Machine Learning and Differential Privacy

Ellen Vitercik

December 5, 2018

Today we'll talk about…

1. **The importance of privacy in machine learning**

2. One way of defining privacy (*differential privacy*)

3. Tools for designing privacy-preserving algorithms

Learning and privacy

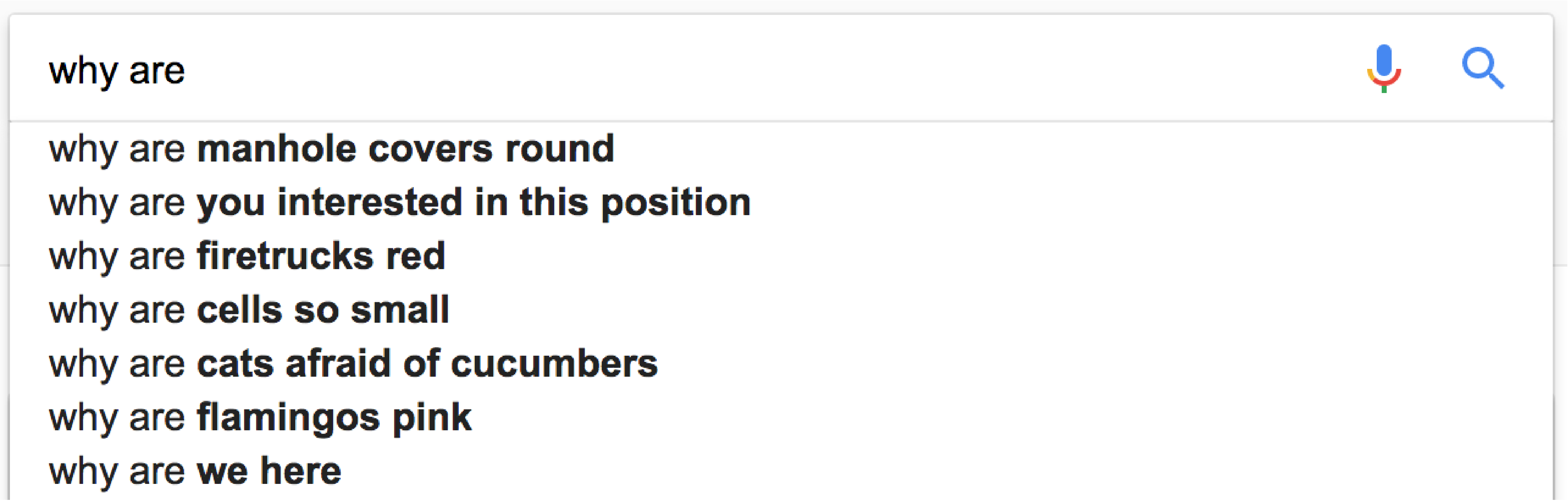
To do machine learning, we need data

What if the data contains sensitive information?

Is it enough to trust the person running the learning algorithm?

No: Perhaps algorithm’s output reveals sensitive information

Example: search query completions



Example: search query completions

What if we use your friends' search logs to suggest completions?

Might be good for accuracy, but…

why are _
why are **my feet so itchy?**

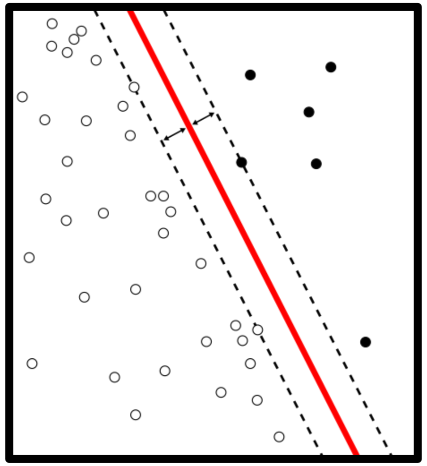

Privacy leaks can be subtle!

Hospital wants to be able to predict who has condition *X*

Collect data from residents, use perceptron algorithm

$\boldsymbol{w} =$

0.1	1.3	1	0



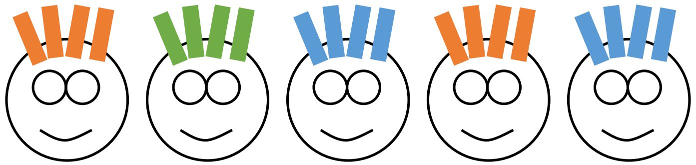
Privacy leaks can be subtle!

