Adapting to Linear Separable Subsets with Large-Margin in Differentially Private Learning

Erchi Wang†, Yuqing Zhu‡, Yu-Xiang Wang†

†*University of California, San Diego* ‡*LinkedIn*

TL;DR: We propose an *efficient* (ε, δ)-DP algorithm for learning linear classifier that adapts to the *best linear separable subsets without separability assumption or margin information.*

Intro & Background

<u>(ε, δ)-DP:</u> An algorithm $\mathcal{M}: \mathcal{X}^* \to \mathcal{O}$ satisfies (ε, δ)-differential privacy if for any dataset X', X differ by at most one point, for any $O \subseteq \mathcal{O}$, $\Pr(\mathcal{M}(X) \in O) \le e^{\varepsilon} \Pr(\mathcal{M}(X') \in O) + \delta$

<u>Linear classifier and Margin:</u>

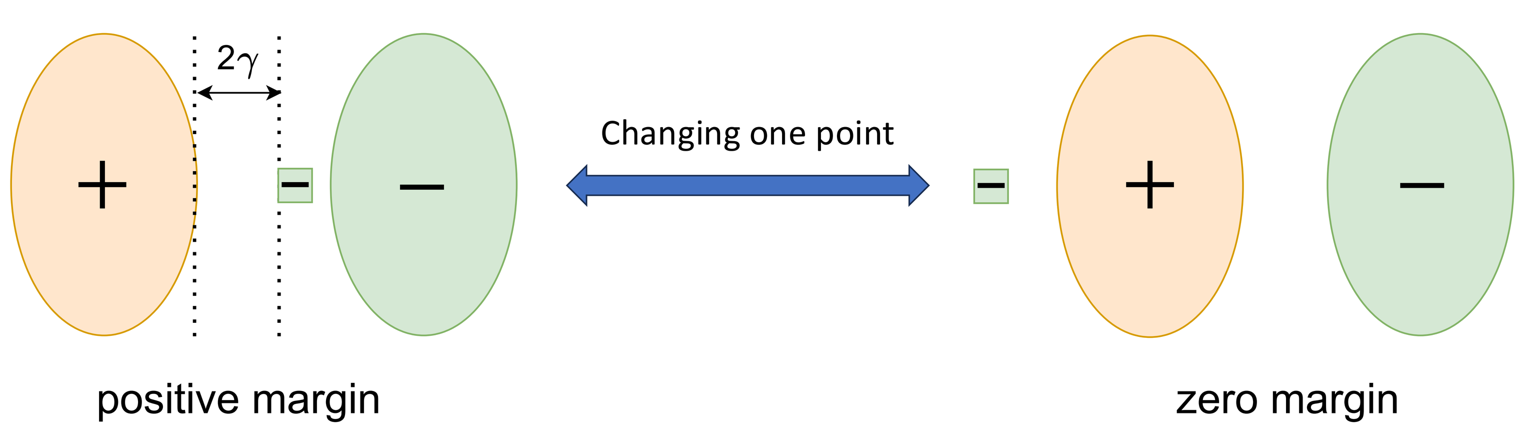
$h_w: (\boldsymbol{x}, y) \to \mathrm{sign}(y\langle \boldsymbol{w}, x\rangle)$ — Linear Classifier

$\mathrm{Margin}(h_w; S) = \max\left\{\min_{(x,y)\in S} \frac{y\langle \boldsymbol{w}, x\rangle}{\|\boldsymbol{w}\|}, 0\right\}$ — Margin for h_w

$\mathrm{Margin}(S) = \max_{\boldsymbol{w}} \mathrm{Margin}(h_w; S)$ — Margin for dataset S

