Differential privacy and machine learning

Kamalika Chaudhuri
Dept. of CSE
UC San Diego

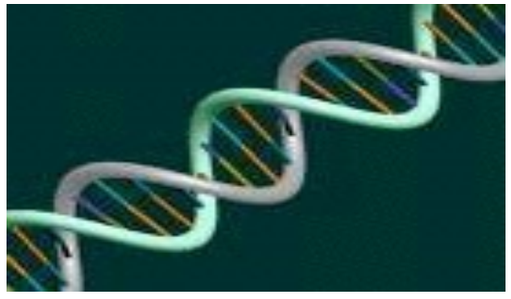
Anand D. Sarwate
Dept. of ECE
Rutgers University

Some Motivation

Sensitive Data

Medical Records

Genetic Data



Search Logs

Google

AOL Violates Privacy

AOL Violates Privacy

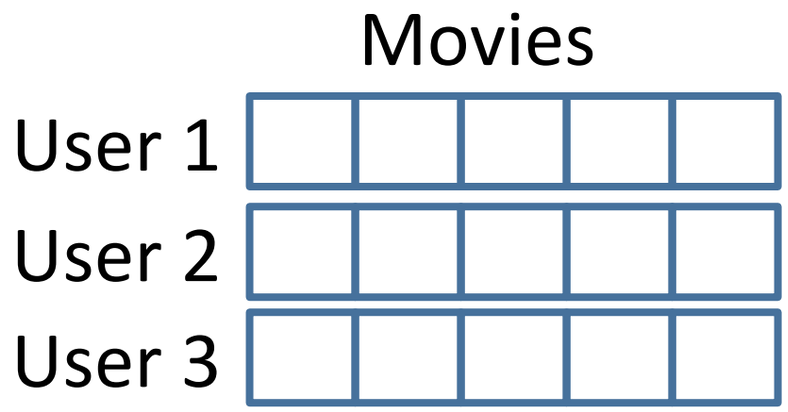
A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on

Netflix Violates Privacy [NS08]



2-8 movie-ratings and dates for Alice reveals:

Whether Alice is in the dataset or not

Alice's other movie ratings

High-dimensional Data is Unique

Example: UCSD Employee Salary Table

Position	Gender	Department	Ethnicity	Salary
Faculty	Female	CSE	SE Asian	-

One employee (Kamalika) fits description!

Simply anonymizing data is **unsafe!**

Disease Association Studies [WLWTZ09]

Cancer

1.00										
.190	1.00									
.216	.251	1.00								
.186	.117	.047	1.00							
.154	.011	.170	.083	1.00						
.190	.140	.102	.095	.139	1.00					
.270	.215	.294	.248	.140	.141	1.00				
.101	.085	.170	.056	.234	.099	.175	1.00			
.239	.071	.163	.111	.161	.093	.199	.157	1.00		
.471	.117	.243	.094	.144	.123	.283	.216	.274	1.00	
.179	.202	.132	.094	.087	.159	.207	.108	.092	.294	1.00

Correlations

Healthy

1.00										
.141	1.00									
.099	.175	1.00								
.093	.199	.157	1.00							
.123	.283	.216	.274	1.00						
.159	.207	.108	.092	.294	1.00					
.088	.152	.075	.163	.156	.220	1.00				
.046	.161	.092	.072	.157	.143	.147	1.00			
.078	.392	.122	.229	.160	.172	.145	.177	1.00		
.045	.155	.135	.139	.110	.048	.126	.104	.169	1.00	
.178	.135	.102	.258	.314	.165	.147	.158	.131	.074	1.00

Correlations

Correlation (R^2 values), Alice's DNA reveals:

If Alice is in the **Cancer** set or **Healthy** set

Simply anonymizing data is **unsafe!**

Simply anonymizing data is **unsafe!**

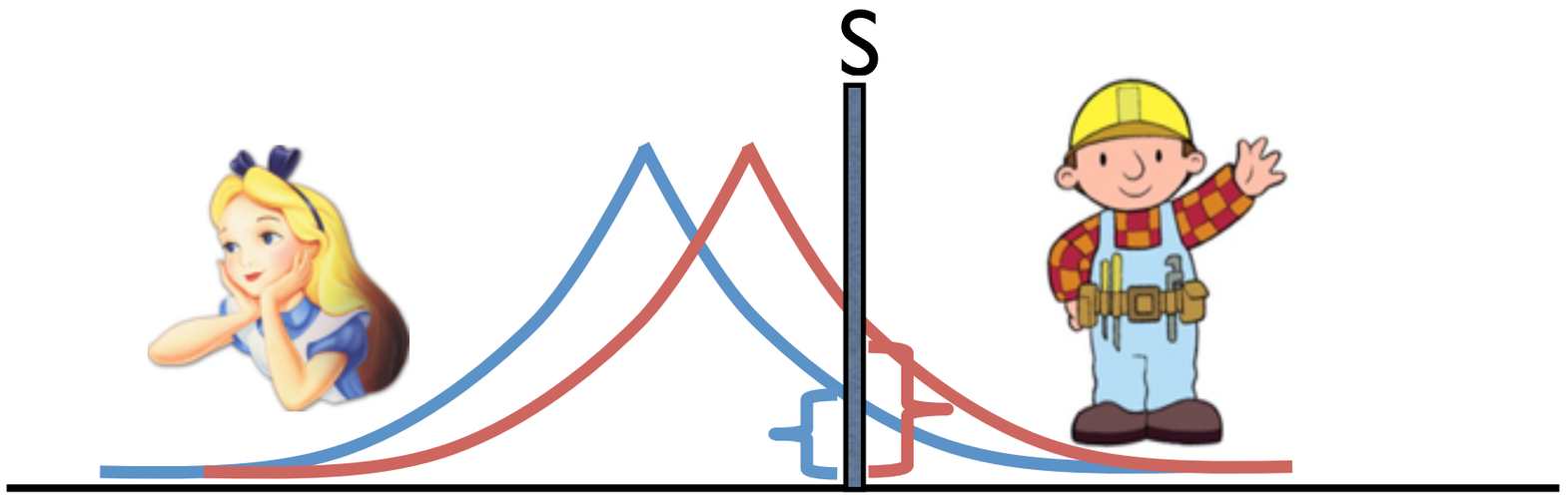
Releasing a lot of statistics based on raw data is **unsafe!**

The schedule

1. Privacy definitions
2. Sensitivity and guaranteeing privacy

— INTERMISSION —

3. Beyond sensitivity
4. Practicalities
5. Applications & Extensions



Formally defining privacy

The Setting

Data
(sensitive)

Sanitizer

Statistics

Data release

Public

Property of Sanitizer

Data
(sensitive)

Sanitizer

Statistics

Data release

Public

Aggregate information computable

Individual information protected
(robust to side-information)

Property of Sanitizer

Statistics

Data release

Data
(sensitive)

Sanitizer

Public

Aggregate information computable

Individual information protected
(robust to side-information)

**Participation of individual
does not change outcome**

Adversary

Prior Knowledge:

A's Genetic profile

A smokes

Case 1: Study

1.00										
.190	1.00									
.216	.251	1.00								
.186	.117	.047	1.00							
.154	.011	.170	.083	1.00						
.190	.140	.102	.095	.139	1.00					
.270	.215	.294	.248	.140	.141	1.00				
.101	.085	.170	.056	.234	.099	.175	1.00			
.239	.071	.163	.111	.161	.093	.199	.157	1.00		
.471	.117	.243	.094	.144	.123	.283	.216	.274	1.00	
.179	.202	.132	.094	.087	.159	.207	.108	.092	.294	1.00

Cancer

A has cancer

[Study violates A's privacy]

Adversary

Prior Knowledge:

A's Genetic profile

A smokes

Adversary

Prior Knowledge:

A's Genetic profile

A smokes

Case 1: Study

1.00										
.190	1.00									
.216	.251	1.00								
.186	.117	.047	1.00							
.154	.011	.170	.083	1.00						
.190	.140	.102	.095	.139	1.00					
.270	.215	.294	.248	.140	.141	1.00				
.101	.085	.170	.056	.234	.099	.175	1.00			
.239	.071	.163	.111	.161	.093	.199	.157	1.00		
.471	.117	.243	.094	.144	.123	.283	.216	.274	1.00	
.179	.202	.132	.094	.087	.159	.207	.108	.092	.294	1.00

Cancer

A has cancer

[Study violates A's privacy]

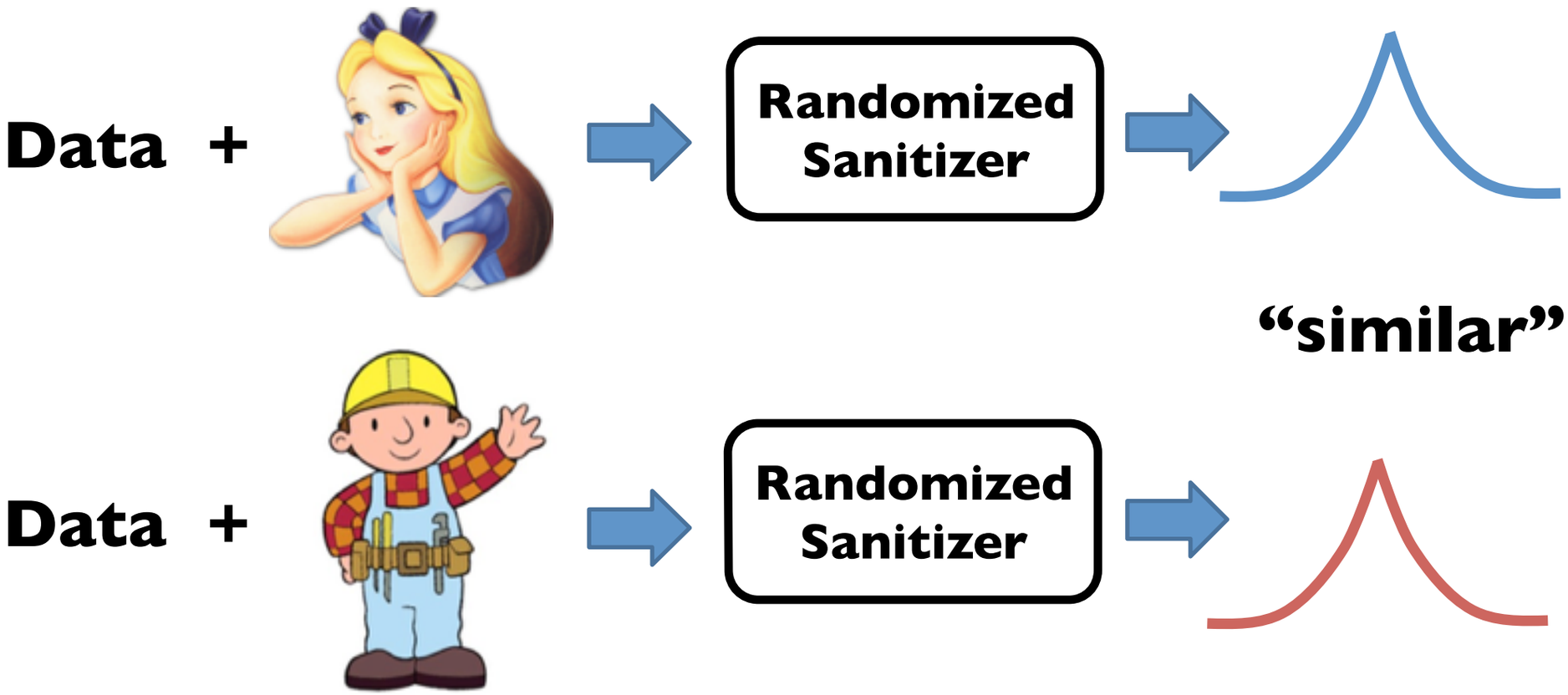
Case 2: Study

Smoking causes cancer

A probably has cancer

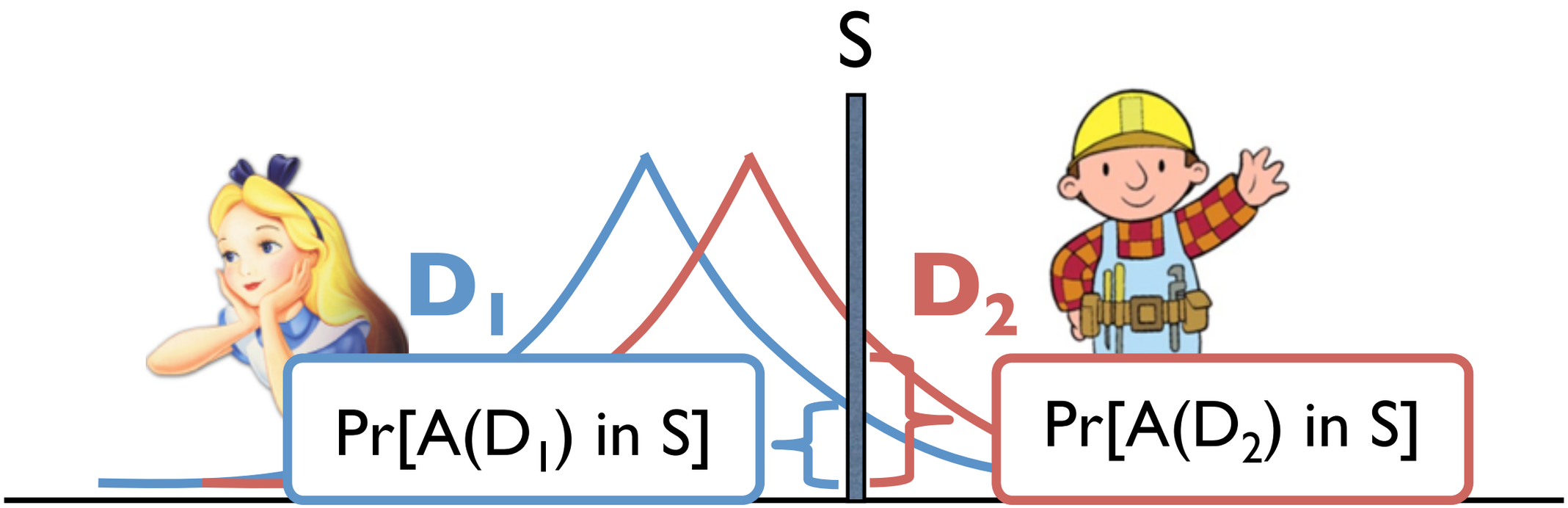
[Study does not violate privacy]

Differential Privacy [DMNS06]



Participation of single person does not change outcome

Differential Privacy [DMNS06]

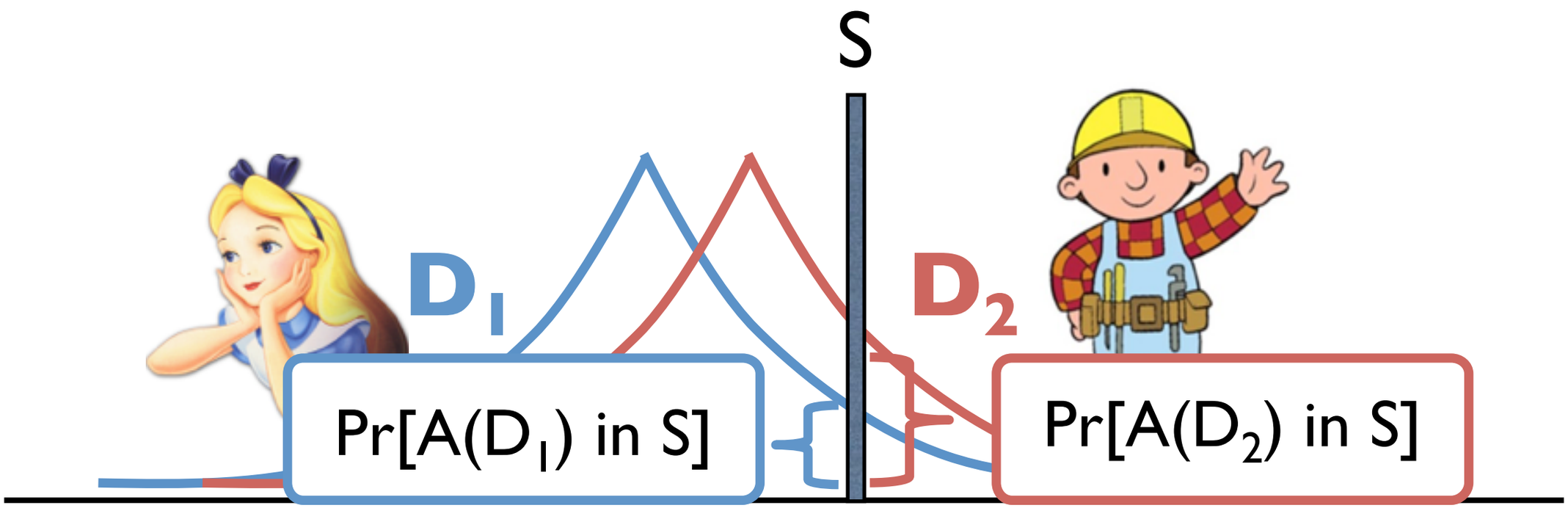


For all D_1, D_2 that differ in one person's value, any set S,

If A = (ϵ, δ)-differentially private randomized algorithm, then:

$$\Pr(A(D_1) \in S) \leq e^{\epsilon} \Pr(A(D_2) \in S) + \delta$$

Differential Privacy [DMNS06]



For all D_1, D_2 that differ in one person's value, any set S,

If A = (ϵ, δ)-differentially private randomized algorithm, then:

$$\Pr(A(D_1) \in S) \leq e^{\epsilon} \Pr(A(D_2) \in S) + \delta$$

Pure differential privacy: $\delta = 0$

Attacker's Hypothesis Test [WZ10, OV13]

H0: Input to algorithm: Data + 

H1: Input to algorithm: Data + 

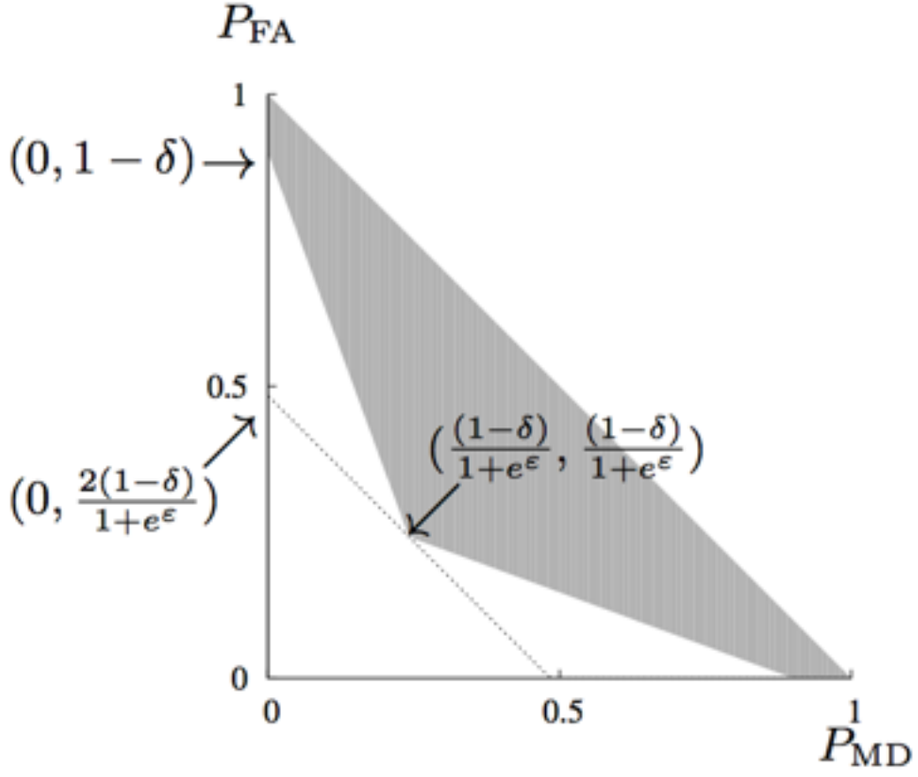
Attacker's Hypothesis Test [WZ10, OV13]

H0: Input to algorithm: Data + 

H1: Input to algorithm: Data + 

Failure Events: False alarm (FA), Missed Detection (MD)

