

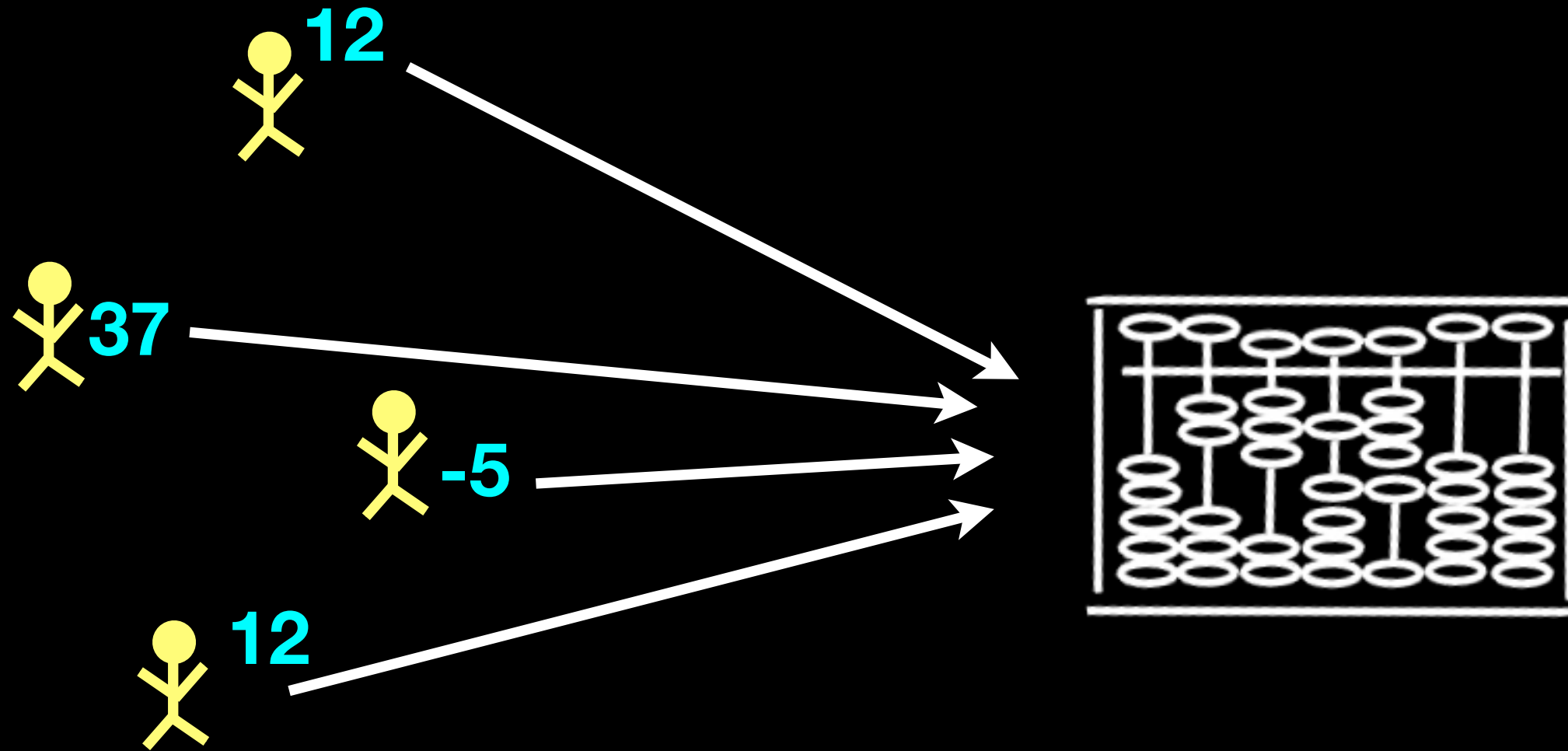
Accuracy First:
Selecting a Differential Privacy Level for
Accuracy-Constrained Empirical Risk
Minimization

Katrina Ligett
HUJI & Caltech

joint with
Seth Neel, Aaron Roth, Bo Waggoner, Steven Wu

NIPS 2017

many of today's most
interesting computations



are on individuals' data

facebook

Google

data privacy



data privacy

data privacy **laws**

data privacy **policy**

data privacy **act**

Google Search

I'm Feeling Lucky

facebook

Google

data privacy



data privacy

data privacy **laws**

data privacy **policy**

data privacy **act**

Google Search

I'm Feeling Lucky

facebook





data privacy



data privacy

data privacy laws

data privacy policy

data privacy act

Google Search

I'm Feeling Lucky

facebook



waze

OUTSMARTING TRAFFIC, TOGETHER.



NETFLIX



Google

data privacy

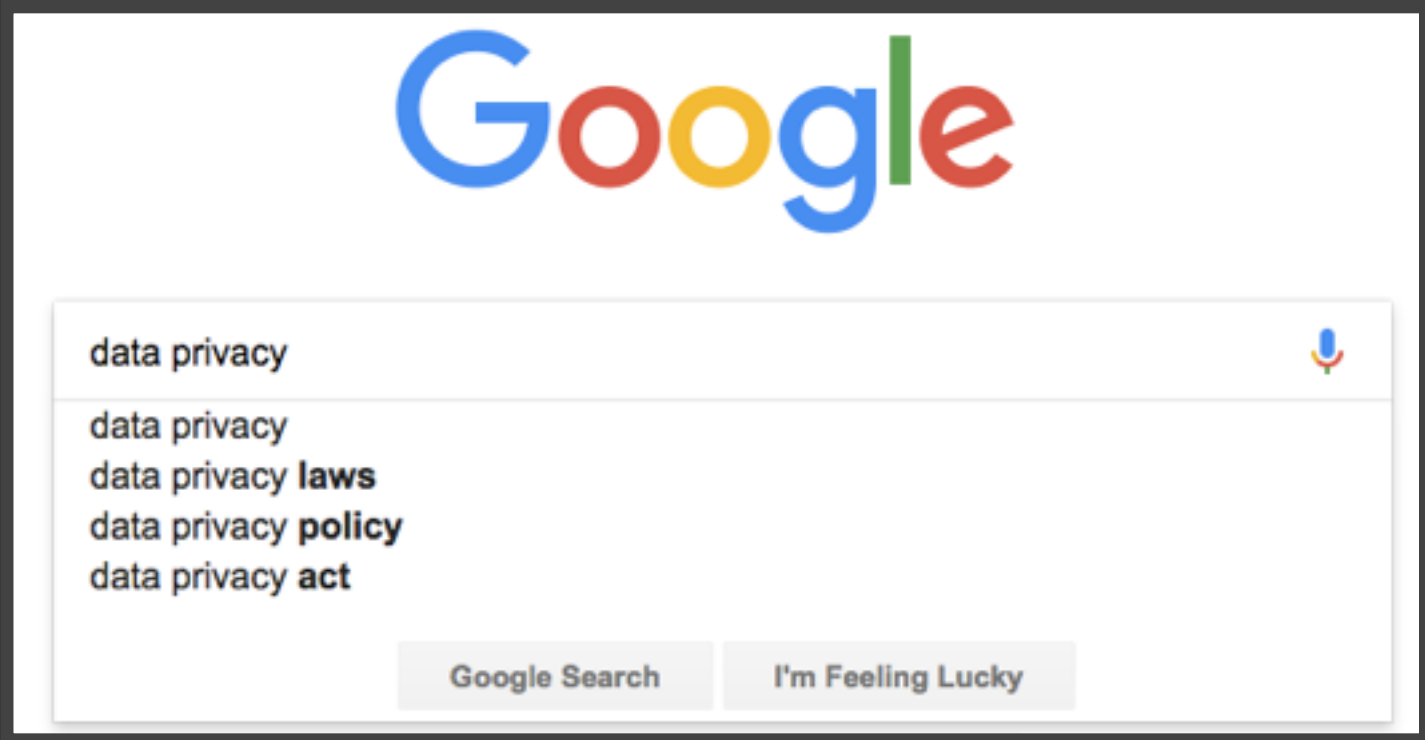
data privacy
data privacy laws
data privacy policy
data privacy act

