CS2080: Applied Privacy for Data Science
Reconstruction Attacks

School of Engineering & Applied Sciences
Harvard University

February 3, 2025

Discussion

Imagine that you and your tablemates are on an advisory committee tasked with making recommendations to lawmakers about data privacy. Specifically, they want to know how seriously to take the evidence found by computer scientists against de-identification. Does it warrant changes to how data are protected? Why or why not?

> If you are at an ODD numbered table, form your strongest arguments *in favor* of de-identification.
>
> If you are at an EVEN numbered table, form your strongest arguments *against* de-identification.

Cohen & Nissim

Linear Program Reconstruction in Practice

> *"The goal of Diffix is to allow data analysts to perform an unlimited number of statistical queries on a sensitive database while protecting the underlying data and while introducing only minimal error. It is advertised as an off-the-shelf, GDPR-compliant privacy solution, and Aircloak reports that CNIL, the French national data protection authority, has already evaluated Diffex against the GDPR anonymity criteria, and have stated that Aircloak delivers GDPR-level anonymity."*

Cohen & Nissim

Linear Program Reconstruction in Practice

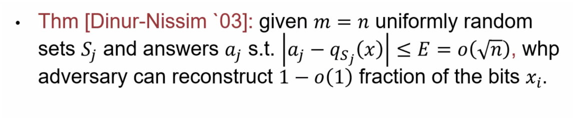
- Use queries of sums over random subsets to reconstruct individual data.
- Importantly, the members of the subset are reported in each sum.
- Received the Aircloak Bounty ($5000) for reidentifying challenge data in the *Diffix* commercial system.

`https://journalprivacyconfidentiality.org/index.php/jpc/article/view/711`

Cohen & Nissim

Linear Program Reconstruction in Practice

- Use queries of sums over random subsets to reconstruct individual data.
- Importantly, the members of the subset are reported in each sum.
- Received the Aircloak Bounty ($5000) for reidentifying challenge data in the *Diffix* commercial system.

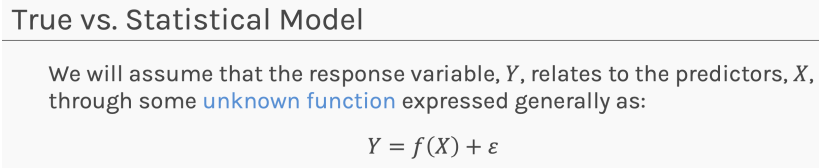


- Thm [Dinur-Nissim `03]: given $m = n$ uniformly random sets S_j and answers a_j s.t. $\left|a_j - q_{S_j}(x)\right| \le E = o(\sqrt{n})$, whp adversary can reconstruct $1 - o(1)$ fraction of the bits x_i.

https://journalprivacyconfidentiality.org/index.php/jpc/article/view/711

Regression Based Reconstruction

From CS109:



True vs. Statistical Model

We will assume that the response variable, Y, relates to the predictors, X, through some unknown function expressed generally as:

$$Y = f(X) + \varepsilon$$

Regression Based Reconstruction

Find $\hat{x_1}, \ldots, \hat{x_N}$ s.t.:

$$\hat{x} = \underset{\hat{x}}{\operatorname{argmin}}[\sum_{j=1}^{m}(a_j - \sum_{i \in S_j} \hat{x}_i)^2]$$

$$\hat{x} = \underset{\hat{x}}{\operatorname{argmin}}[\sum_{j=1}^{m}(a_j - \sum_{i=1}^{n} \hat{x}_i \; s_{j,i})^2]$$

$$\hat{x} = \underset{\hat{x}}{\operatorname{argmin}}[\sum_{j=1}^{m}(a_j - \hat{a}_j)^2]$$

In R see:

`lm()`

In Python see for example:

`linear_model.LinearRegression()`

from scikit-learn.

