Privacy-Preserving Machine Learning in Practice

Honors Thesis Defense

Aneesh Patel

Introduction

"Agencies shall use available… privacy-enhancing technologies (PETs)… to protect privacy and to combat the broader… societal risks that result from the improper use of people's data." [1]

Motivation: The increasing prevalence of machine learning and subsequent concerns about data breaches and privacy violations highlights the critical need for robust privacy safeguards

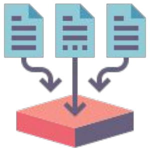
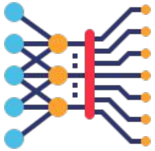
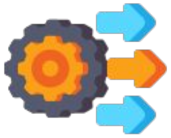
Current State: Existing privacy measures are often inadequate, due in part because privacy research is largely theoretical, with limited practical implementation

Goal: Provide guidance for practitioners and researchers in building and applying secure and privacy-compliant machine learning models in the real world

Contribution: This research conducts a comprehensive analysis of available privacy-preserving methods, assessing their relative performance, efficiency, security and shortcomings on real-world sensitive datasets

[1] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

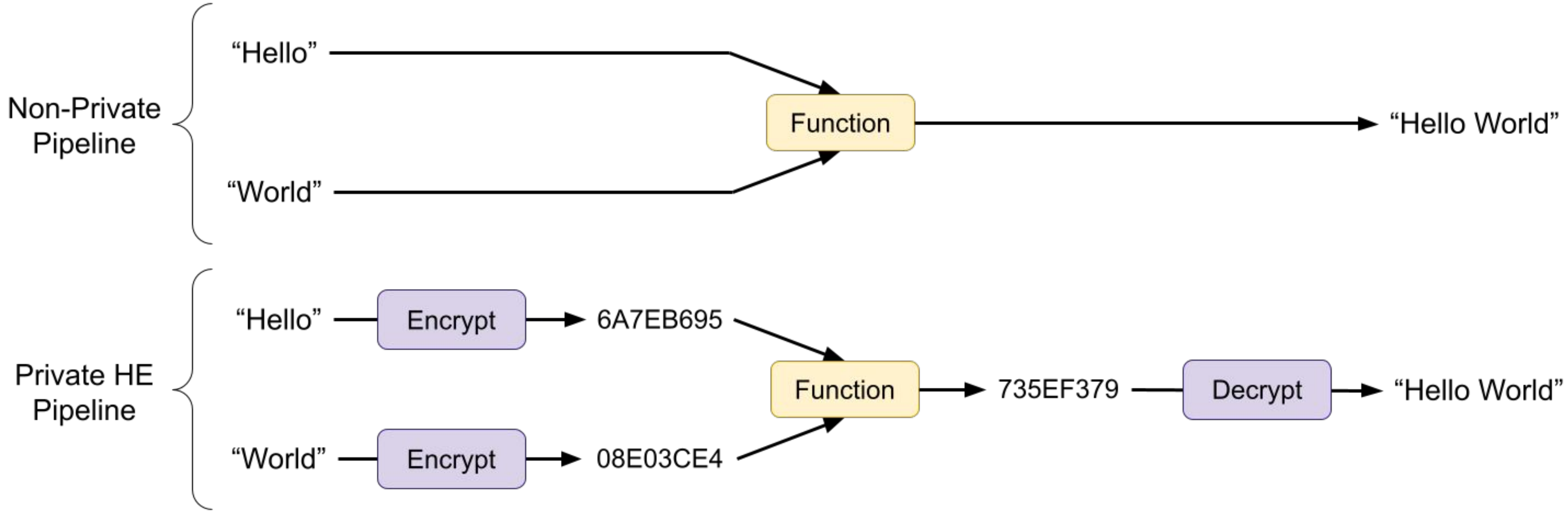
Privacy in Machine Learning

Stage	**Data Collection**	**Data Storage**	**Model Training**	**Model Output**
Privacy Concern	Exposing information during acquisition	Ensuring security and integrity of data at rest	Protecting raw data from central server	Leaking information about the training data
Attacks	–	Privacy Breaches, Unauthorized Access	Privacy Breaches, Data Interception	Reconstruction, Linkage, & Membership Inference
Privacy From	–	Untrustworthy Storage Platform, Curator	Untrustworthy Sharing Channel, Collaborators	Untrustworthy Client

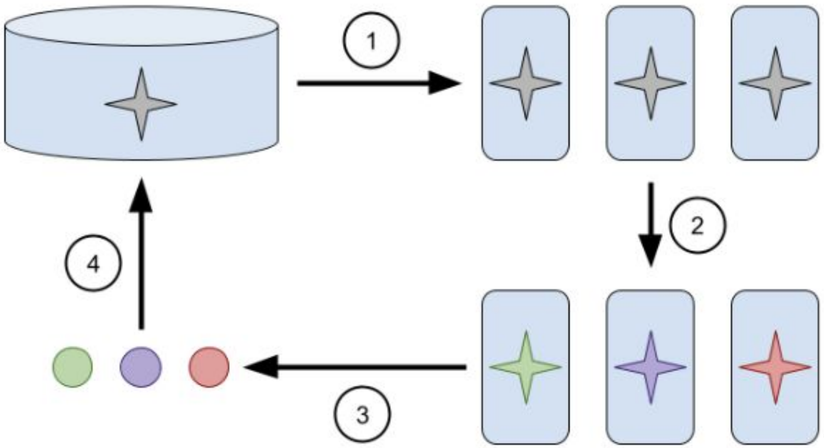
Homomorphic Encryption

Protecting Data at Rest and In Use



Federated Learning

Protecting Data via Decentralized Learning

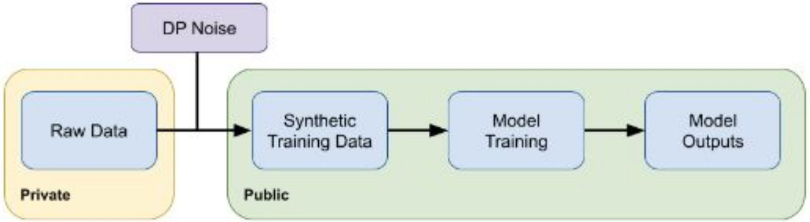


(1) Central server transmits central model to devices
(2) Devices train their models using local sensitive data
(3) Model updates are aggregated
(4) Aggregated model updates are sent to the central server and applied to the central model