Attacker's Hypothesis Test [WZ10, OV13]



If algorithm is (ϵ, δ)-DP, then

$$\Pr(FA) + e^{\epsilon} \Pr(MD) \geq 1 - \delta$$

$$e^{\epsilon} \Pr(FA) + \Pr(MD) \geq 1 - \delta$$

images: [OV13]

An Example Privacy Mechanism

Privacy from Perturbation

Example: Mean of $x_1, \ldots, x_n$, where x_i in [0, 1]

Privacy from Perturbation

Example: Mean of $x_1, \ldots, x_n$, where x_i in [0, 1]

Mechanism:

1. Calculate mean: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

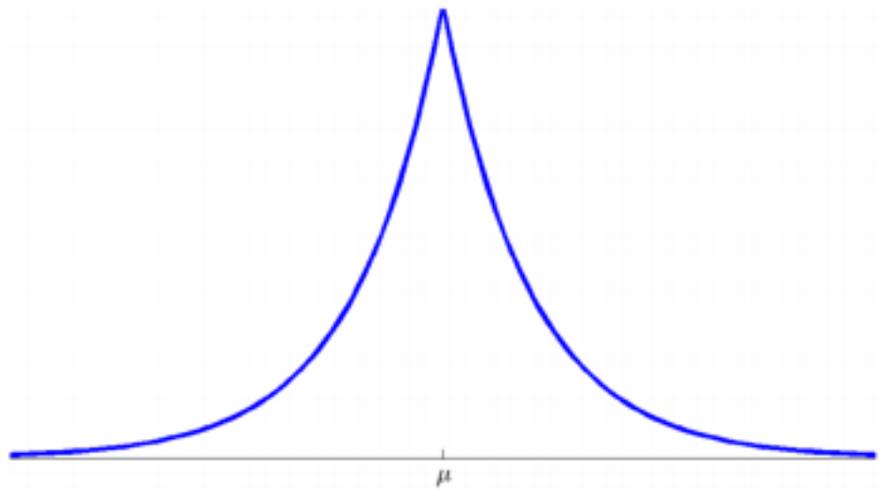
Privacy from Perturbation

Example: Mean of $x_1, \ldots, x_n$, where x_i in [0, 1]

Mechanism:

1. Calculate mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. Output:

$$\bar{x} + \frac{1}{n\epsilon} Z, \text{ where } Z \sim \text{Lap}(0, 1)$$



Laplace density

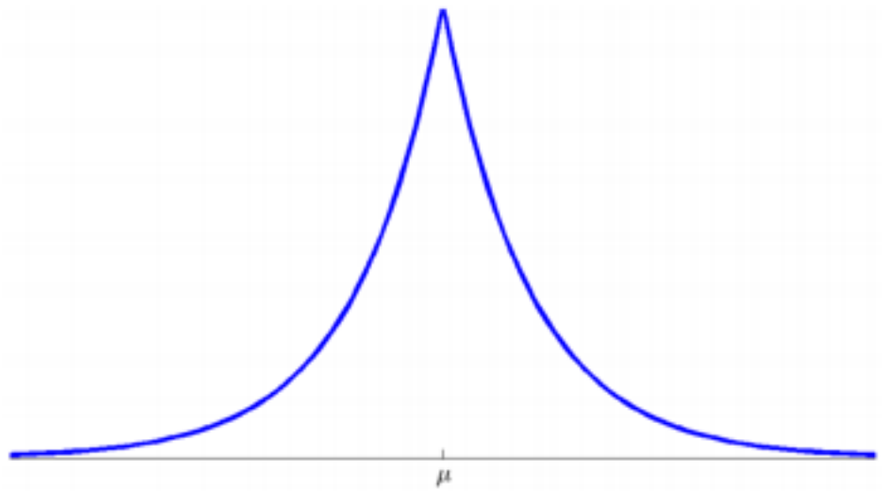
Privacy from Perturbation

Example: Mean of $x_1, \ldots, x_n$, where x_i in [0, 1]

Mechanism:

1. Calculate mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. Output:

$$\bar{x} + \frac{1}{n\epsilon} Z, \text{ where } Z \sim \text{Lap}(0, 1)$$

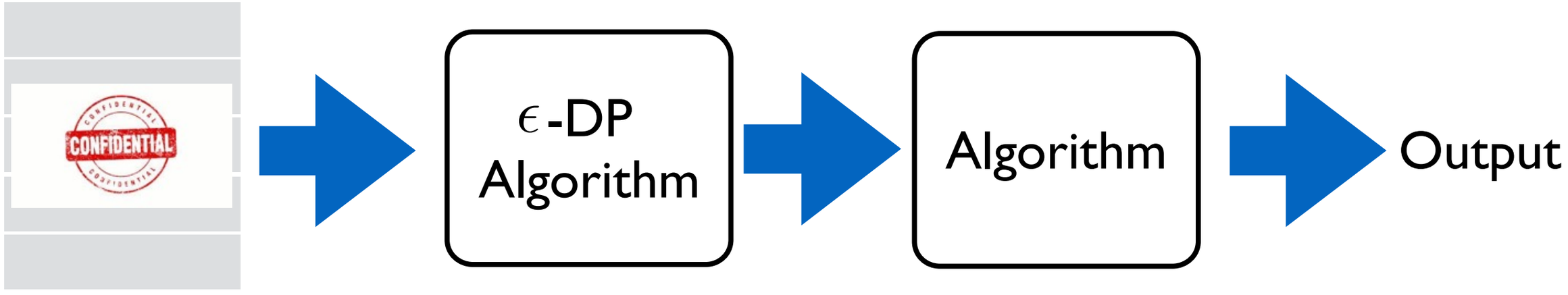


Laplace density

More Examples Coming Up!

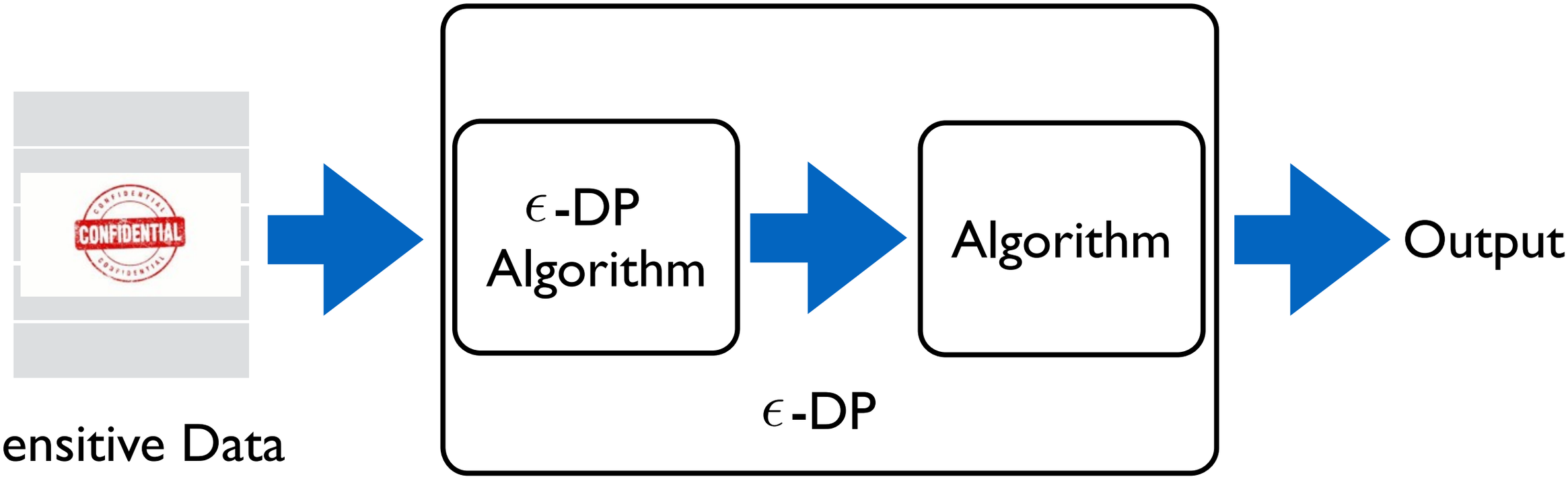
Properties of Differential Privacy

Property 1: Postprocessing Invariance

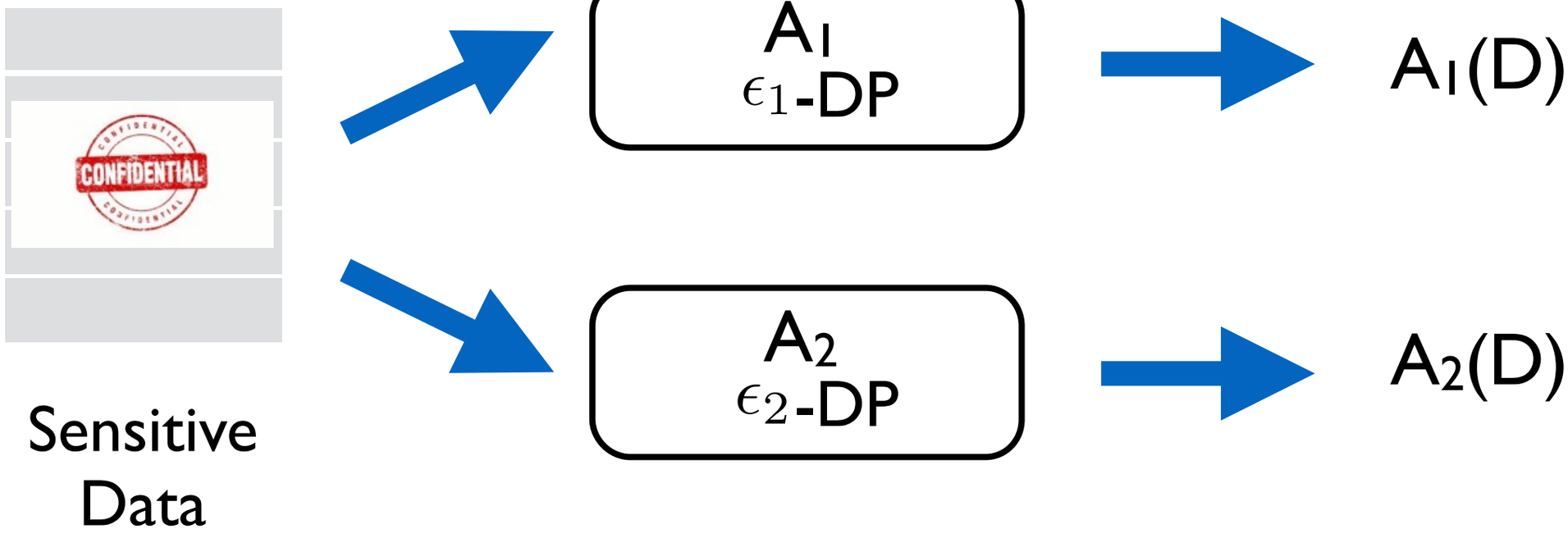


Sensitive Data

Property 1: Postprocessing Invariance



Property 2: Composition



If A_1 is ϵ_1-DP and A_2 is ϵ_2-DP, then the union $(A_1(D), A_2(D))$ is $(\epsilon_1 + \epsilon_2)$-DP

More Advanced Composition Theorems: [DRV09, OV13]

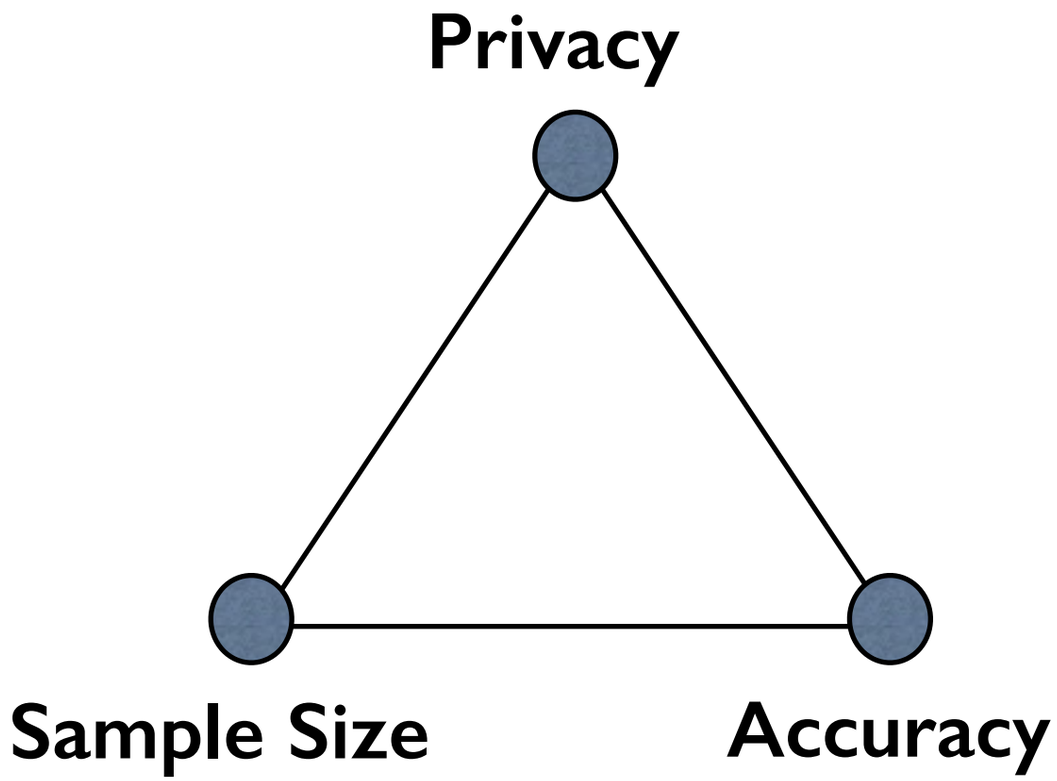
Property 3: Quantifiability

Amount of perturbation to get privacy is quantifiable

Properties of Differential Privacy

1. Postprocessing invariance
2. Composition
3. Quantifiability

The Price of Privacy


Privacy
Sample Size
Accuracy

How to Ensure Differential Privacy?

Private Data Release

[DMNS06] Data release faithful to **all query classes**: difficult

[BLR13, HLM12] Data release faithful to specific query classes

See tutorial by Miklau (2013) for details

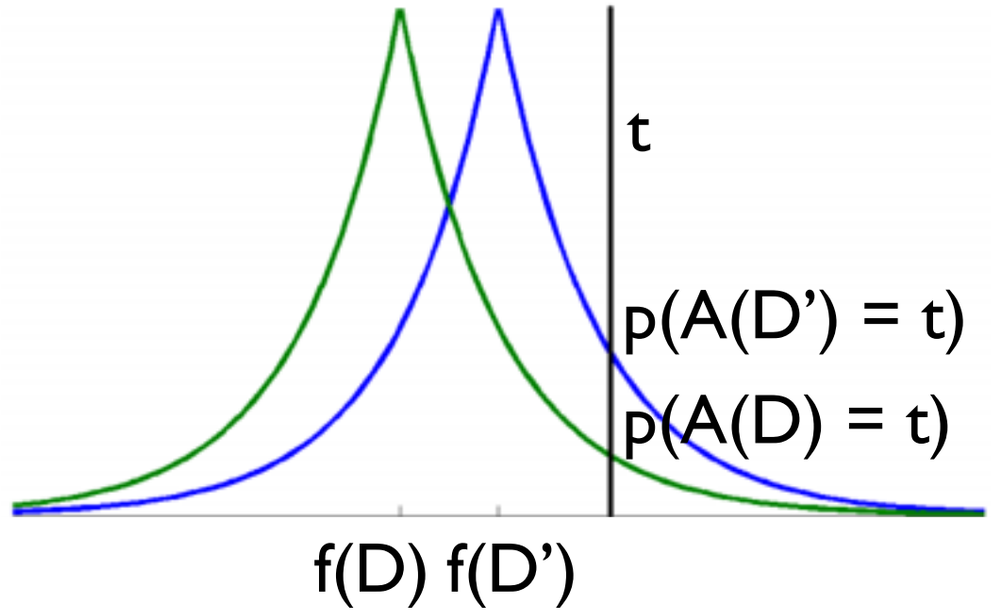
This Talk: Answering Queries on Sensitive Data

The schedule

1. Privacy definitions
2. Sensitivity and guaranteeing privacy

— INTERMISSION —

3. Beyond sensitivity
4. Practicalities
5. Applications & Extensions



Differential privacy and sensitivity

The Problem

Given: Function f, Sensitive Data D

Find: Differentially private approximation to f(D)

Goal: Good privacy-accuracy-sample size tradeoff

The Global Sensitivity Method [DMNS06]

Given: A function f, sensitive dataset D

The Global Sensitivity Method [DMNS06]

Given: A function f, sensitive dataset D

Define: dist(D, D') = #individual records D, D' differ by

The Global Sensitivity Method [DMNS06]

Given: A function f, sensitive dataset D

Define: dist(D, D') = #individual records D, D' differ by

Global Sensitivity of f:

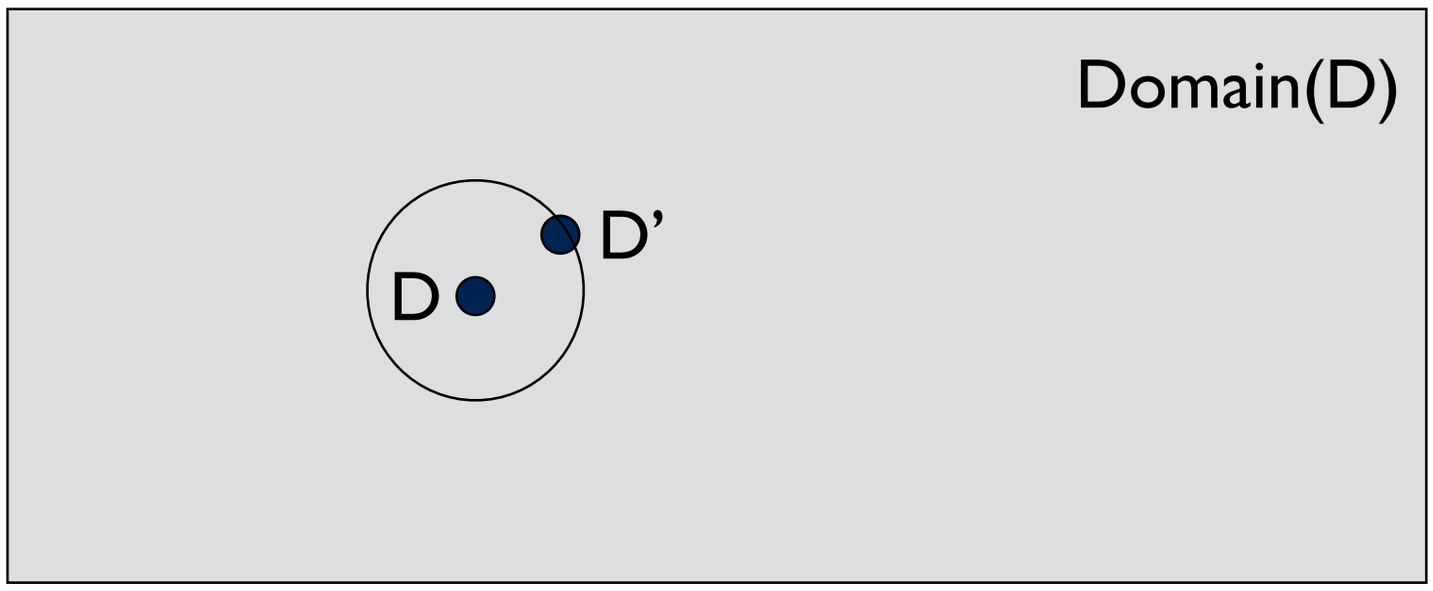
The Global Sensitivity Method [DMNS06]

