Privacy

Christos Dimitrakakis

May 14, 2019

Just because they're the problem,
doesn't mean we aren't.

Privacy in statitical disclosure.

- Public analysis of sensitive data.
- Publication of "anonymised" data.

Not about cryptography

- Secure communication and computation.
- Authentication and verification.

An issue of trust

- Who to trust and how much.
- With what data to trust them.
- What you want out of the service.

Databases

Example 1 (Typical relational database in a tax office)

ID	Name	Salary	Deposits	Age	Postcode	Prof
1959060783	Mike Pence	150,000	1e6	60	1001	Poli
1946061408	Donald Trump	300,000	-1e9	72	1001	Ren
2100010101	A. B. Student	10,000	100,000	40	1001	Tim

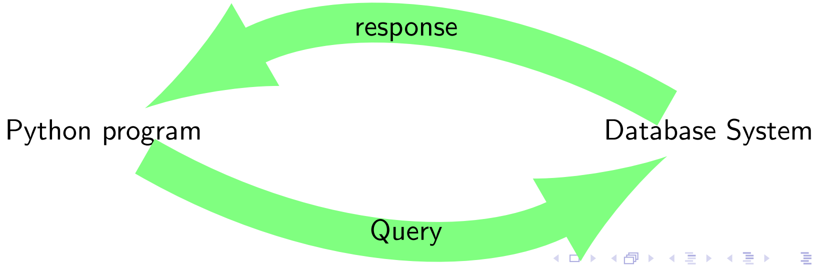
Database access

- When owning the database: Direct look-up.
- When accessing a server etc: Query model.

Databases

Example 1 (Typical relational database in a tax office)

ID	Name	Salary	Deposits	Age	Postcode	Prof
1959060783	Mike Pence	150,000	1e6	60	1001	Poli
1946061408	Donald Trump	300,000	-1e9	72	1001	Ren
2100010101	A. B. Student	10,000	100,000	40	1001	Tim



Queries in SQL

The SELECT statement

- `SELECT column1, column2 FROM table;`
- `SELECT * FROM table;`

Selecting rows

`SELECT * FROM table WHERE column = value;`

Arithmetic queries

- `SELECT COUNT(column) FROM table WHERE condition;`
- `SELECT AVG(column) FROM table WHERE condition;`
- `SELECT SUM(column) FROM table WHERE condition;`

Introduction

Database access models

Privacy in databases

k-anonymity

Differential privacy

Anonymisation

Example 2 (Typical relational database in Tinder)

Birthday	Name	Height	Weight	Age	Postcode	Professio
06/07	Li Pu	190	80	60-70	1001	Politicia
06/14	Sara Lee	185	110	70+	1001	Rentier
01/01	A. B. Student	170	70	40-60	6732	Time Tra

Anonymisation

Example 2 (Typical relational database in Tinder)

Birthday	Name	Height	Weight	Age	Postcode	Profession
06/07		190	80	60-70	1001	Politician
06/14		185	110	70+	1001	Rentier
01/01		170	70	40-60	6732	Time Traveller

The simple act of hiding or using random identifiers is called anonymisation.

Record linkage

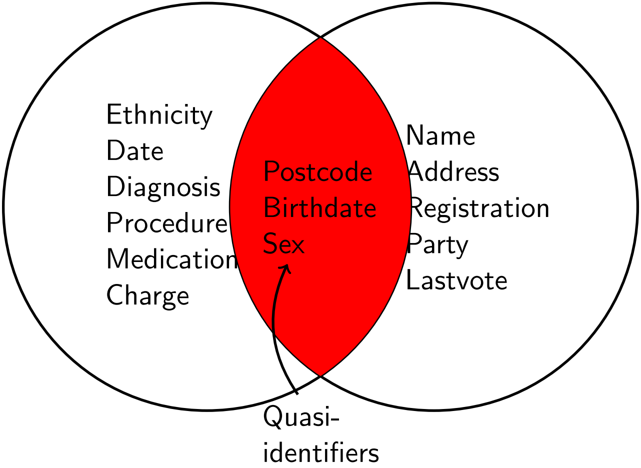


Figure : An example of two datasets, one containing sensitive and the other public information. The two datasets can be linked and individuals identified through the use of quasi-identifiers.

k-anonymity

(a) Samarati

(b) Sweeney

Definition 5 (*k*-anonymity)

A database provides *k*-anonymity if for every person in the database is indistinguishable from $k - 1$ persons with respect to *quasi-identifiers*.