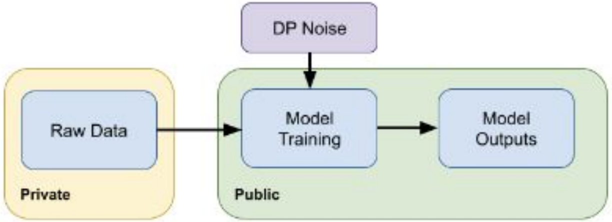
Differential Privacy

Protecting Information about Training Data in Model Outputs

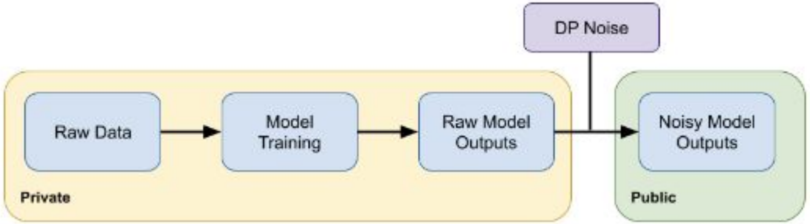
1) Add noise to raw data → private synthetic data



2) Add noise during model training → private model

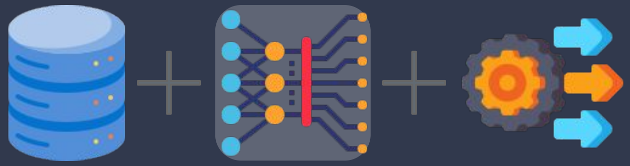


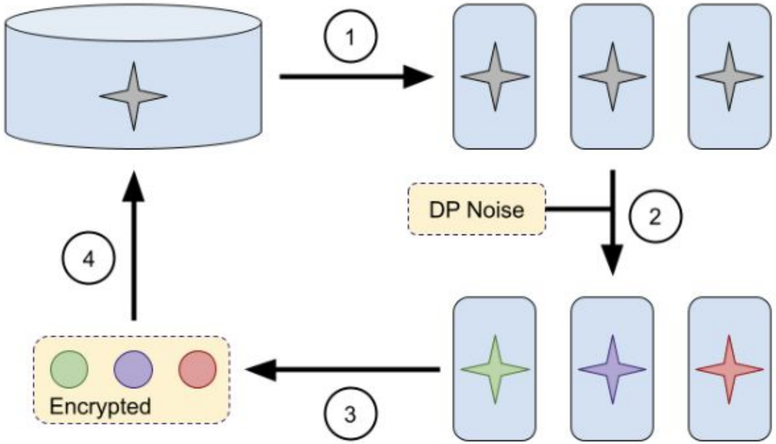
3) Add noise to model outputs → private outputs



Hybrid (FL + DP + HE)

Protects Sensitive Information at All Steps of the ML Pipeline





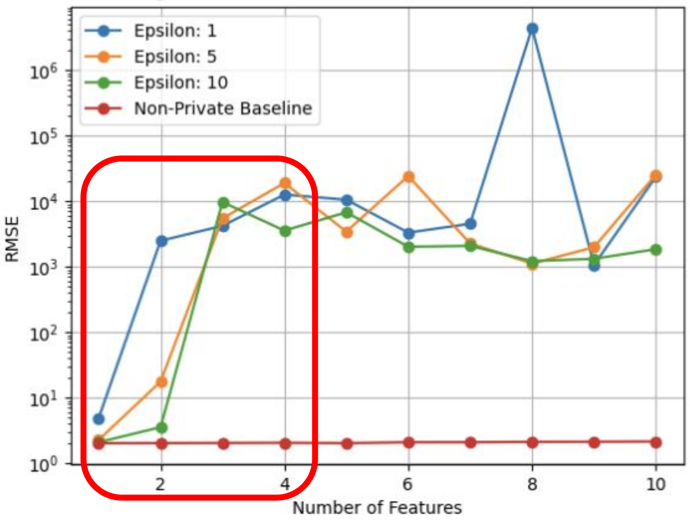
(1) Central server transmits central model to devices
(2) Devices train their models on local sensitive data in a DP manner
(3) Model updates are aggregated and encrypted according to a HE scheme
(4) Encrypted aggregated model updates are sent to the central server and applied to the central model

Summary of Privacy-Preserving Approaches

Approach	What?	Why?	How?	Cost?
Homomorphic Encryption	Computation on encrypted data	Protects data at rest and during computation	Cryptographic techniques and number theory	Reduced efficiency
Federated Learning	Decentralized model training	Preserves privacy from untrustworthy server	Local model training and secure aggregation	Reduced efficiency
Differential Privacy	Privacy in data analysis	Ensures privacy of the training data in the output	Noisy model training to mask individual data points	Reduced utility
MMs for Synthetic Data Generation	Generate synthetic private data	Enables data sharing without privacy concerns	Train a model on raw data to label synthetic data	Reduced utility*

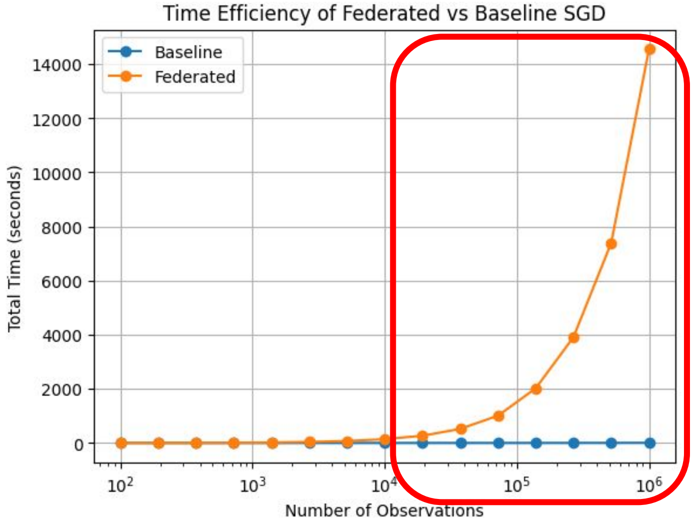
Highlighted Example

Differential Privacy



1
Curse of Dimensionality

Federated Learning



2
Time Scales Linearly With Observations

3 Federated Learning with Differential Privacy yields similar results

Key Takeaways and Next Steps

Key Takeaways

- Navigating the tradeoffs between privacy, utility, and efficiency is a critical aspect of PPML in practice
- Each privacy-preserving approach offers unique protections against different threat models
- Combining multiple approaches in a hybrid manner can enhance overall privacy
- Practical implementations often require unique and counterintuitive solutions