Given: A function f, sensitive dataset D

Define: dist(D, D') = #individual records D, D' differ by

Global Sensitivity of f:

$$S(f) = \max_{dist(D,D')=1} |f(D) - f(D')|$$



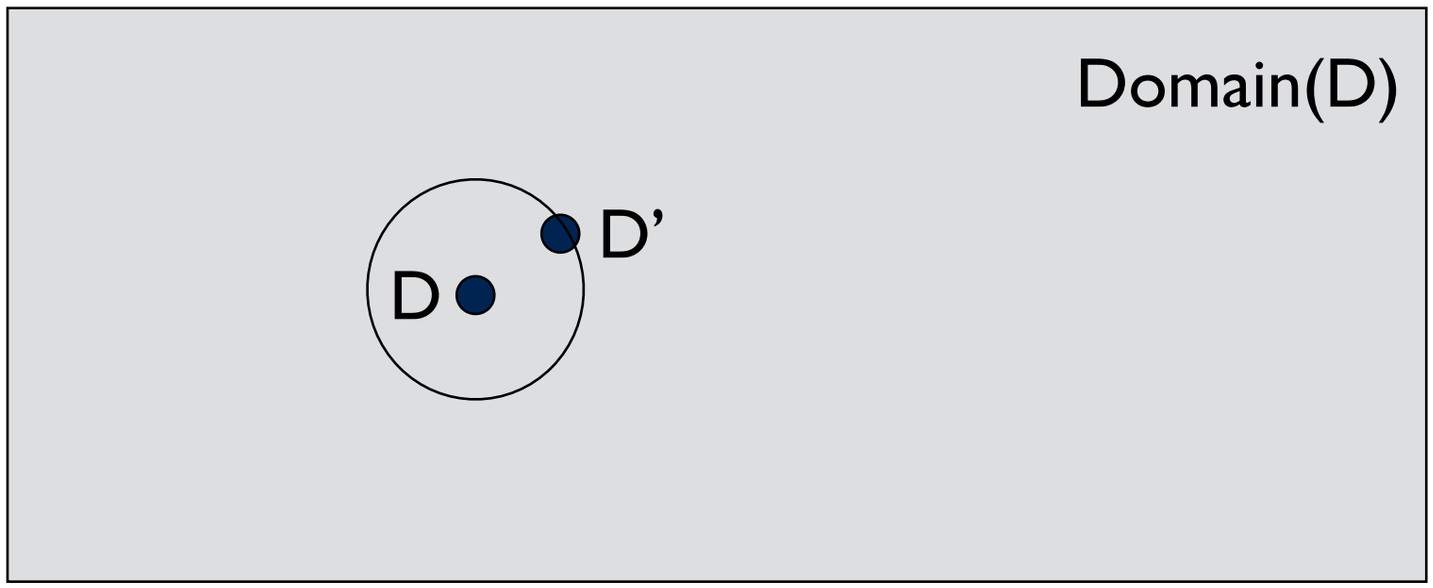
The Global Sensitivity Method [DMNS06]

Given: A function f, sensitive dataset D

Define: dist(D, D') = #individual records D, D' differ by

Global Sensitivity of f:

$$S(f) = \max_{dist(D, D') = 1} |f(D) - f(D')|$$



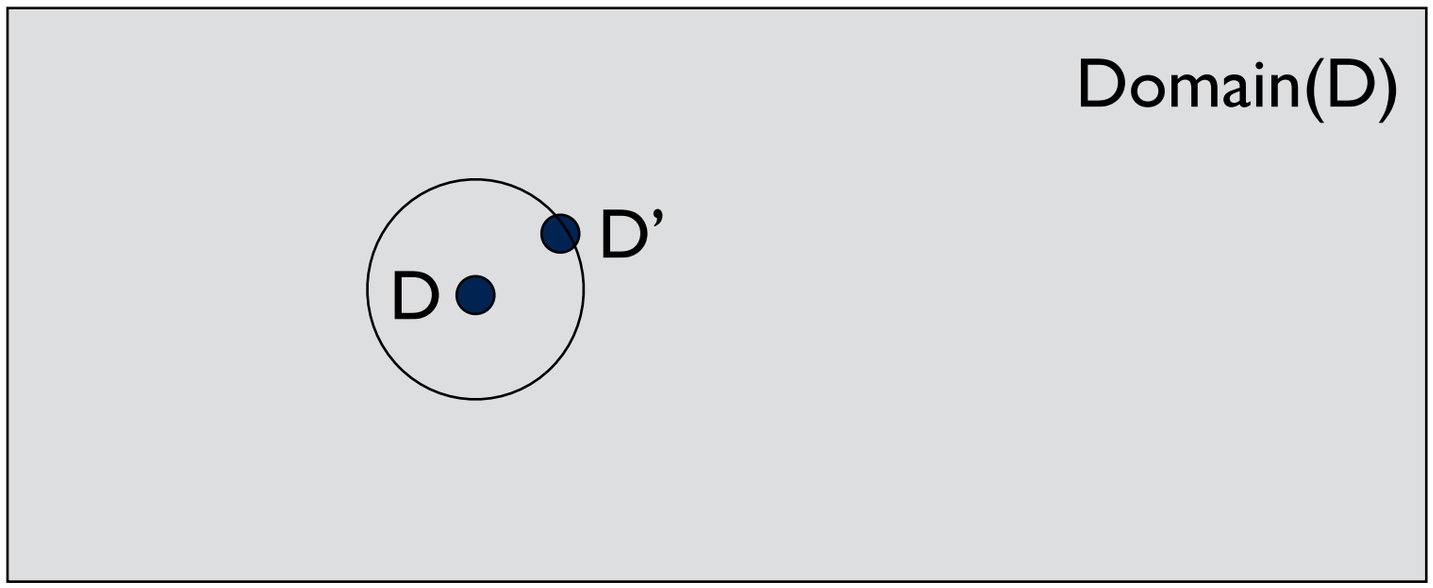
The Global Sensitivity Method [DMNS06]

Given: A function f, sensitive dataset D

Define: dist(D, D') = #individual records D, D' differ by

Global Sensitivity of f:

$$S(f) = \max_{dist(D, D') = 1} |f(D) - f(D')|$$



The Global Sensitivity Method [DMNS06]

Global Sensitivity of f:

$$S(f) = \max_{\text{dist}(D, D') = 1} |f(D) - f(D')|$$

Global Sensitivity Method:

Output $f(D) + Z$, where $Z \sim \frac{S(f)}{\epsilon}\text{Lap}(0, 1)$ (Privacy ϵ)

The Global Sensitivity Method [DMNS06]

Global Sensitivity of f:

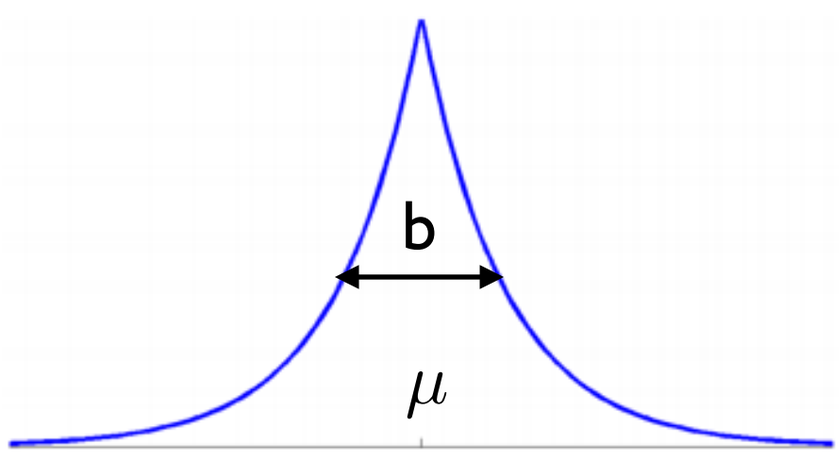
$$S(f) = \max_{dist(D, D') = 1} |f(D) - f(D')|$$

Global Sensitivity Method:

Output $f(D) + Z$, where $Z \sim \frac{S(f)}{\epsilon}\text{Lap}(0, 1)$ (Privacy ϵ)

Laplace Distribution:

$$p(z|\mu, b) = \frac{1}{2b}\exp\left(-\frac{|z - \mu|}{b}\right)$$



Privacy Proof

Global Sensitivity Method:

Output $f(D) + Z$, where $Z \sim \frac{S(f)}{\epsilon}\mathrm{Lap}(0,1)$ (Privacy ϵ)

Privacy Proof:

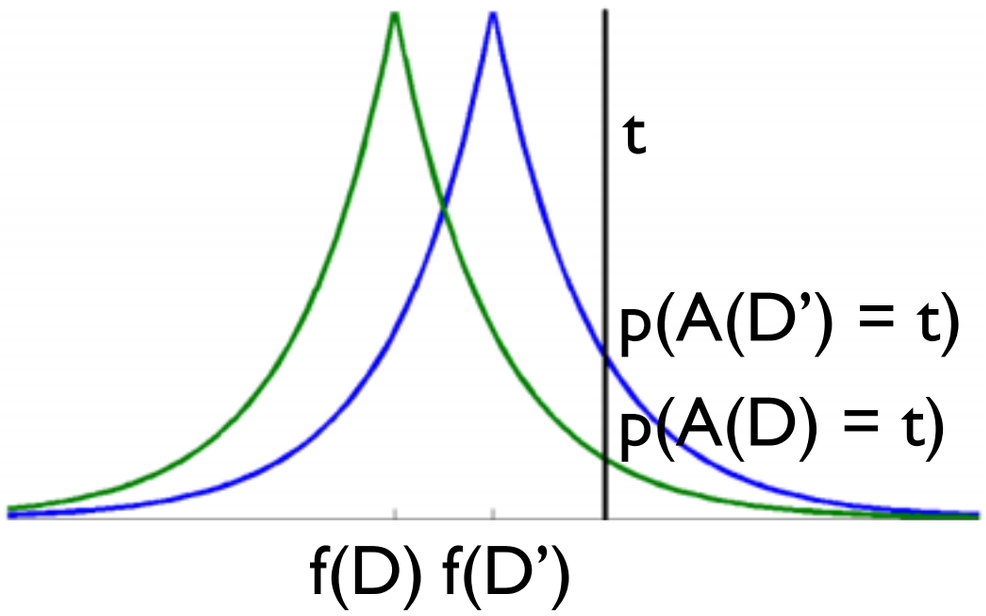
Privacy Proof

Global Sensitivity Method:

Output $f(D) + Z$, where $Z \sim \frac{S(f)}{\epsilon}\mathrm{Lap}(0,1)$ (Privacy ϵ)

Privacy Proof: For any t, any D, D' s.t dist(D, D') = 1,

$$\frac{p(A(D) = t)}{p(A(D') = t)} =$$



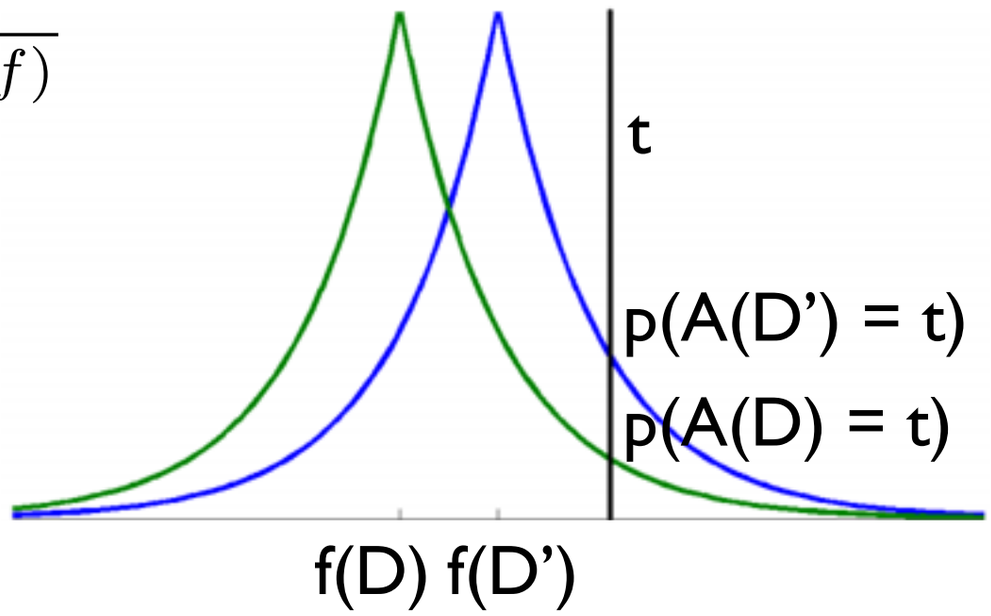
Privacy Proof

Global Sensitivity Method:

Output $f(D) + Z$, where $Z \sim \frac{S(f)}{\epsilon}\text{Lap}(0,1)$ (Privacy ϵ)

Privacy Proof: For any t, any D, D' s.t dist(D, D') = 1,

$$\frac{p(A(D) = t)}{p(A(D') = t)} = \frac{e^{-\epsilon|f(D)-t|/S(f)}}{e^{-\epsilon|f(D')-t|/S(f)}}$$



Privacy Proof

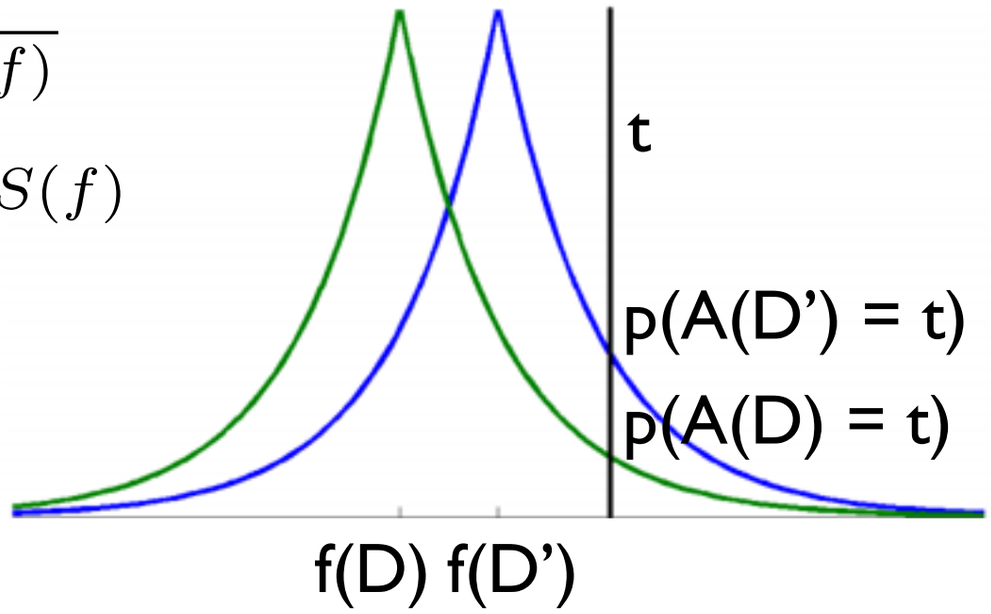
Global Sensitivity Method:

Output $f(D) + Z$, where $Z \sim \frac{S(f)}{\epsilon}\mathrm{Lap}(0,1)$ (Privacy ϵ)

Privacy Proof: For any t, any D, D' s.t dist(D, D') = 1,

$$\frac{p(A(D)=t)}{p(A(D')=t)} = \frac{e^{-\epsilon|f(D)-t|/S(f)}}{e^{-\epsilon|f(D')-t|/S(f)}}$$

$$\leq e^{\epsilon|f(D)-f(D')|/S(f)}$$



Privacy Proof

Global Sensitivity Method:

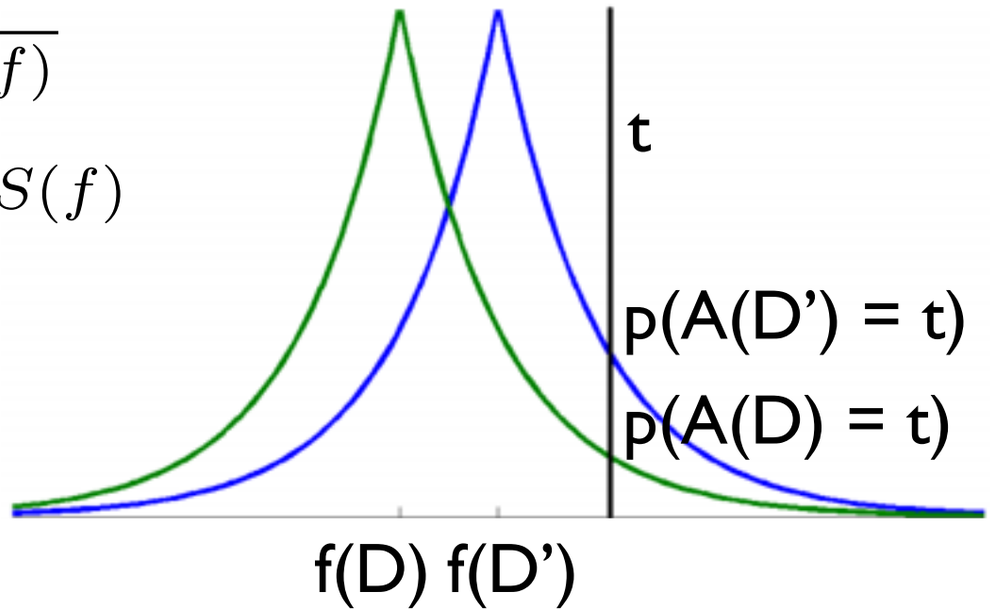
Output $f(D) + Z$, where $Z \sim \frac{S(f)}{\epsilon}\text{Lap}(0,1)$ (Privacy ϵ)

Privacy Proof: For any t, any D, D' s.t dist(D, D') = 1,

$$\frac{p(A(D)=t)}{p(A(D')=t)} = \frac{e^{-\epsilon|f(D)-t|/S(f)}}{e^{-\epsilon|f(D')-t|/S(f)}}$$

$$\leq e^{\epsilon|f(D)-f(D')|/S(f)}$$

$$\leq e^{\epsilon}$$



Example 1: Mean

f(D) = Mean(D), where each record is a scalar in [0,1]

Example 1: Mean

f(D) = Mean(D), where each record is a scalar in [0,1]

Global Sensitivity of f = 1/n

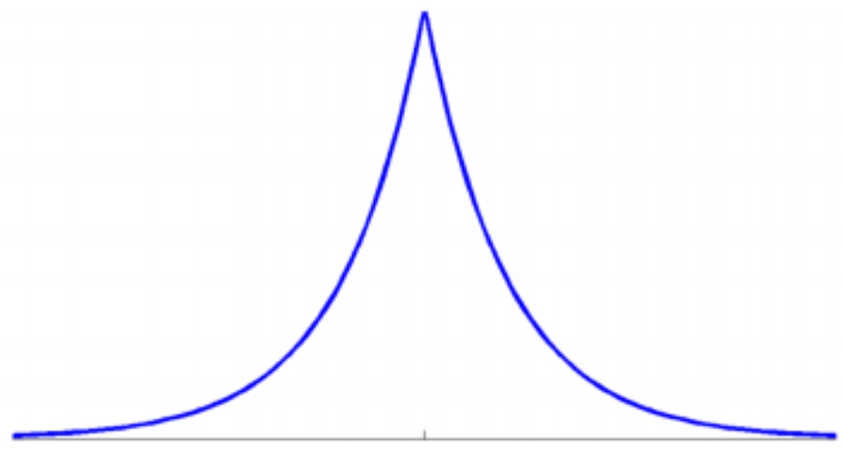
Example 1: Mean

f(D) = Mean(D), where each record is a scalar in [0,1]

