Privacy Preserving Matrix Factorization For Recommendation Systems

Masavir Khliq(19314), Advisor: Arpita Patra
Department of Computer Science and Automation
Cryptography and Information Security(CrIS) Lab.

Introduction

As machine learning becomes increasingly prevalent in various industries, the reliance on personal data raises significant privacy concerns. Privacy-Preserving Machine Learning (PPML) aims to balance between maximizing the utility of machine learning models while safeguarding personal information.

Our project focuses on identifying privacy risks in machine learning systems, specifically in recommendation systems that utilize matrix factorization for collaborative filtering.

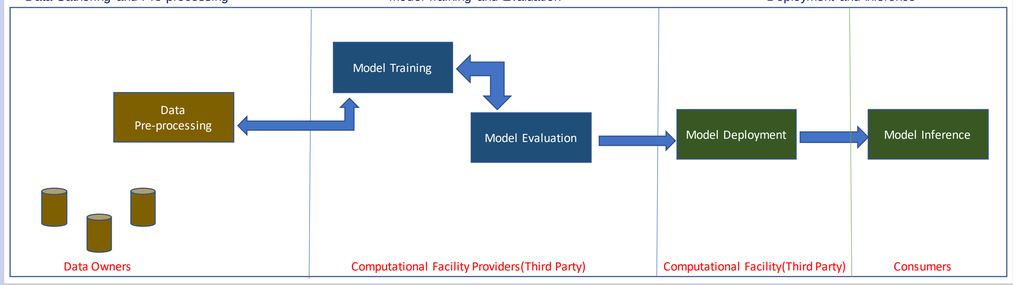
We conducted a comprehensive literature review of various PPML methods, such as MPC, garbled circuits, mask-based approaches, HE,FE and differential privacy, to learn the most effective ways to protect personal data. We have constructed a privacy setting in which we will implement SOTA matrix factorization in a privacy-preserving manner for the recommendation systems.

Machine Learning Pipeline and Threats

ML Pipeline

- Data gathering and pre-processing.
- Model training and evaluation.
- Deployment and inference.

Different parties often manage these stages(MLaaS,IaaS), making privacy preservation crucial throughout the machine-learning pipeline.



Threats

- Gathering sensitive data in a clear form can make it easy for adversaries to access and misuse it.
- Transformed data containing features, can also pose a threat it can be reversed through reconstruction attacks, allowing to reconstruct the original data from the extracted features.
- Trained machine-learning models, can be attacked via reconstruction attacks, model inversion attacks, and membership inference attacks.

PPML Methods

STAGE OF ML PIPELINE	PPML METHODS EMPLOYED
Data Gathering and Pre-Processing	• k-anonymity, t-closeness, and l-diversity to anonymize. • surrogate datasets, sketch techniques, differential privacy.
Model Training and Evaluation	• Secret Sharing-based approaches. • mask-based approaches. • Garbled circuits-based approaches. • Homomorphic encryption. • Functional encryption.
Model Deployment and inference	• Limiting the number of queries. • Private aggregation. • Model transformation. • Model compression.

Problem Statement

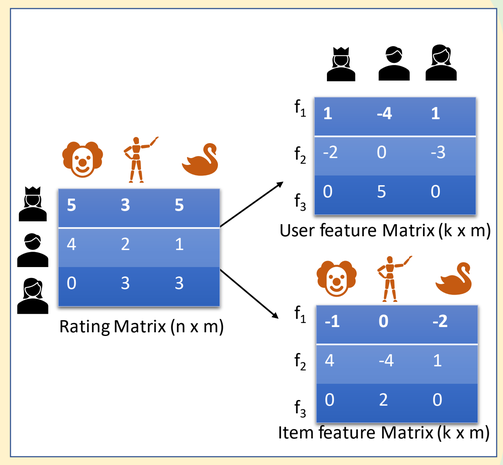
- Matrix factorization the workhorse of today's recommendation systems.
- The goal is to predict ratings for the entire matrix A=[n]x[m] without revealing the ratings to the recommendation system from subset of ratings given by users.
- The objective is to find matrices U and V such that A ≈ U.V by solving the following least square minimization.

$$||A - U.V||_F^2 + R(U) + R(V)$$

- We fix one matrix and update the other matrix and do it alternatively.