Google Search I'm Feeling Lucky

facebook



NETFLIX

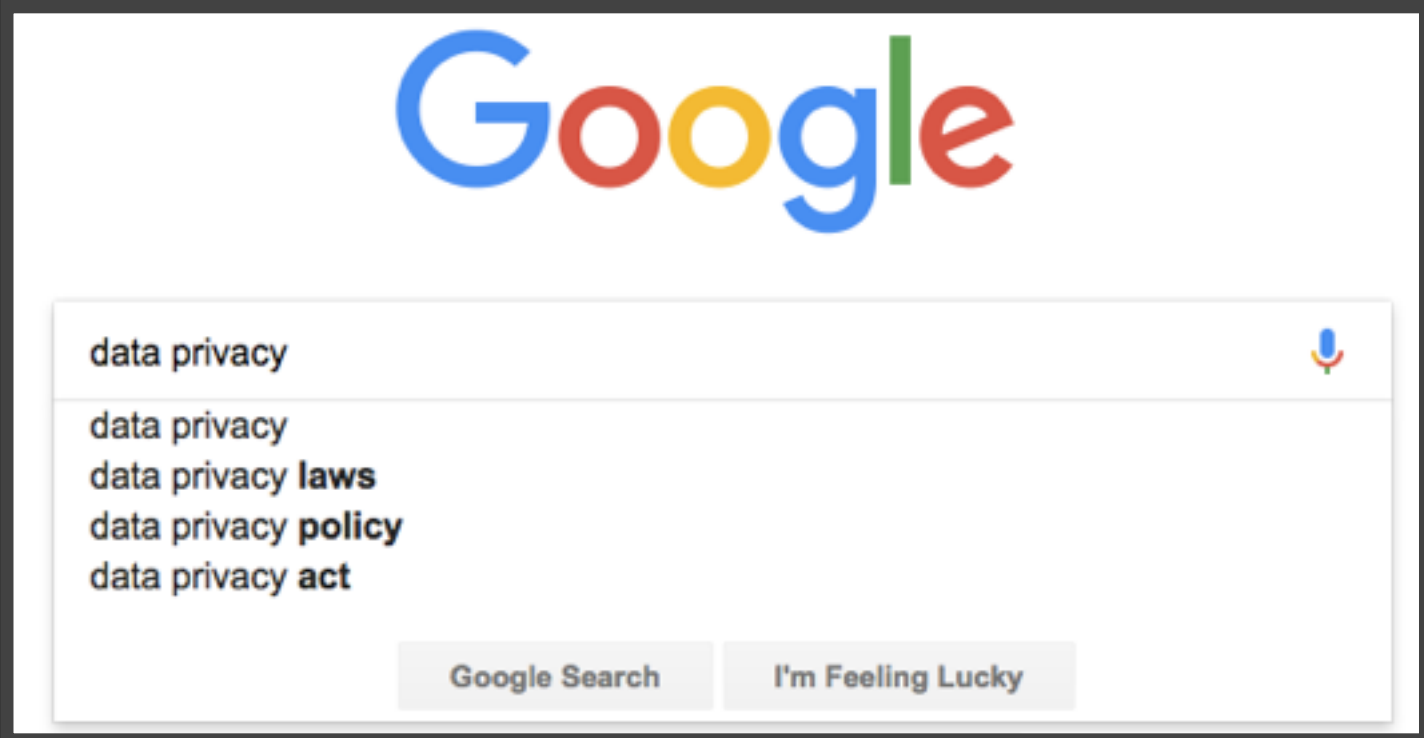


facebook



amazon

NETFLIX



Google

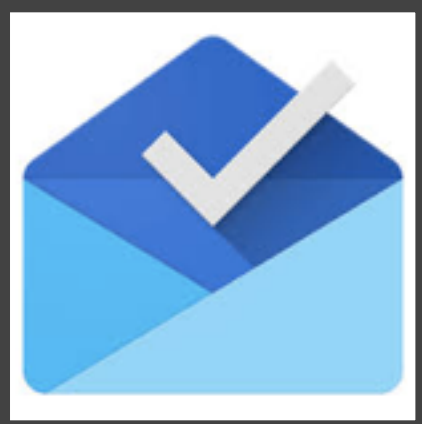
data privacy

data privacy
data privacy laws
data privacy policy
data privacy act

Google Search I'm Feeling Lucky

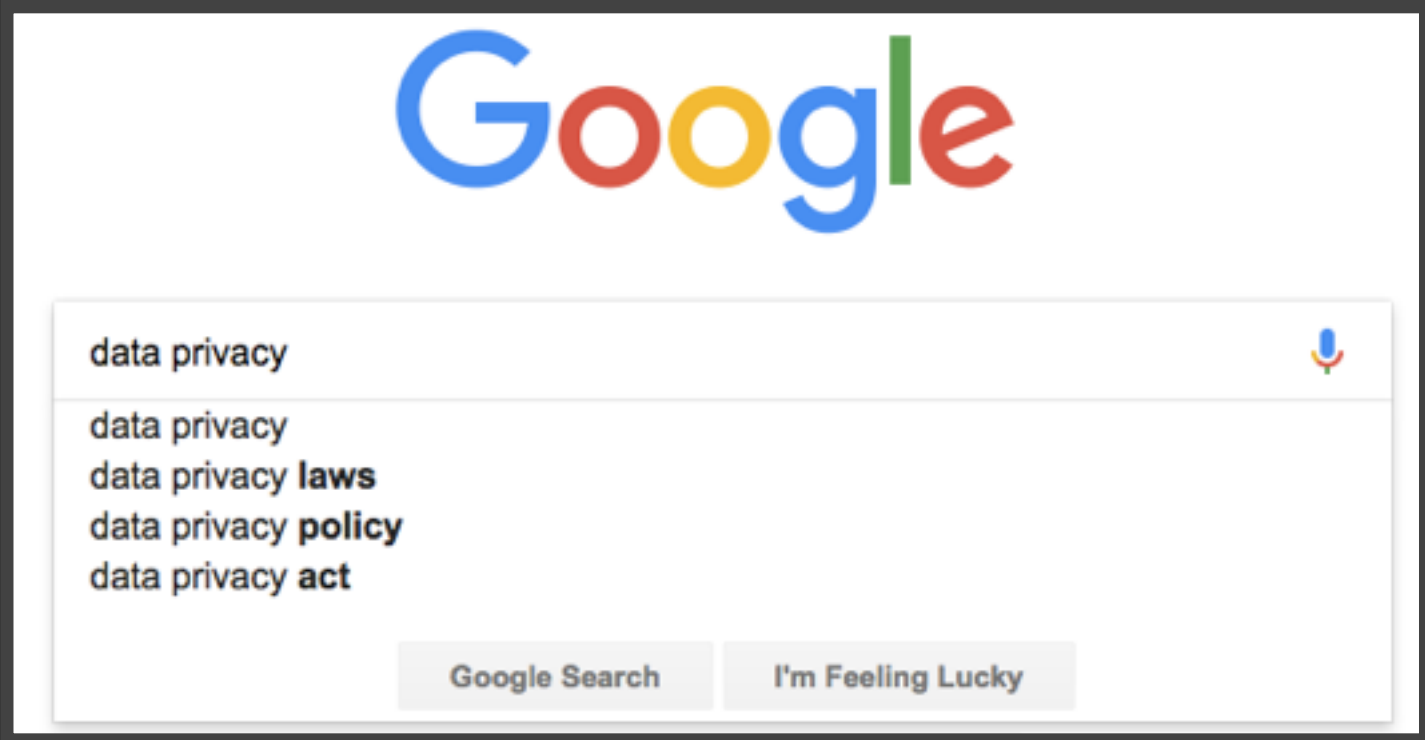


facebook



amazon

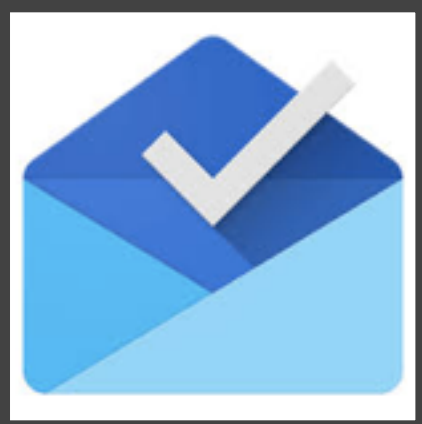
NETFLIX



facebook



amazon



many times, want to do this subject to privacy

- legal/regulatory requirements
- avoid subpoena
- public perception
- brand identity

Apple will not
see your data



sometimes, privacy requirements are clear and binding