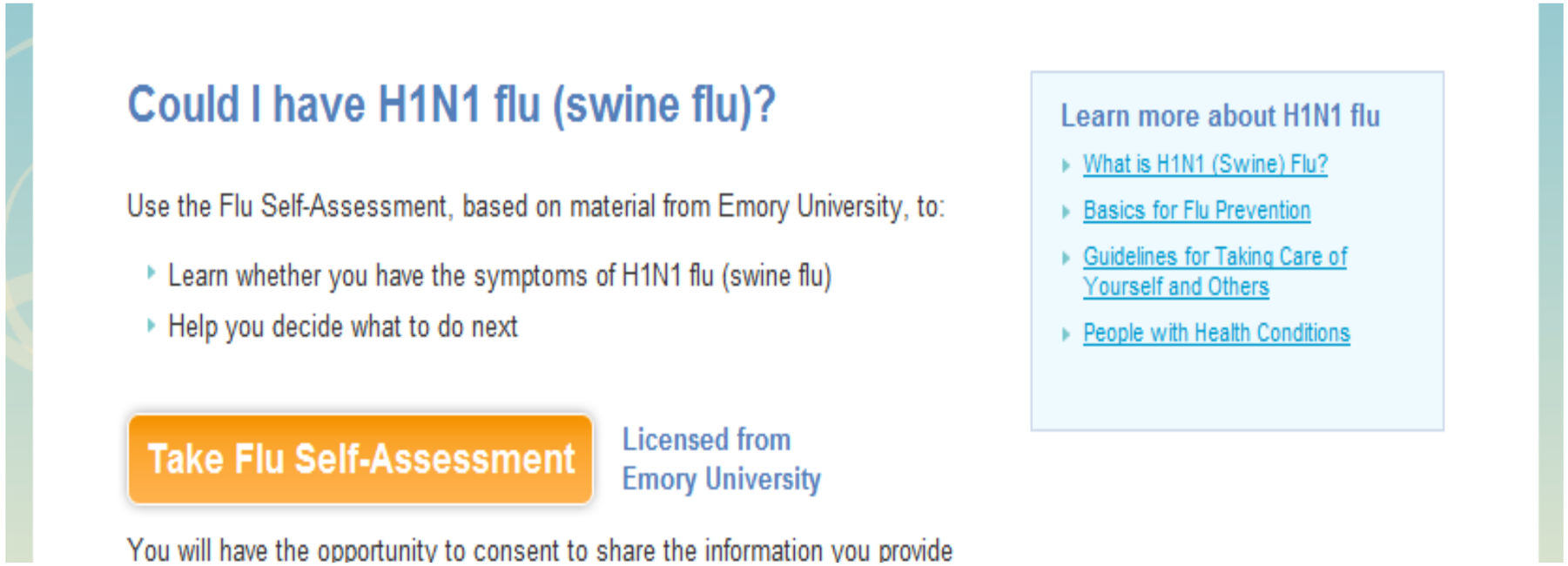
Global Sensitivity of f = 1/n

Global Sensitivity Method:

Output $f(D) + Z$, where $Z \sim \frac{1}{n\epsilon}\mathrm{Lap}(0,1)$ (Privacy ϵ)



Example 2: Classification

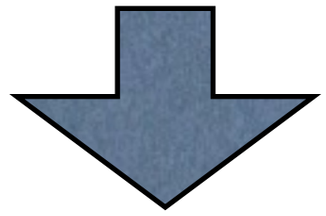


Predicts flu or not, based on patient symptoms

Trained on sensitive patient data

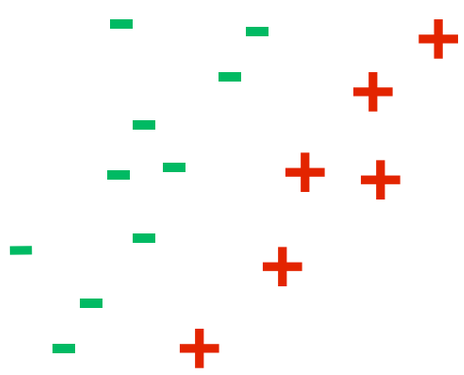
From Attributes to Labeled Data

Yes	No	99F	No
Sore Throat	Fever	Temperature	Flu?

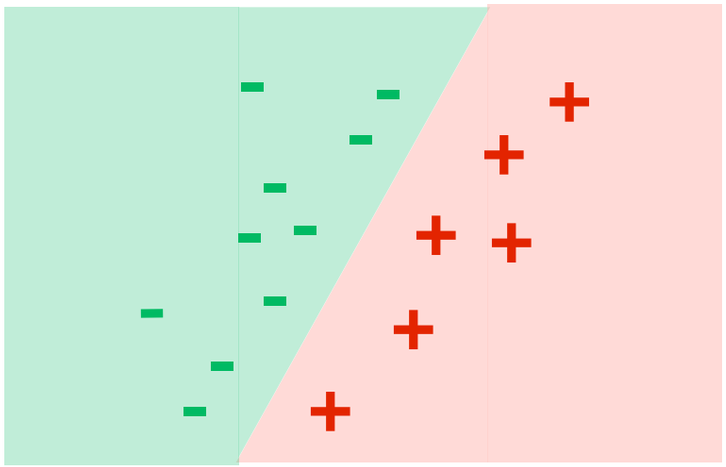


1	0	99	-
Data			Label

Linear Classification

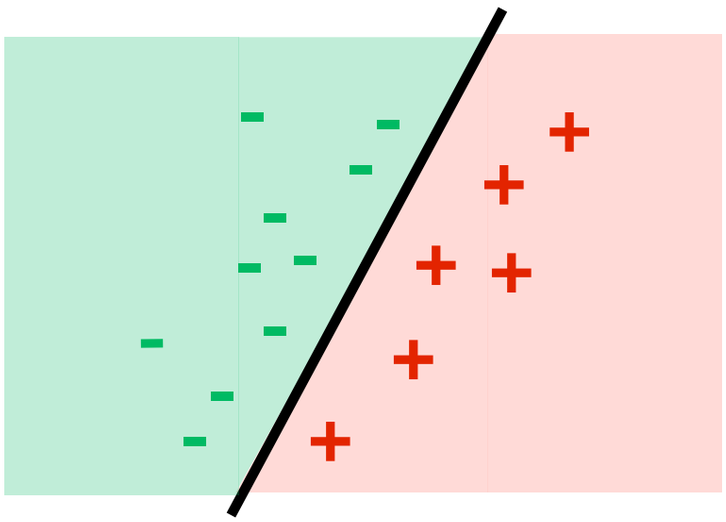


Linear Classification



Distribution P over
labelled examples

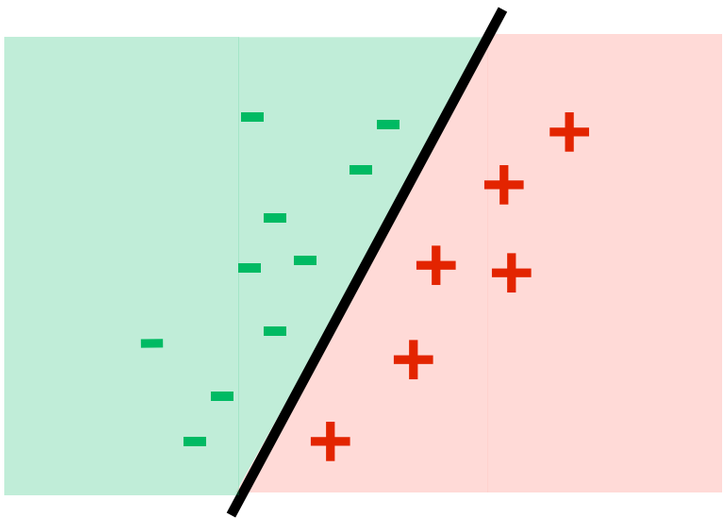
Linear Classification



Distribution P over labelled examples

Goal: Find a vector w that separates + from - for most points from P

Linear Classification



Distribution P over labelled examples

Goal: Find a vector w that separates + from - for most points from P

Key: Find a simple model to fit the samples

Empirical Risk Minimization (ERM)

Given: Labeled data D = {(x_i, y_i)}, find w minimizing:

$$\frac{1}{2}\lambda\|w\|^2 \quad + \quad \frac{1}{n}\sum_{i=1}^{n} L(y_i w^T x_i)$$

Regularizer
(Model Complexity)

Risk
(Training Error)

Empirical Risk Minimization (ERM)

Given: Labeled data D = $\{(x_i, y_i)\}$, find w minimizing:

$$\frac{1}{2}\lambda\|w\|^2 \quad + \quad \frac{1}{n}\sum_{i=1}^{n} L(y_i w^T x_i)$$

Regularizer
(Model Complexity)

Risk
(Training Error)

L = Logistic Loss ⇒ **Logistic Regression**

L = Hinge Loss ⇒ **SVM**

Global Sensitivity of ERM [CMS11]

Goal: Labeled data D = {(x_i, y_i)}, find:

$$f(D) = \text{argmin}_w \frac{1}{2}\lambda\|w\|^2 + \frac{1}{n}\sum_{i=1}^{n} L(y_i w^T x_i)$$

Global Sensitivity of ERM [CMS11]

Goal: Labeled data D = $\{(x_i, y_i)\}$, find:

$$f(D) = \text{argmin}_w \frac{1}{2}\lambda\|w\|^2 + \frac{1}{n}\sum_{i=1}^{n} L(y_i w^T x_i)$$

Theorem [CMS11, BIPR12]:
If $\|x_i\| \leq 1$ and L is 1-Lipschitz, then, for any D, D' with dist(D, D') = 1,

$$\|f(D) - f(D')\|_2 \leq \frac{2}{\lambda n}$$

Global Sensitivity of ERM [CMS11]

Goal: Labeled data D = $\{(x_i, y_i)\}$, find:

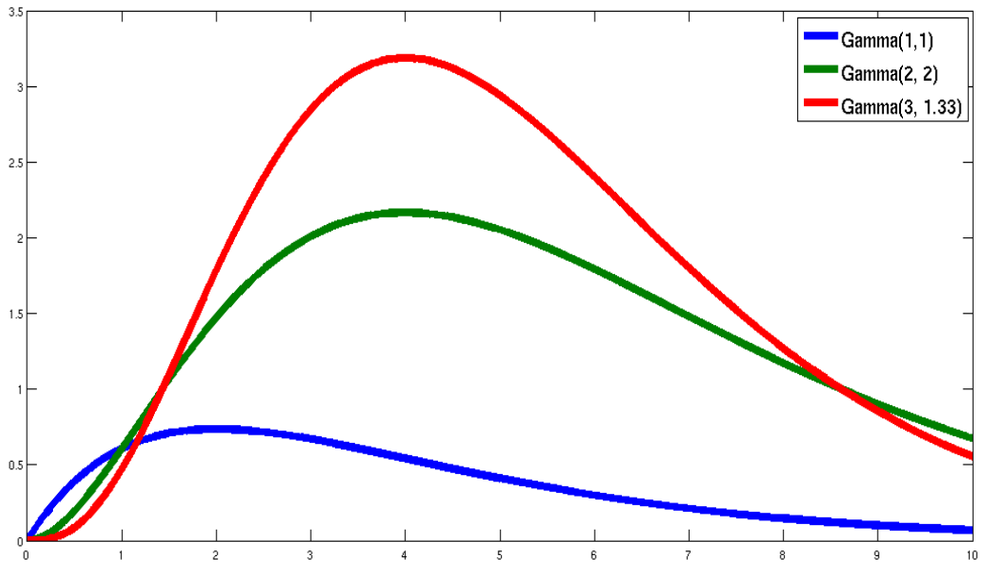
$$f(D) = \text{argmin}_w \frac{1}{2}\lambda\|w\|^2 + \frac{1}{n}\sum_{i=1}^{n} L(y_i w^T x_i)$$

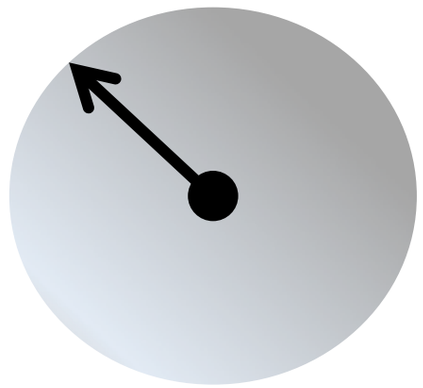
Theorem [CMS11, BIPR12]:
If $\|x_i\| \leq 1$ and L is 1-Lipschitz, then, for any D, D' with dist(D, D') = 1,

$$\|f(D) - f(D')\|_2 \leq \frac{2}{\lambda n}$$

Apply vector version of Global Sensitivity Method

Global Sensitivity Method for ERM





Output: f(D) + Z, where f(D) = non-private classifier

Perturbation Z drawn from:

Magnitude: Drawn from $\Gamma(d, 2/\lambda n\epsilon)$

Direction: Uniformly at random

Smoothed Sensitivity [NRS07]

Smoothed Sensitivity: Relaxes Global Sensitivity

(details in [NRS07])

image: Wikipedia

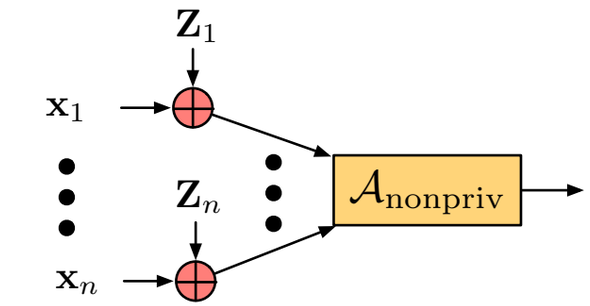
Intermission

The schedule

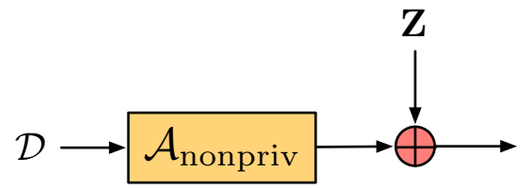
1. Privacy definitions
2. Sensitivity and guaranteeing privacy

— INTERMISSION —

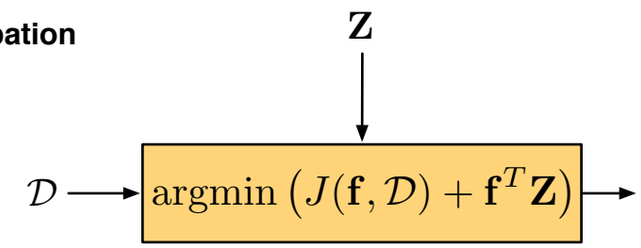
3. Beyond sensitivity
4. Practicalities
5. Applications & Extensions



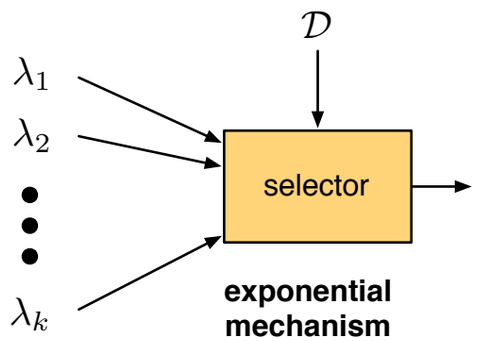
input perturbation



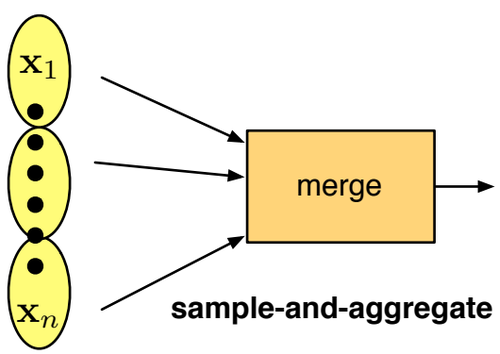
output perturbation



objective perturbation



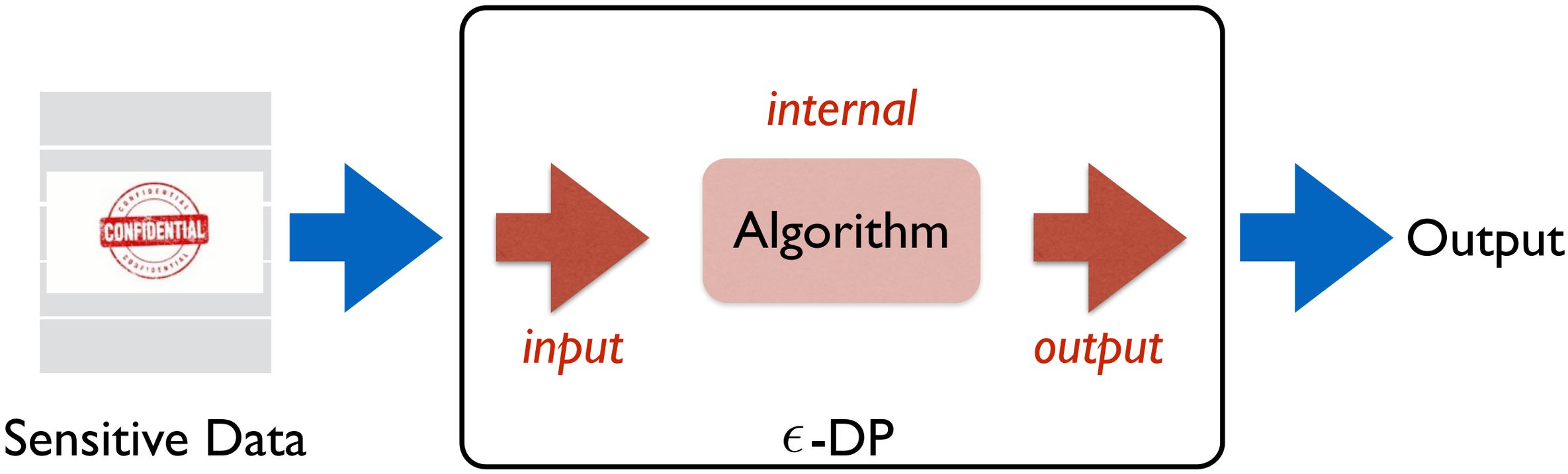
exponential mechanism



sample-and-aggregate

Privacy beyond sensitivity

Where should we add the noise?



- *input perturbation*: add noise to the input before running algorithm
- *output perturbation*: run algorithm, then add noise (sensitivity)
- *internal perturbation*: randomize the internals of the algorithm

Input perturbation and randomized response

Randomized response [W65] is a classical privacy protection:

- example: want avg. # of drug users in population
- surveyor allows subjets to lie randomly with certain probability
- correct for systematic errors due to lying

This guarantees a stronger form of differential privacy known as *local privacy*.

The Exponential Mechanism [MT07]

Suppose we have a measure of quality *q(r, D)* that tells us how good a response *r* is on database *D*. The *exponential mechanism* selects a random output biased towards ones with high quality:

$$p(r) \propto \exp\left(\frac{\epsilon}{2\Delta_q} q(r, D)\right)$$

Where Δ_q is the sensitivity of the quality measure.

Example: parameter selection in ERM

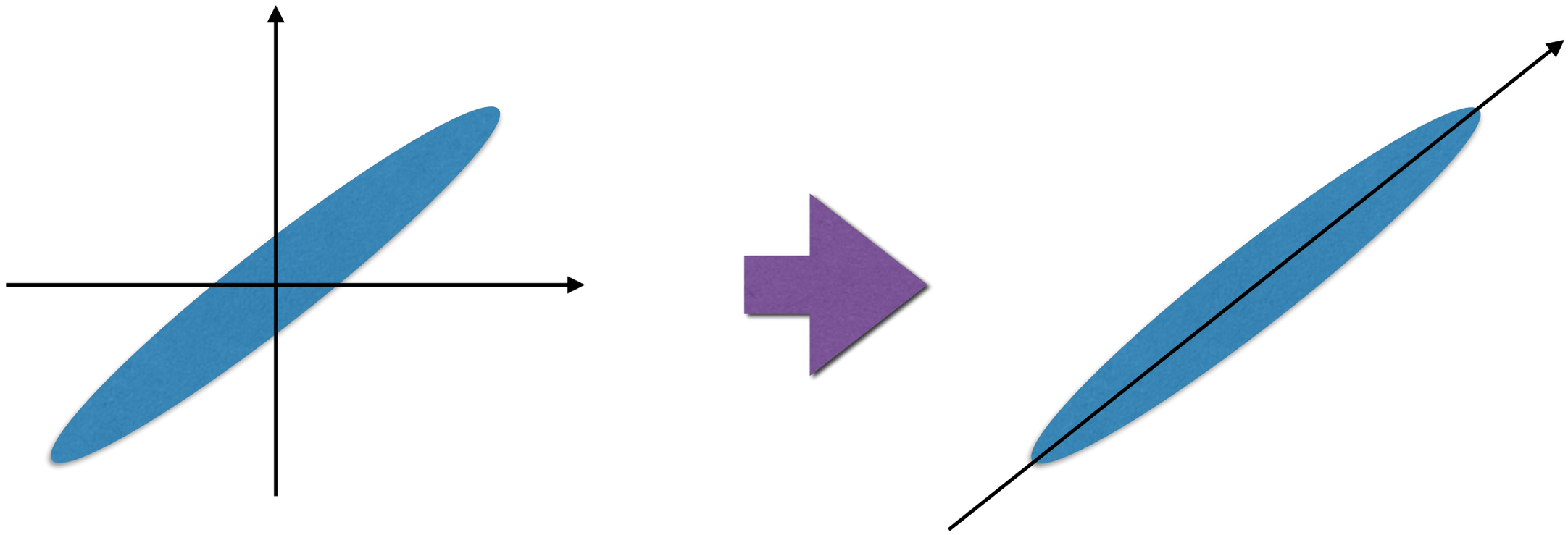
Recall *empirical risk minimization*:

$$\operatorname*{argmin}_w \frac{1}{n} \sum_{i=1}^{n} L(y_i w^\top x_i) + \frac{1}{2}\lambda \|w\|^2$$

We have to pick a value of λ — we can do this using a *validation set* of additional private data:

- $q(\lambda, D)$ is the number of correct predictions made by the output of the algorithm run with parameter λ.
- Use the exponential mechanism to select a λ from some finite set of candidates.