Hospital wants to be able to predict who has condition *X*

Collect data from residents, use perceptron algorithm

$$\boldsymbol{w} = \begin{array}{|c|c|c|c|}\hline 0.1 & 1.3 & 1 & 0 \\ \hline \end{array}$$

Has green hair?



Only one person in town has green hair.

We now know the green-haired person has condition *X*!

How can we be confident that this won't happen?

Today we’ll talk about…

1. The importance of privacy in machine learning

2. **One way of defining privacy (*differential privacy*)**

3. Tools for designing privacy-preserving algorithms

What is privacy?

?

What isn’t privacy?

Privacy isn’t **restricting questions to large populations**.

- “What is the average salary of CMU faculty?”
- “What is the average salary of CMU faculty not named Nina Balcan?”

Adapted from Aaron Roth’s 2011 course slides

What isn’t privacy?

Privacy isn’t **restricting to “ordinary” facts**

Statistics on Alice’s bread buying habits:

For 20 years she regularly buys bread, then stops.

Type 2 diabetes?

What isn’t privacy?

Privacy isn’t **“Anonymization”**

Case study: Publicly available “anonymized” hospitalization data

- Latanya Sweeney re-identified patients by name

The New York Times

Bits
Business, Innovation, Technology, Society

PRIVACY

With a Few Bits of Data, Researchers Identify ‘Anonymous’ People

By Natasha Singer January 29, 2015 2:01 pm

Adapted from Aaron Roth’s 2011 course slides

What is privacy?

Attempt 1:

Analysis of dataset D is private if:

Analyst knows no more about Alice after analysis than before.

Problematic example:

Analysis of dataset D ⇒ West Virginians have high obesity rates

Alice, whose information **isn't** in dataset D, lives in WV

Insurance agency knows Alice lives in WV ⇒ they raise her rates!

Was Alice's privacy violated?

Yes, under this definition…

What is privacy?

Attempt 2:

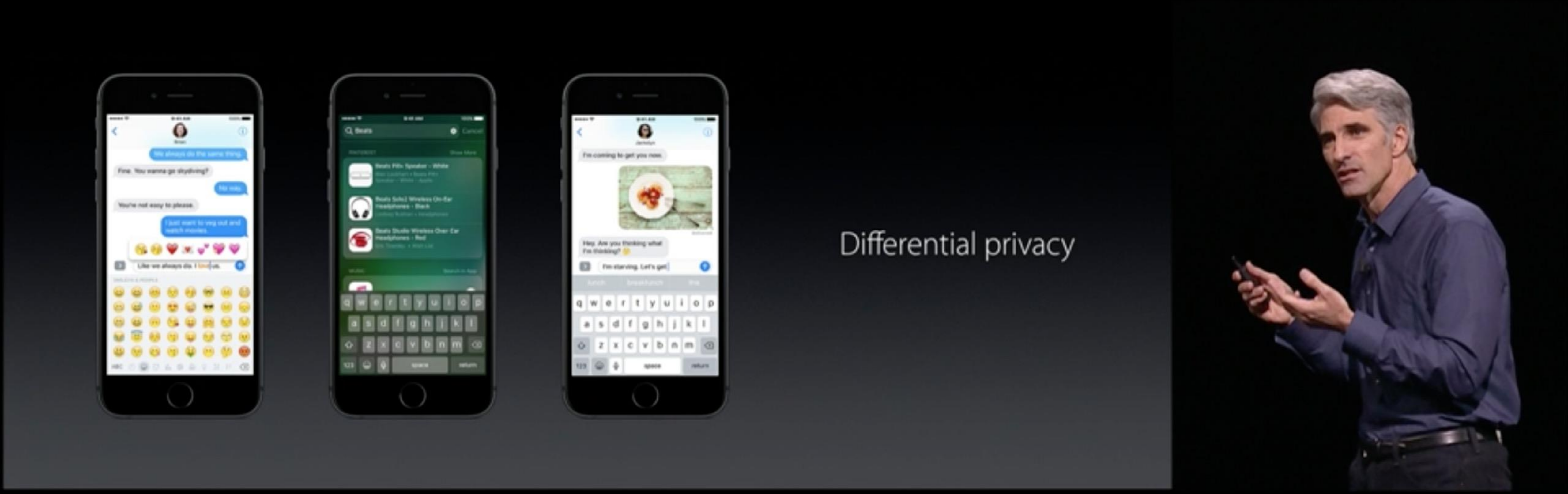
Analysis of dataset D is private if:
analyst knows **almost** no more about Alice after analysis
than he **would have**
had he conducted the **same analysis** on
an **identical** dataset w/ **Alice’s data removed**

