Private Stochastic Convex Optimization with Optimal Rate

Raef Bassily
Ohio State University

Vitaly Feldman
Google Brain

Kunal Talwar
Google Brain

Abhradeep Guha Thakaurta
UC Santa Cruz
Google Brain

This work

Differentially private (DP) algorithms for
stochastic convex optimization with **optimal excess population risk**

Stochastic Convex Optimization (SCO)

Unknown distribution (population) $\mathcal{D}$ over data universe $\mathcal{Z}$

Convex parameter space $\mathcal{C} \subset \mathbb{R}^d$

Convex loss function $\ell: \mathcal{C} \times \mathcal{Z} \to \mathbb{R}$

L_2/L_2 setting:
$\mathcal{C}$ and $\partial\ell$ are bounded in L_2-norm

Dataset $S = (z_1, \dots, z_n) \sim \mathcal{D}^n$

Stochastic Convex Optimization (SCO)

Unknown distribution (population) $\mathcal{D}$ over data universe $\mathcal{Z}$

Convex parameter space $\mathcal{C} \subset \mathbb{R}^d$

Convex loss function $\ell: \mathcal{C} \times \mathcal{Z} \to \mathbb{R}$

L_2/L_2 setting:
$\mathcal{C}$ and $\partial\ell$ are bounded in L_2-norm

Dataset $S = (z_1, \dots, z_n) \sim \mathcal{D}^n$

A SCO algorithm, given S, outputs $\hat{\theta} \in \mathcal{C}$ s.t.

$$\textbf{Excess Pop. Risk} \triangleq \mathbb{E}_{z\sim\mathcal{D}}\left[\ell\left(\hat{\theta}, z\right)\right] - \min_{\theta\in\mathcal{C}} \mathbb{E}_{z\sim\mathcal{D}}[\ell(\theta, z)]$$

is as small as possible

Well-studied problem: *optimal rate* $\approx \frac{1}{\sqrt{n}}$

Private *Stochastic Convex Optimization* (*P*SCO)

Unknown distribution (population) $\mathcal{D}$ over data universe $\mathcal{Z}$

Convex parameter space $\mathcal{C} \subset \mathbb{R}^d$

Convex loss function $\ell: \mathcal{C} \times \mathcal{Z} \rightarrow \mathbb{R}$

L_2/L_2 setting:
$\mathcal{C}$ and $\partial\ell$ are bounded in L_2-norm

Dataset $S = (z_1, \ldots, z_n) \sim \mathcal{D}^n$

Goal: (ϵ, δ)-DP algorithm $\mathcal{A}_{priv}$ that, given S, outputs $\hat{\theta} \in \mathcal{C}$ s.t.

Excess Pop. Risk $\triangleq \mathbb{E}_{z\sim\mathcal{D}}\left[\ell\left(\hat{\theta}, z\right)\right] - \min_{\theta\in\mathcal{C}} \mathbb{E}_{z\sim\mathcal{D}}[\ell(\theta, z)]$

is as small as possible

Main Result

Optimal excess population risk for PSCO is $\approx \max\left(\frac{1}{\sqrt{n}}, \frac{\sqrt{d}}{\epsilon\, n}\right)$

Optimal non-private population risk

Optimal private empirical risk [BST14]

Main Result

Optimal excess population risk for PSCO is $\approx \underbrace{\max\left(\frac{1}{\sqrt{n}}, \frac{\sqrt{d}}{\epsilon\, n}\right)}$

When $d = \Theta(n)$ (common in modern ML)

Opt. risk for **PSCO** $\approx \frac{1}{\sqrt{n}} =$ opt. risk for **SCO**

⇓

asymptotically no cost of privacy

Algorithms

Two algorithms under *mild smoothness assumption on* ℓ *:*

- A variant of **mini-batch noisy SGD:**

- **Objective Perturbation** (entails rank assumption on $\nabla^2 \ell$)

Algorithms

Two algorithms under *mild smoothness assumption on* ℓ *:*

- ➢ A variant of **mini-batch noisy SGD:**

- ➢ **Objective Perturbation** (entails rank assumption on $\nabla^2 \ell$)

- The objective function in **both** algorithms is the **empirical risk**.

Algorithms

Two algorithms under *mild smoothness assumption on ℓ :*

- A variant of **mini-batch noisy SGD:**

- **Objective Perturbation** (entails rank assumption on $\nabla^2 \ell$)

- The objective function in **both** algorithms is the **empirical risk**.
- **Generalization error** is bounded via **uniform stability**:
  - For the first algorithm: uniform stability of SGD [HRS15, **F**V19].
  - For the second algorithm: uniform stability due to regularization.

Algorithms

- *General non-smooth loss:*
  - A *new, efficient,* **noisy stochastic proximal gradient algorithm:**
    - Based on **Moreau-Yosida** *smoothing*
    - A **gradient step w.r.t. the *smoothed loss*** is equivalent to a **proximal step w.r.t. the *original loss*.**

Results vs. Prior Work on DP-ERM

This work

- Optimal excess population risk for PSCO is $\approx \max\left(\frac{1}{\sqrt{n}}, \frac{\sqrt{d}}{\epsilon\, n}\right)$

Previous work

- Focused on the empirical version (DP-ERM): [CMS11, KST12, BST14, TTZ15, …]
- Optimal **empirical** risk is previously known [BST14], but ***not*** optimal **population** risk.
- Best known **population risk** using *DP-ERM algorithms* $\approx \max\left(\frac{d^{1/4}}{\sqrt{n}}, \frac{\sqrt{d}}{\epsilon\, n}\right)$ [BST14].

Poster #163

Full version: arXiv:1908.09970