Differentially Private Matrix Completion, Revisited
Prateek Jain\textsuperscript{1}, Om Thakkar\textsuperscript{2}, Abhradeep Thakurta\textsuperscript{3}
\textsuperscript{1}Microsoft Research, \textsuperscript{2}The Pennsylvania State University, \textsuperscript{3}University of California-Santa Cruz

INTRODUCTION

Collaborative Filtering: To provide personalized recommendations via crowdsourcing.

Examples:

Google Play

Amazon

Netflix

Matrix Completion: Given an incomplete matrix \(X \in \mathbb{R}^{m \times n}\), output \(Y\), such that \(Y \approx X\).

 THE NEED FOR PRIVACY

Users may not prefer to reveal what movies they saw.
- Or how much they liked them!

Our Goal: To provide personalized recommendations using crowdsourced data while ensuring user-level differential privacy (DP).

PRIVACY MODEL

Joint Differential Privacy (JDP): Mechanism \(A: \mathbb{D}^m \rightarrow T \in (\varepsilon, \delta)\)-Joint DP if for all neighboring datasets \(x, x^* \in \mathbb{D}^m\), and for all \(i \in [m]\), for all sets of outcomes \(S_i \subseteq T_i\),

\[ P(A_i(x) \in S_i) \leq e^\varepsilon P(A_i(x^*) \in S_i) + \delta. \]

Here, \(A_i(x) = \text{output of } A \text{ on input } X \text{ without user } i \text{'s output}.

Distinction from standard DP (\textsuperscript{DMNS}'06):
- Under Joint DP, A's output for user \(i\) can depend \textit{arbitrarily} on \(i\)'s input.
Consequence: Better personalized recommendations!

THEORETICAL RESULTS

An iterative process, having two major steps in every iteration \(t\):
1. Local (user-side) computation: Each user \(i \in [m]\) computes the error of her incomplete row \(X_i\) from her predicted row \(Y_i\), and sends the covariance of the error to a central server.

2. Global (server-side) computation: The server adds Gaussian noise \(N(0, \sigma^2(1/\varepsilon))\) to the sum of the error covariances, computes a global rank-1 update via SVD, and releases it publicly so that each user can update her own prediction row.

Error due to Privacy

Standard Frank-Wolfe convergence error

1. Privacy guarantee:
- If \(\varepsilon = \frac{c}{\sqrt{m}} \cdot (2\ln(\frac{m}{\delta}))\), then the Frank-Wolfe algorithm above is \((\varepsilon, \delta)\)-Joint DP

2. Utility guarantee:
- If \(\|Y\|_2^2 \leq k \cdot \max_{i \in [m]} \|X_i\|_2^2 \leq L^2\) and \(\varepsilon, \delta\)-Joint DP Frank-Wolfe (FW) algorithm for \(T\) iterations, then with high probability:

\[ \text{Empirical Risk} = \frac{1}{m} \sum_{i=1}^{m} \langle Y_i, X_i \rangle = \delta \left( \frac{1}{m} \sum_{i=1}^{m} \langle X_i, X_i \rangle^{1/2} \right). \]

Here, \(\Omega\) is the set of non-zero indices in \(X\).

CONCLUSIONS

- We design a variant of the Frank-Wolfe algorithm for matrix completion.
- We make it amenable for user-level Joint DP by splitting the iterative update step into 2 parts, local (user-side) computation and global (server-side) computation.
- We provide the privacy and utility guarantees for it.
- We demonstrate its performance on a variety of benchmark datasets, showing that:
  - It provides nearly the same accuracy as the state-of-the-art non-private algorithm, and
  - It outperforms the existing state-of-the-art private matrix completion method [MM'09] by as much as 30%.

REFERENCES


EXPERIMENTAL RESULTS

- \(m = \text{number of users, } n = \text{number of items}\)
- Unless specified, we sample \(= 80\) ratings per user, and each rating \(\in [0, 1]\)

- Netflix (Top 400)\n  \(m = 474k, n = 400\) most rated movies

- Yahoo (Top 400)\n  \(m = 905k, n = 400\) most rated songs

Legend for all the plots