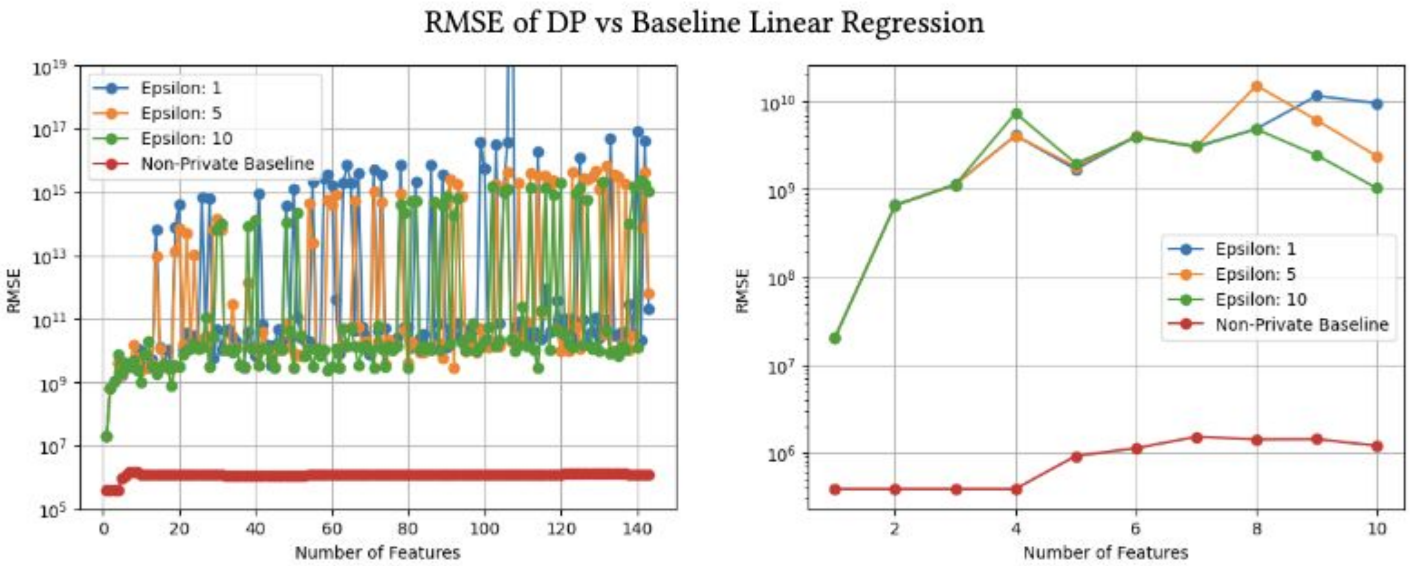
Next Steps

- Develop streamlined methods and tools for implementing PPML in real-world applications
- Further investigate the tradeoffs between privacy, utility, and efficiency in different scenarios
- Explore additional privacy-preserving approaches and novel combinations of them in practice

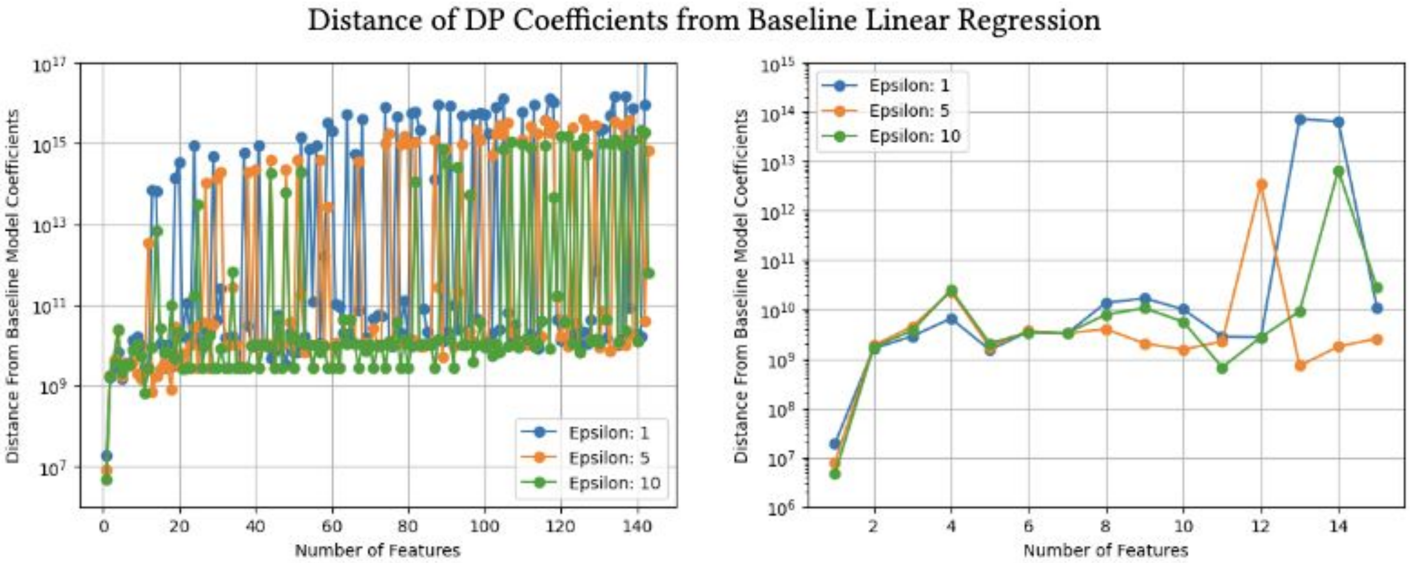
Appendix

[DP] House Prices Linear Regression - Slide 13: House Prices - RMSE
[DP] Student Performance Linear Regression - Slide 17: Student Performance - RMSE
[DP] Census Income Logistic Regression - Slide 20: Census Income - Accuracy
[DP] Census Income Gaussian Naive Bayes - Slide 23: Census Income - Accuracy
[DP] Household Electric - Slide 24: Household Electric - RMSE
[DP + FL] Household Electric - Slide 25: Household Electric - RMSE