- legal/regulatory requirements
- avoid subpoena
- public perception
- brand identity

sometimes, privacy requirements
are clear and binding

- legal/regulatory requirements

well-studied optimization problem:
meet privacy requirements while
minimally impacting accuracy

sometimes, privacy requirements
are clear and binding

...but often not

- avoid subpoena
- public perception
- brand identity

this paper: how to understand the
best privacy we can give, subject to
accuracy constraints?

today

→ Formalizing privacy: differential privacy

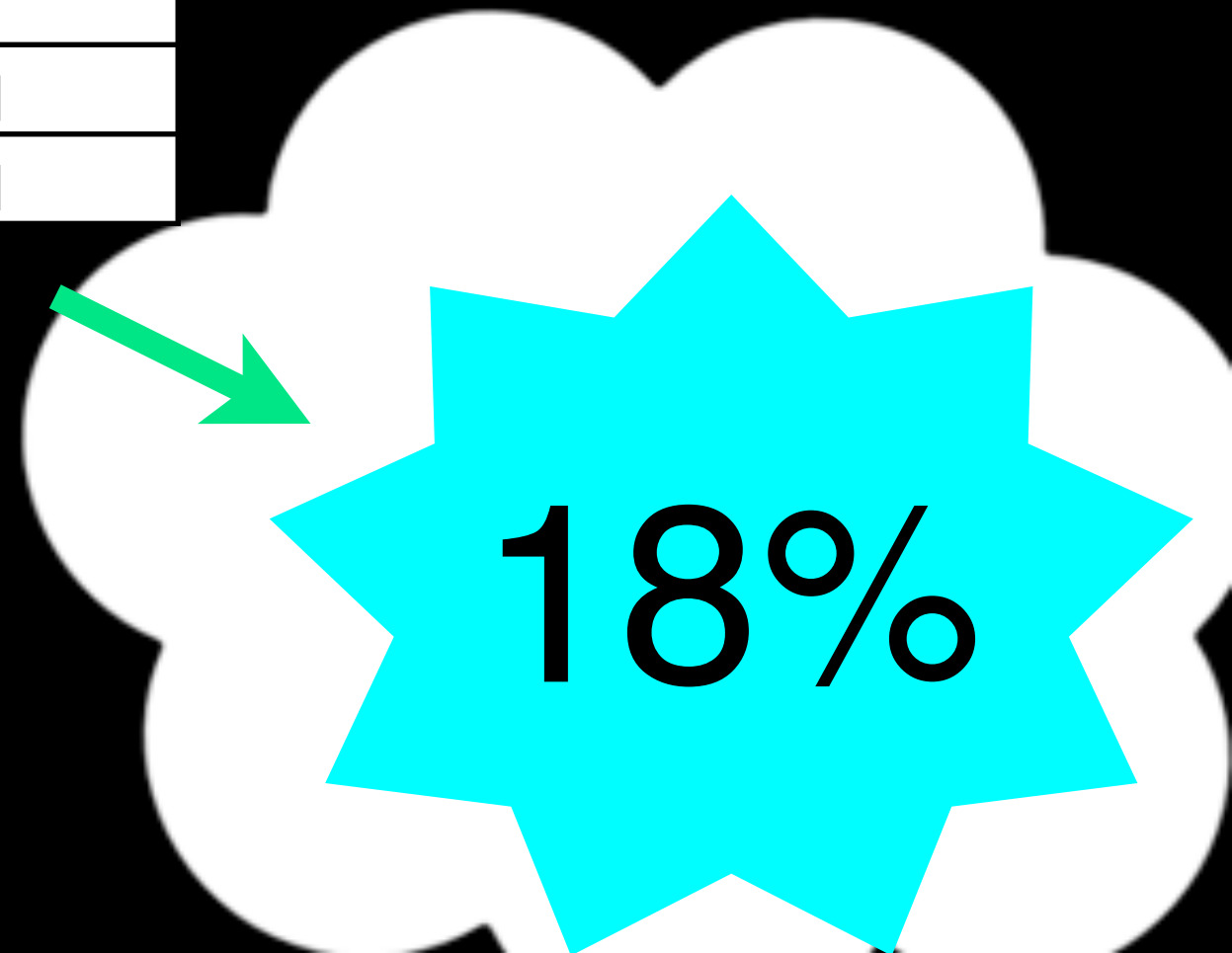
- (private) Empirical risk minimization (ERM)
- Accuracy-First Private ERM