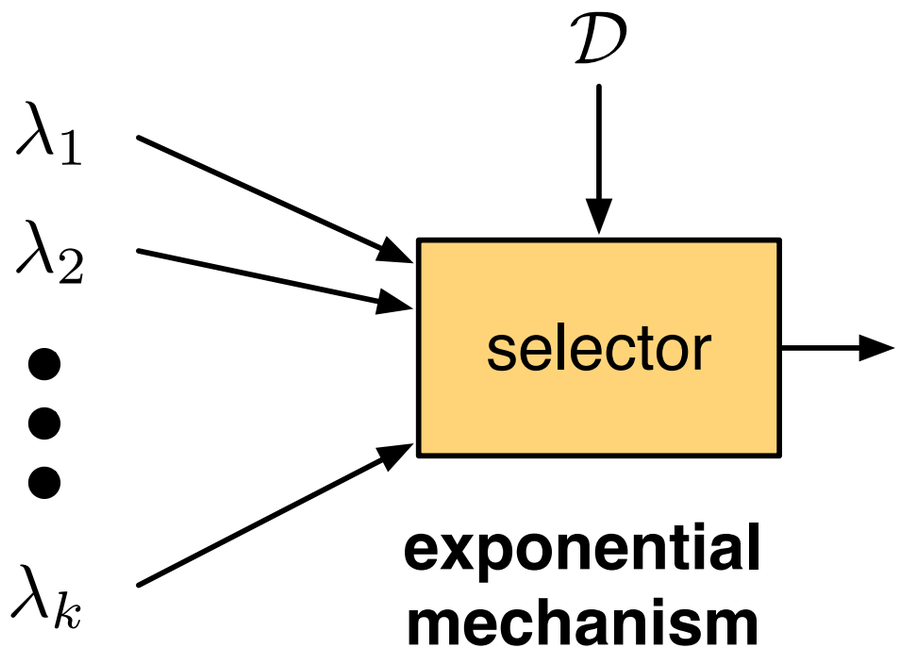
Example: private PCA [CSS13]



In *principal components analysis*, given a *d x n* data matrix *X* we want to find a low-dimensional subspace in which the data lie. Output a *d x k* matrix *V* with orthogonal columns using the exponential mechanism and score function:

$$q(V, X) = \mathrm{tr}(V^\top (XX^\top) V)$$

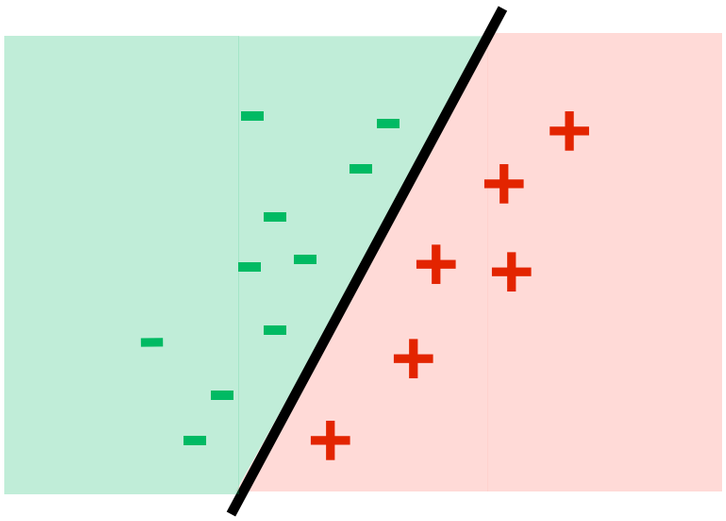
Other uses for the exponential mechanism



Lots of other examples (just a few):

- other PCA [KT13]
- auctioning goods [MT07]
- classification [BST14]
- generating synthetic data [XXY10]
- recommender systems [MKS11]

Linear Classification Revisited

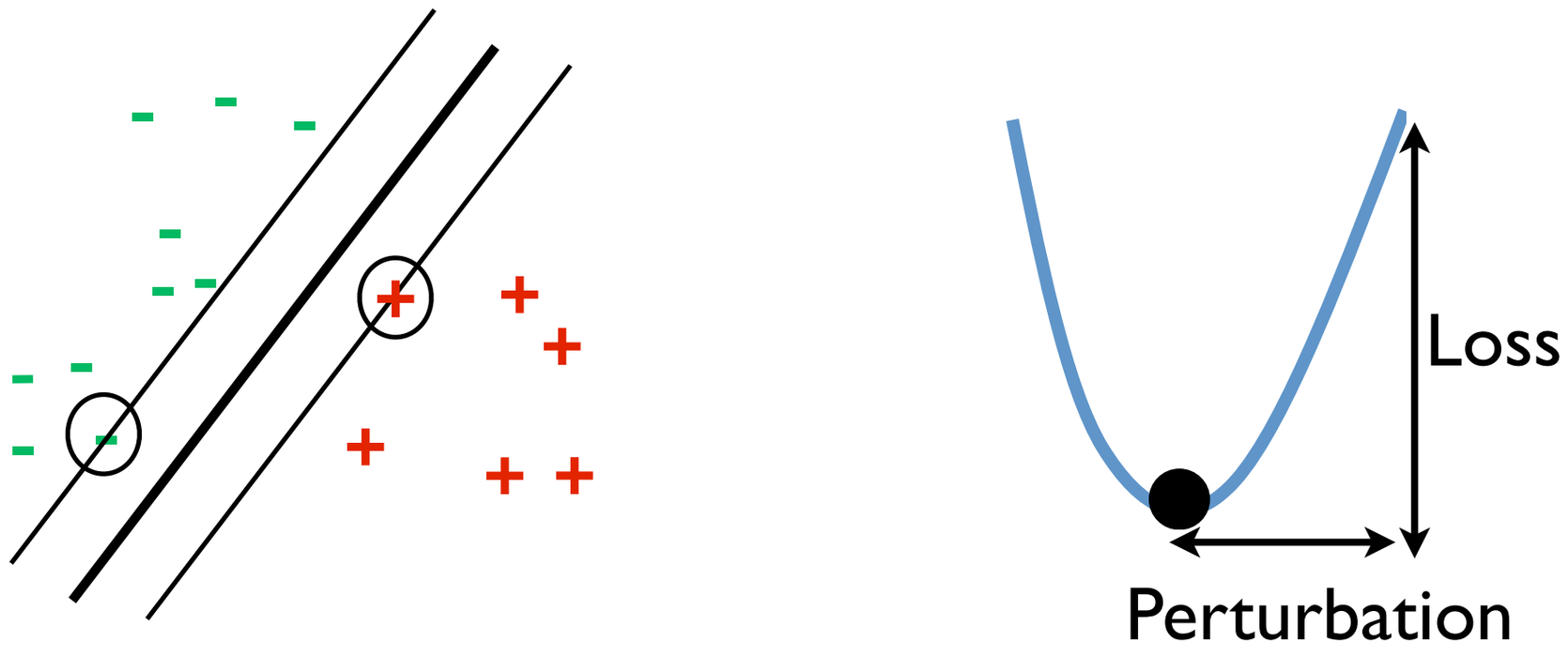


Distribution P over labelled examples

Goal: Find a vector w that separates + from - for most points from P

Key: Find a simple model to fit the samples

Properties of Real Data



Optimization surface is very steep in some directions

High loss if perturbed in those directions

Insight: Perturb optimization surface and then **optimize**

Objective perturbation

$$\underset{w}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} L(y_i w^\top x_i) + \frac{1}{2} \lambda \|w\|^2 + \text{noise} \right\}$$

Main idea: *add noise as part of the computation*:

- Regularization already changes the objective to protects against overfitting.
- Change the objective a little bit more to protect privacy.

Empirical Risk Minimization

Goal: Labeled data (x_i, y_i), find w minimizing:

$$\frac{1}{2}\lambda\|w\|^2 \quad + \quad \frac{1}{n}\sum_{i=1}^{n} L(y_i w^T x_i) \quad + \quad \frac{1}{n} b^\top w$$

Regularizer

(Model Complexity)

Risk

(Training Error)

Perturbation

(Privacy)

Here, the vector b is a noise vector.

Privacy Guarantees

Algorithm: Given labeled data (x_i, y_i), find w to minimize:

$$\frac{1}{2}\lambda\|w\|^2 \quad + \quad \frac{1}{n}\sum_{i=1}^{n} L(y_i w^T x_i) \quad + \quad \frac{1}{n} b^\top w$$

Theorem: If L is convex and doubly-differentiable with $|L'(z)| \leq 1$ and $|L''(z)| \leq c$ then Algorithm is $\alpha + 2\log\left(1 + \frac{c}{n\lambda}\right)$-differentially private

Privacy Guarantees

Algorithm: Given labeled data (x_i, y_i), find w to minimize:

$$\frac{1}{2}\lambda\|w\|^2 \quad + \quad \frac{1}{n}\sum_{i=1}^{n} L(y_i w^T x_i) \quad + \quad \frac{1}{n} b^\top w$$

L = Logistic Loss ⇒ **Private Logistic Regression**

L = Huber Loss ⇒ **Private SVM**

(Hinge Loss is not differentiable)

Sample Requirement

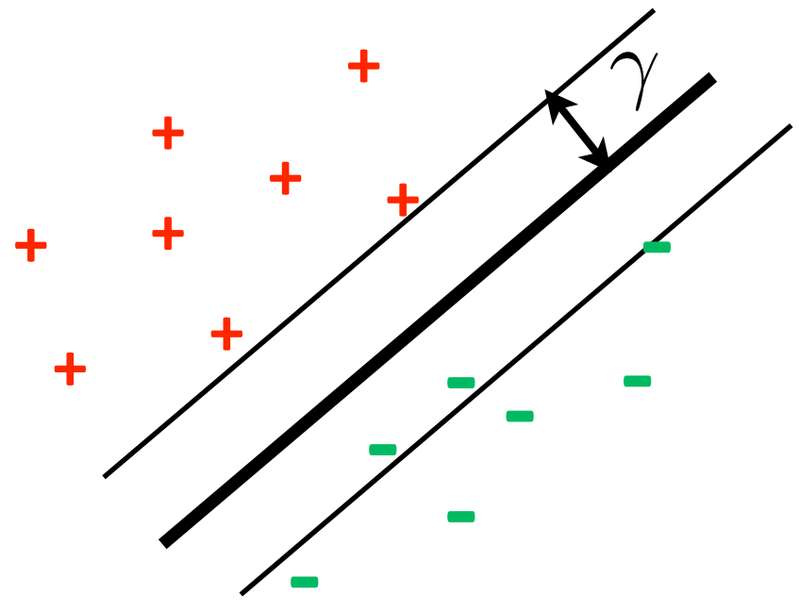
d : #dimensions

γ : margin

α : privacy

ϵ : error

$\gamma, \alpha, \epsilon < 1$

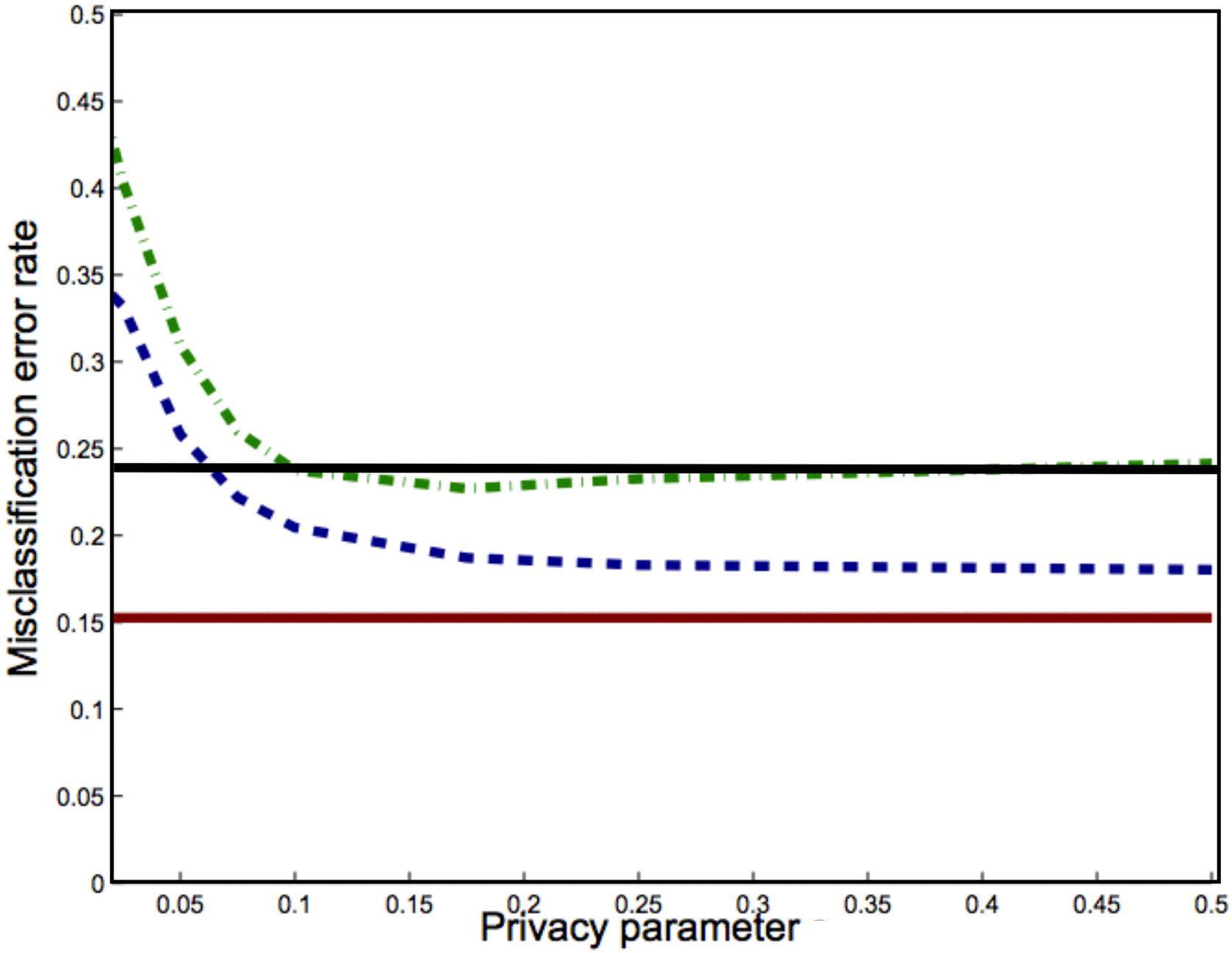


Normal SVM: $1/\epsilon^2\gamma^2$

Our Algorithm: $1/\epsilon^2\gamma^2 + d/\epsilon\alpha\gamma$

Standard Method: $1/\epsilon^2\gamma^2 + d/\epsilon^{3/2}\alpha\gamma$

Results: SVM



Non-private

Standard Method

Our Method

Predict majority label

Example: sparse regression [KST12]

Improved analysis of objective perturbation can include hard constraints and non-differentiable regularizers, including the LASSO:

$$\operatorname*{argmin}_w \left\{ \frac{1}{n} \sum_{i=1}^{n} L(y_i w^\top x_i) + \frac{\lambda}{2n} r(w) + \frac{\Delta}{2n} \|w\|^2 + \frac{1}{n} b^\top w \right\}$$

Relaxing the requirement to (ϵ, δ) improves the dependence on *d* to $\sqrt{d \log(1/\delta)}$. A further improvement [JT14] shows that for a particular choice of λ we may avoid the dependence on *d*.

Other methods for convex optimization

We can use objective perturbation for more general convex optimization problems beyond ERM. There are also other ways to change the objective function:

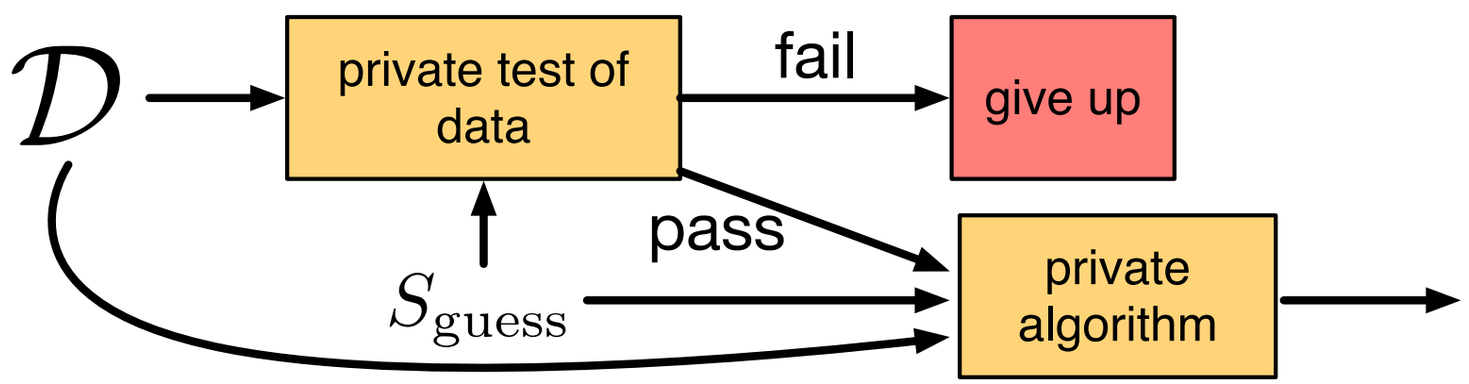
- Functional approximation of the objective with noise [ZZXYW12]
- Kernels [CMS11] [HRW13] [JT13]
- other optimization problems [HTP14]
- stochastic optimization (later in this tutorial…)

The schedule

1. Privacy definitions
2. Sensitivity and guaranteeing privacy

— INTERMISSION —

3. Beyond sensitivity
4. Practicalities
5. Applications & Extensions



Dealing with (some) practical issues

Some practical issues

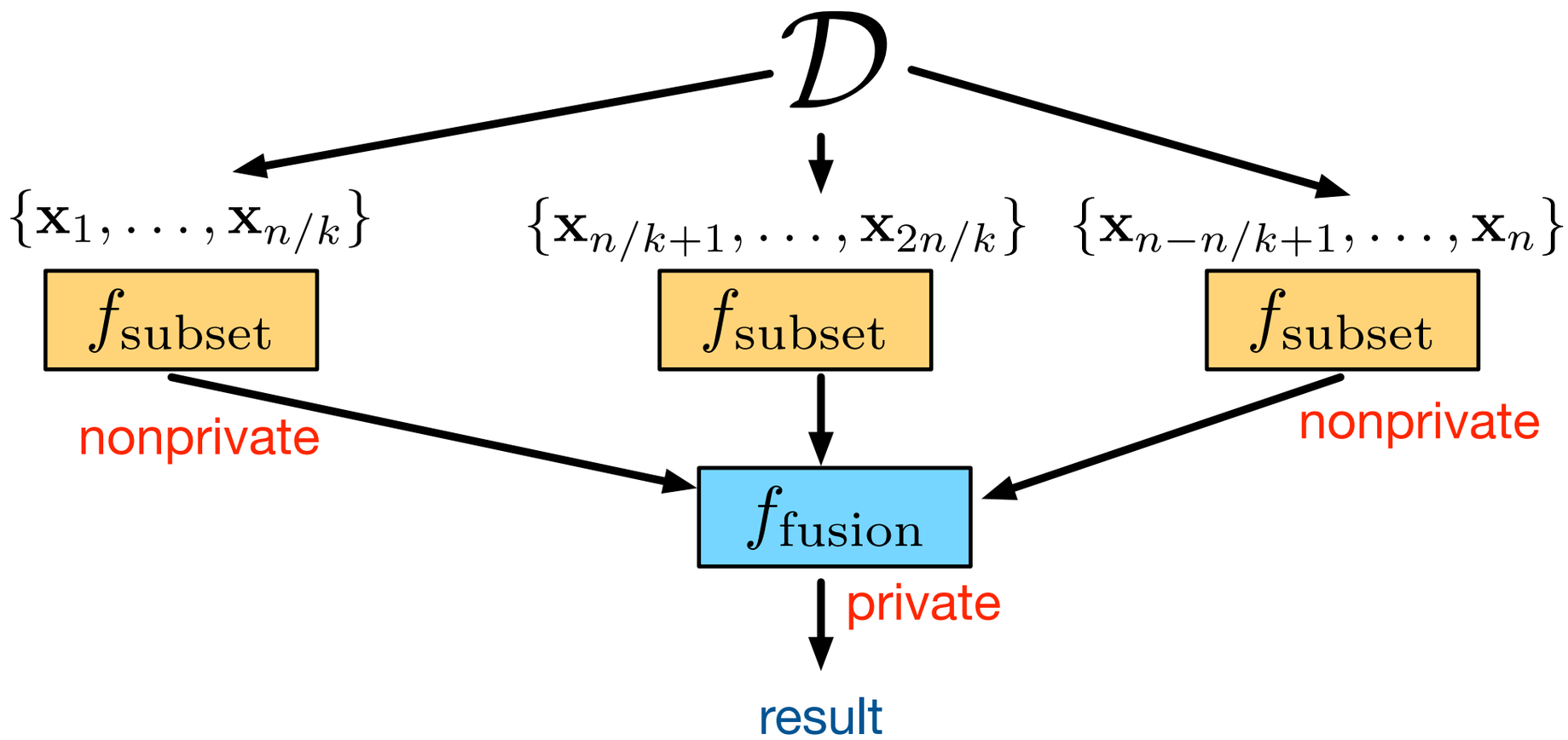
Someone gives you a private data set and asks you do to an analysis. What do you do?