$$U_{t+1} \leftarrow \arg\min_U ||A - UV||_F^2 + R(U)$$

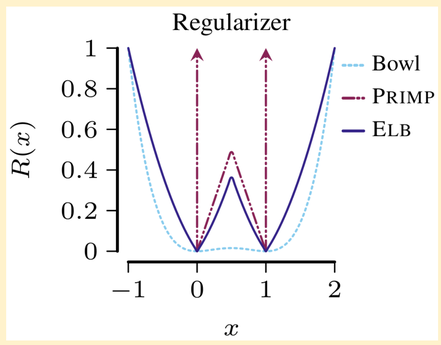
$$V_{t+1} \leftarrow \arg\min_V ||A - U_{t+1}V||_F^2 + R(V)$$



- Here $R(X) = \sum_{x \in X}\{r(x), r(x-1)\}$ is the regularization used. Where $r(x) = \kappa||x||_1 + \lambda||x||_2^2$ is called elastic net Regularization.
- Gradients: $\nabla_U f = UVV^T - AV^T \; and \; \nabla_V f = U^TUV - U^TA$
- This allows the use of proximal gradient descent approach.

$$prox_R(X - \eta\nabla f)$$

- Elastic Net regularization has great advantages.



PALM Algorithm

Algorithm for Matrix Factorization

Input: Matrix A ∈ $R^{m \times n}$
Output: U ∈ $R^{n \times k}$, V ∈ $R^{k \times m}$

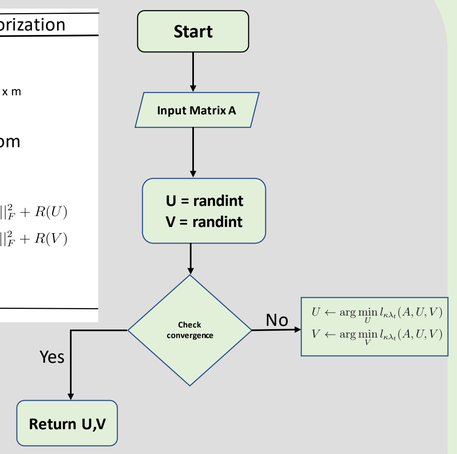
Initialize U, and V at random

Until Convergence do:

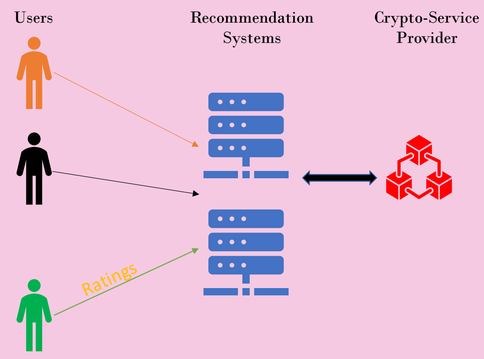
$U_{t+1} \leftarrow \arg\min_U ||A - UV||_F^2 + R(U)$

$V_{t+1} \leftarrow \arg\min_V ||A - U_{t+1}V||_F^2 + R(V)$

end
return U,V



Privacy Settings



Actors

The Protocol for recommendation systems involves the users, the recommendation system, and the crypto-service provider.

Goals

Each user desires to keep their ratings private. The recommendation system performs the matrix factorization in a privacy-preserving manner, while the crypto-service provider facilitates private computation.

Future Work

The project aims to address the problem of performing matrix factorization in a privacy-preserving manner for recommendation systems.

The plan for future work includes:

- Implementing a standard version of matrix factorization as a benchmark for comparison(Done)
- Developing a privacy-preserving variant of matrix factorization using the proximal gradient descent approach and providing its security proof
- Implementing a privacy-preserving variant of the matrix factorization algorithm utilizing techniques from Secure Multi-Party Computation and using it in recommendation systems
- Evaluating the variants on diverse datasets, comparing computation and communication overhead and identifying opportunities for optimization.

Choudhury, Ashish, and Arpita Patra. *Secure Multi-Party Computation Against Passive Adversaries*. Springer Nature, 2022.
Dalleiger, Sebastian, and Jilles Vreeken. "Efficiently Factorizing Boolean Matrices using Proximal Gradient Descent." *Advances in Neural Information Processing Systems* (2022).
Nikolaenko, Valeria, et al. "Privacy-preserving matrix factorization." *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. 2013.
Kim, Jinsu, et al. "Efficient privacy-preserving matrix factorization for recommendation via fully homomorphic encryption." *ACM Transactions on Privacy and Security (TOPS)* 21.4 (2018): 1-30.