It's the analyst's job to define quasi-identifiers

Birthday	Name	Height	Weight	Age	Postcode	Pr
06/07	Li Pu	190	80	60+	1001	Po
06/14	Sara Lee	185	110	60+	1001	Re
06/12	Nikos Papadopoulos	170	82	60+	1243	Po
01/01	A. B. Student	170	70	40-60	6732	Ti
05/08	Li Yang	175	72	30-40	6910	Ti

Table : 1-anonymity.

Birthday	Name	Height	Weight	Age	Postcode	Profession
06/07		190	80	60+	1001	Politician
06/14		185	110	60+	1001	Rentier
06/12		170	82	60+	1243	Politician
01/01		170	70	40-60	6732	Time Traveller
05/08		175	72	30-40	6910	Policeman

1-anonymity

Birthday	Name	Height	Weight	Age	Postcode	Profession
06/07		180-190	80+	60+	1*	
06/14		180-190	80+	60+	1*	
06/12		170-180	60+	60+	1*	
01/01		170-180	60-80	20-60	6*	
05/08		170-180	60-80	20-60	6*	

1-anonymity

Birthday	Name	Height	Weight	Age	Postcode	Profession
		180-190	80+	60+	1*	
		180-190	80+	60+	1*	
		170-180	60-80	69+	1*	
		170-180	60-80	20-60	6*	
		170-180	60-80	20-60	6*	

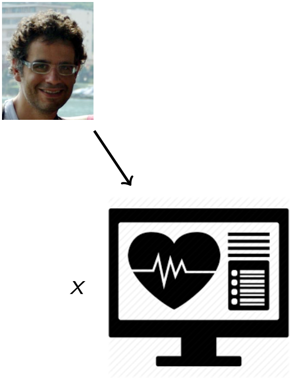
Table : 2-anonymity: the database can be partitioned in sets of at least 2 records