privacy: what to promise?

access to the output should not enable one to learn much more about an individual than could be learned via the same analysis omitting that individual from the database

what to promise?

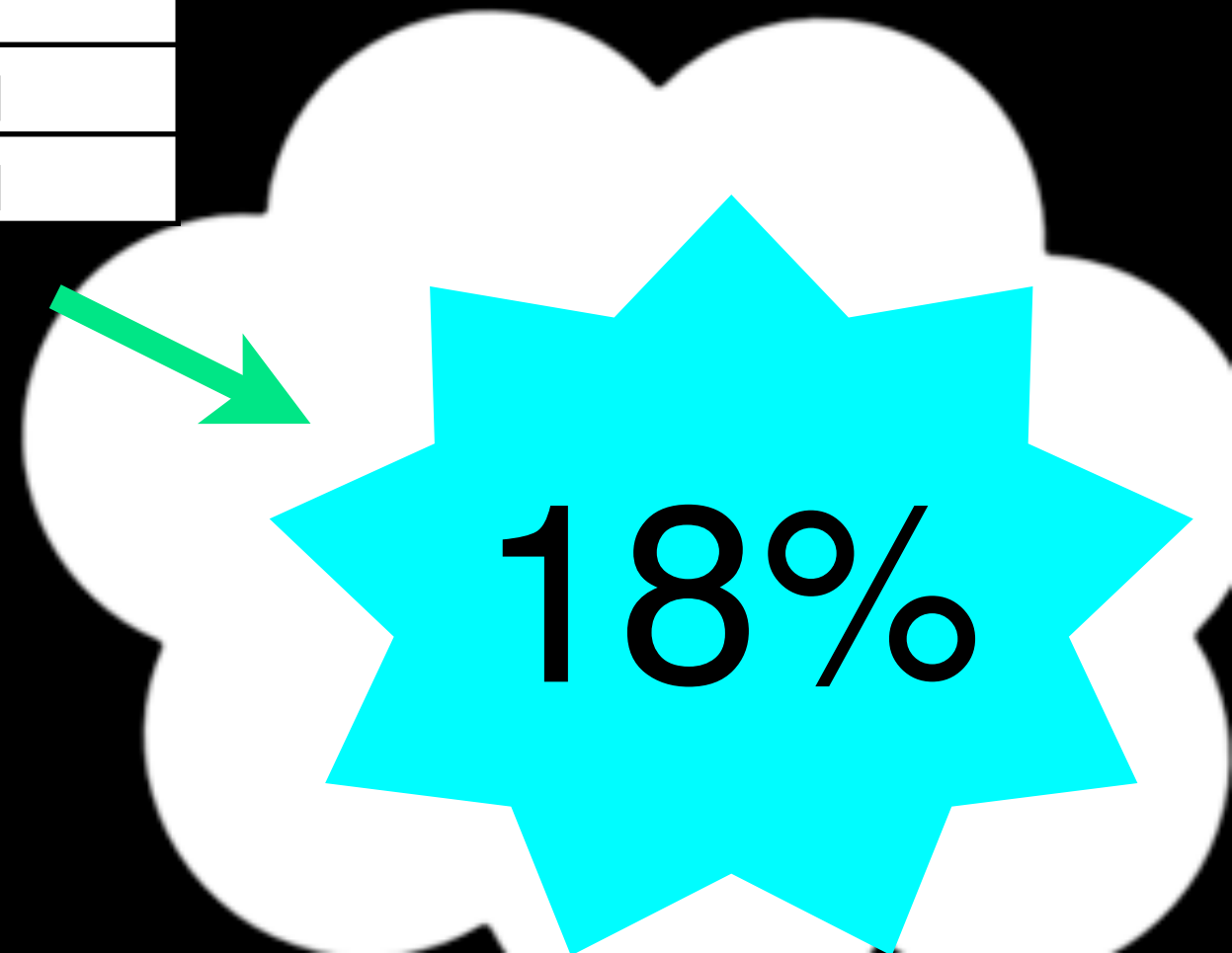
name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N



what to promise?