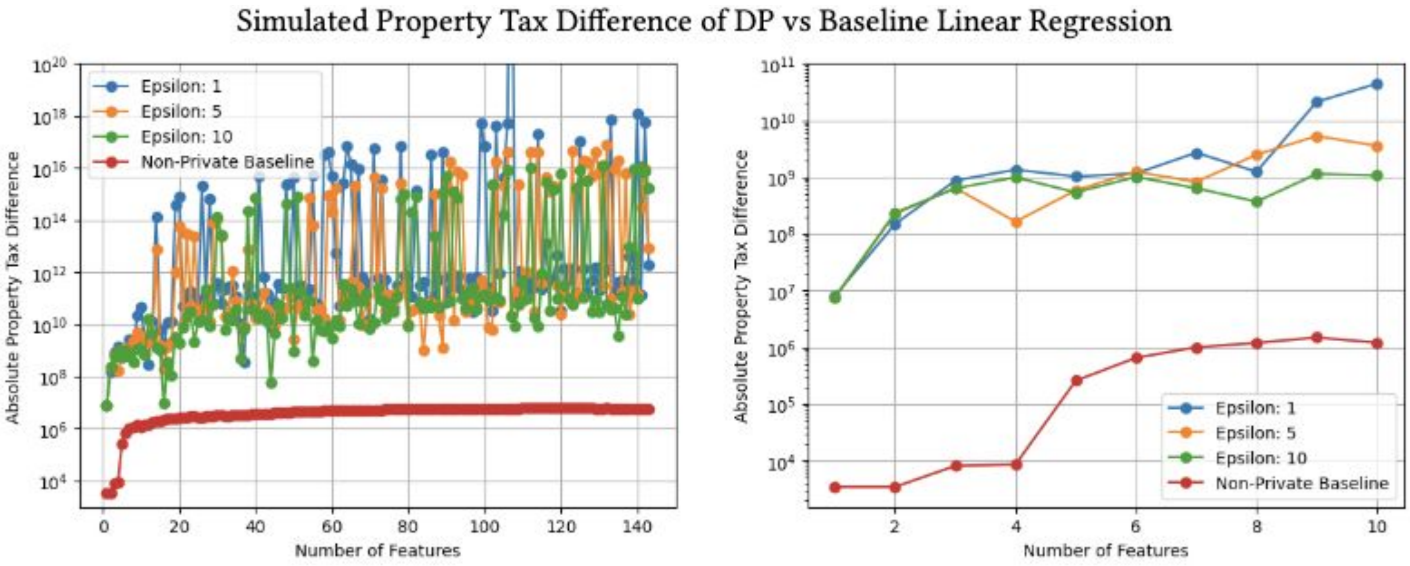
House Prices – RMSE



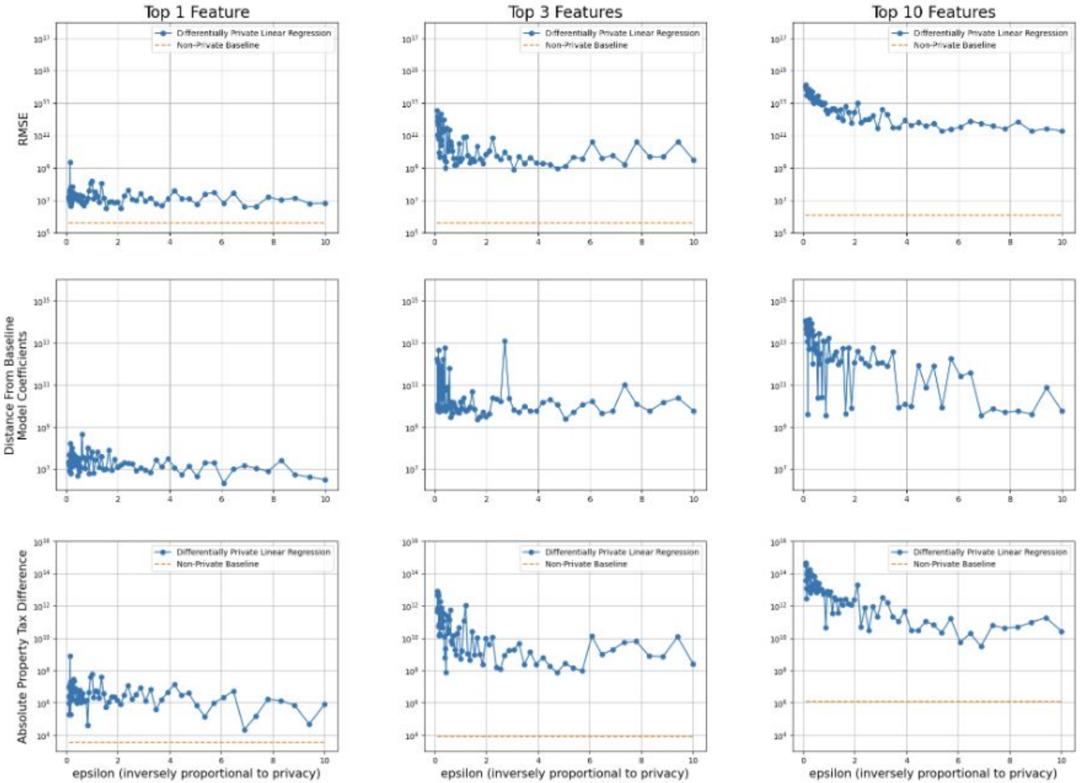
House Prices – Coefficient Distances



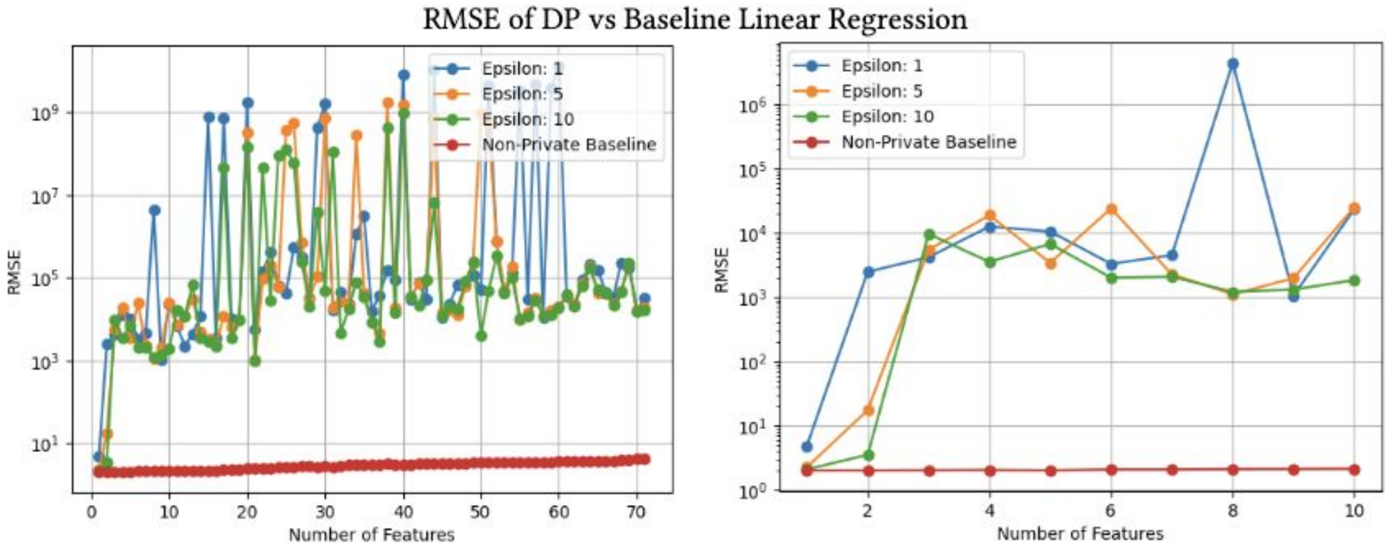