x_1

x

Figure : If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

x_1



x

Figure : If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

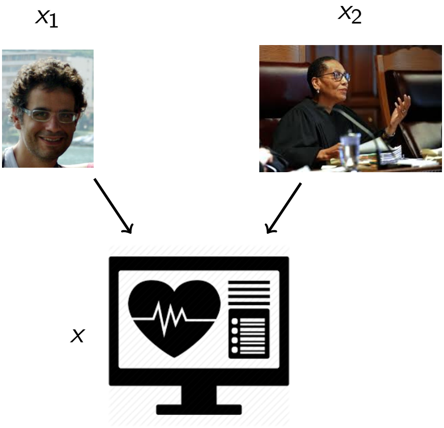


Figure : If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

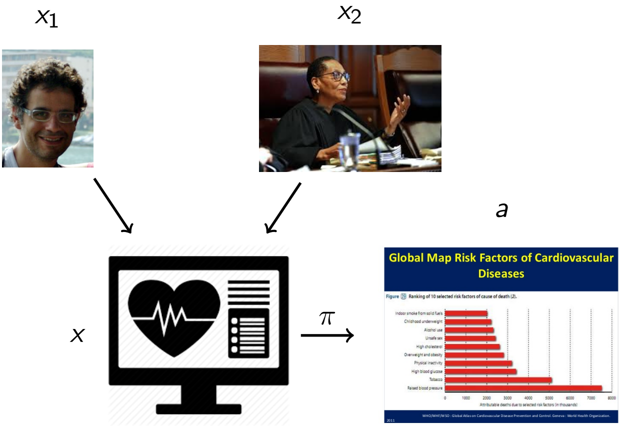


Figure : If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

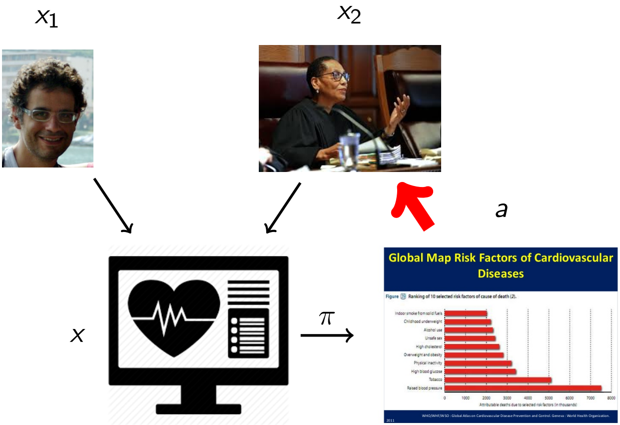


Figure : If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

Privacy desiderata

We wish to calculate something on some private data and publish a privacy-preserving, but useful, version of the result.

- Anonymity: Individual participation remains hidden.
- Secrecy: Individual data x_i is not revealed.
- Side-information: Linkage attacks are not possible.
- Utility: The calculation remains useful.

Example: The prevalence of drug use in sport

- n athletes
- Ask whether they have doped in the past year.
- Aim: calculate % of doping.
- How can we get truthful / accurate results?

Example: The prevalence of drug use in sport

- n athletes
- Ask whether they have doped in the past year.
- Aim: calculate % of doping.
- How can we get truthful / accurate results?

Algorithm for randomising responses about drug use

1. Flip a coin.
2. If it comes heads, respond truthfully.
3. Otherwise, flip another coin and respond `yes` if it comes heads and `no` otherwise.

Exercise 1

Assume that the observed rate of positive responses in a sample is p, that everybody follows the protocol, and the coin is fair. Then, what is the true rate q of drug use in the population?

Example: The prevalence of drug use in sport

- n athletes
- Ask whether they have doped in the past year.
- Aim: calculate % of doping.
- How can we get truthful / accurate results?

Solution.
Since the responses are random, we will deal with expectations first

$$\mathbb{E}\, p = \frac{1}{2} \times \frac{1}{2} + q \times \frac{1}{2}$$

□

Example: The prevalence of drug use in sport

- n athletes
- Ask whether they have doped in the past year.
- Aim: calculate % of doping.
- How can we get truthful / accurate results?

Solution.

Since the responses are random, we will deal with expectations first

$$\mathbb{E}\, p = \frac{1}{2} \times \frac{1}{2} + q \times \frac{1}{2} = \frac{1}{4} + \frac{q}{2}$$

Example: The prevalence of drug use in sport

- n athletes
- Ask whether they have doped in the past year.
- Aim: calculate % of doping.
- How can we get truthful / accurate results?

Solution.

Since the responses are random, we will deal with expectations first

$$\mathbb{E}\, p = \frac{1}{2} \times \frac{1}{2} + q \times \frac{1}{2} = \frac{1}{4} + \frac{q}{2}$$
$$q = 2\,\mathbb{E}\, p - \frac{1}{2}.$$

□

The randomised response mechanism

Definition 6 (Randomised response)

The i-th user, whose data is $x_i \in \{0, 1\}$, responds with $a_i \in \{0, 1\}$ with probability

$$\pi(a_i = j \mid x_i = k) = p, \qquad \pi(a_i = k \mid x_i = k) = 1 - p,$$

where $j \neq k$.

The randomised response mechanism

Definition 6 (Randomised response)

The i-th user, whose data is $x_i \in \{0, 1\}$, responds with $a_i \in \{0, 1\}$ with probability

$$\pi(a_i = j \mid x_i = k) = p, \qquad \pi(a_i = k \mid x_i = k) = 1 - p,$$

where $j \neq k$.

Given the complete data x, the mechanism's output is $a = (a_1, \ldots, a_n)$.

The randomised response mechanism

Definition 6 (Randomised response)

The i-th user, whose data is $x_i \in \{0, 1\}$, responds with $a_i \in \{0, 1\}$ with probability

$$\pi(a_i = j \mid x_i = k) = p, \qquad \pi(a_i = k \mid x_i = k) = 1 - p,$$

where $j \neq k$.

