Accuracy First: Selecting a DP Level for Accurate ERM

BIRS 2018, NIPS 2017, TPDP 2017

Seth V. Neel

May 3, 2018
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 $\epsilon$, Alg achieves accuracy $\alpha$. 

Engineer at Data Corp:

How should I choose $\epsilon$?

What if accuracy is critical to the system?
From theory to practice

Theorem (Generic ML privacy theorem)

When run with privacy level $\epsilon$, Alg achieves accuracy $\alpha$.

e.g., $\alpha = \frac{10000}{\epsilon}$. 
From theory to practice

Theorem (Generic ML privacy theorem)

*When run with privacy level $\epsilon$, Alg achieves accuracy $\alpha$. *

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

e.g., $\alpha = O \left( \frac{1}{\epsilon} \right)$.
Theorem (Generic ML privacy theorem)

When run with privacy level $\epsilon$, Alg achieves accuracy $\alpha$.

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

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

Engineer at Data Corp: How should I choose $\epsilon$?
From theory to practice

Theorem (Generic ML privacy theorem)

When run with privacy level $\epsilon$, Alg achieves accuracy $\alpha$.

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

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

Engineer at Data Corp: How should I choose $\epsilon$?

What if accuracy is critical to the system?
This work

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:

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

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

- Given accuracy tolerance \( \alpha \), find the most private \( \theta_{\text{priv}} \):

\[
L(\theta_{\text{priv}}, D) \leq L(\theta^*, D) + \alpha
\]
Private ERM


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

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

How to go beyond worst-case analysis?
Naive Search: Doubling...

- For $t \in [T]$ generate $\epsilon_t$-private hypothesis $\theta_t$

- Check privately if $L(\theta_t, D) \leq L(\theta^*, D) + \alpha$
  - if yes: stop, output $(\theta_1, \ldots, \theta_t)$
  - if no: double $\epsilon_t$

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

**Definition (ex-post privacy loss)**

\[
\text{Loss}(o) = \max_{D, D': D \sim D'} \log \frac{P[A(D) = o]}{P[A(D') = o]}.
\]
We say that $A$ satisfies $E(o)$-ex-post differential privacy if for all $o \in O$, $\text{Loss}(o) \leq E(o)$.

- 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

To privately evaluate the error of each $\theta^t$ use AboveThreshold (Trick: Ex-post AboveThreshold)

Generate $\{\theta_i\}^t_{i=1}$ such that publishing any prefix $(\theta^1, \ldots, \theta^k)$ released incurs only privacy loss $\epsilon_k$ (Trick: Noise Reduction)
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!
4. Use InteractiveAboveThreshold to check accuracy
Our framework: example

1. Compute “true” output non-privately

True (non-private) $\theta$

\[
\begin{pmatrix}
1.0 \\
4.0 \\
\vdots \\
2.0
\end{pmatrix}
\]
Our framework: example

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

True (non-private) \( \theta \)

\[
\begin{pmatrix}
1.0 \\
4.0 \\
\ldots \\
2.0
\end{pmatrix}
\]
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*

---

**True (non-private) $\theta$**

```
1.0
4.0
...
2.0
```

**Random walk noise**

![Random walk noise chart]

**Private output $\theta'$**

```
1.35
4.29
...
1.53
```
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*
We want to publish the most private query $\theta_t \in \{\theta_i\}_{i=1}^T$ whose accuracy is above the threshold $\alpha$

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 \textit{ex-post} analysis

\[\textbf{Algorithm 2} \text{ InteractiveAboveThreshold: IAT}(D, \varepsilon, W, \Delta, M)\]

\begin{align*}
\textbf{Input:} & \text{ Dataset } D, \text{ privacy loss } \varepsilon, \text{ threshold } W, \ell_1 \text{ sensitivity } \Delta, \text{ algorithm } M \\
\text{Let } \hat{W} = W + \text{Lap}\left(\frac{2\Delta}{\varepsilon}\right) & \\
\text{for each query } t = 1, \ldots, T \text{ do} & \\
\quad \text{Query } f_t \leftarrow M(D)_t & \\
\quad \text{if } f_t(D) + \text{Lap}\left(\frac{4\Delta}{\varepsilon}\right) \geq \hat{W}: \text{ then Output } (t, f_t); \text{ Halt.} & \\
\text{Output } (T, \bot). & 
\end{align*}
Ex-post Above Threshold II

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

**Proof.**

\[
\begin{align*}
\Pr[\text{IAT}(D) = t, f_1, \ldots, f_t] & \leq \Pr[\text{IAT}(D) = t | f_1, \ldots, f_t] \Pr[M(D) = f_1, \ldots, f_t] \\
& \leq e^{\epsilon \Lambda} \cdot e^{\epsilon_t} = e^{\epsilon \Lambda + \epsilon_t},
\end{align*}
\]

- \( \epsilon_0 \approx O\left(\frac{\log(T/\gamma)}{\alpha n}\right) \); \( \epsilon_t \) 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) $\theta$

\[
\begin{pmatrix}
? \\
? \\
\ldots \\
? \\
\end{pmatrix}
\]

Random walk noise

\[
\text{Private output } \theta'
\]

\[
\begin{pmatrix}
1.73 \\
4.26 \\
\ldots \\
1.41 \\
\end{pmatrix}
\]
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!

True (non-private) $\theta$

\[
\begin{array}{c}
? \\
? \\
\ldots \\
? \\
\end{array}
\]

Random walk noise

Private output $\theta'$

\[
\begin{array}{c}
1.09 \\
4.28 \\
\ldots \\
1.81 \\
\end{array}
\]
Instead of generating private hypothesis \( \{\theta_t\} \) independently via the Laplace Mechanism, use correlated noise technique.

Each \( \theta_t \) is a post-processing of every \( \theta_s, s < t \).

Publishing the prefix \( \{\theta_1, \ldots, \theta_t\} \) incurs only loss \( \epsilon_t \) 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

Logistic regression.
Classify network activity in KDDCup99 dataset, n = 100k.
Experiments: vs using theorems

Logistic regression. Classify network activity in KDDCup99 dataset, \( n = 100k \).
Experiments: vs using theorems (2)

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

![Comparison to theory approach graph]

- CovarPert theory
- OutputPert theory
- NoiseReduction
Experiments: vs using Doubling

Comparison to Doubling

- Doubling
- NoiseReduction

Input $\alpha$ (excess error guarantee)

ex-post privacy loss $\epsilon$
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?