think of output as randomized

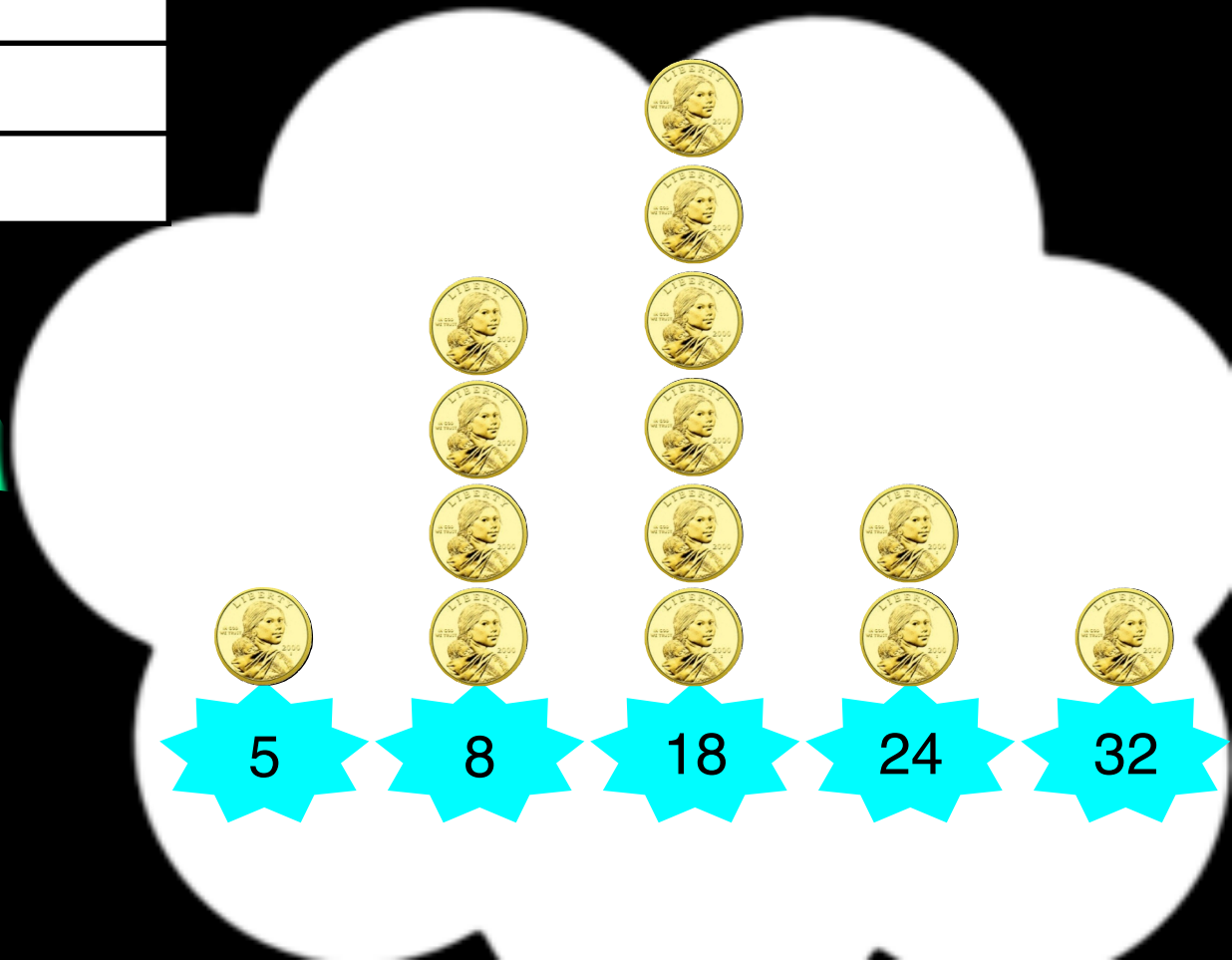
name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N



what to promise?

think of output as randomized

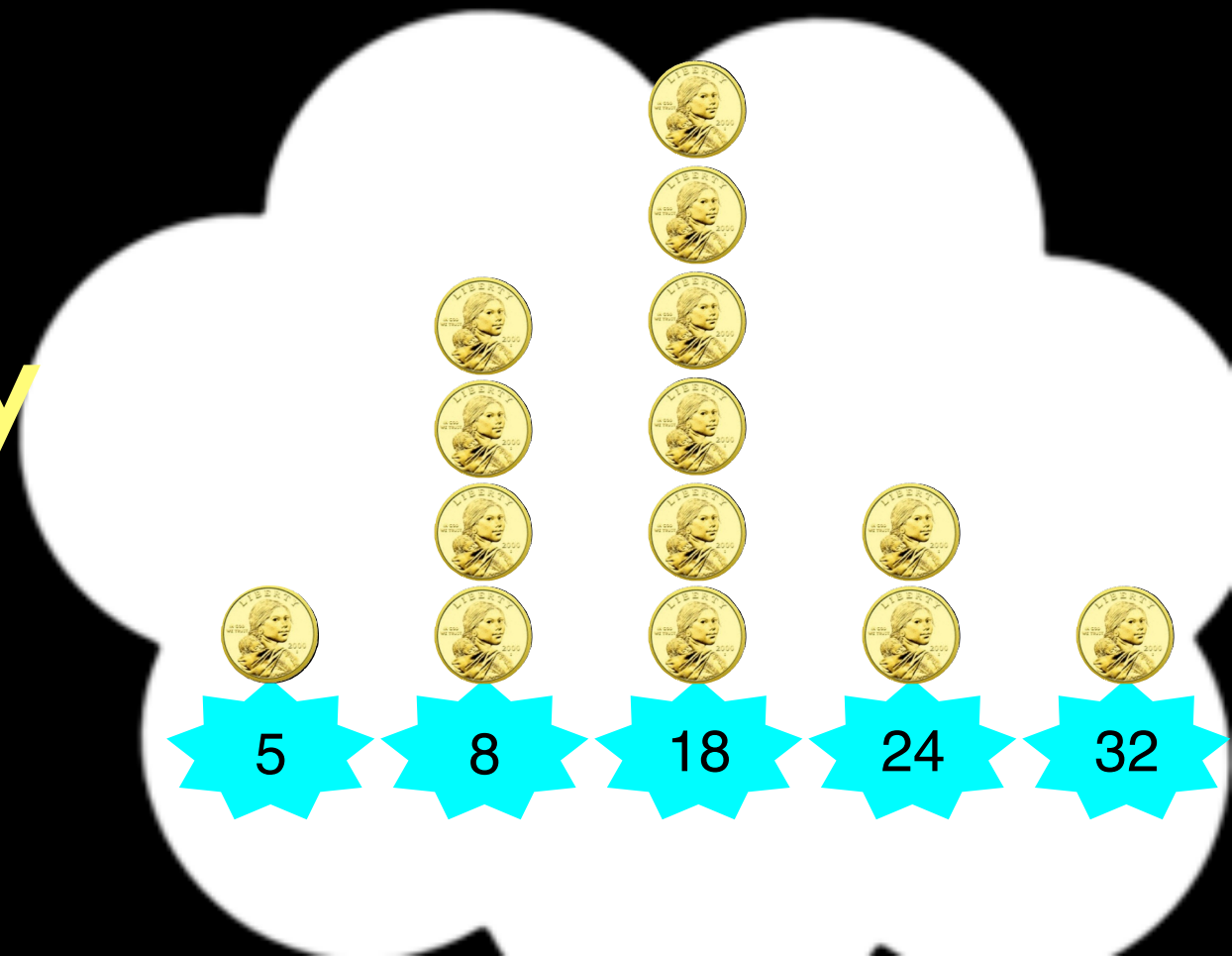
name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N



what to promise?

