Google Research LabelDP-Pro: Learning with Label Differential Privacy via Projections

PRINCETON UNIVERSITY

Badih Ghazi[†], Yangsibo Huang[†‡], Pritish Kamath[†], Ravi Kumar[†], Pasin Manurangsi[†], Chiyuan Zhang[†]

[†]Google Research, [‡]Princeton University

Background & motivation

Differential privacy (DP; Dwork et al., 2006) gives formal privacy guarantee

- Privacy budget ε - lower value: higher privacy guarantee

Label differential privacy (Label DP; Chaudhuri & Hsu, 2011)

- In certain applications, x (input) is public, only y (label) has privacy concern
- Example: online advertising

Challenge: achieving Label DP in **high privacy regime** (i.e., low ε) with **good utility**

- Randomized response (RR; Warner, 1965) achieves label DP but with a huge utility loss

User	Private label
Alice	1
Bob	0
Charlie	0
…	…

Label for Alice? → RR

1, ✅ $\frac{e^\epsilon}{e^\epsilon + K - 1}$ $(K = 2)$

0, ❌ Otherwise

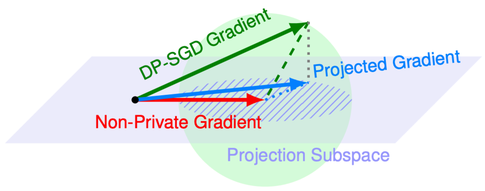
ε=1, RR returns a wrong label in 54% cases!

Main contributions

- We propose **LabelDP-Pro**, a new family of methods which achieves a much better privacy-utility tradeoff for labelDP in high privacy regime
- LabelDP-Pro improves the state-of-the-art for LabelDP training in the high-privacy regime on 4 benchmarks & for user-level DP applications
- Theoretical justification of LabelDP-Pro (in paper)

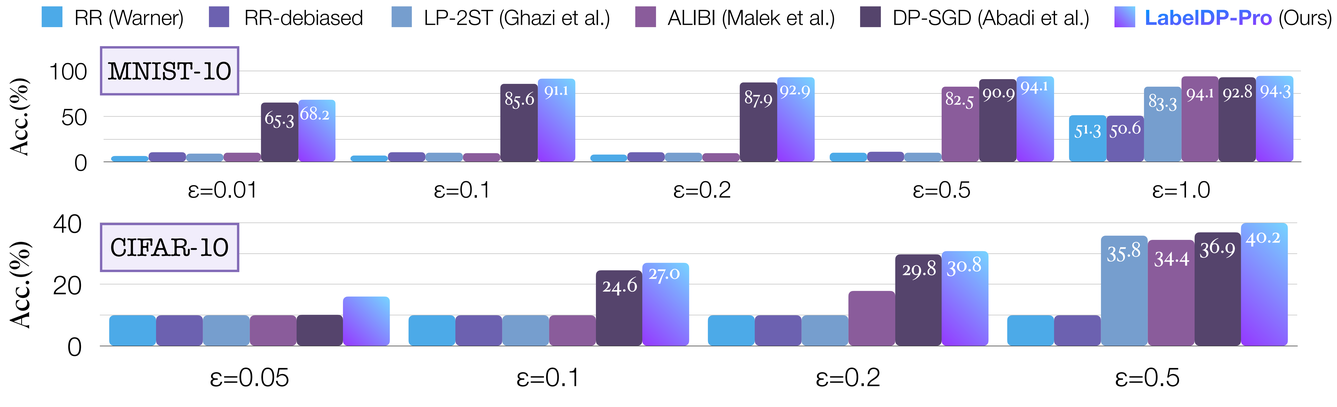
Our proposal

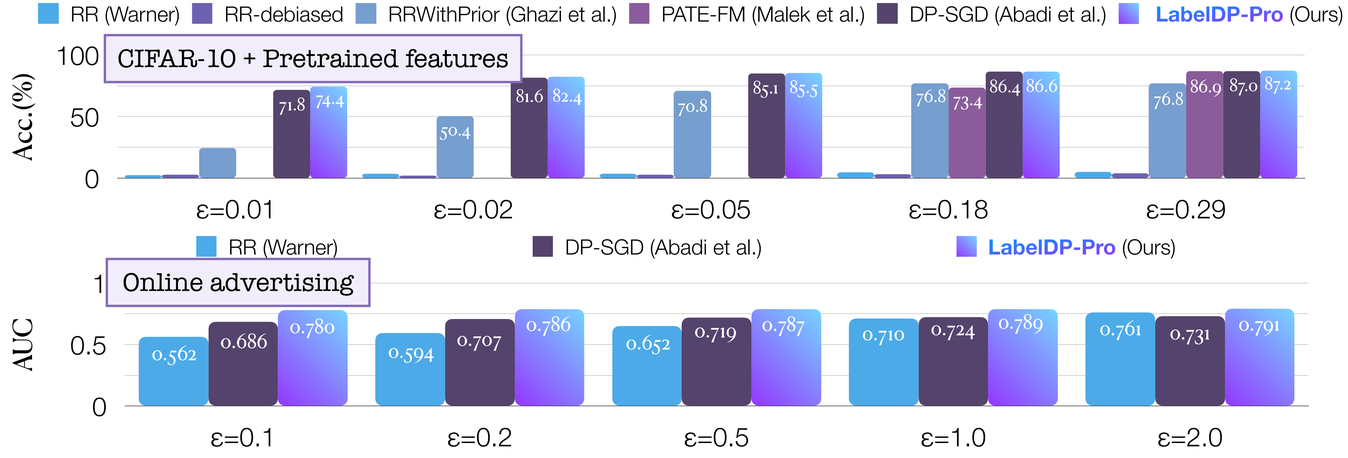
- DP-SGD (Abadi et al., 2016) achieves better utility for high privacy regime, but it targets full-DP: **overly strict for LabelDP**
- We propose LabelDP-Pro ('Pro' for projection): **"denoise" DP-SGD gradient by projecting it** onto a subspace constructed from public information



Denoiser	Output
NoOp	$\tilde{\mathbf{g}}_t$
SelfSpan	$\text{Proj}_A(\tilde{\mathbf{g}}_t)$ for $A = \text{span}\left(\{\nabla_{\theta_t}\ell(\theta_t,(x_i,\kappa)\}_{i\in I_t^G,\kappa\in[K]}\right)$
SelfConv	$\text{Proj}_A(\tilde{\mathbf{g}}_t)$ for $A = \text{conv}\left(\{\nabla_{\theta_t}\ell(\theta_t,(x_i,\kappa)\}_{i\in I_t^G,\kappa\in[K]}\right)$
AltConv	$\text{Proj}_A(\tilde{\mathbf{g}}_t)$ for $A = \text{conv}\left(\{\nabla_{\theta_t}\ell(\theta_t,(x_i,\kappa)\}_{i\in I_t^P,\kappa\in[K]}\right)$

Results





More details (in paper)

- Efficient implementation of the projection-based denoiser
- More experimental results
- Theoretical analysis: Why LabelDP-Pro is better than RR?

Future work

- More efficient denoisers
- Extension of LabelDP-Pro to regression tasks