Adapted from Aaron Roth’s 2011 course slides

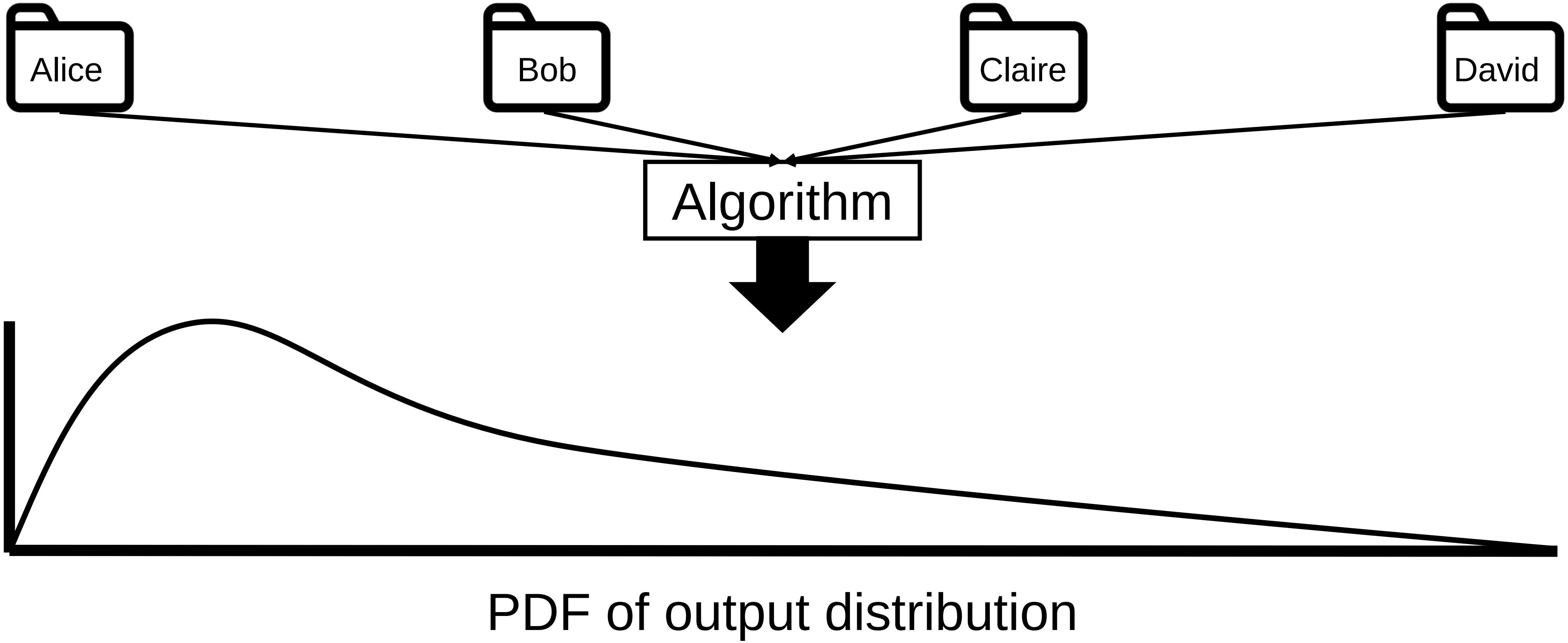
Differential privacy

“Calibrating Noise to Sensitivity in Private Data Analysis.” Dwork, McSherry, Nissim, and Smith. *TCC*. 2006.