House Prices – Simulated Property Tax



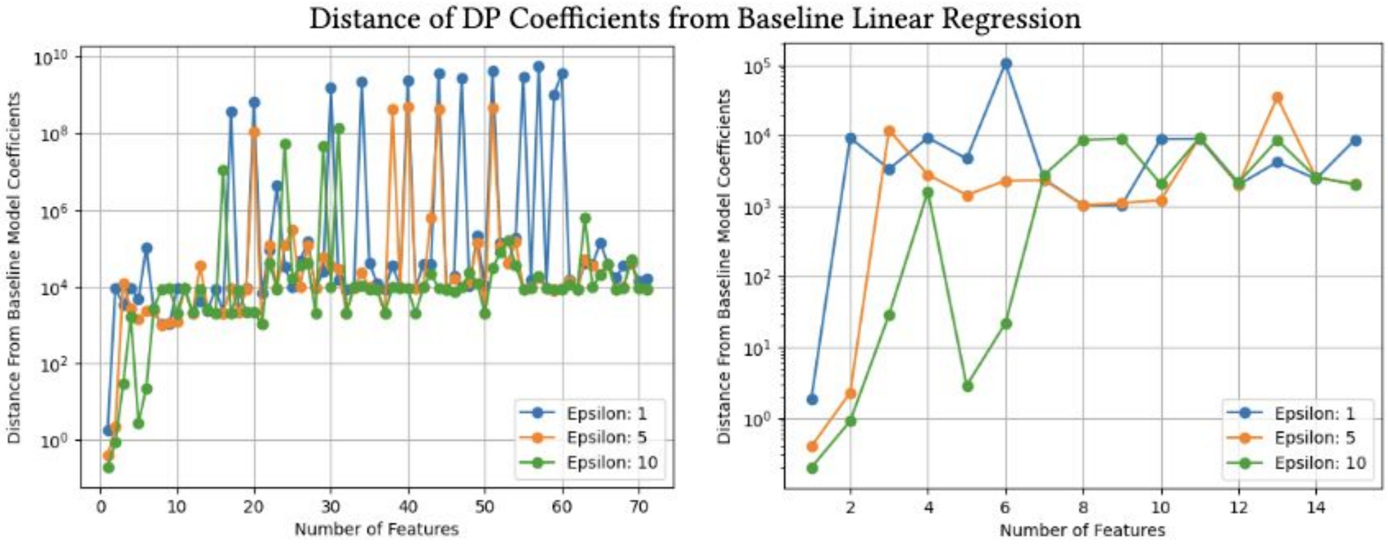
House Prices – Top 1,3,5 Features



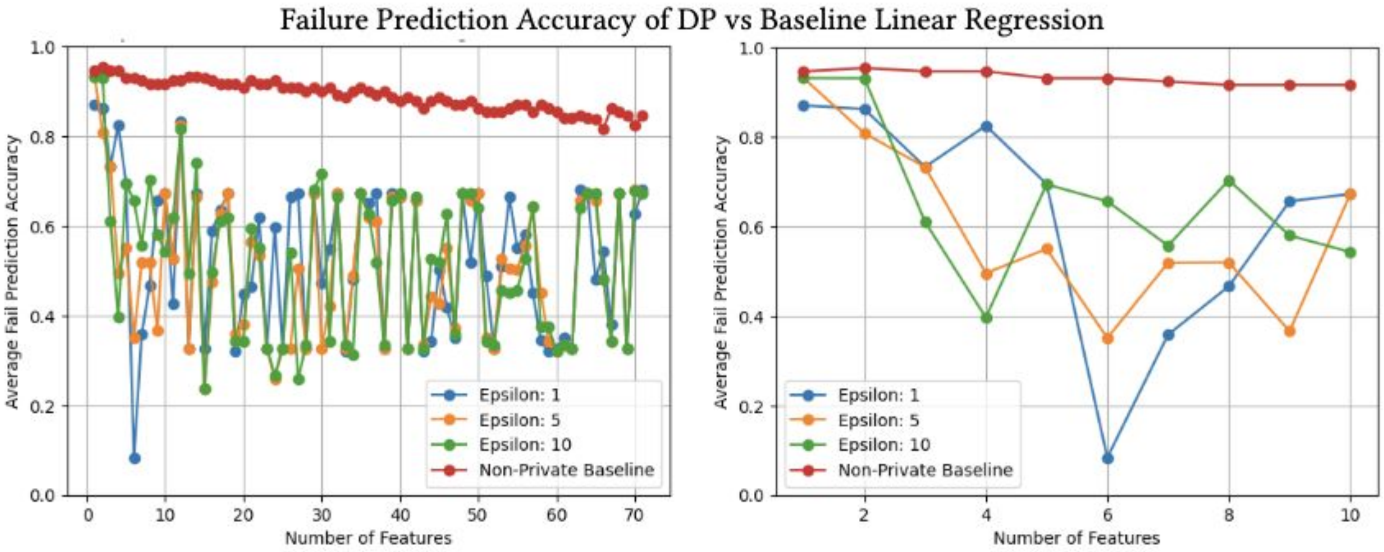
Student Performance – RMSE