Given the complete data x, the mechanism's output is $a = (a_1, \ldots, a_n)$. Since the algorithm independently calculates a new value for each data entry, the output is

$$\pi(a \mid x) = \prod_i \pi(a_i \mid x_i)$$

Exercise 1

Let the adversary have a prior $\xi(x = 0) = 1 - \xi(x = 1)$ over the values of the true response of an individual. we use the randomised response mechanism with p and the adversary observes the randomised data $a = 1$ for that individual, then what is $\xi(x = 1 \mid a = 1)$?

The local privacy model

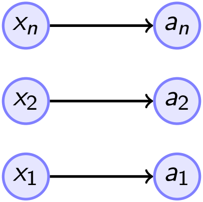
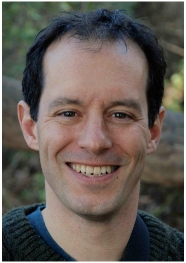


Figure : The local privacy model

Differential privacy.

Definition 7 (ϵ-Differential Privacy)

A stochastic algorithm $\pi : \mathcal{X} \to \mathcal{A}$, where $\mathcal{X}$ is endowed with a neighbourhood relation N, is said to be ϵ-differentially private if

$$\left| \ln \frac{\pi(a \mid x)}{\pi(a \mid x')} \right| \leq \epsilon, \qquad \forall x N x'. \tag{5.1}$$

The definition of differential privacy

- First rigorous mathematical definition of privacy.
- Relaxations and generalisations possible.
- Connection to learning theory and reproducibility.

Current uses

- Apple.
- Google.
- Uber.
- US 2020 Census.

Open problems

- Complexity of differential privacy.
- Verification of implementations and queries.

Remark 1

The randomised response mechanism with $p \leq 1/2$ is $(\ln \frac{1-p}{p})$-DP.

Proof.

Consider $x = (x_1, \ldots, x_j, \ldots, x_n)$, $x' = (x_1, \ldots, x_j', \ldots, x_n)$. Then

$$\pi(a \mid x)$$

$\square$

Remark 1

The randomised response mechanism with $p \leq 1/2$ is $(\ln \frac{1-p}{p})$-DP.

Proof.

Consider $x = (x_1, \ldots, x_j, \ldots, x_n)$, $x' = (x_1, \ldots, x'_j, \ldots, x_n)$. Then

$$\pi(a \mid x) = \prod_i \pi(a_i \mid x_i)$$

□

Remark 1

The randomised response mechanism with $p \leq 1/2$ is $(\ln \frac{1-p}{p})$-DP.

Proof.

Consider $x = (x_1, \ldots, x_j, \ldots, x_n)$, $x' = (x_1, \ldots, x'_j, \ldots, x_n)$. Then

$$\begin{aligned}
\pi(a \mid x) &= \prod_i \pi(a_i \mid x_i) \\
&= \pi(a_j \mid x_j) \prod_{i \neq j} \pi(a_i \mid x_i)
\end{aligned}$$

$\square$

Remark 1

The randomised response mechanism with $p \leq 1/2$ is $(\ln \frac{1-p}{p})$-DP.

Proof.

Consider $x = (x_1, \ldots, x_j, \ldots, x_n)$, $x' = (x_1, \ldots, x'_j, \ldots, x_n)$. Then

$$\begin{aligned}
\pi(a \mid x) &= \prod_i \pi(a_i \mid x_i) \\
&= \pi(a_j \mid x_j) \prod_{i \neq j} \pi(a_i \mid x_i) \\
&\leq \frac{p}{1-p} \pi(a_j \mid x'_j) \prod_{i \neq j} \pi(a_i \mid x_i)
\end{aligned}$$

$\pi(a_j = k \mid x_j = k) = 1 - p$ so the ratio is $\max\{(1-p)/p, p/(1-p)\} \leq (1-p)/p$ for $p \leq 1/2$. □

Remark 1

The randomised response mechanism with $p \leq 1/2$ is $(\ln \frac{1-p}{p})$-DP.

Proof.

