Differentially Private Linear Regression and Synthetic Data Generation with Statistical Gaurantees

Shurong Lin

Joint work with *Aleksandra (Seša) Slavković* (Penn State)

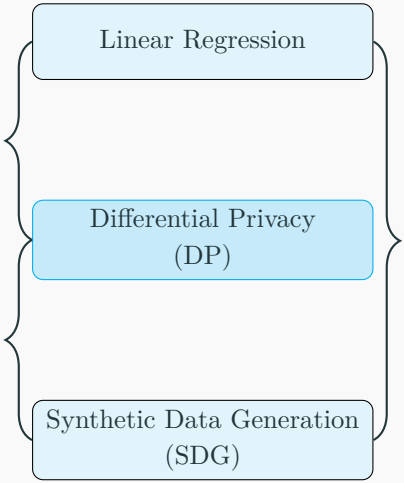
Department of Statistics
Pennsylvania State University
shurong@psu.edu

Overview

Existing Work

- Poor performance in small-sample settings
- Absence of statistical inference in most methods
- Lack of support for SDG

- Require large datasets for training
- Discretize continuous data, losing data continuity
- Lack of theoretical guarantees for downstream regression



Our Method

- Improved/comparable accuracy in smaller-d regimes
- Support for statistical inference
- Support for SGD at no extra privacy cost
- No need for large datasets for SDG
- (Some limitations)

Overview

We propose a novel binning–aggregation approach that provides:

- Algorithm 1: **Differentially private linear regression** with valid confidence intervals
- Algorithm 2: **Differentially private synthetic data generation**, consistent with Algorithm 1
- **Theoretical guarantees** for both privacy and statistical inference
- Extensive **empirical comparisons** with state-of-the-art DP linear regression and DP synthetic data generation methods

Available on *arXiv:2510.16974*; accepted to *AISTATS* 2026; open-source software available on GitHub.