Accuracy First: Selecting a DP Level for Accurate ERM

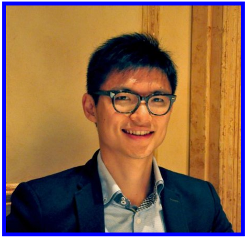
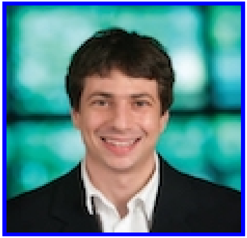
NIPS 2017, TPDP 2017

Seth V. Neel

October 26, 2017

Authors

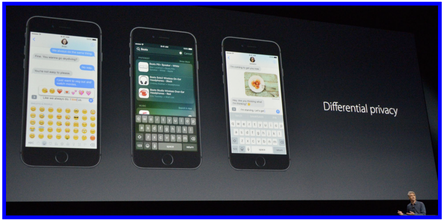


Seth Neel, Bo Waggoner, Katrina Ligett, Steven Wu, Aaron Roth

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?*

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 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?

This Paper

- Formalizes a notion of *ex-post* privacy
- 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

An adaptive definition of differential privacy...

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

Definition (ex-post privacy loss)

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

Ex-post differential privacy

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 2016] which analyzes compositions of private mechanisms with adaptive stopping time.

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)

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

Ex-post Above Threshold II

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

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

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

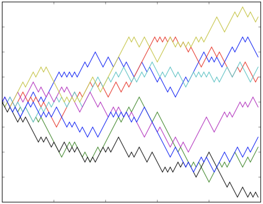
Proof.

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

□

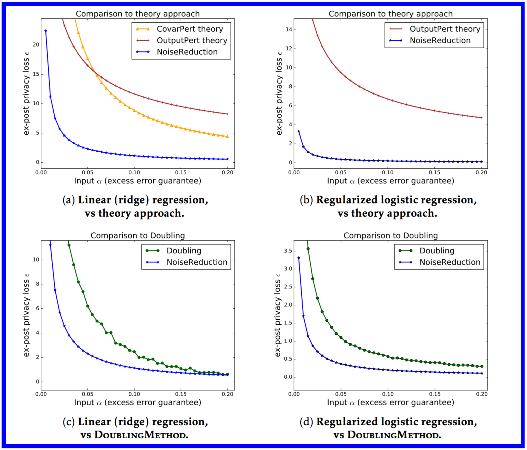
Noise Reduction [Koufogiannis 2017]

- 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 $\{\theta_1, \ldots \theta_t\}$ 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

Experiments



(a) **Linear (ridge) regression, vs theory approach.**

(b) **Regularized logistic regression, vs theory approach.**

(c) **Linear (ridge) regression, vs DoublingMethod.**

(d) **Regularized logistic regression, vs DoublingMethod.**

Datasets: (Twitter, $p = 77, n = 100,000$), (KDD-Cup99, $p = 28, n = 100,000$)

Some References I

Ryan Rogers, Aaron Roth, Jonathan Ullman, Salil Vadhan
Privacy Odometers and Filters: Pay-as-you-Go Composition
NIPS 2016.

Raef Bassily, Adam D. Smith, Abhradeep Thakurta
Private Empirical Risk Minimization
CoRR 2014.

Kamalika Chaudhuri and Claire Monteleoni
Privacy-Preserving Logistic Regression
NIPS 2008.

F. Koufogiannis, Shuo Han, and George J. Pappas.
Gradual Release of sensitive data under differential privacy.
Journal of Privacy and Confidentiality 2017.

A. Smith, J. Upadhyay, and A. Thakurta.
Is interaction necessary for distributed private learning?
IEEE Symposium on Security and Privacy 2017.

Some References II

O. Williams, F. McSherry.
Probabilistic inference and differential privacy?
NIPS 2010.