Consider $x = (x_1, \ldots, x_j, \ldots, x_n)$, $x' = (x_1, \ldots, x'_j, \ldots, x_n)$. Then

$$
\begin{aligned}
\pi(a \mid x) &= \prod_i \pi(a_i \mid x_i) \\
&= \pi(a_j \mid x_j) \prod_{i \neq j} \pi(a_i \mid x_i) \\
&\leq \frac{p}{1-p} \pi(a_j \mid x'_j) \prod_{i \neq j} \pi(a_i \mid x_i) \\
&= \frac{1-p}{p} \pi(a \mid x')
\end{aligned}
$$

□

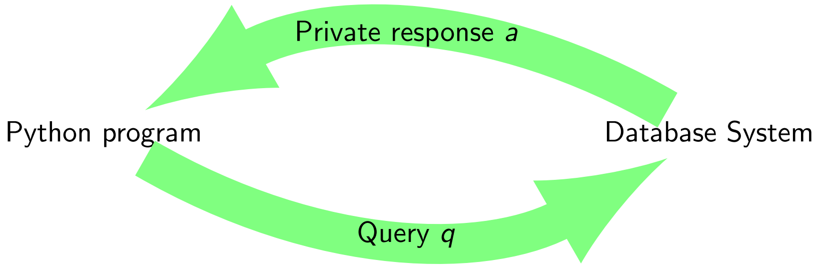


Figure : Private database access model

Response policy

The policy defines a distribution over responses a given the data x and the query q.

$$\pi(a \mid x, q)$$

Differentially private queries

The DP-SELECT statement

- `DP-SELECT` ϵ `column1, column2 FROM table;`
- `DP-SELECT` ϵ `* FROM table;`

Selecting rows

`DP-SELECT` ϵ `* FROM table WHERE column = value;`

Arithmetic queries

- `DP-SELECT` ϵ `COUNT(column) FROM table WHERE condition;`
- `DP-SELECT` ϵ `AVG(column) FROM table WHERE condition;`
- `DP-SELECT` ϵ `SUM(column) FROM table WHERE condition;`

Composition

If we answer T queries with an ϵ-DP mechanism, then our cumulative privacy loss is ϵT.

Exercise 2

Adversary knowledge

$$\boldsymbol{x} = (x_1, \ldots, x_j = 0, \ldots, x_n)$$

$$\boldsymbol{x}' = (x_1, \ldots, x_j = 1, \ldots, x_n).$$

$$\xi(\boldsymbol{x}) = 1 - \xi(\boldsymbol{x}')$$

What can we say about the posterior distribution of the adversary $\xi(\boldsymbol{x} \mid a, \pi)$ after having seen the output, if π is ϵ-DP?

Exercise 2

Adversary knowledge

$$\boldsymbol{x} = (x_1, \ldots, x_j = 0, \ldots, x_n)$$

$$\boldsymbol{x}' = (x_1, \ldots, x_j = 1, \ldots, x_n).$$

$$\xi(\boldsymbol{x}) = 1 - \xi(\boldsymbol{x}')$$

$$a_t, \qquad \pi(a_t \mid \boldsymbol{x}_t) \Rightarrow \begin{cases} \pi(a_t \mid \boldsymbol{x}_t = \boldsymbol{x}) \\ \pi(a_t \mid \boldsymbol{x}_t = \boldsymbol{x}') \end{cases}$$

What can we say about the posterior distribution of the adversary $\xi(\boldsymbol{x} \mid a, \pi)$ after having seen the output, if π is ϵ-DP?

Dealing with multiple attributes.

Independent release of multiple attributes.

For n users and k attributes, if the release of each attribute i is ϵ-DP then the data release is $k\epsilon$-DP. Thus to get ϵ-DP overall, we need ϵ/k-DP per attribute.

The Laplace mechanism.

Definition 8 (The Laplace mechanism)

For any function $f : \mathcal{X} \to \mathbb{R}$,

$$\pi(a \mid x) = \mathit{Laplace}(f(x), \lambda), \tag{5.2}$$

where the Laplace density is defined as

$$p(\omega \mid \mu, \lambda) = \frac{1}{2\lambda} \exp\left(-\frac{|\omega - \mu|}{\lambda}\right).$$

and has mean μ and variance $2\lambda^2$.