Motivation

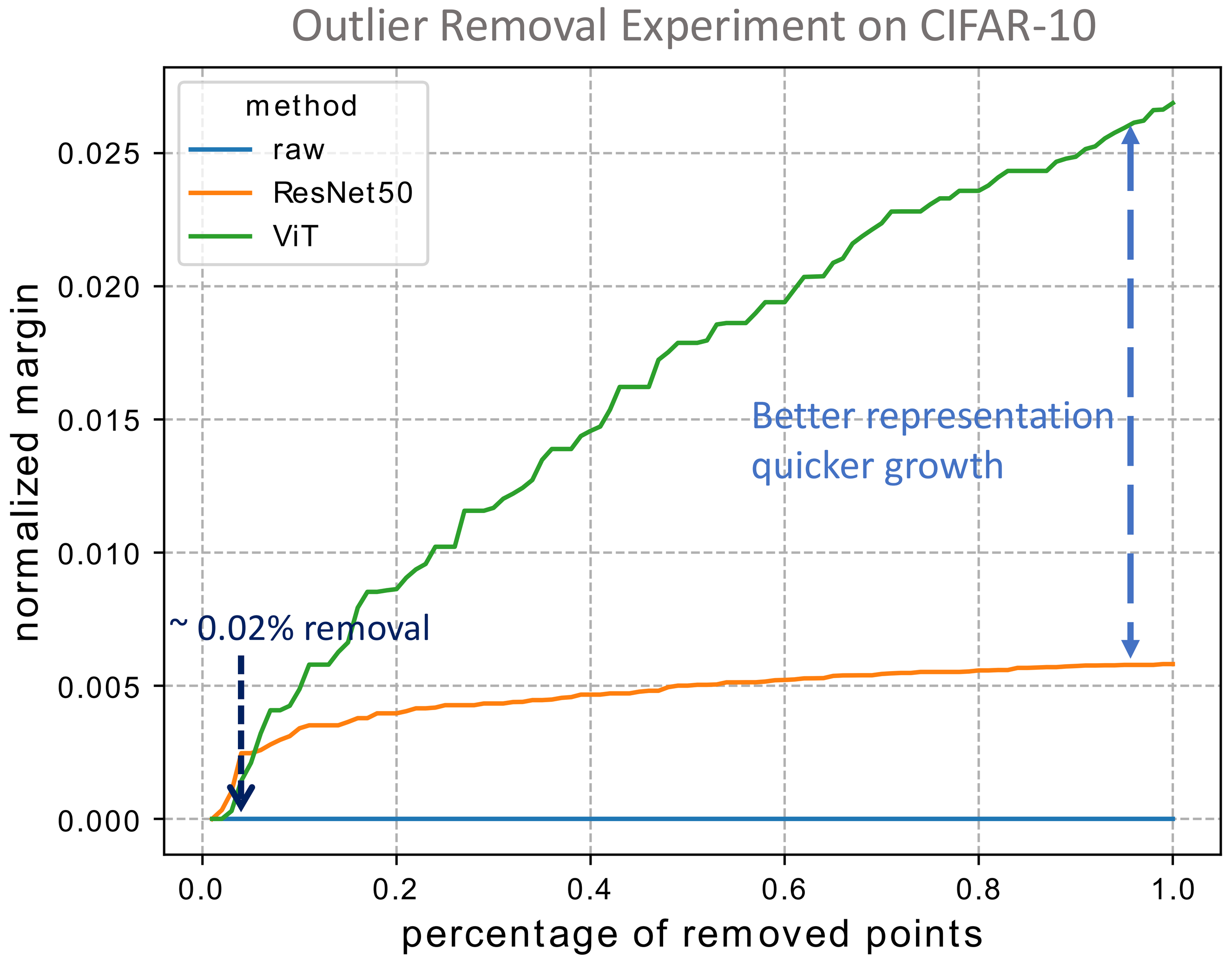
Margin is unstable: A single data point can significantly alter the margin



Hard for stability-based DP mechanism to work ☹

"Robust Margin": real-world data is not always linearly separable, but it *can become separable* after removing a few "outliers".

Outlier Removal Experiment on CIFAR-10



Question: Can we adapt to "robust margin" ?