- Test the data to see which algorithm to use.
- Cross-validation and parameter tuning.
- Bootstrapping and evaluating performance.
- Picking an epsilon (and delta) to give good results.

We have to do all of these things while *preserving privacy*.

This is called *end-to-end privacy*.

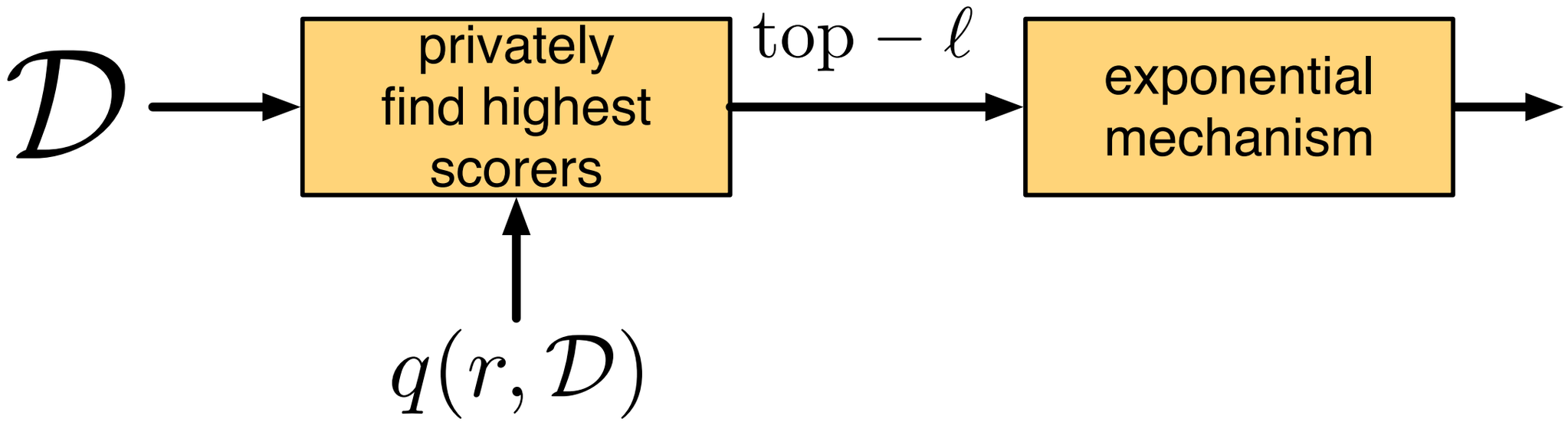
Sample-and-aggregate [NRS07]



Example: *evaluate the maximum value of a query.*

- compute query on subsets of the data
- use a differentially private method to select max

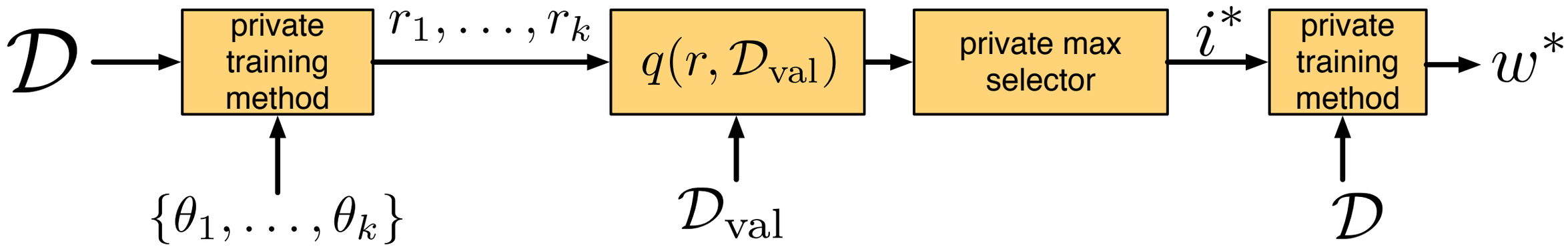
Large-margin mechanism [CHS14]



Goal: produce the maximizer of a function of the data. Use a two-step procedure:

- reduce the set of candidates by approximately (privately) finding the almost-maximizers
- the value of l depends on the data set
- select from the candidates using the exponential mechanism

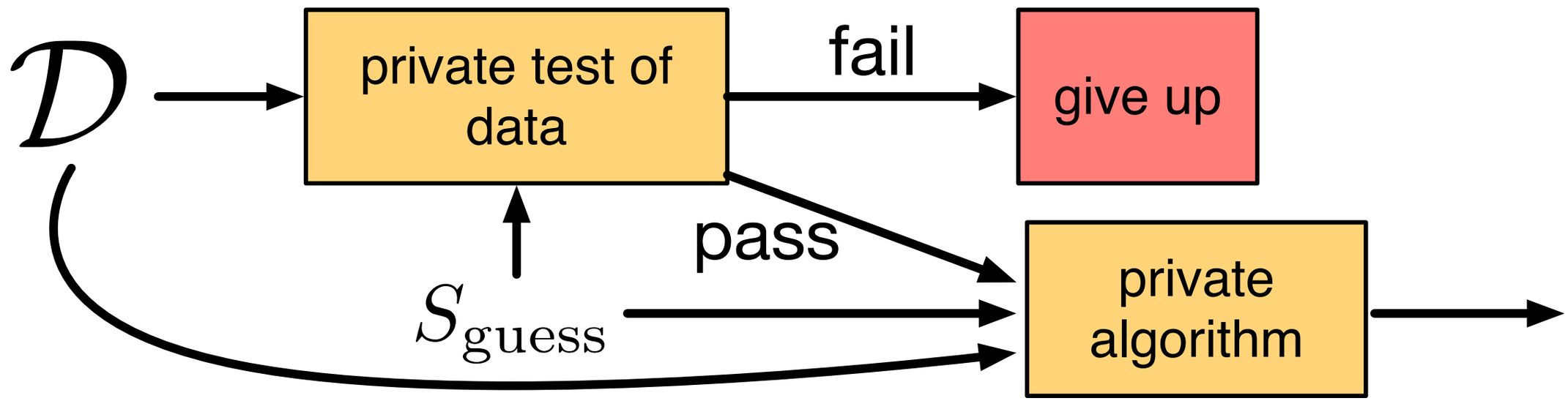
Cross-validation [CV13]



Idea: validation performance should be stable between *D* and *D'* when using same random bits in the training algorithm:

- exploit this to reduce the overall privacy risk
- outperforms parameter tuning using the exponential mechanism

Propose-test-release [DL09]



Idea: test if the data set has some "nice" property that we can exploit.

- propose a bound/value for the property
- run a differentially private test to see if it holds
- run the private algorithm tuned to the nice property

Example: *testing the sensitivity of the target function.*

Setting epsilon and delta [KS08]

Privacy-sensitive

- pick an epsilon and delta based on cost of privacy violations (e.g. lawsuits)
- run algorithms with this epsilon and delta

Utility-sensitive

- run extensive tests to see the privacy-utility tradeoff
- run algorithms targeting a given utility loss

Caveats:

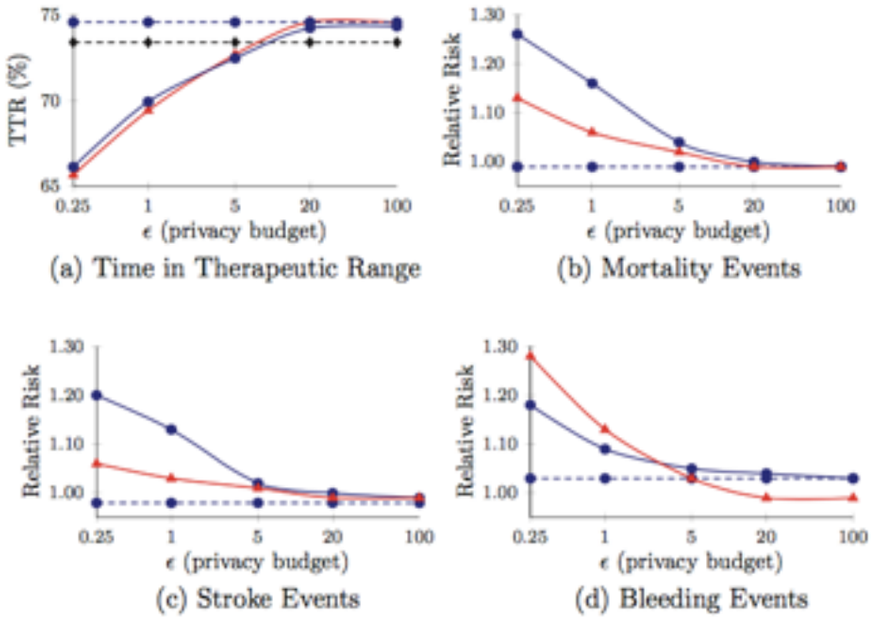
- Theory suggests we can set $\epsilon < 1, \delta = o(n^2)$
- In practice the story is much more complicated
- Commandment: *know thy data*.

The schedule

1. Privacy definitions
2. Sensitivity and guaranteeing privacy

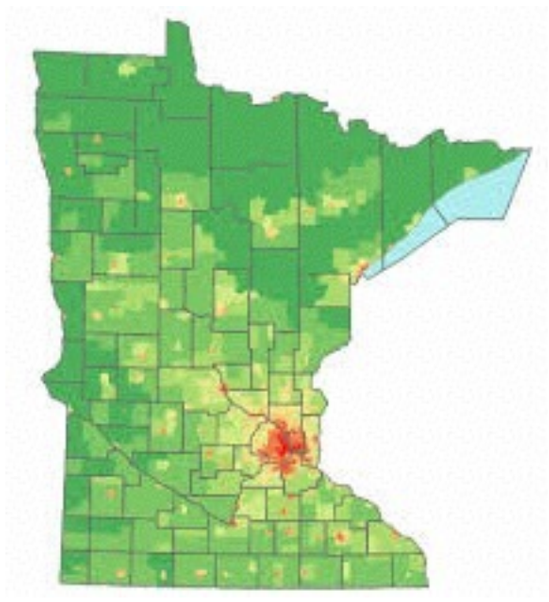
— INTERMISSION —

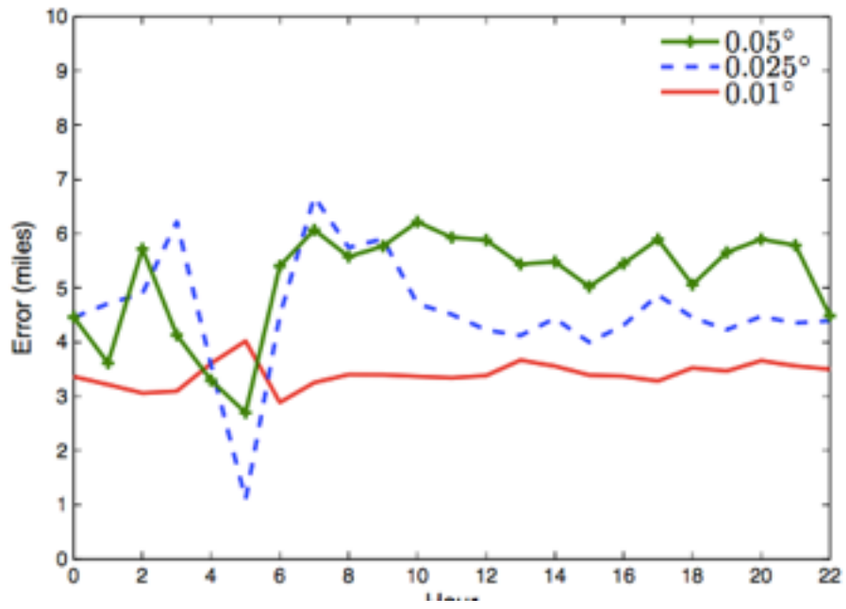
3. Beyond sensitivity
4. Practicalities
5. Applications & Extensions



(a) Time in Therapeutic Range (b) Mortality Events

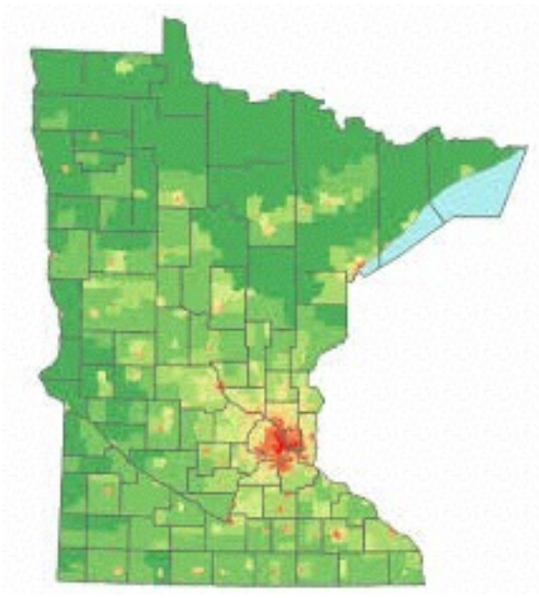
(c) Stroke Events (d) Bleeding Events

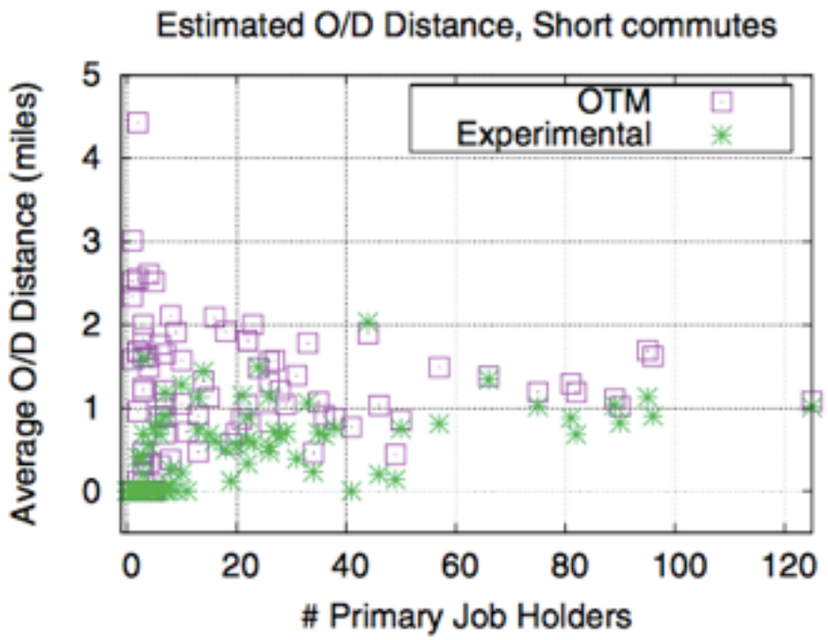




Applications

Privacy on the Map [MKAKV08]



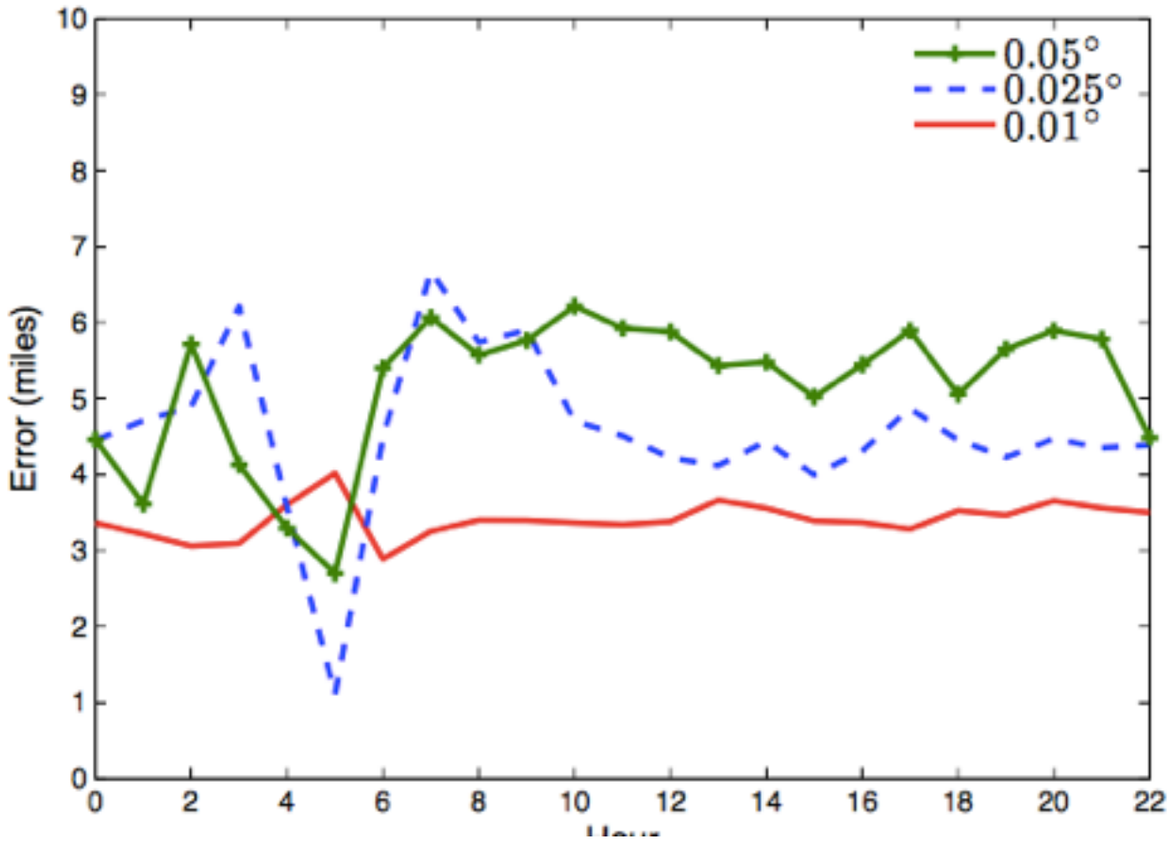


Goal: synthetic data for estimating commute distances

- for each "workplace" block, plot points on the map representing the "home" blocks
- ~ 233,726 locations in MN: large domain
- ~ 1.5 million data points (pairs of home/work locations)
- epsilon = 8.6, delta = 0.00001

images: Wikimedia, MKAKV08

Human Mobility [MICMW13]