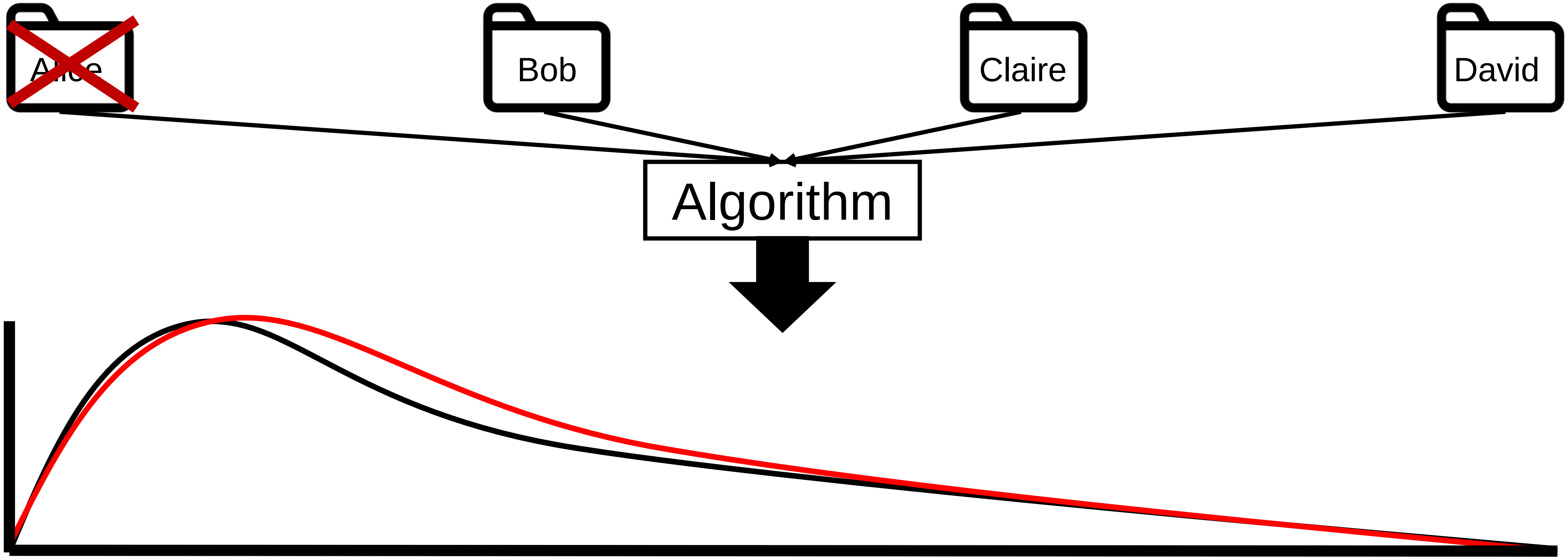
“The Algorithmic Foundations of Differential Privacy”. Dwork and Roth. Foundations and Trends in Theoretical Computer Science, NOW Publishers. 2014.


Differential privacy

Differential privacy



Differential privacy

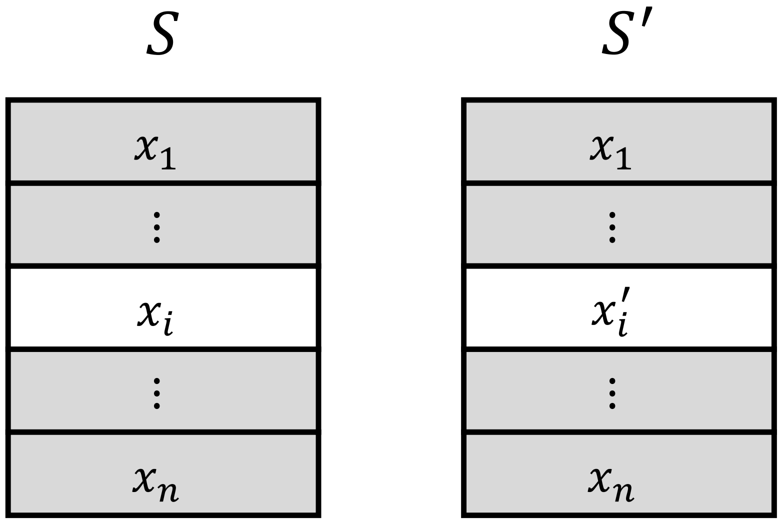


Can’t tell if used Alice’s data, let alone what her data was!