.

Example 9 (Calculating the average salary)

- The i-th person receives salary x_i
- We wish to calculate the average salary in a private manner.

Local privacy model

- Obtain $y_i = x_i + \omega$, where $\omega \sim \mathcal{L}aplace(\lambda)$.
- Return $a = n^{-1} \sum_{i=1}^{n} y_i$.

Centralised privacy model

Return $a = n^{-1} \sum_{i=1}^{n} x_i + \omega$, where $\omega \sim \mathcal{L}aplace(\lambda')$.

How should we add noise in order to guarantee privacy?

The centralised privacy model

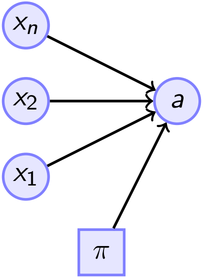


Figure : The centralised privacy model

Assumption 1

The data x is collected and the result a is published by a trusted curator

DP properties of the Laplace mechanism

Definition 10 (Sensitivity)

The sensitivity of a function f is

$$\mathbb{L}(f) \triangleq \sup_{xNx'} |f(x) - f(x')|$$

Example 11

If $f : \mathcal{X} \to [0, B]$, e.g. $\mathcal{X} = \mathbb{R}$ and $f(x) = \min\{B, \max\{0, x\}\}$, then

DP properties of the Laplace mechanism

Definition 10 (Sensitivity)

The sensitivity of a function f is

$$\mathbb{L}(f) \triangleq \sup_{xNx'} |f(x) - f(x')|$$

Example 11

If $f : \mathcal{X} \to [0, B]$, e.g. $\mathcal{X} = \mathbb{R}$ and $f(x) = \min\{B, \max\{0, x\}\}$, then $\mathbb{L}(f) = B$.

DP properties of the Laplace mechanism

Definition 10 (Sensitivity)

The sensitivity of a function f is

$$\mathbb{L}(f) \triangleq \sup_{xNx'} |f(x) - f(x')|$$

Example 11

If $f : \mathcal{X} \to [0, B]$, e.g. $\mathcal{X} = \mathbb{R}$ and $f(x) = \min\{B, \max\{0, x\}\}$, then $\mathbb{L}(f) = B$.

Example 12

If $f : [0, B]^n \to [0, B]$ is $f = \frac{1}{n} \sum_{t=1}^{n} x_t$, then

DP properties of the Laplace mechanism

Definition 10 (Sensitivity)

The sensitivity of a function f is

$$\mathbb{L}(f) \triangleq \sup_{xNx'} |f(x) - f(x')|$$

Example 11

If $f : \mathcal{X} \to [0, B]$, e.g. $\mathcal{X} = \mathbb{R}$ and $f(x) = \min\{B, \max\{0, x\}\}$, then $\mathbb{L}(f) = B$.

Example 12

If $f : [0, B]^n \to [0, B]$ is $f = \frac{1}{n} \sum_{t=1}^n x_t$, then $\mathbb{L}(f) = B/n$.

Theorem 13

The Laplace mechanism on a function f with sensitivity $\mathbb{L}(f)$, ran with $\mathcal{L}aplace(\lambda)$ is $\mathbb{L}(f)/\lambda$-DP.

Proof.

$$\frac{\pi(a \mid x)}{\pi(a \mid x')} = \frac{e^{|a-f(x')|/\lambda}}{e^{|a-f(x)|/\lambda}} \leq \frac{e^{|a-f(x)|/\lambda + \mathbb{L}(f)/\lambda}}{e^{|a-f(x)|/\lambda}} = e^{\mathbb{L}(f)/\lambda}$$

□

So we need to use $\lambda = \mathbb{L}(f)/\epsilon$ for ϵ-DP. What is the effect of applying the Laplace mechanism in the local versus centralised model?

Interactive queries

- System has data x.
- User asks query q.
- System responds with a.
- There is a common utility function $U : \mathcal{X}, \mathcal{A}, \mathcal{Q} \to \mathbb{R}$.

We wish to maximisation U with our answers, but are constrained by the fact that we also want to preserve privacy.

The Exponential Mechanism.