Regression Based Reconstruction

$$a_j = x_1 s_{1,j} + x_2 s_{2,j} + \ldots + x_n s_{n,j} + e_j$$

Here:

n	is the number of people in the database
m	is the number of queries
i	is a person index
j	is query index
a_j	is j-th query release
$s_{i,j}$	is a $\{0,1\}$-indicator $i \in S_j$
x_h	is h's sensitive data
e_i	is the residual/error of the i-th prediction

Regression Based Reconstruction

$$
\begin{aligned}
a_j &= x_1 s_{1,j} + x_2 s_{2,j} + \ldots + x_n s_{n,j} + e_j \\
7 &= 1{\cdot}1 + 0{\cdot}1 + 1{\cdot}0 + 0{\cdot}0 + \ldots + 0{\cdot}1 + 2 \\
4 &= 1{\cdot}0 + 0{\cdot}1 + 1{\cdot}1 + 0{\cdot}1 + \ldots + 0{\cdot}1 + (-1) \\
6 &= 1{\cdot}0 + 0{\cdot}0 + 1{\cdot}0 + 0{\cdot}1 + \ldots + 0{\cdot}0 + 3
\end{aligned}
$$

Here:

n is the number of people in the database

m is the number of queries

i is a person index

j is query index

a_j is j-th query release

$s_{i,j}$ is a $\{0,1\}$-indicator $i \in S_j$

x_h is h's sensitive data

e_i is the residual/error of the i-th prediction

Regression Based Reconstruction

$$
\begin{aligned}
a_j &= x_1 s_{1,j} + x_2 s_{2,j} + \ldots + x_n s_{n,j} + e_j \\
7 &= 1 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 + 0 \cdot 0 + \ldots + 0 \cdot 1 + 2 \\
4 &= 1 \cdot 0 + 0 \cdot 1 + 1 \cdot 1 + 0 \cdot 1 + \ldots + 0 \cdot 1 + (-1) \\
6 &= 1 \cdot 0 + 0 \cdot 0 + 1 \cdot 0 + 0 \cdot 1 + \ldots + 0 \cdot 0 + 3
\end{aligned}
$$

Here:

n is the number of people in the database

m is the number of queries

i is a person index

j is query index

a_j is j-th query release

$s_{i,j}$ is a $\{0,1\}$-indicator $i \in S_j$

x_h is h's sensitive data

e_i is the residual/error of the i-th prediction

Regression Based Reconstruction

$$a_j = \hat{x}_1 s_{1,j} + \hat{x}_2 s_{2,j} + \ldots + \hat{x}_n s_{n,j} + e_j$$

Here:

n is the number of people in the database
m is the number of queries
i is a person index
j is query index
a_j is j-th query release
$s_{i,j}$ is a $\{0, 1\}$-indicator $i \in S_j$
x_h is h's sensitive data
e_i is the residual/error of the i-th prediction

Regression Based Reconstruction

$$
\begin{aligned}
a_j &= \hat{x}_1 s_{1,j} + \hat{x}_2 s_{2,j} + \ldots + \hat{x}_n s_{n,j} + e_j \\
7 &= 0.92\cdot 1 + 0.11\cdot 1 + 1.07\cdot 0 + -0.08\cdot 0 + \ldots + 0.07\cdot 1 + 5.71 \\
4 &= 0.92\cdot 0 + 0.11\cdot 1 + 1.07\cdot 1 + -0.08\cdot 1 + \ldots + 0.07\cdot 1 + 2.31 \\
6 &= 0.92\cdot 0 + 0.11\cdot 0 + 1.07\cdot 0 + -0.08\cdot 1 + \ldots + 0.07\cdot 0 - 1.04
\end{aligned}
$$

Here:

n is the number of people in the database

m is the number of queries

i is a person index

j is query index

a_j is j-th query release

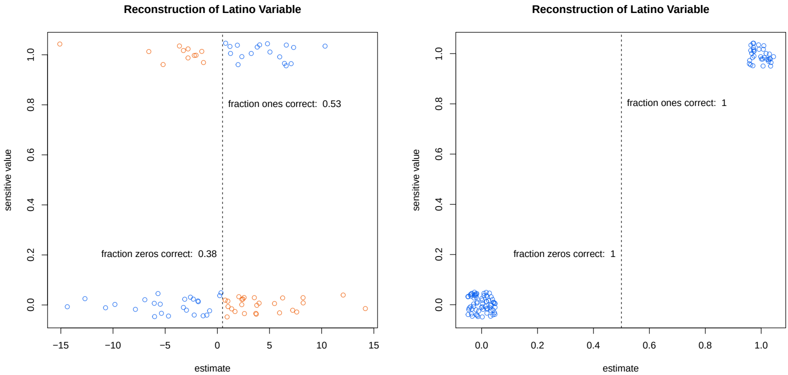
$s_{i,j}$ is a $\{0,1\}$-indicator $i \in S_j$