Differential privacy

Def: Two datasets S, S' are **neighboring** if differ on ≤1 entry

1 entry ≡ 1 person



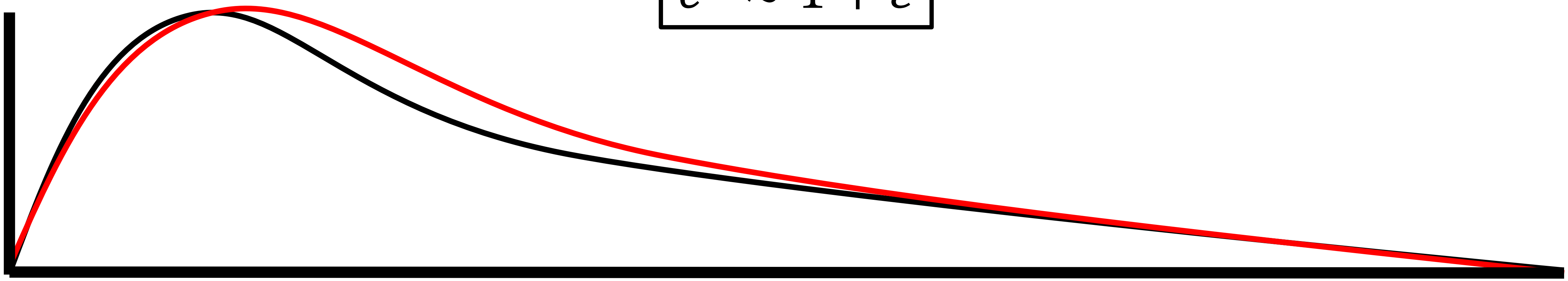
Differential privacy

Algorithm $\mathcal{A}$ is **$\boldsymbol{\varepsilon}$-differentially private** if:

For all pairs of neighboring sets S, S' and all sets $\mathcal{O}$ of outputs,

$$\mathbb{P}[\mathcal{A}(S) \in \mathcal{O}] \leq e^{\varepsilon}\mathbb{P}[\mathcal{A}(S') \in \mathcal{O}]$$

$$\boxed{e^{\varepsilon} \approx 1 + \varepsilon}$$



DP protects against additional harm

$\mathcal{A}$:= DP algorithm

$f\colon \mathrm{Range}(\mathcal{A}) \to W$ maps $\mathcal{A}$'s output to a future world state $w \in W$

Suppose I have a utility function $u\colon W \to \mathbb{R}$

E.g., $u(w)$ = "how happy am I if the world is w"

DP guarantees that

$$\mathbb{E}_{w\sim f(\mathcal{A}(S))}[u(w)] \approx e^{\pm\epsilon} \cdot \mathbb{E}_{w\sim f(\mathcal{A}(S'))}[u(w)]$$

Today we'll talk about…

1. The importance of privacy in machine learning

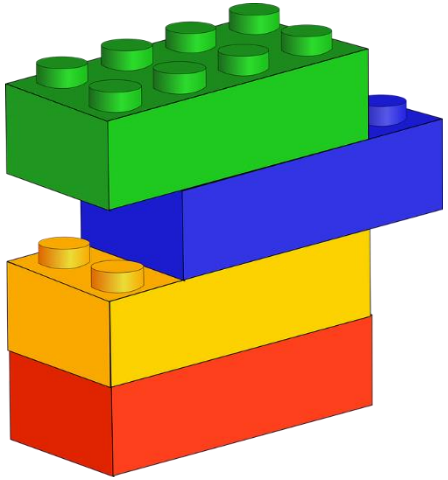
2. One way of defining privacy (*differential privacy*)

3. Tools for designing privacy-preserving algorithms
    a) **Laplace mechanism**
    b) Exponential mechanism
    c) Composing private algorithms
    d) Examples of differentially-private ML tools

Laplace mechanism

Very useful building block for designing private algorithms.

“Calibrating Noise to Sensitivity in Private Data Analysis.” Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. TCC, 2006.