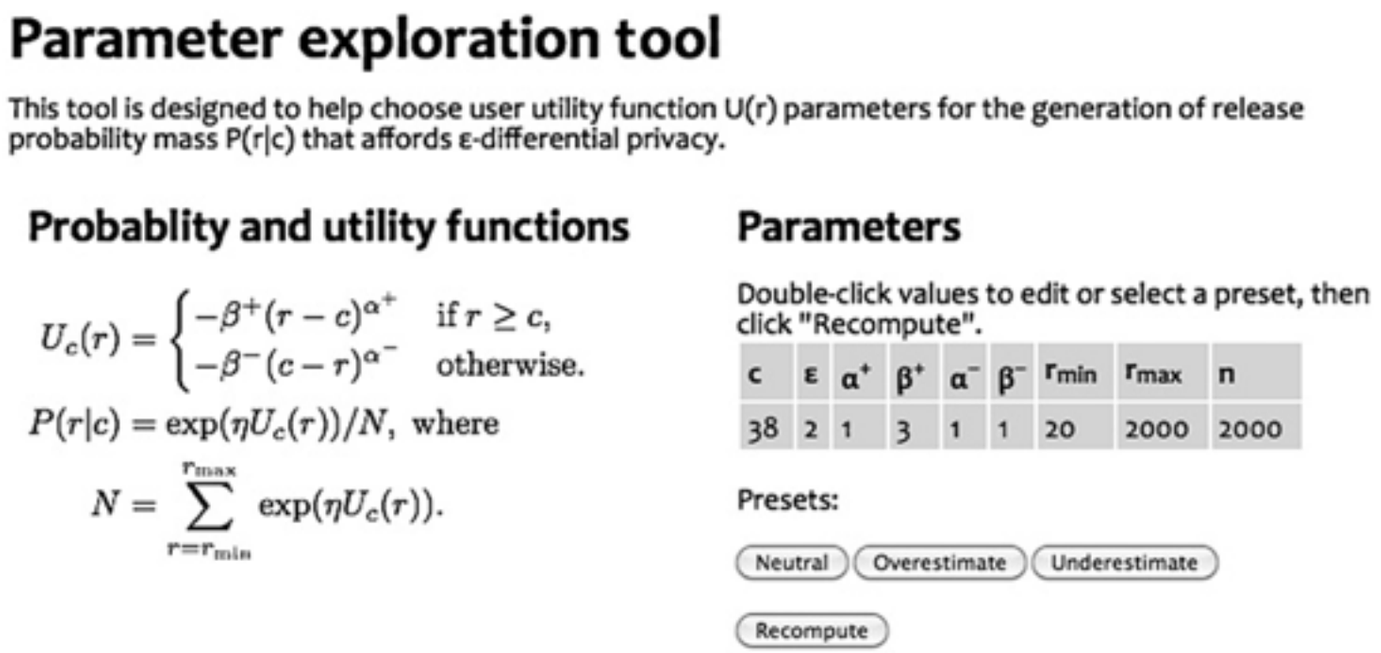


Goal: synthetic data to estimate commute patterns from call detail records

- 1 billion records
- ~ 250,000 phones
- epsilon = 0.23

images: AT&T research, MICMW13

Towards exploratory analysis [VSB12]

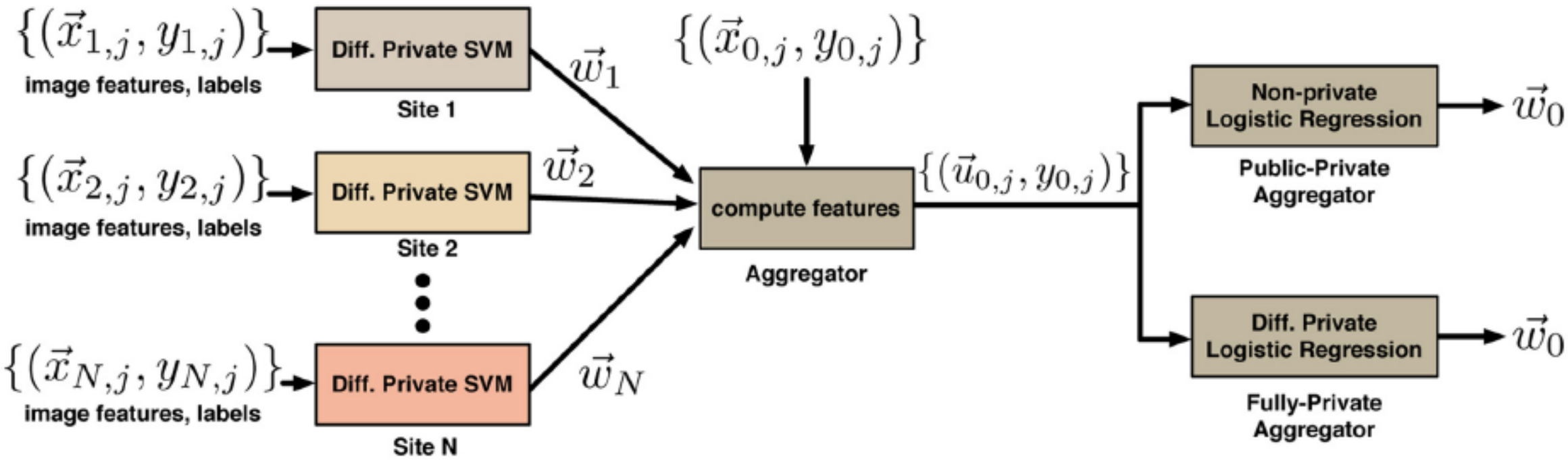


Goal: differentially private exploratory data analysis for clinical research

- modify existing methodologies (return approximate counts) to quantify and track privacy
- allow for user-tuned preferences

images: UCSD, [VSB12]

Neuroimaging [SPTAC14]

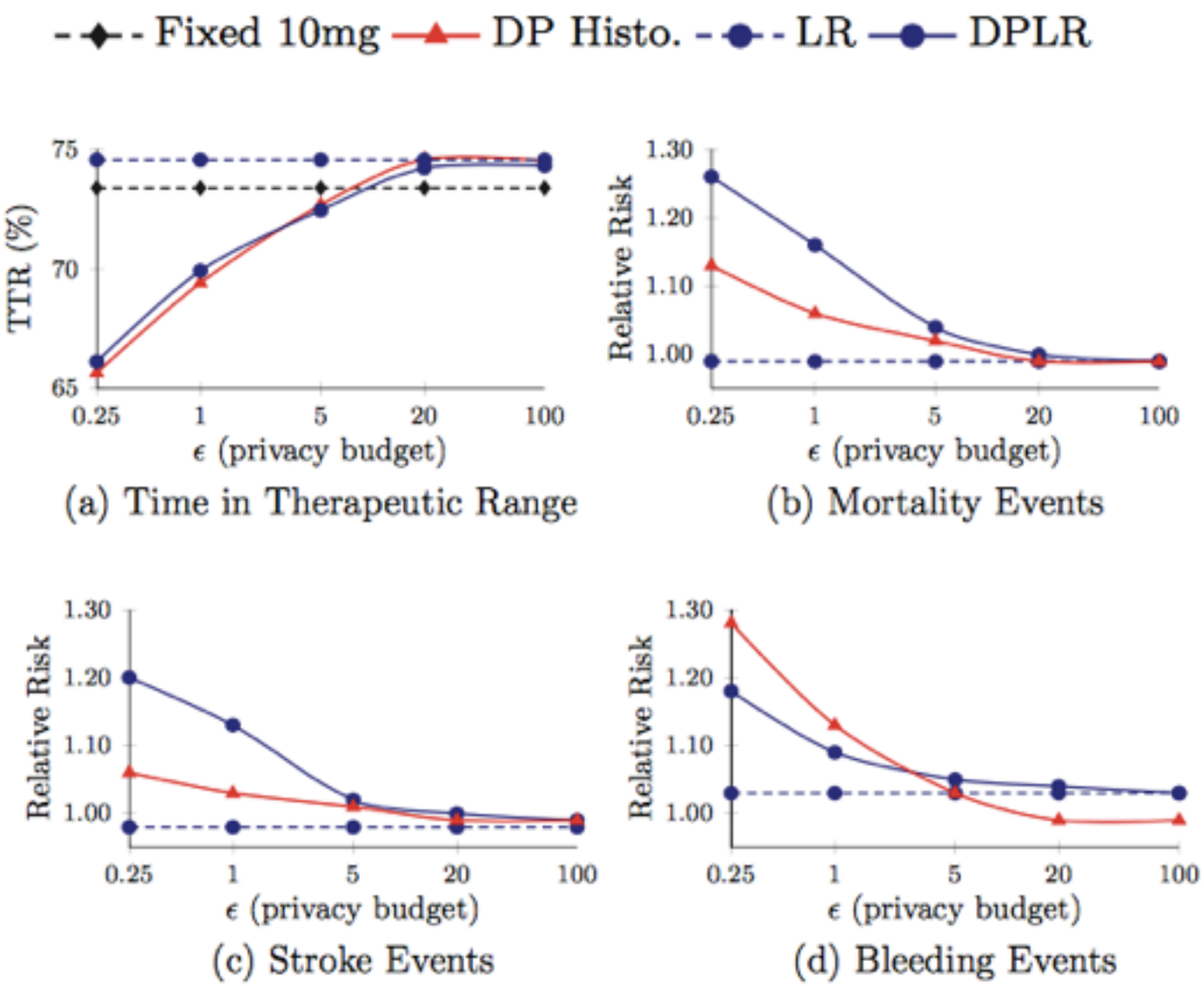


Goal: merge DP classifiers for schizophrenia from different study sites / locations.

- each site learns a classifier from local data
- fusion rule achieves significantly lower error

images: [SPTAC14]

Pharmacogenetics [FLJLPR14]

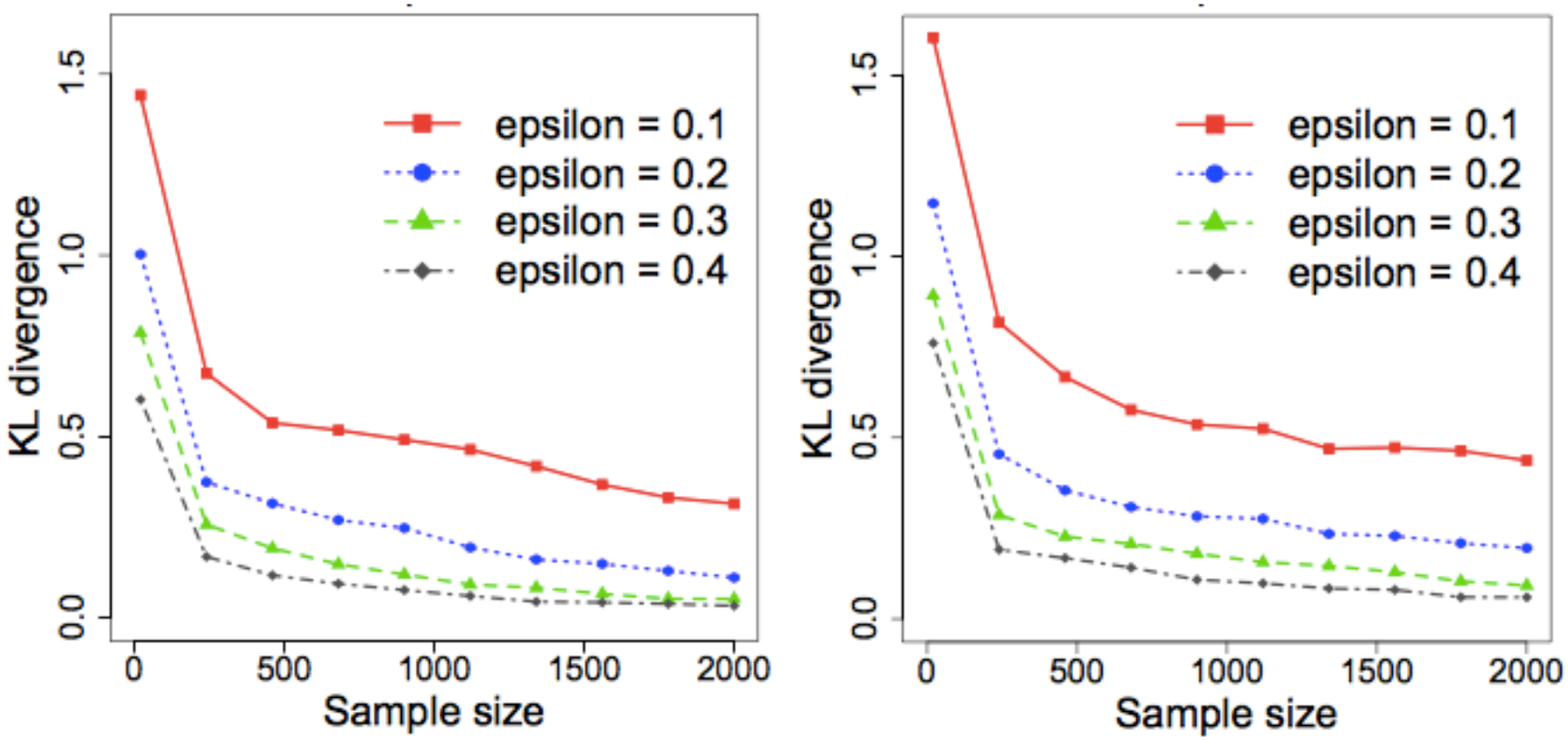


Goal: personalized dosing for warfarin

- see if genetic markers can be predicted from DP models
- small epsilon (< 1) does protect privacy but even moderate epsilon (< 5) leads to increased risk of fatality

images: [FLJLPR14]

Genomic studies [FSU12]

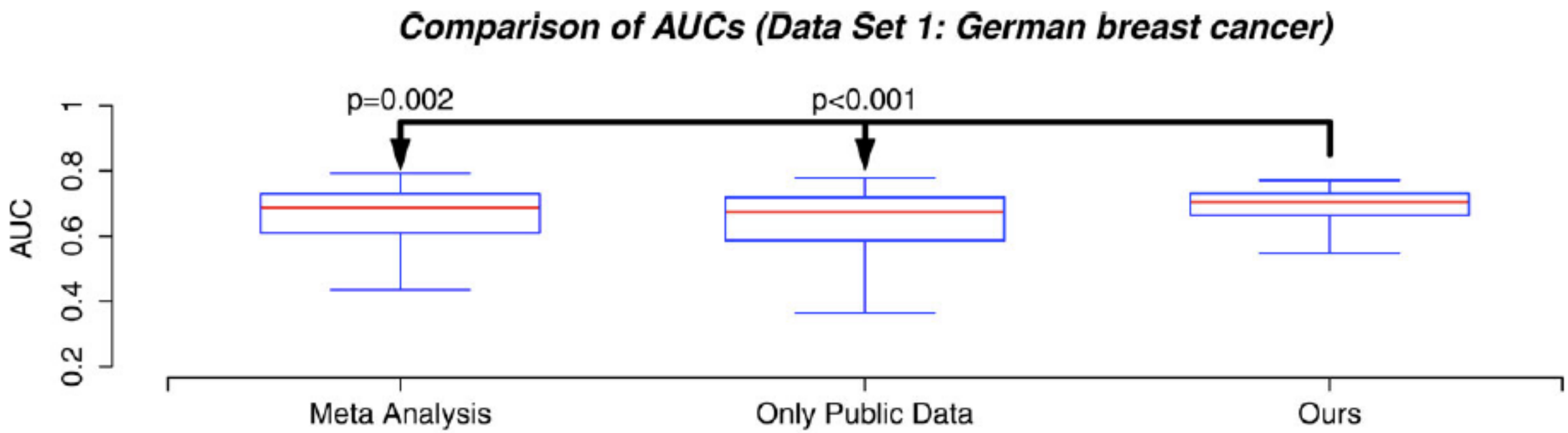


Goal: release aggregate statistics for single nucleotide polymorphisms (SNPs) (3x2 contingency table for each location)

- 40842 locations
- 685 individuals (dogs)
- want to find the most relevant SNPs for predicting a feature (long hair)

images: [FSU12]

Meta-analysis [JJWXO14]



Goal: perform a meta analysis from

- 9 attributes, 686 individuals, split into 5-20 sites
- epsilons range from 0.5 to 5

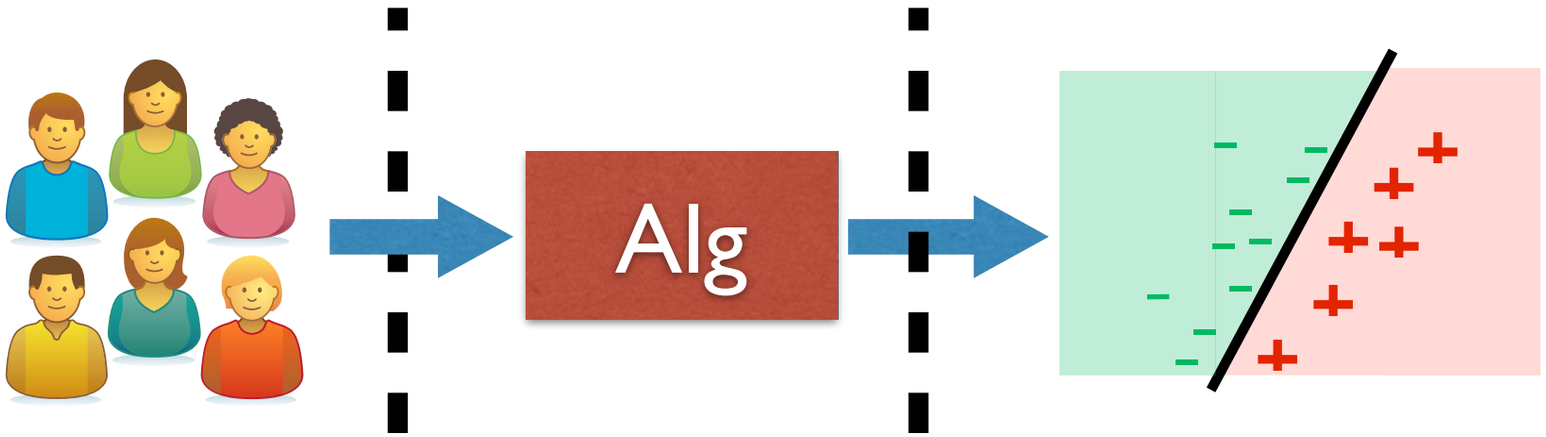
images: [JJWXO14]

