Privacy Preserving Machine Learning in Practice

Aneesh Patel

Background

Motivation

The rising prevalence of machine learning and related data privacy concerns necessitate robust privacy safeguards

Goal

Guide practitioners and researchers in building and applying privacy-compliant ML models in the real world

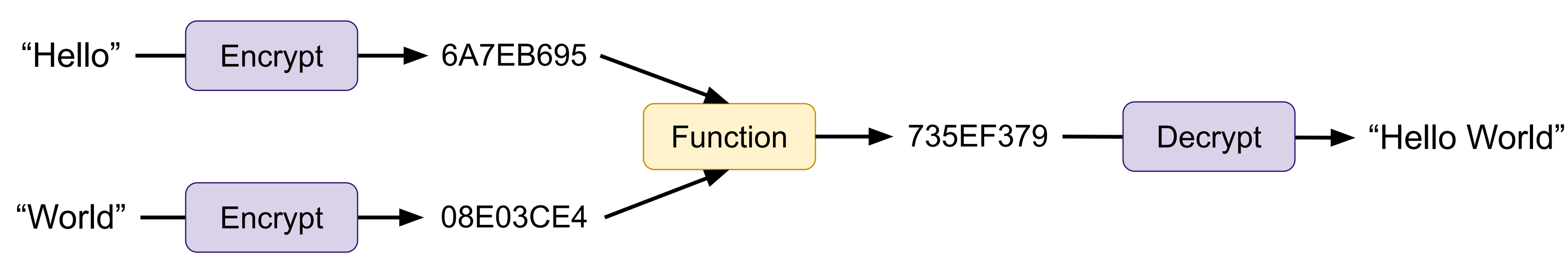
Contribution

Comprehensive analysis of the performance, efficiency, and limitations of PPML approaches on real-world datasets

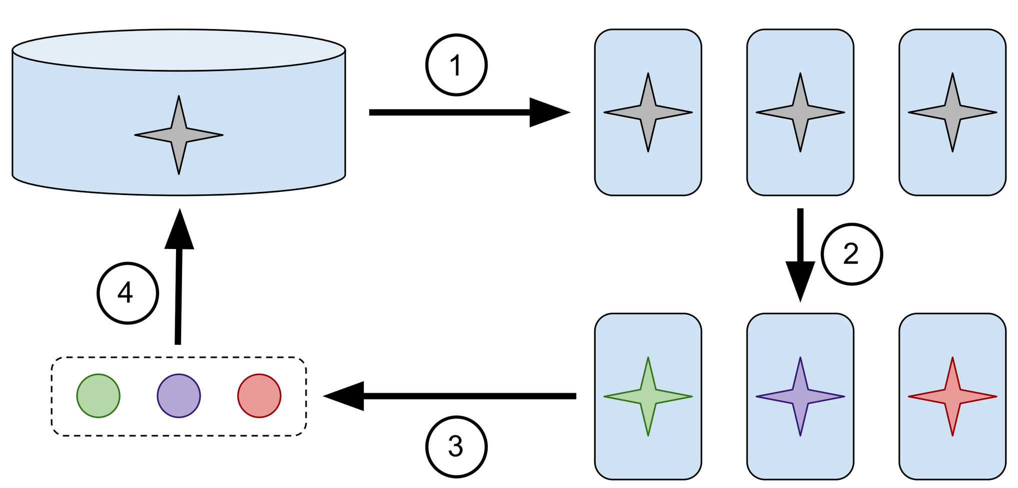
Privacy-Preserving Approaches

Approach	What?	Why?	How?	Cost?
Homomorphic Encryption	Computation on encrypted data	Protects data at rest and in use	Cryptography and number theory	Reduced efficiency
Federated Learning	Decentralized model training	Privacy from central server	Local training and secure aggregation	Reduced efficiency
Differential Privacy	Privacy in data analysis	Privacy of training data in output	Add noise to mask individual data points	Reduced utility