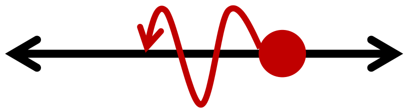


Laplace mechanism

Goal: Evaluate $f: D \to \mathbb{R}$ mapping datasets to $\mathbb{R}$; preserve ϵ-DP

Ex., $f(S) :=$ mean weight of people in S

Idea: Compute $f(S)$ and $\underbrace{\text{add noise}}_{\text{How little can we get away with?}}$ to hide any individual's info



Laplace mechanism

Goal: Evaluate $f: D \to \mathbb{R}$ mapping datasets to $\mathbb{R}$; preserve ϵ-DP

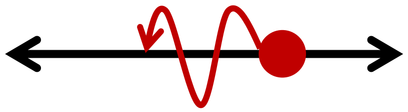
Ex., $f(S) \coloneqq$ mean weight of people in S

Idea: Compute $f(S)$ and add noise to hide any individual's info

Def: Sensitivity of f is $\Delta_f = \max_{S,S' \text{ neighboring}} |f(S) - f(S')|$

Laplace Mechanism outputs $Z_S \sim \mathrm{Lap}\left(f(S), \frac{\Delta_f}{\epsilon}\right)$

PDF $p_{Z_S}(z) = \frac{\Delta_f}{2\epsilon} \exp\left(-\frac{\epsilon}{\Delta_f} |z - f(S)|\right)$



Laplace mechanism: Privacy guarantees

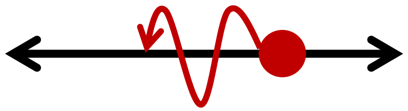
Def: Sensitivity of f is $\Delta_f = \max_{S,S' \text{ neighboring}} |f(S) - f(S')|$

Laplace Mechanism outputs $Z_S \sim \mathrm{Lap}\left(f(S), \frac{\Delta_f}{\epsilon}\right)$

PDF $p_{Z_S}(z) = \frac{\Delta_f}{2\epsilon} \exp\left(-\frac{\epsilon}{\Delta_f} |z - f(S)|\right)$

Privacy: The Laplace mechanism preserves ϵ-DP.

We'll see why on the board.



Laplace mechanism: Utility guarantees

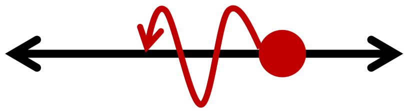
Def: Sensitivity of f is $\Delta_f = \max_{S,S' \text{ neighboring}} |f(S) - f(S')|$

Laplace Mechanism outputs $Z_S \sim \text{Lap}\left(f(S), \frac{\Delta_f}{\epsilon}\right)$

PDF $p_{Z_S}(z) = \frac{\Delta_f}{2\epsilon} \exp\left(-\frac{\epsilon}{\Delta_f}|z - f(S)|\right)$

Utility: With probability at least $\boldsymbol{1 - \delta}$, $\boldsymbol{|Z_S - f(S)| \le \frac{\Delta_f}{\epsilon} \log \frac{1}{\delta}}$.

Proof idea: analyze Laplace distribution's CDF.



Laplace mechanism: Computing means

Given set $S = \{x_1, \dots, x_n\} \subset [0,1]$, privately compute $f(S) = \frac{1}{n}\sum x_i$

Question: What is $\Delta_f = \max_{S,S' \text{ neighboring}} |f(S) - f(S')|$?

Answer: $\Delta_f = \frac{1}{n}$

Laplace mechanism: Computing means

Given set $S = \{x_1, \ldots, x_n\} \subset [0,1]$, privately compute $f(S) = \frac{1}{n}\sum x_i$

Recall: Laplace mech. outputs $Z_S \sim \mathrm{Lap}\left(f(S), \frac{1}{n\epsilon}\right)$

Laplace mechanism: Computing means

Given set $S = \{x_1, \ldots, x_n\} \subset [0,1]$, privately compute $f(S) = \frac{1}{n}\sum x_i$

Utility: With probability at least $1 - \delta$, $|Z_S - f(S)| \leq \frac{1}{n\epsilon} \log \frac{1}{\delta}$.

If $S \sim P^n$ and goal is to estimate $\mathbb{E}_{x\sim P}[x]$ using $f(S)$, w.p. $1 - \delta$,

$$|\mathbb{E}_{x\sim P}[x] - f(S)| \leq \sqrt{\frac{1}{2n} \ln \frac{1}{\delta}}.$$

Error due to privacy negligible compared to sampling error!

