Erchi Wang

La Jolla, CA erw011@ucsd.edu 217-693-8227 erchiw.github.io erchi-wang erchiw

Summary

Ph.D. student with over five years of experience in differentially private machine learning. Research focuses on the practical integration of differential privacy techniques into machine learning and addressing safety concerns in generative models.

Education

University of California, San Diego, San Diego, US — *Jul. 2024 – Estimated 2026*
Ph.D. in Data Science, GPA: 3.90/4.0 — Advised by *Prof. Yu-Xiang Wang*

University of California, Santa Barbara, Santa Barbara, US — *Aug. 2021 – Jul. 2024*
Ph.D. in Statistics (transferred to UCSD), GPA: 3.91/4.0 — Advised by *Prof. Yu-Xiang Wang*

University of Illinois at Urbana-Champaign, Urbana, US — *Aug. 2018 – Dec. 2020*
M.S. in Statistics, GPA: 3.82/4.0

Ocean University of China, Qingdao, China — *Aug. 2013 – Jul. 2018*
B.S. in Applied Math and Biological Science, GPA: 3.76/4.0

Publications & Manuscripts

[1] **Adapting to Linear Separable Subsets with Large Margin in Differentially Private Learning.** Erchi Wang, Yuqing Zhu, Yu-Xiang Wang. *Accepted to ICML 2025; chosen for oral presentation at TPDP 2025*

[2] **Purifying Approximate Differential Privacy with Randomized Post-processing.** Yingyu Lin*, Erchi Wang*, Yi-An Ma, Yu-Xiang Wang. *In Submission; chosen for oral presentation at TPDP 2025*

[3] **The Staining Efficiency of Cyanine Dyes for Single-Stranded DNA is Enormously Dependent on Nucleotide Composition.** Xutiange Han*, Erchi Wang*, Yixiao Cui, et al., *Electrophoresis, 2019*

[4] **Differentially Private XGBoost Revisit: Is Random Decision Trees Really Better than Greedy Ones?** Erchi Wang, Arinbjörn Kolbeinsson, Luca Foschini, Yu-Xiang Wang. *Manuscript.*

Related Projects

(Ongoing) Quantifying Per-instance Memorization in Large Language Model

- Systematically reviewed various concepts of LLM memorization and proposed a per-instance memorization framework analyzed through the lens of a data reconstruction attack.
- Developing algorithmic tools for auditing per-instance memorization.

Converting Approximate DP Mechanisms into Pure DP Mechanisms

- Developed an approximate-to-pure differential privacy converter and used it to design efficient pure DP optimization and data-dependent algorithms, which are previously challenging to design.

Differential Private Adaptive Margin Learning

- Designed a computationally efficient differentially private algorithm for classification problems. Implemented advanced private hyperparameter tuning methods and refined the analysis of DP-SGD, allowing the algorithm to adapt to large margins without requiring prior knowledge of the margin value. Theoretically, the proposed method guarantees utility adaptation to both separable and non-separable cases.

Differential Private Greedy XGBoost on Tabular Data

- Designed and implemented an enhanced differentially private greedy XGBoost algorithm, leveraging modern privacy accounting techniques, including Rényi Differential Privacy-based composition and bounded range analysis for the exponential mechanism. code
- Conducted extensive empirical studies on 18 UCI tabular datasets, achieving state-of-the-art performance with DP-XGBoost by reducing the number of trees by 30% to 50%, thereby enhancing model explainability

and accelerating inference speed.

Work & Research Experience

Machine Learning Researcher *Santa Barbara & San Diego, CA*
S2ML Lab, Advised by *Prof. Yu-Xiang Wang* *Jun. 2022 – Now*

- Research and develop practical differentially private machine learning models and algorithms with provable guarantees. My projects include approximate DP to pure DP conversion, DP large-margin learning, and DP-XGBoost.

Data Scientist *Santa Barbara, CA*
Evidation Health, Advised by *Dr. Arinbjörn Kolbeinsson* *Mar. 2022 – Sep. 2022*

- Led a collaborative healthcare data analysis project between UCSB and Evidation. Developed a differentially private neural network using noisy gradient descent and explored federated learning with private FedSGD and FedAvg. Enhanced prediction performance by 10% through public data integration.

Programming Skills

Languages: Python, R, Bash, Git, SQL

Libraries & Frameworks: Pytorch, Pandas, SciPy, Scikit-learn, Opacus, AutoDP,

Service

Reviewers for NeurIPS (2024, 2025), ICLR (2025), AISTAT (2025), ICML (2025)