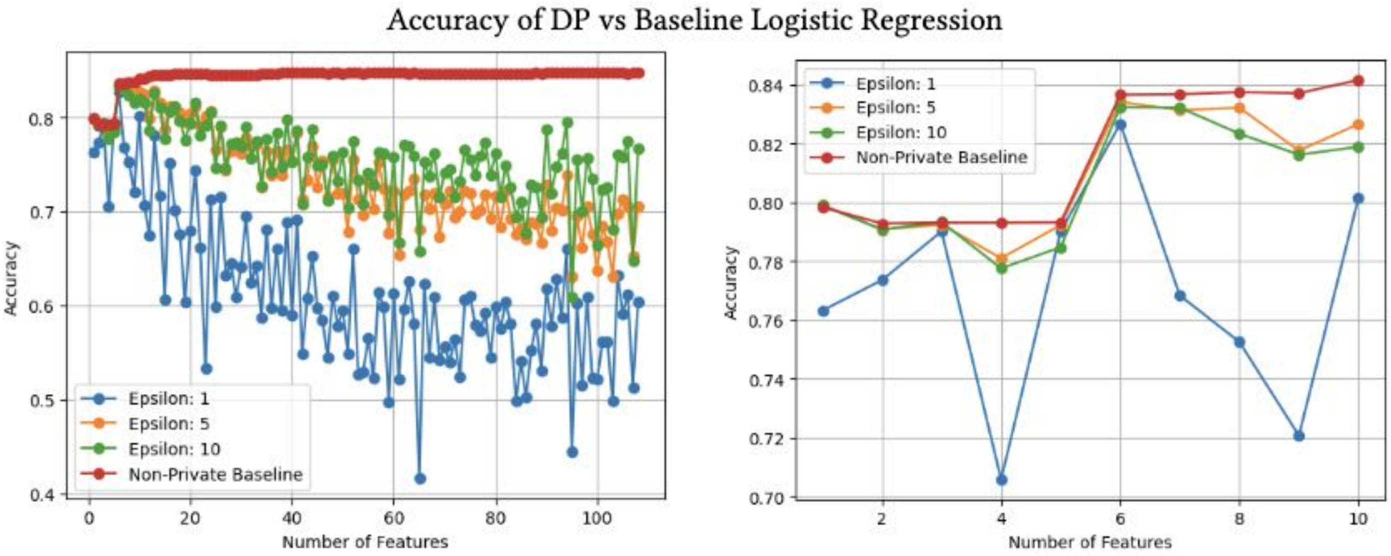
Student Performance – Coefficient Distances



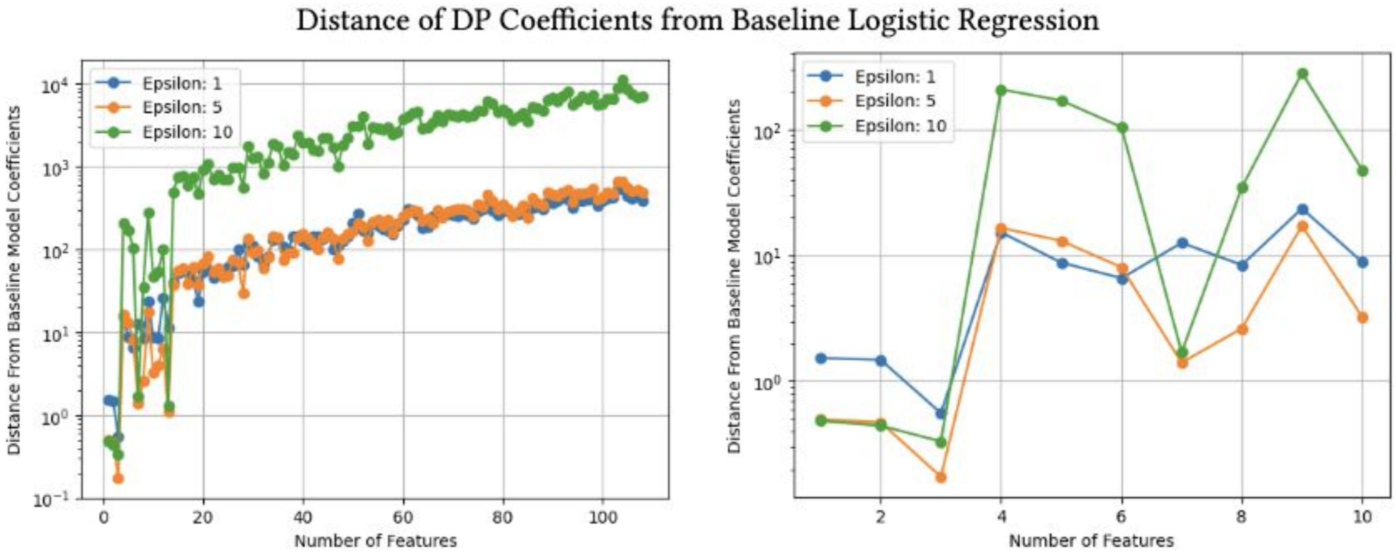
Student Performance – Fail Prediction Accuracy