Laplace mechanism: Multi-dim functions

What if function f maps to $\mathbb{R}^d$? I.e., $f: D \to \mathbb{R}^d$

Example: $f(S) = \langle \text{mean weight in } S, \text{mean height in } S \rangle$

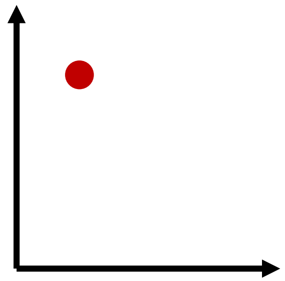
Def: The sensitivity of f is $\Delta_f = \max_{S,S' \text{ neighboring}} \|f(S) - f(S')\|_1$.

Def: The Laplace Mechanism outputs $f(S) + \boldsymbol{Z}$

$\boldsymbol{Z} \in \mathbb{R}^d$ has components drawn from $\mathrm{Lap}\left(0, \frac{\Delta_f}{\epsilon}\right)$ distribution

Privacy: The Laplace mechanism preserves ϵ-DP

Utility: With probability at least $1 - \delta$, $\|\boldsymbol{Z}\|_\infty \leq \frac{\Delta_f}{\epsilon} \log \frac{d}{\delta}$



Today we'll talk about…

1. The importance of privacy in machine learning

2. One way of defining privacy (*differential privacy*)

3. Tools for designing privacy-preserving algorithms
   a) Laplace mechanism
   b) **Exponential mechanism**
   c) Examples of differentially-private ML tools

Exponential mechanism

Goal: Choose the “best” item from a finite set Y of items

E.g., voting in a local election

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In Foundations of Computer Science. 2007.

Exponential mechanism

Given utility function $u(S, y) =$ “utility of y for dataset S”
Goal: Find $y \in Y$ maximizing $u(S, y)$

Question: Why can’t we use the Laplace Mechanism?

Answer: E.g., $Y = \{$town welder, town farmer,…$\}$
How do we add noise to “town mechanic”?

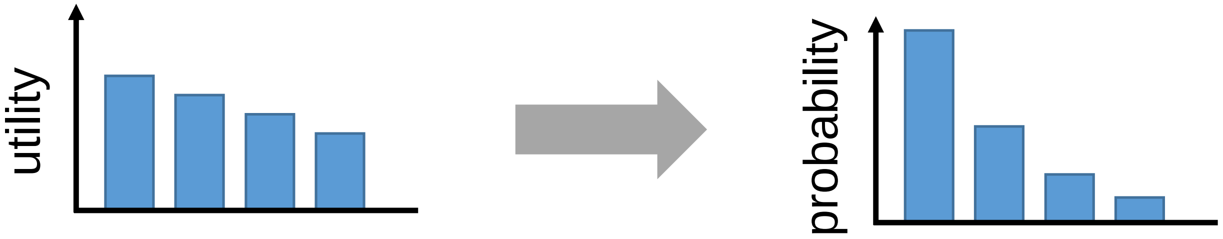
Exponential mechanism

Given utility function $u(S, y)$ = "utility of y for dataset S"

Goal: Find $y \in Y$ maximizing $u(S, y)$

Def: The sensitivity of u is $\Delta_u = \max_{S,S',y} |u(S, y) - u(S', y)|$

Exponential Mechanism outputs y with w.p. $\propto \exp\left(\frac{\epsilon}{2\Delta_u} u(S, y)\right)$



Exponential mechanism

Given utility function $u(S, y)$ = "utility of y for dataset S"

Goal: Find $y \in Y$ maximizing $u(S, y)$

Def: The sensitivity of u is $\Delta_u = \max_{S,S',y} |u(S, y) - u(S', y)|$

Exponential Mechanism outputs y with w.p. $\propto \exp\left(\frac{\epsilon}{2\Delta_u} u(S, y)\right)$

Privacy: The exponential mechanism preserves ϵ-DP.

Proof follows from algebraic manipulations of density function.

Exponential mechanism

Given utility function $u(S, y) =$ "utility of y for dataset S"

Goal: Find $y \in Y$ maximizing $u(S, y)$

Def: The sensitivity of u is $\Delta_u = \max_{S,S',y} |u(S,y) - u(S',y)|$

Exponential Mechanism outputs y with w.p. $\propto \exp\left(\frac{\epsilon}{2\Delta_u} u(S,y)\right)$

Privacy: The exponential mechanism preserves ϵ-DP.

We'll see why on the board.

Database sanitization

Given dataset S, produce synthetic dataset $\hat{S}$, preserve DP

Ideally: $\hat{S}$ behaves basically the same as S (for our purposes)

Based on "A learning theory approach to noninteractive database privacy.“ Avrim Blum, Katrina Ligett, Aaron Roth. *Journal of the ACM (JACM)* 60.2 (2013): 12.