Given dataset S and margin parameter $\gamma \in [0, 1]$

Margin Inliers $\mathcal{S}_{\text{in}}(\gamma) = \{S' \subseteq S \mid \mathrm{Margin}(S') \ge \gamma\}$

Margin Outliers $\mathcal{S}_{\text{out}}(\gamma) = \{S \backslash S' \mid S' \in \mathcal{S}_{\text{in}}(\gamma)\}$

- *Extension of the definition of Margin*

Main Result

There exists an *efficient* (ε, δ)-DP algorithm $\mathcal{M}^*$, for any input dataset S, privacy budgets ε, δ satisfying $\delta \in (0, 1)$ and $\varepsilon \in (0, 8\log(1/\delta))$, with high probability, for any $S_{\text{out}} \in 2^S$ with $\gamma := \gamma(S \setminus S_{\text{out}}) > 0$, simultaneously:

$$\tilde{\mathcal{R}}_S(\mathcal{M}^*(S, \varepsilon, \delta)) \le \tilde{\mathcal{O}}\left(\frac{1}{n\gamma^2 \min\{\varepsilon, 1\}} + \frac{|S_{\text{out}}|}{\gamma n}\right) \wedge 1$$

$$\mathcal{R}_D(\mathcal{M}^*(S, \varepsilon, \delta)) \le \tilde{\mathcal{O}}\left(\frac{1}{n\gamma^2 \min\{\varepsilon, 1\}} + \frac{|S_{\text{out}}|}{\gamma n}\right) \wedge 1$$

$\tilde{\mathcal{R}}_S(\cdot)$ averaged empirical zero-one loss $\mathcal{R}_D(\cdot)$ population zero-one loss

- *Doesn't need data to be separable*
- *Dimension-independent rate*
- *No needing to know which S_{out} to remove*

Comparison with Related Work (Population 0-1 Risk)

Source	Realizable case	Agnostic case	poly-time?
[NUZ20] Thm. 6, Thm. 11	$\frac{1}{n\gamma^2\varepsilon}$	NA	✓
[BMS22] Thm. 3.1	$\frac{1}{n\gamma^2\varepsilon}$	$\frac{1}{n\gamma^2\varepsilon} + \min_{w\in\mathcal{B}^d(1)}\left(\tilde{\mathcal{R}}_S^\gamma(w) + \sqrt{\left(\frac{1}{n^2\gamma^2} + \frac{1}{n}\right)\cdot \tilde{\mathcal{R}}_S^\gamma(w)}\right)$	✗
[BMS22] Thm. 3.2	$\frac{1}{n^{1/2}\gamma\varepsilon^{1/2}}$	$\frac{1}{n^{1/2}\gamma\varepsilon^{1/2}} + \min_{w\in\mathcal{B}^d(1)} \tilde{L}_S^\gamma(w)$	✓
Ours	$\frac{1}{n\gamma^2\varepsilon}$	$\min_{\substack{S_{\text{out}}\subset S \\ \gamma:=\text{margin}(S\backslash S_{\text{out}})}} \left(\frac{\|S_{\text{out}}\|}{n\gamma} + \frac{1}{n\gamma^2\varepsilon}\right)$	✓

(Logarithmic factors are ignored)

- Recover realizable case in [NUZ20]
- $\sqrt{n}$ improvement over Thm 3.2 [BMS22] if $|S_{\text{out}}| = o(\sqrt{n})$

<u>Reference:</u>

[NUZ20] Lê Nguyễn, Huy, Jonathan Ullman, and Lydia Zakynthinou. "Efficient private algorithms for learning large-margin halfspaces." ALT 2020

[BMS22] Bassily, Raef, Mehryar Mohri, and Ananda Theertha Suresh. "Differentially private learning with margin guarantees." Neurips 2022

Algorithms

Main algorithm DP Adaptive Margin consists of two parts:

- *Base Algorithm $\mathcal{A}_{JLGD}$*
  Random projection based noisy GD [NUZ20, BMS22]
- *Hyperparameter selector $\mathcal{A}_{Iter}$*
  (1) Run $\mathcal{A}_{JLGD}$ for each hyperparameter configuration
  (2) report the best configuration through noisy zero-one loss

DP Adaptive Margin $\mathcal{M}^*(S, \varepsilon, \delta)$

Input: dataset $S = \{\mathbf{x}_i, y_i\}_{i=1}^n$, Privacy budget ε, δ
Set: Margin grid $\Gamma = \{\frac{1}{n}, \frac{2}{n}, \frac{4}{n}, ..., \frac{2^{\lfloor \log_2 n\rfloor}}{n}, 1\}$,
  GDP budget $\mu = \frac{\varepsilon}{2\sqrt{2\log(1/\delta)}}$,
  failure probability for JL projection β
Initialize hyperparameter set $\Theta = \{\phi\}$ ▷ *empty set*
for $\gamma \in \Gamma$ **do**
  $k_\gamma = \mathcal{O}\left(\frac{1}{\gamma^2}\log(\frac{|\Gamma|(n+2)(n+1)}{\beta})\right)$
  $\Phi_\gamma \sim (\mathrm{Rad}(\frac{1}{2})/\sqrt{k_\gamma})^{k_\gamma \times d}$
  $\Theta = \Theta \cup \{(\gamma, \Phi_\gamma)\}$
$(\tilde{\mathbf{w}}_{\text{out}}, \gamma_{\text{out}}, \Phi_{\gamma_{\text{out}}}) = \mathcal{A}_{\text{Iter}}(\mathcal{A}_{\text{JLGD}}(\cdot), \Theta, S, \mu)$
Output: $(\tilde{\mathbf{w}}_{\text{out}}, \gamma_{\text{out}}, \Phi_{\gamma_{\text{out}}})$

Proof Sketch

TL;DR *JL projection reduces dimensionality, and hyperparameter search adapts to the margin parameter, which is coupled with the reduced dimension.*

- JL Projection preserves margin with high probability
- Error decomposition using margin Inlier/Outliers

$$\hat{L}_c(\mathbf{w}; S) \le \hat{L}_c(\mathbf{w}^*; S) + \text{EER} \quad \leftarrow \text{Excess empirical risk}$$

$$\le \underbrace{\hat{L}_c(\mathring{\mathbf{w}}_{\text{in}}; S_{\text{in}}(\gamma))}_{\text{(a) zero if } c \le \gamma} + \underbrace{\hat{L}_c(\mathring{\mathbf{w}}_{\text{in}}; S_{\text{out}}(\gamma))}_{(b)\le\|\mathbf{x}\|\cdot|S_{\text{out}}(\gamma)|/c} + \text{EER} \quad (\mathring{\mathbf{w}}_{\text{in}}\text{: Separator on } S_{\text{in}})$$

$$\le \mathcal{O}(|S_{\text{out}}(\gamma)|/c) + \text{EER}$$

- Adapting to unknown margin via geometric grid

Construct geometric grid $\Gamma = \{\frac{1}{n}, \frac{2}{n}, ..., \frac{2^{\lfloor \log_2 n\rfloor}}{n}, 1\}$:

$$\tilde{\mathcal{R}}_S(\mathcal{M}^*(\varepsilon, \delta, S)) \lesssim \min_{\substack{\gamma\in\Gamma \\ S_{\text{out}}\in\mathcal{S}_{\text{out}}(\gamma)}} \left(\frac{|S_{\text{out}}|}{n\gamma} + \frac{1}{n\gamma^2\varepsilon}\right)$$

$$\le \min_{\substack{S_{\text{out}}\subset S \\ \gamma:=\gamma(S\backslash S_{\text{out}})>0}} \mathcal{O}\left(\frac{|S_{\text{out}}|}{n\gamma} + \frac{1}{n\gamma^2\varepsilon}\right)$$