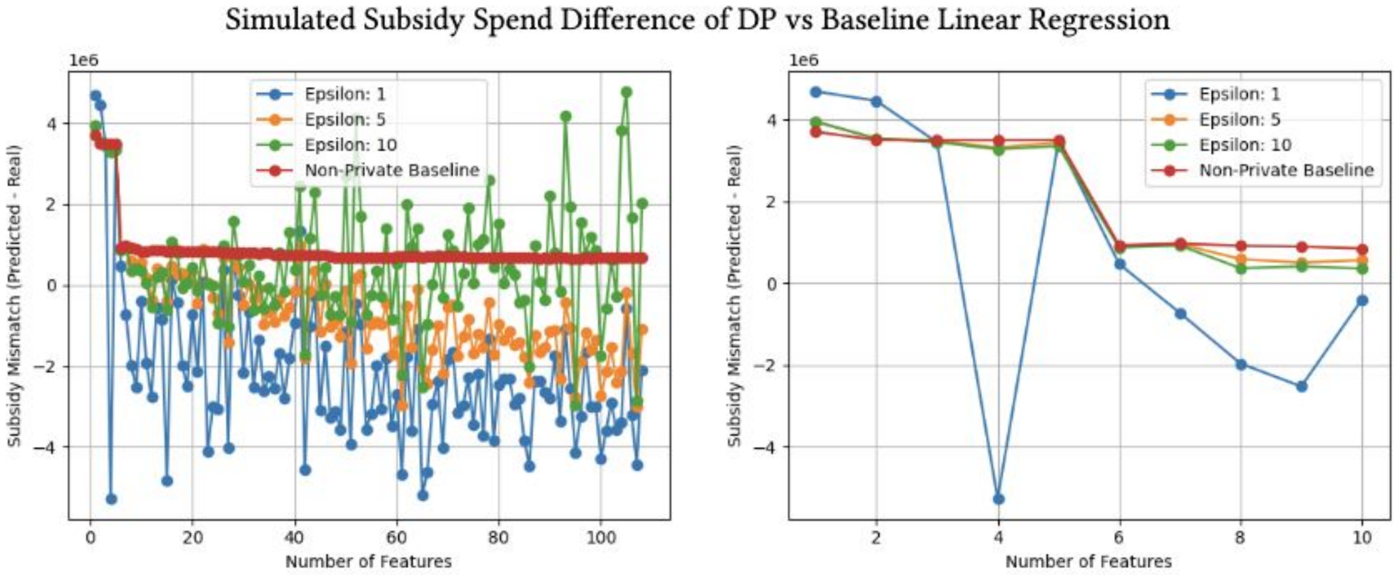
Census Income – Accuracy



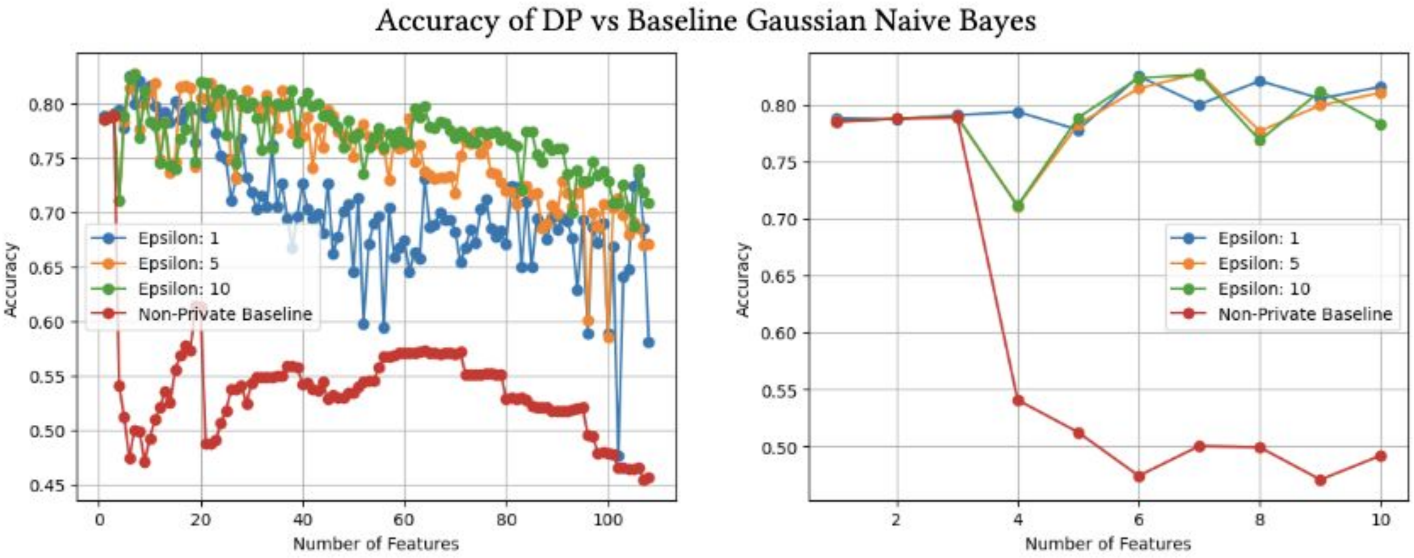
Census Income – Coefficient Distances



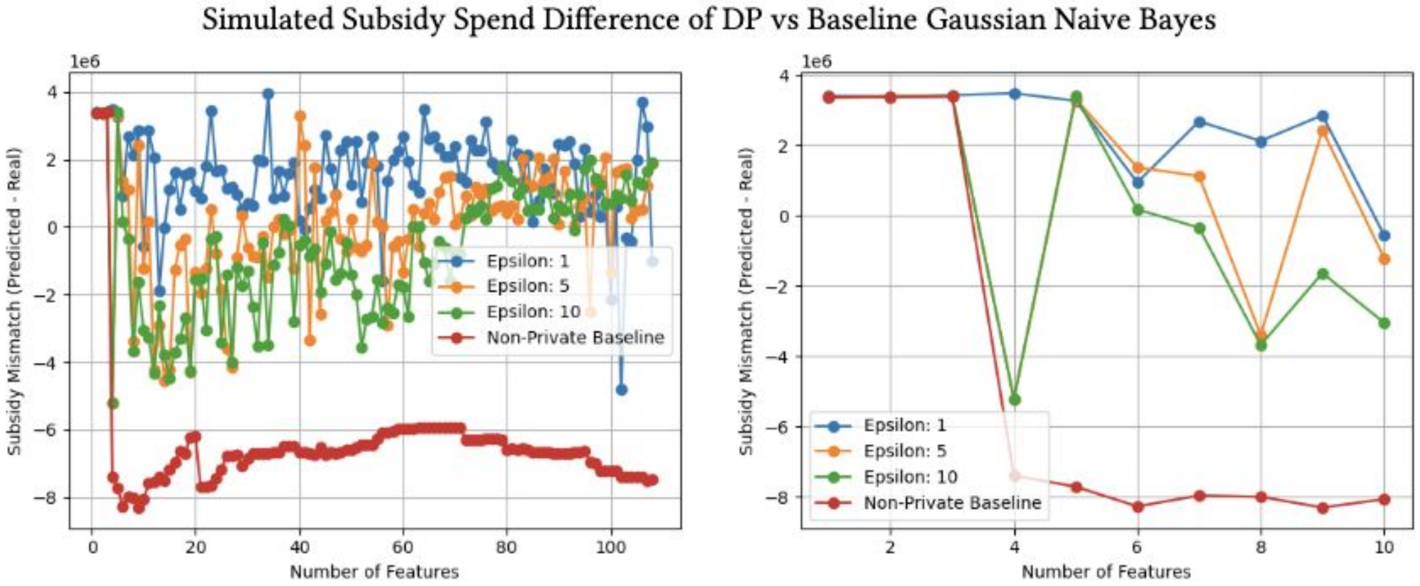
Census Income – Simulated Subsidy Difference