Database sanitization

More formally:

- Let $S \subseteq \{0,1\}^d$ be a dataset of d-dimensional binary vectors
- Let H be a set of functions $h: \{0,1\}^d \to \{0,1\}$ with VC-dim D
- Let $h(S) = \frac{1}{|S|}\sum_{\boldsymbol{x}\in S} h(\boldsymbol{x})$ be the fraction of $\boldsymbol{x} \in S$ with $h(\boldsymbol{x}) = 1$

If $|\boldsymbol{S}| \geq \widetilde{\boldsymbol{O}}\left(\frac{\boldsymbol{dD}}{\boldsymbol{\alpha^3 \epsilon}}\right)$, can find $\hat{\boldsymbol{S}} \subset \{\boldsymbol{0},\boldsymbol{1}\}^{\boldsymbol{d}}$ while preserving $\boldsymbol{\epsilon}$-DP s.t. w.h.p., for all $\boldsymbol{h} \in \boldsymbol{H}$, $\left|\boldsymbol{h}(\boldsymbol{S}) - \boldsymbol{h}(\hat{\boldsymbol{S}})\right| \leq \boldsymbol{\alpha}$.

Proof uses VC dimension guarantees and probabilistic method

Today we'll talk about…

1. The importance of privacy in machine learning

2. One way of defining privacy (*differential privacy*)

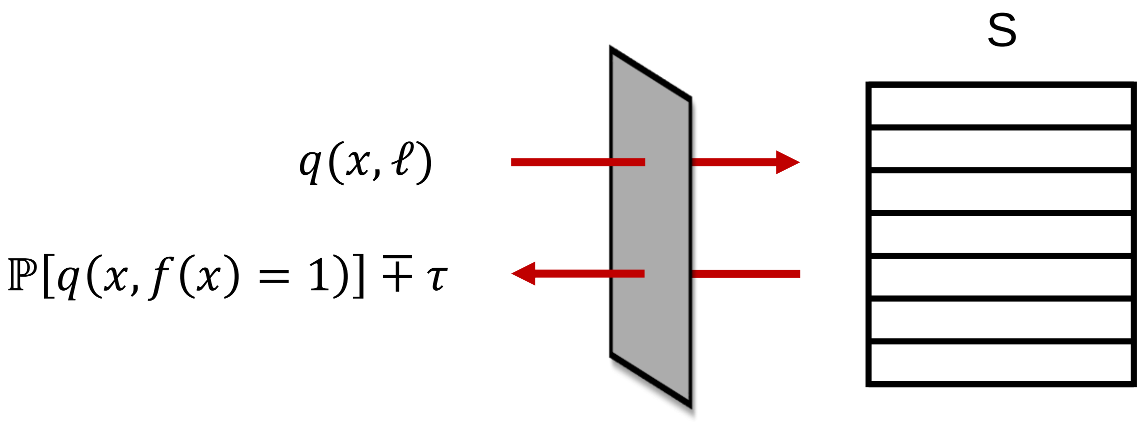
3. Tools for designing privacy-preserving algorithms
    a) Laplace mechanism
    b) Exponential mechanism
    c) **Examples of differentially-private ML tools**

DP + ML using statistical queries

Anything learnable using *statistical queries* is privately learnable.

Practical Privacy: The SuLQ Framework. Blum, Dwork, McSherry, Nissim. *PODS.* 2005.

Statistical query model [Kearns, ’98]**:**

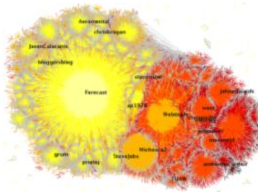


Many algorithms (e.g., ID3, Perceptron, SVM, PCA) can be re-written to interface via statistical queries.

DP+ML more generally

Active learning

Balcan and Feldman. "Statistical active learning algorithms." *NeurIPS*. 2013.



Clustering

Balcan, Dick, Liang, Mou, and Zhang. “Differentially Private Clustering in High-Dimensional Euclidean Spaces.” *ICML*. 2017.

Distributed learning

Blacan, Blum, Fine, and Mansour. “Distributed Learning, Communication Complexity and Privacy.” *COLT*. 2012.