think of output as randomized

promise: if you leave
the database, no
outcome will
change probability
by very much



statistical database model

- X set of possible entries/rows

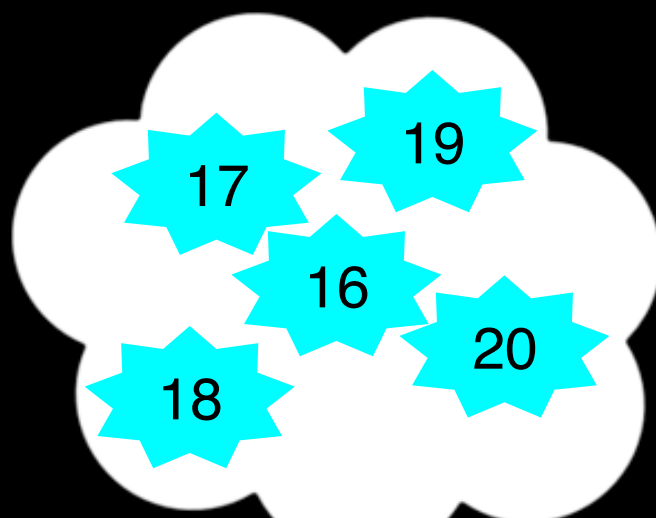
one row per person

- database D a set of rows; $D \in X^*$

name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N

analyst objective

- wishes to compute on $D \in X^*$
 - fit a model, compute a statistic, share “sanitized” data
- preserve privacy of individuals
- design randomized algorithm A mapping D into outcome space, that masks small changes in D



name	DOB	sex	weight	smoke	lung_cancer
John Doe	12/1/51	M	185	Y	N
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/2/45	F	140	N	N
Jane Smith	3/2/45	F	140	N	N
Ethan Jones	4/24/55	F	180	Y	Y
Jennifer Kim	3/1/72	F	130	N	N
Rachel	10/2/43	F	140	N	N



neighboring databases

what's a small change?

require nearly identical behavior on **neighboring databases** differing by the addition or removal of a single row:

$$\|D - D'\|_1 \leq 1$$

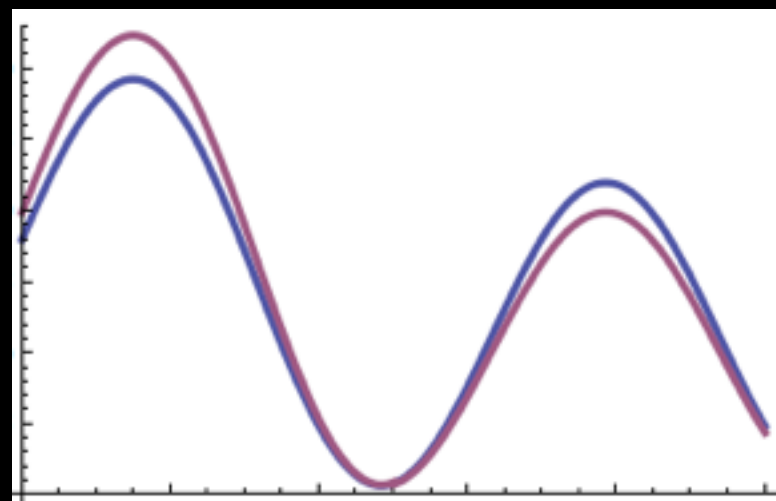
for $D, D' \in X^*$

differential privacy

[DinurNissim03, DworkNissimMcSherrySmith06, Dwork06]

A randomized alg $A : X^* \rightarrow O$ is ϵ -**differentially private** if for every pair of neighboring data sets D, D' and for every event $S \subseteq O$:

$$\Pr[A(D) \in S] \leq e^\epsilon \Pr[A(D') \in S]$$

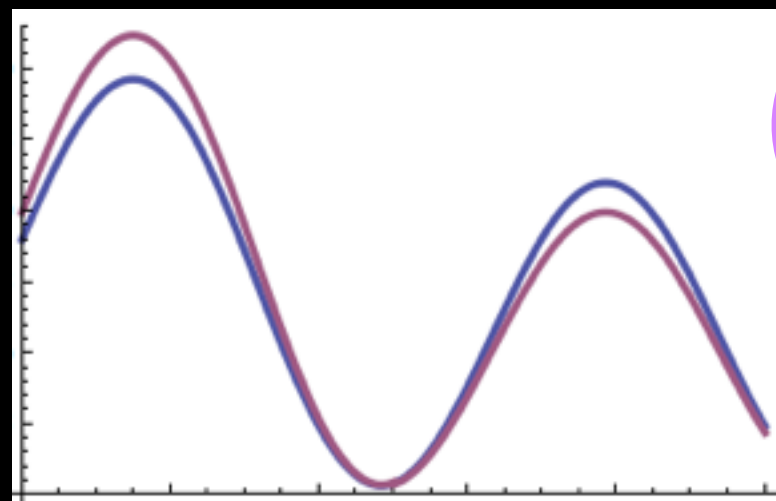


differential privacy

[DinurNissim03, DworkNissimMcSherrySmith06, Dwork06]

A randomized alg $A : X^* \rightarrow O$ is ϵ -**differentially private** if for every pair of neighboring data sets D, D' and for every event $S \subseteq O$:

$$\Pr[A(D) \in S] \leq e^\epsilon \Pr[A(D') \in S]$$

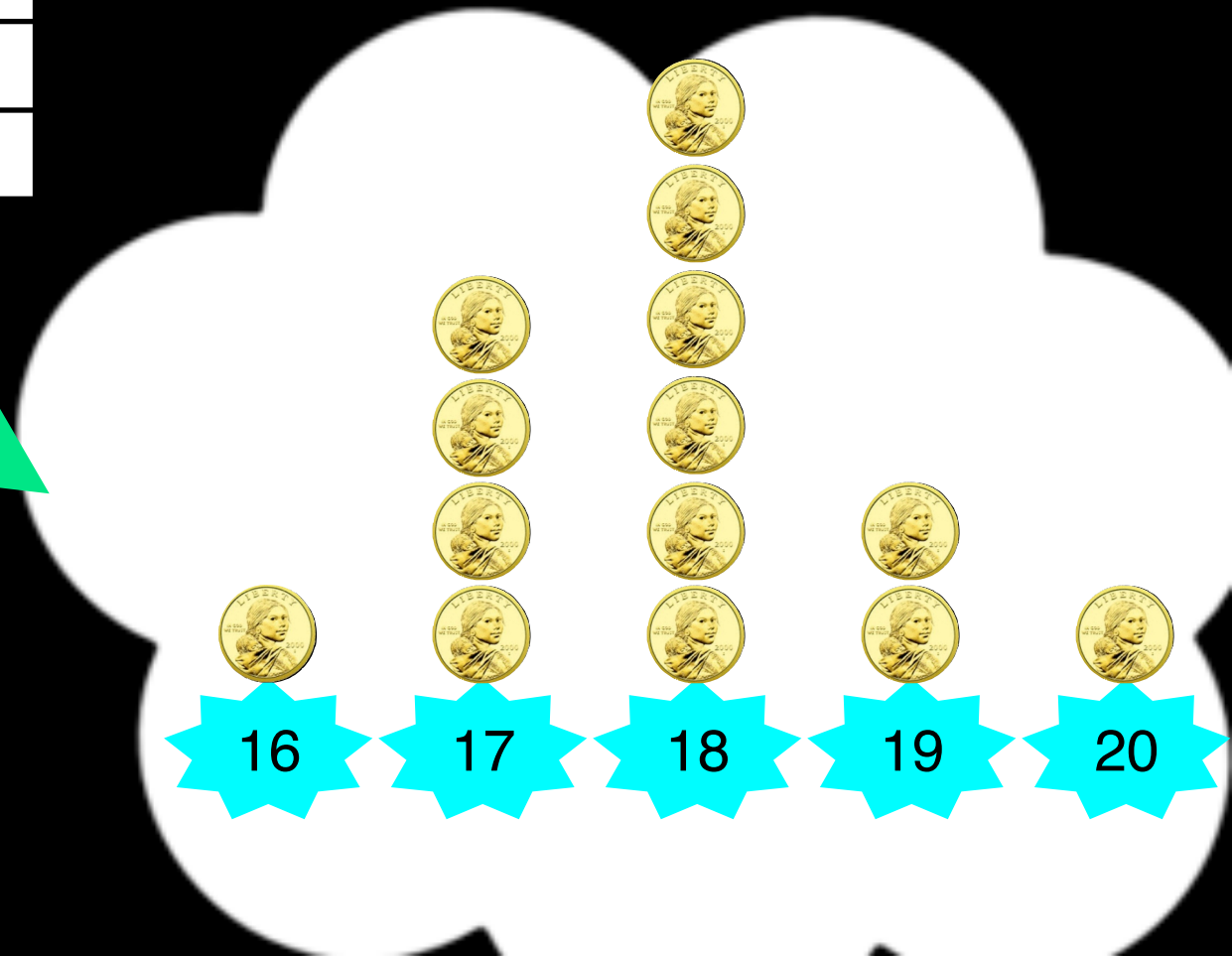


$$e^\epsilon \sim (1 + \epsilon)$$

differential privacy

$$\Pr[A(D) \in \mathcal{S}] \leq e^\epsilon \Pr[A(D') \in \mathcal{S}]$$

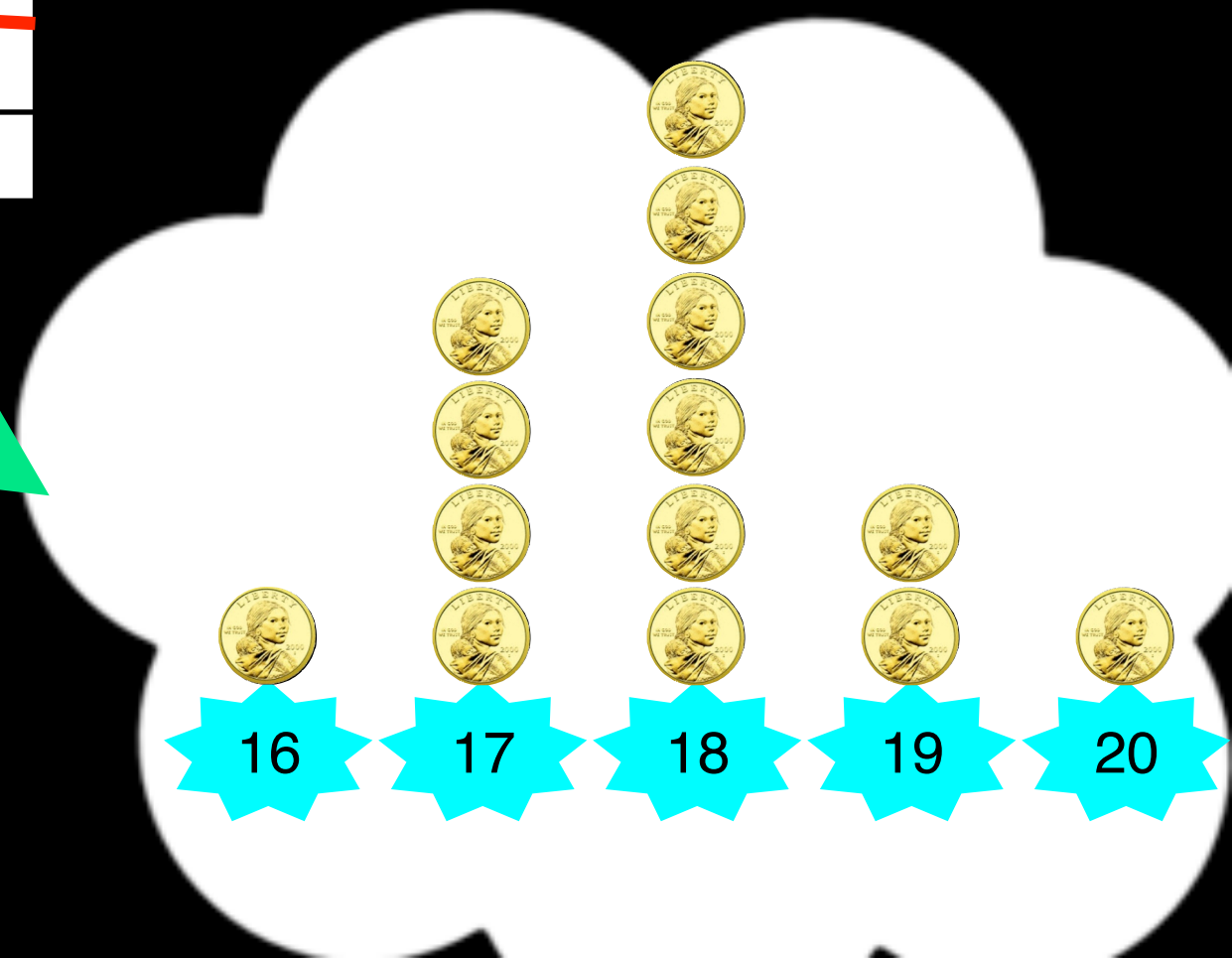
name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N