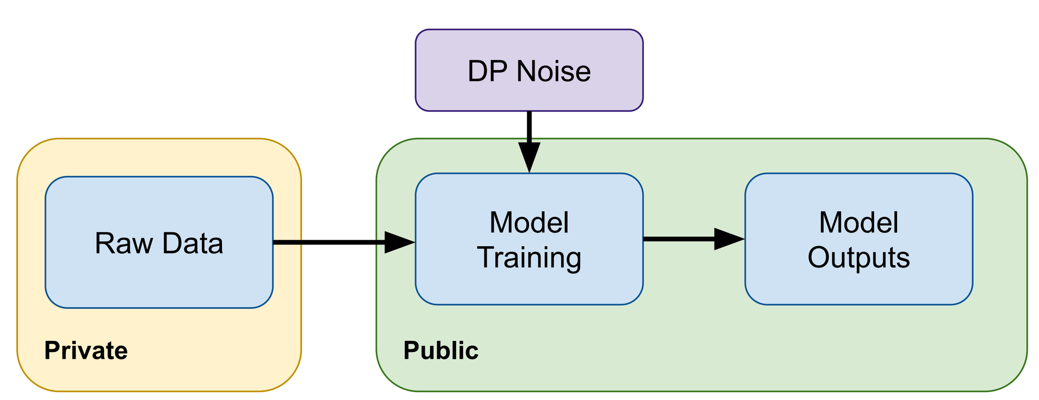
Homomorphic Encryption



Federated Learning



Differential Privacy

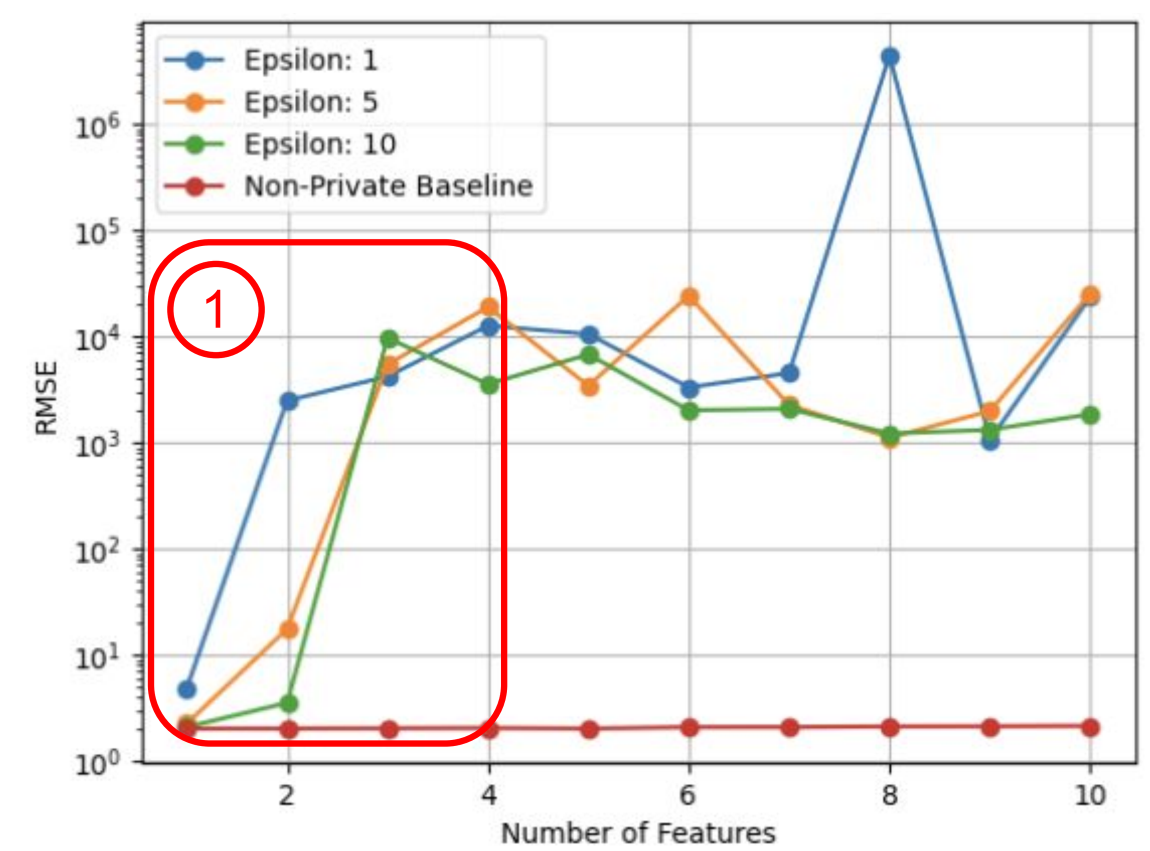


Tested Models

Linear Regression	Decision Tree
Logistic Regression	Random Forest
Gaussian Naive Bayes	Stochastic Gradient Descent