x_h is h's sensitive data

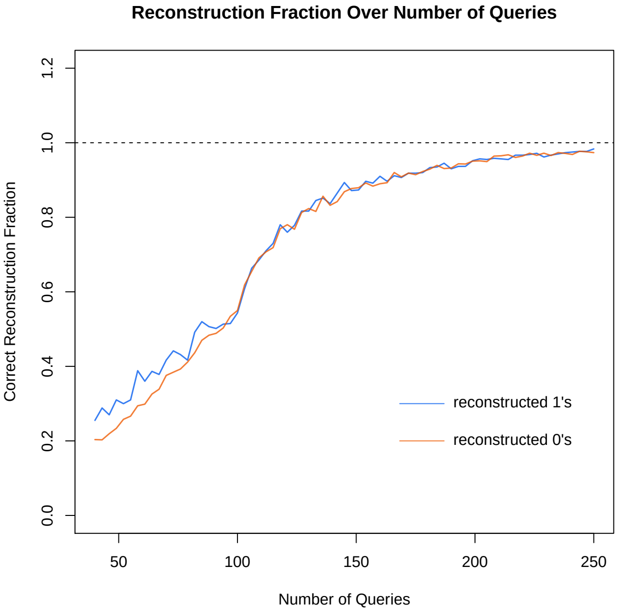
e_i is the residual/error of the i-th prediction

Example

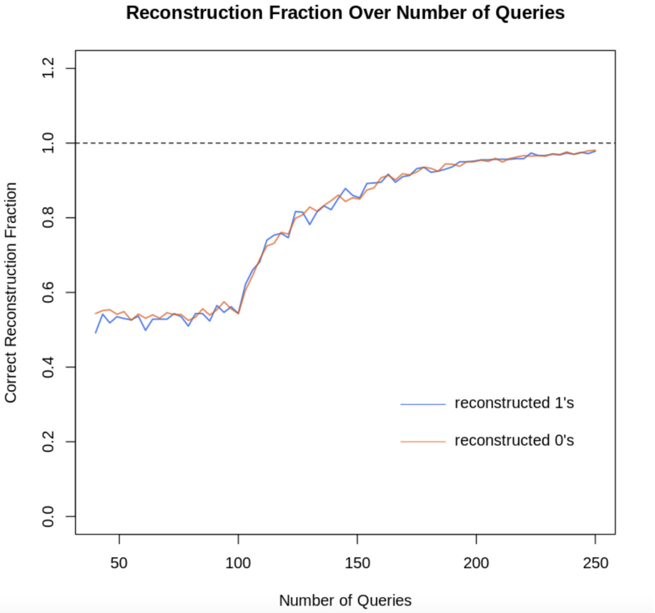
From `wk2_regression_attack.ipynb`:



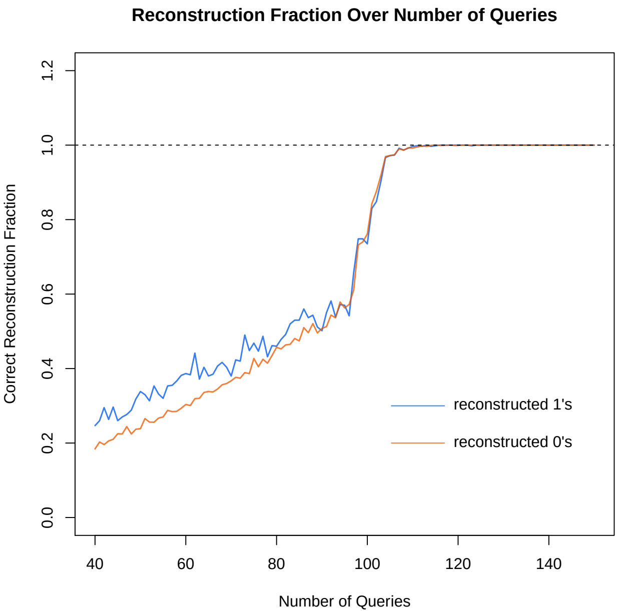
Example: Rounding to Nearest 5



Example: Rounding to Nearest 5 w/ Priors



Example: Normal Errors


Reconstruction Fraction Over Number of Queries
Correct Reconstruction Fraction
Number of Queries
reconstructed 1's
reconstructed 0's