Definition 14 (The Exponential mechanism)

For any utility function $U : \mathcal{Q} \times \mathcal{A} \times \mathcal{X} \to \mathbb{R}$, define the policy

$$\pi(a \mid x) \triangleq \frac{e^{\epsilon U(q,a,x)/\mathbb{L}(U(q))}}{\sum_{a'} e^{\epsilon U(q,a',x)/\mathbb{L}(U(q))}} \tag{5.3}$$

What happens when $\epsilon \to \infty$? What about when $\epsilon \to 0$?

The unfortunate practice of adaptive analysis

Prior

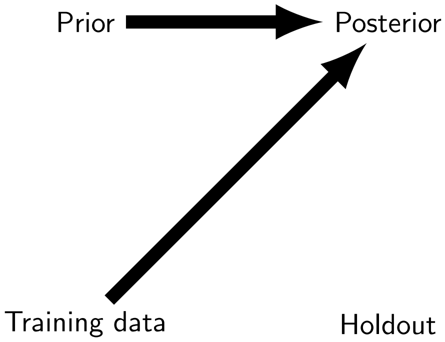
The unfortunate practice of adaptive analysis

Prior

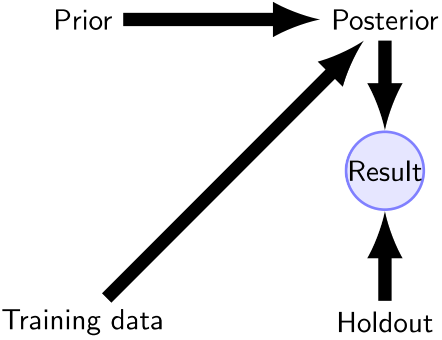
Training data

Holdout

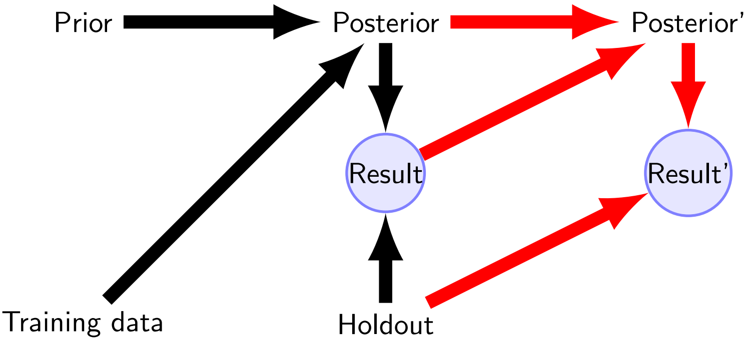
The unfortunate practice of adaptive analysis



The unfortunate practice of adaptive analysis



The unfortunate practice of adaptive analysis



The reusable holdout? [1]

Algorithm parameters

- Performance measure f.
- Threshold τ.
- Noise σ.
- Budget B.

Algorithm idea

Run algorithm λ on data D_T and get e.g. classifier parameters θ.
Run a DP version of the function
$f(\theta, D_H) = \mathbb{I}\left\{U(\theta, D_T) \geq \tau U(\theta, D_H)\right\}$.

[1] Also see
https://ai.googleblog.com/2015/08/the-reusable-holdout-preserving.html

Available privacy toolboxes

k-anonymity

- `https://github.com/qiyuangong/Mondrian` Mondrian k-anonymity

Differential privacy

- `https://github.com/bmcmenamin/thresholdOut-explorations` Threshold out
- `https://github.com/steven7woo/Accuracy-First-Differential-Privacy` Accuracy-constrained DP
- `https://github.com/menisadi/pydp` Various DP algorithms
- `https://github.com/haiphanNJIT/PrivateDeepLearning` Deep learning and DP

Learning outcomes

Understanding

- Linkage attacks and k-anonymity.
- Inferring data from summary statistics.
- The local versus global differential privacy model.
- False discovery rates.

Skills

- Make a dataset satisfy k-anonymity with respect to identifying attributes.
- Apply the randomised response and Laplace mechanism to data.
- Apply the exponential mechanism to simple decision problems.
- Use differential privacy to improve reproducibility.

Reflection