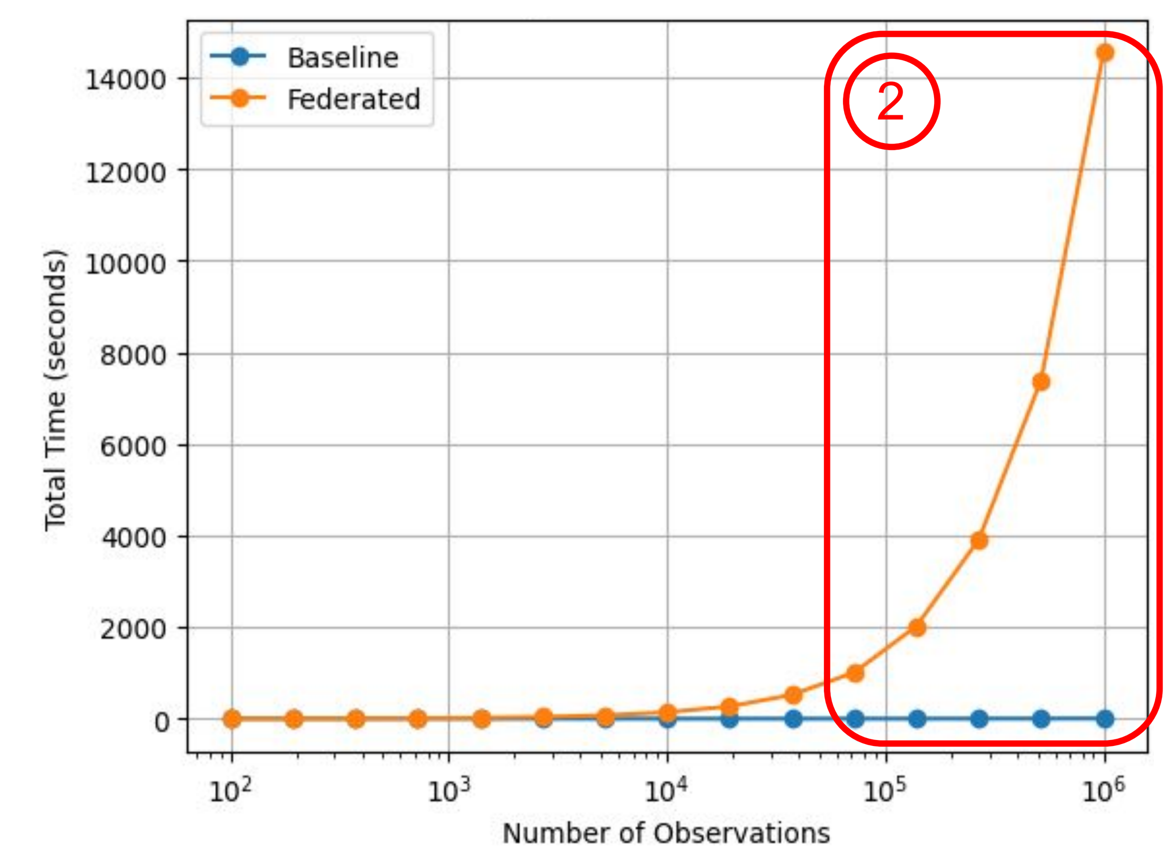
Highlighted Results

Differential Privacy



① Curse of Dimensionality

Federated Learning



② Time Scales Linearly with Number of Observations

~Federated Learning with Differential Privacy yields Similar Results~

Discussion

Key Takeaways

- PPML in practice requires thoughtful navigation of tradeoffs
- Counterintuitive approaches may be necessary for successful implementations

Next Steps

- Explore additional approaches and novel combinations of them
- Develop streamlined methods and tools for implementing PPML in real-world applications

Berkeley UNIVERSITY OF CALIFORNIA

https://github.com/AneeshPatel/Privacy-Preserving-ML

Berkeley College of Computing, Data Science, and Society