshold (accuracy checks) and **NoiseReduction** (random-walk) techniques

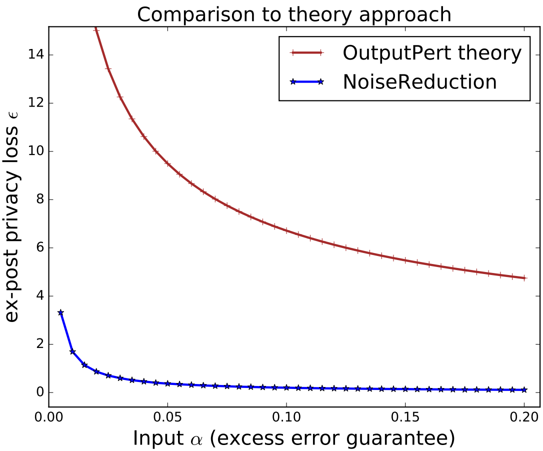
⬇

learning algorithms that are "as private as possible"

Experiments: vs using theorems

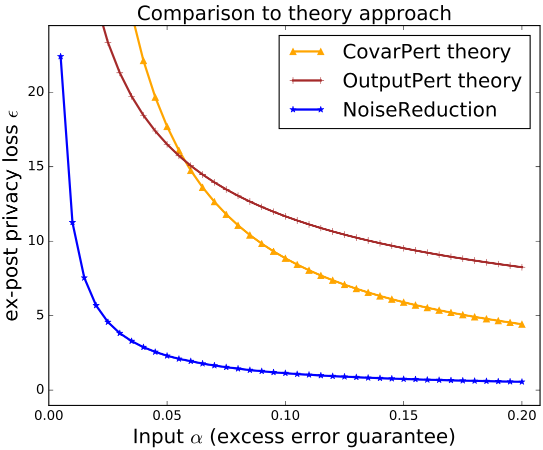
Experiments: vs using theorems

Logistic regression. Classify network activity in KDDCup99 dataset, n =100k.

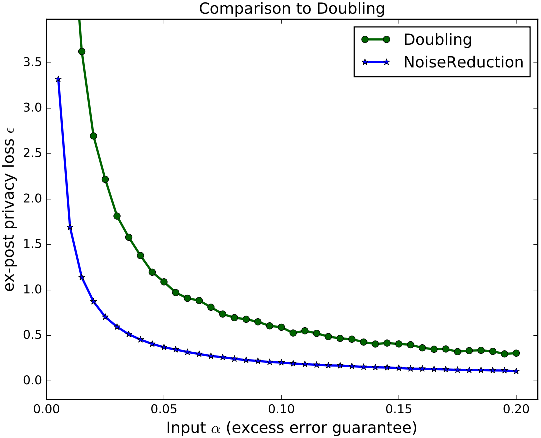


Experiments: vs using theorems (2)

Linear (ridge) regression. Predict $\log$(retweets) on Twitter dataset, $n =$100k.



Experiments: vs using Doubling



References

Privacy Odometers and Filters: Pay-as-you-Go Composition

Private Empirical Risk Minimization

Privacy-Preserving Logistic Regression

Gradual Release of sensitive data under differential privacy.

Is interaction necessary for distributed private learning?