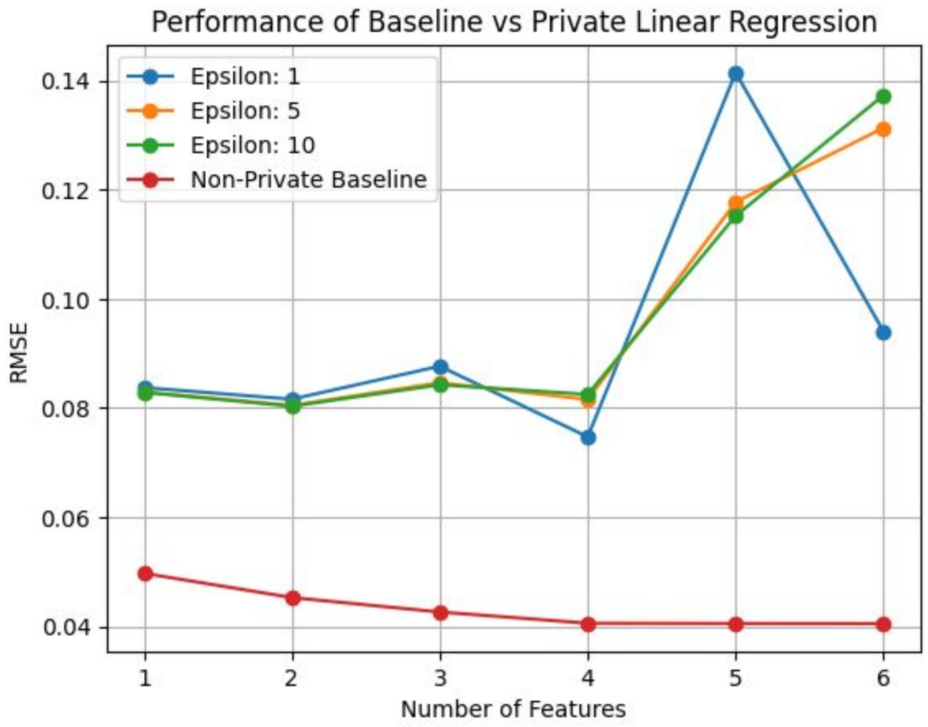
Census Income – Accuracy



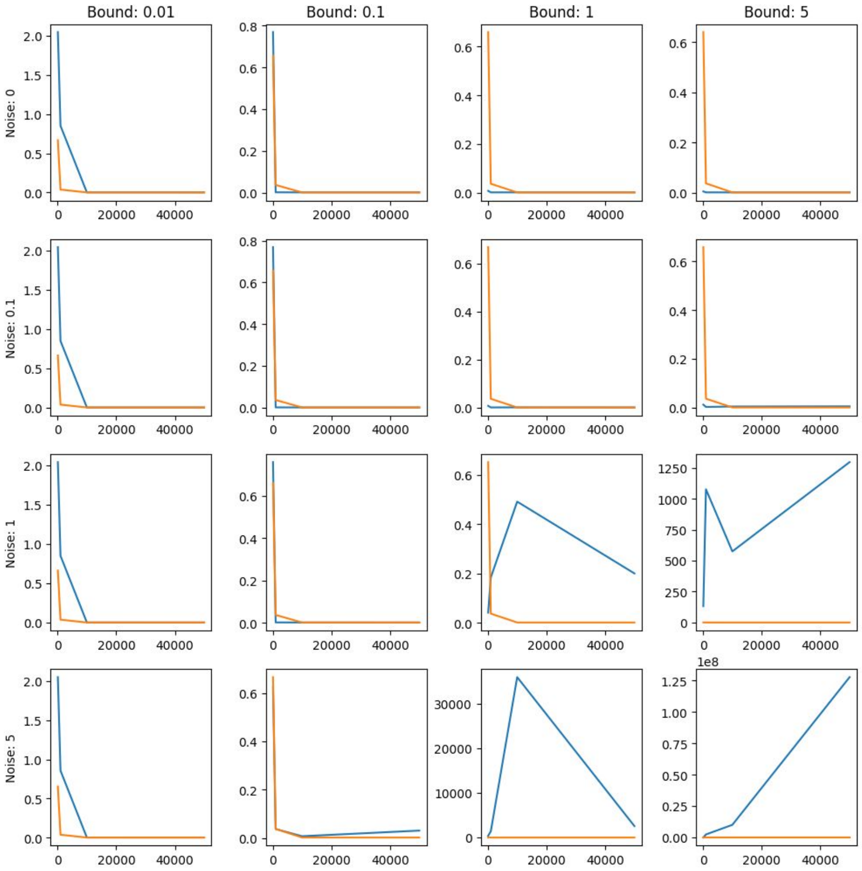
Census Income – Simulated Subsidy Difference



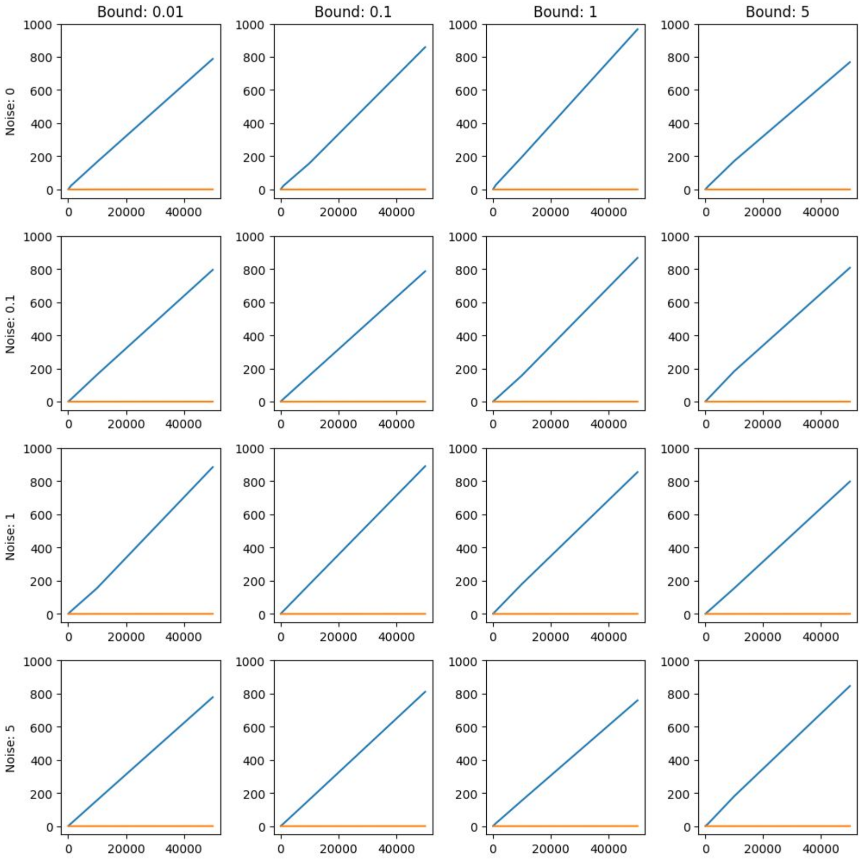
Household Electric – RMSE



Household Electric – RMSE



Household Electric – Time



Novel Approach

Leverage Memorizing Models to Generate Private, Synthetic Data