The schedule

1. Privacy definitions
2. Sensitivity and guaranteeing privacy

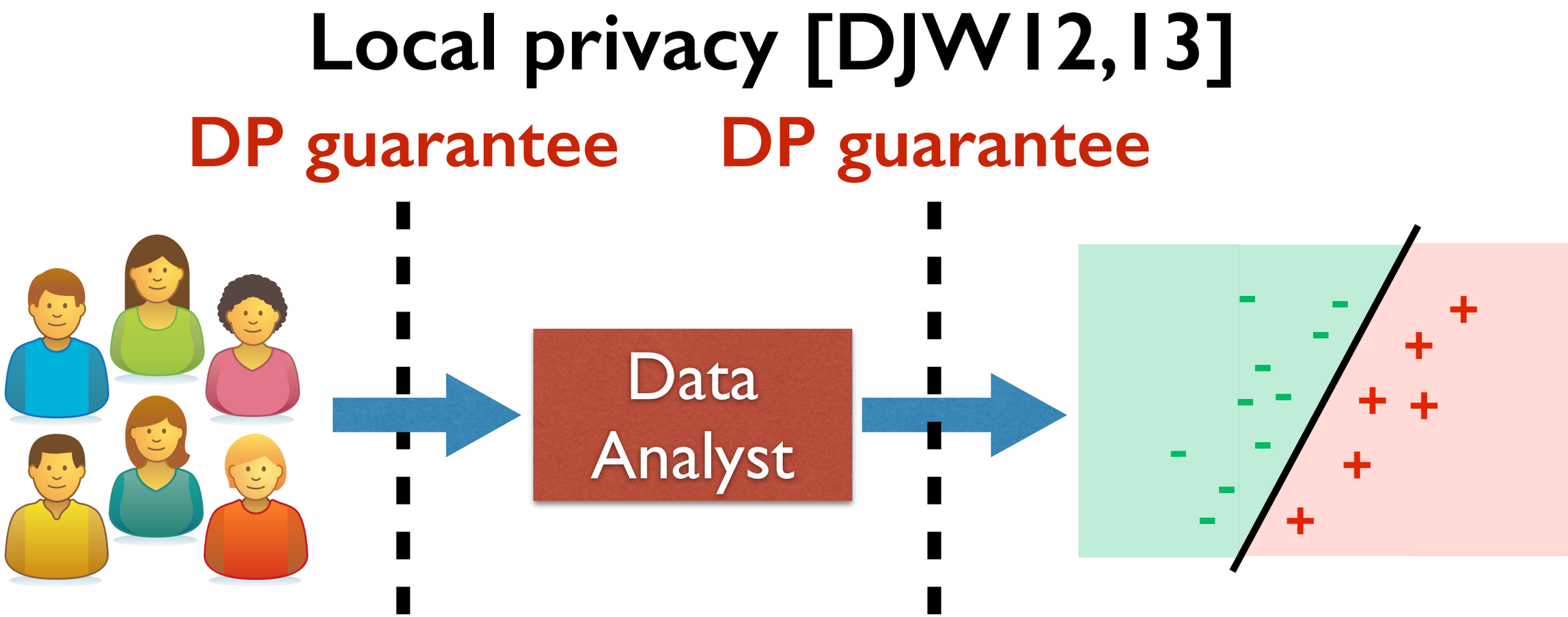
— INTERMISSION —

3. Beyond sensitivity
4. Practicalities
5. Applications & Extensions



Extensions

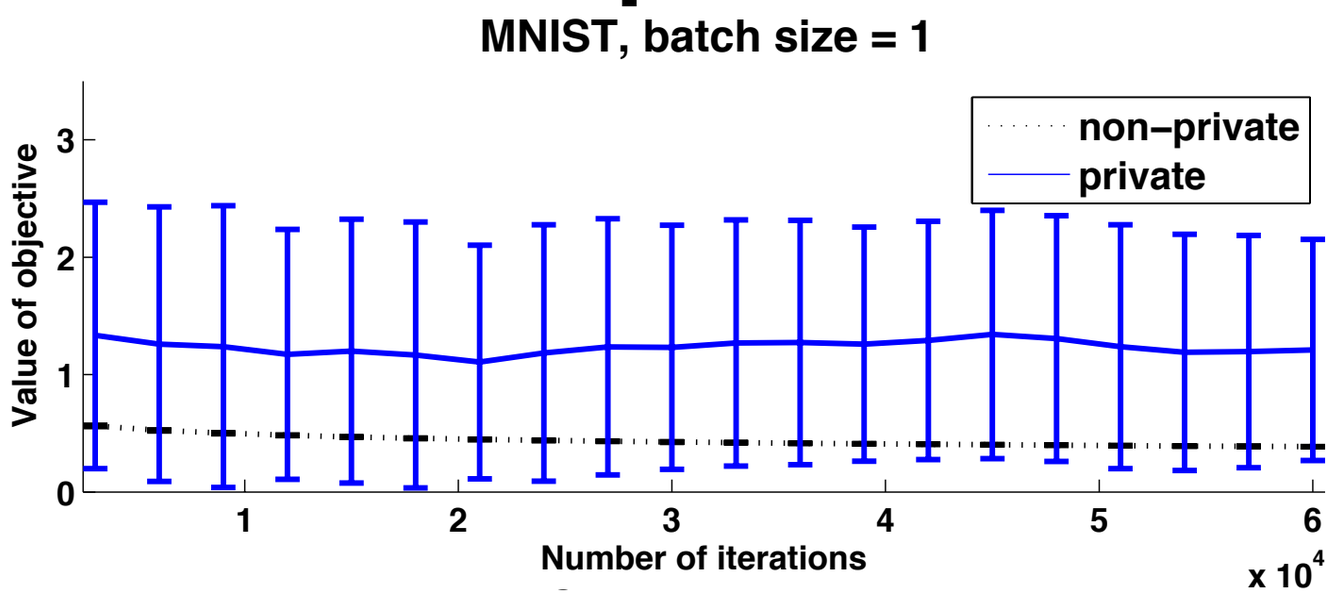
Local privacy [DJW12,13]



Randomized response provides privacy to individuals *while collecting the data*. This is the notion of local privacy. Users provide $(U_1, \ldots, U_n)$ such that

$$\mathbb{P}(U_i \in S | X_i = x) \leq e^{\epsilon} \mathbb{P}(U_i \in S | X_i = x').$$

Example: SGD



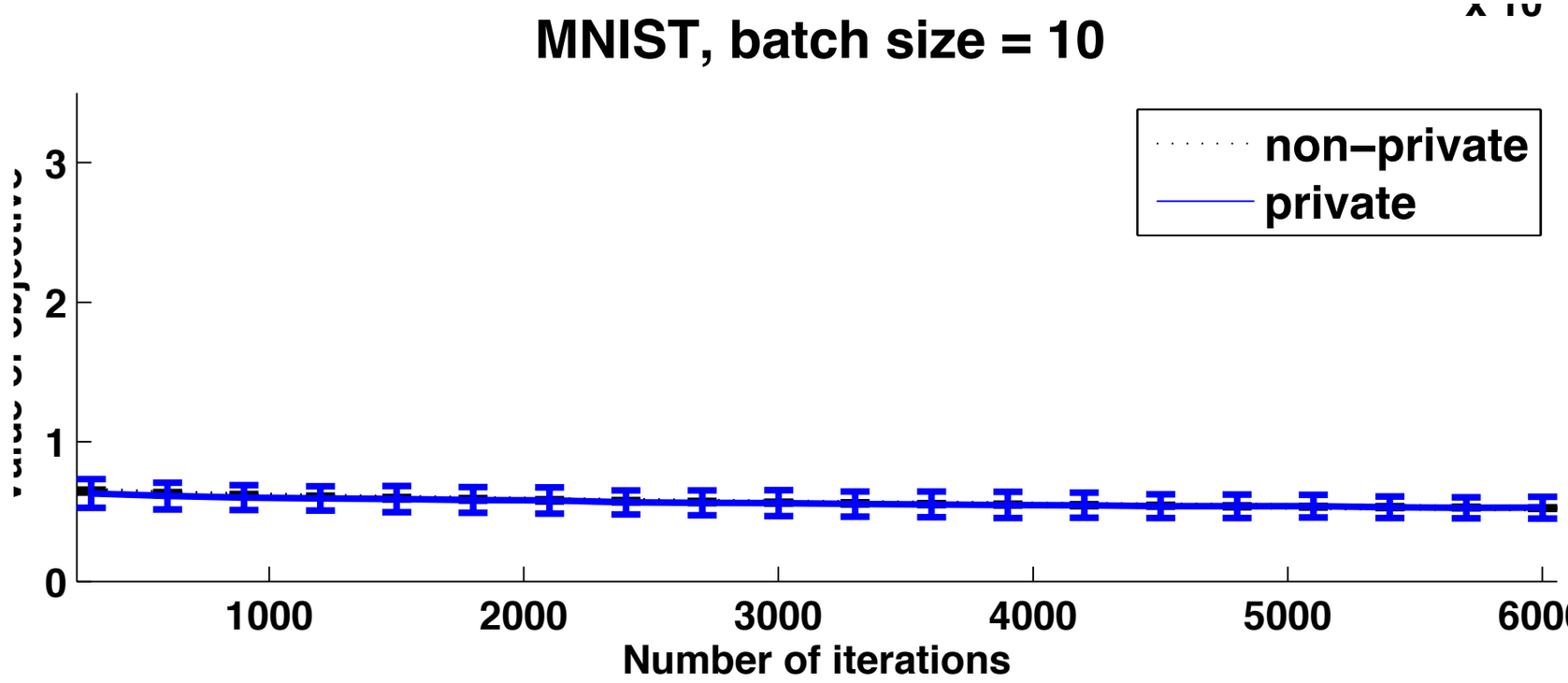
Stochastic gradient descent is a stochastic optimization method widely used for learning from large data sets:

$$w_{t+1} = w_t - \eta_t(\nabla L(y_t w_t^\top x_t) + \lambda w_t)$$

Processes the data points one at a time and takes incremental gradient steps. We can run SGD on the perturbed objective:

$$w_{t+1} = w_t - \eta_t(\nabla L(y_t w_t^\top x_t) + \lambda w_t + Z_t)$$

SGD with minibatching [SCS13]



Performance of SGD can be bad with *pure* local privacy: improved performance dramatically by processing small batches of points at a time (minibatching):

$$w_{t+1} = w_t - \eta_t(\nabla \sum_{i \in B_t} L(y_i w_t^\top x_i) + \lambda w_t + Z_t)$$

Extending SGD

Several works extending and analyzing stochastic gradient approaches for inference and learning under privacy:

- Optimal rates for parameter estimation under local privacy [DJW14]
- Improved analysis to take advantage of random sampling of data points in SGD [BST14]
- Learning from multiple data sets with different privacy requirements [SCS14]

Pufferfish framework [KM14]

Goal: privacy framework that accounts for prior knowledge of the adversary and specifies statements that are to be kept secret.

- Bayesian privacy framework (see also [KS14])
- New privacy definition: hedging privacy
- Bayesian semantics for differential privacy

Provides less absolute guarantees than *pure* differential privacy, but allows an analysis that moves away from the worst case.

Pan-privacy and online monitoring [DNPRY10]

Goal: privacy for streaming algorithms.

- Stronger definition which allows the adversary to view the internal state of the algorithm.
- Applications to online density estimation and other monitoring applications.

Stronger guarantees but little empirical evaluation.

Building differentially private systems

Several attempts to build differentially private systems:

- Airavat [RSKS10]
- GUPT [MTSSC12]

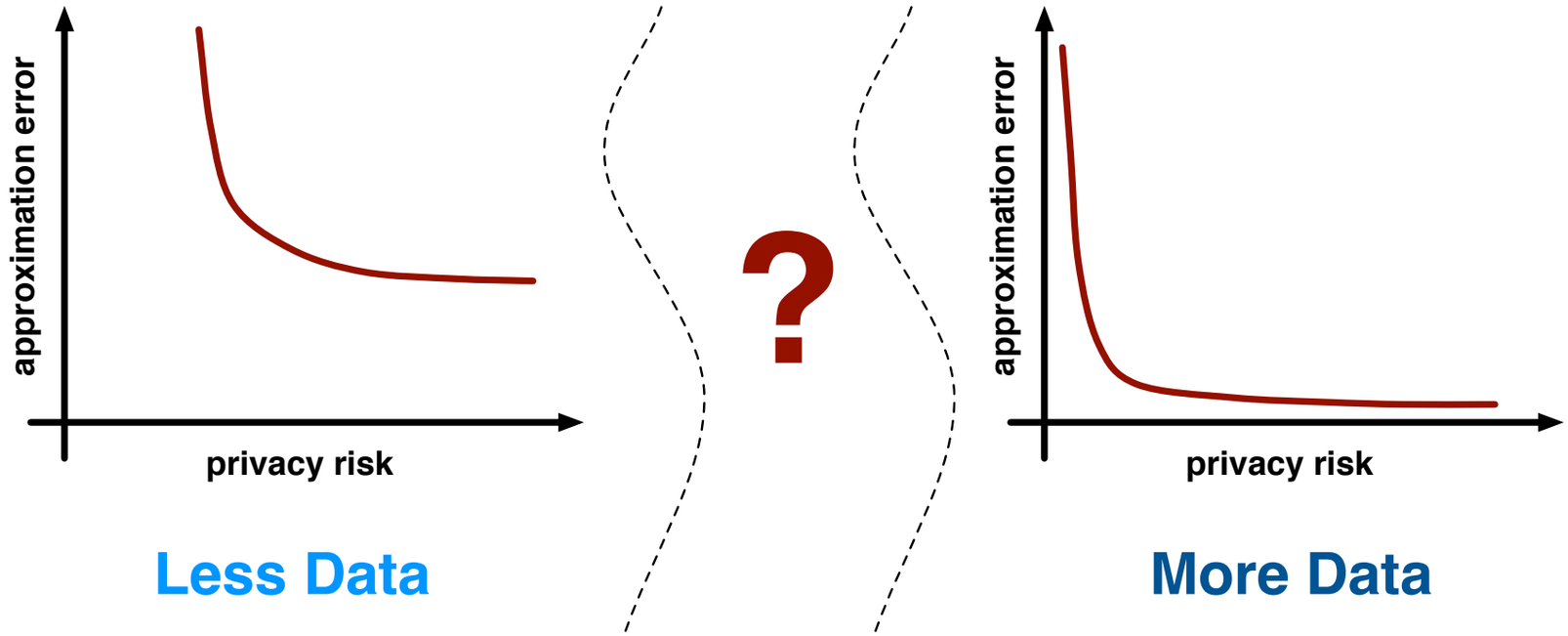
Programming languages

- PINQ [MTSSC12]
- DFuzz [GHHNP13]

Mechanism design and economics

Extensive work on economics and mechanism design for differential privacy [PR13]:

- designing truthful mechanisms for auctions [NST12] [CCKMV13] [GR11]
- equilibrium strategies for games [KPRU12]
- survey design [R12] [GR11] [FL12]



Looking forward

Summary

Differential privacy is a rigorous model for privacy that provides guarantees on the additional risk of re-identification from disclosures:

- nice formal properties
- many mechanisms for designing algorithms
- algorithms and applications in several domains

Ongoing and future work

A lot more work is needed for the future:

- better insight into picking epsilon
- more evaluations on real data
- standard implementations and toolboxes for practitioners
- modified definitions and foundations for different application domains

Other surveys and materials

- Sarwate and Chaudhuri, *Signal processing and machine learning with differential privacy: theory, algorithms, and challenges*, IEEE SP Magazine, September 2013.
- Dwork and Roth, *The Algorithmic Foundations of Differential Privacy*, Foundations and Trends in Theoretical Computer Science, Vol. 9., 2014.
- Dwork and Smith, *Differential privacy for statistics: What we know and what we want to learn*. Journal of Privacy and Confidentiality, 1(2), 2009.
- Simons Workshop on Big Data and Differential Privacy: http://simons.berkeley.edu/workshops/bigdata2013-4

More References

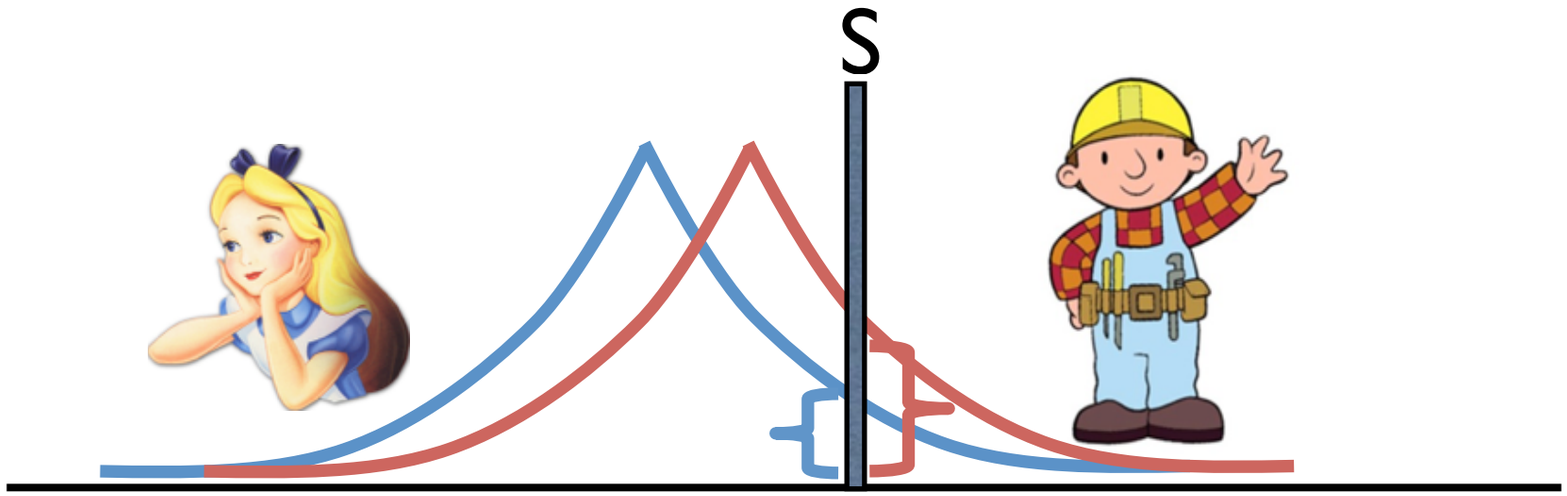
[BST14] Bassily et al., FOCS 2014
[BCDKMT07] Blum et al. PODS 2007
[BDMN05] Blum et al. PODS 2005
[BLR13] Blum et al. JACM 2013
[CHS14] Chaudhuri et al., NIPS 2014
[CMS11] Chaudhuri et al., JMLR 2011
[CSS13] Chaudhuri et al., JMLR 2013
[CV13] Chaudhuri and Vinterbo, NIPS 2013
[CCKMV13] Chen et al., EC 2013
[DJW12] Duchi et al. NIPS 2012
[DJW13] Duchi et al. NIPS 2013
[DL09] Dwork et al. STOC 2009
[DMNS06] Dwork et al., TCC 2006
[DNPRY10] Dwork et al., ICS 2010
[DRV09] Dwork et al. STOC 2009
[FSU12] Fienberg et al.
[FLJLPR14] Fredrikson et al., USENIX Security 2014

Even More References

[GHHNP] Gaboardi et al., POPL 2013
[HRW13] Hall et al. JMLR 2013
[HTP14] Han et al. Allerton 2014
[HLM12] Hardt et al., NIPS 2012
[JT13] Jain and Thakurta, COLT 2013
[JJXWO14] Ji et al, BMC Med. Genomics 2014
[KT13] Kapralov and Talwar, SODA 2013
[KS08] Kasiviswanathan and Smith, 2008
[KS14] Kasiviswanathan and Smith, J. Priv. and Confidentiality 2014
[KPRU12] Kearns et al, 2012
[KST12] Kifer et al, COLT 2012
[KM14] Kifer and Machanavajjhala., TODS 2014
[LR12] Ligett and Roth, Internet and Network Econ. 2012
[MKAGV08] Machanavajjhala et al., ICDE 2008
[MICMW13] Mir et al., IEEE BigData 2013
[MT07] McSherry and Talwar, FOCS 2007
[MTSSC12] Mohan et al., SIGMOD 2012

Yet Even More References

[NRS07] Nissim et al. STOC 2007
[NS08] Narayanan and Shmatikov, Oakland S&P 2008
[OV13] Oh and Viswanath, ArXiV 2013
[PR13] Pai and Roth, tutorial, 2013
[R12] Roth, SIGecom 2012
[RBHT12] Rubinstein et al, J. Priv. and Confidentiality 2012
[RSKS] Roy et al. NSDI 2010
[SC13] Sarwate and Chaudhuri, SP Mag. 2013
[SCS13] Song et al., GlobalSIP 2013
[SCS14] Song et al., preprint 2013
[SPTAC14] Sarwate et al., Frontiers in Neuroinformatics, 2014
[VSB12] Vinterbo et al. JAMIA 2012
[W65] Warner, JASA 1965
[WLWTZ09] Wang et. al, CCS 2009
[WZ10] Wasserman and Zhou, JASA 2010
[ZZXYW12] Zhang et al., VLDB 2012



Thanks!