differential privacy

$$\Pr[A(D) \in \mathcal{S}] \leq e^\epsilon \Pr[A(D') \in \mathcal{S}]$$

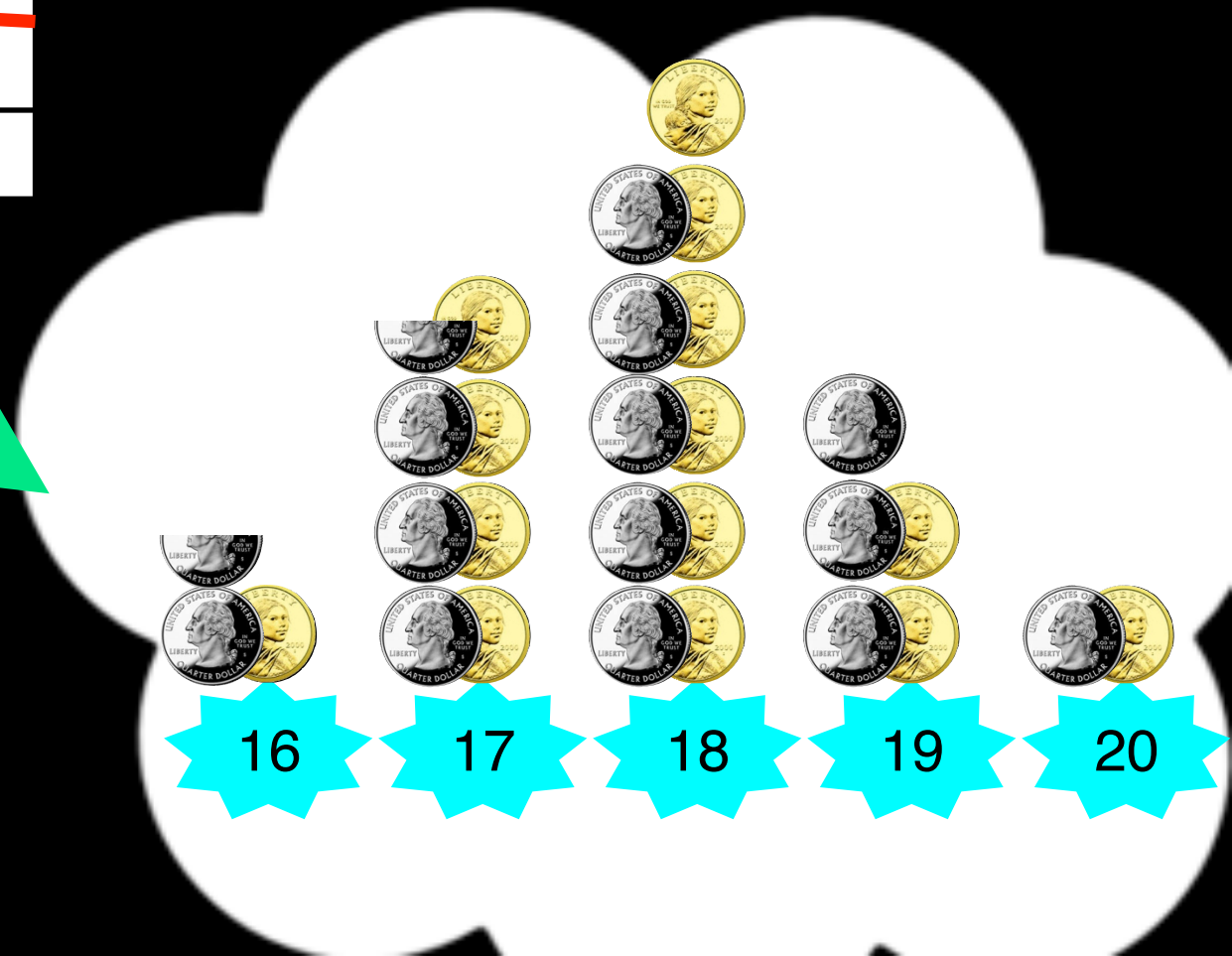
name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N



differential privacy

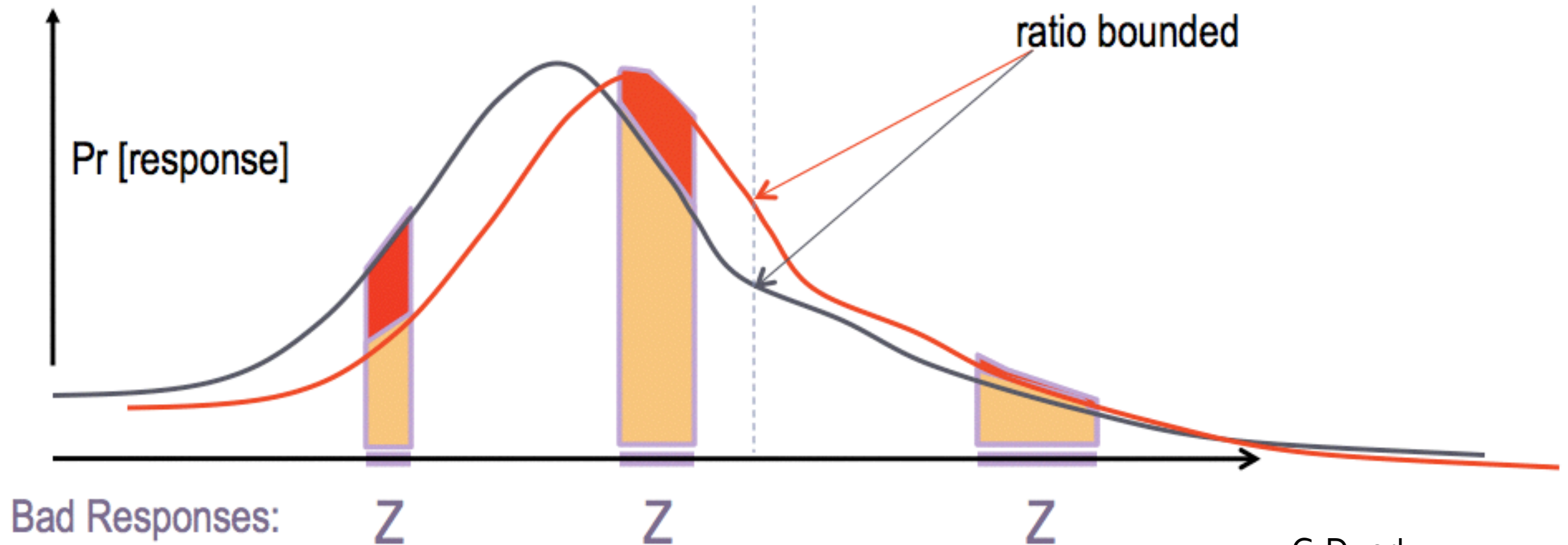
$$\Pr[A(D) \in \mathcal{S}] \leq e^\epsilon \Pr[A(D') \in \mathcal{S}]$$

name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N



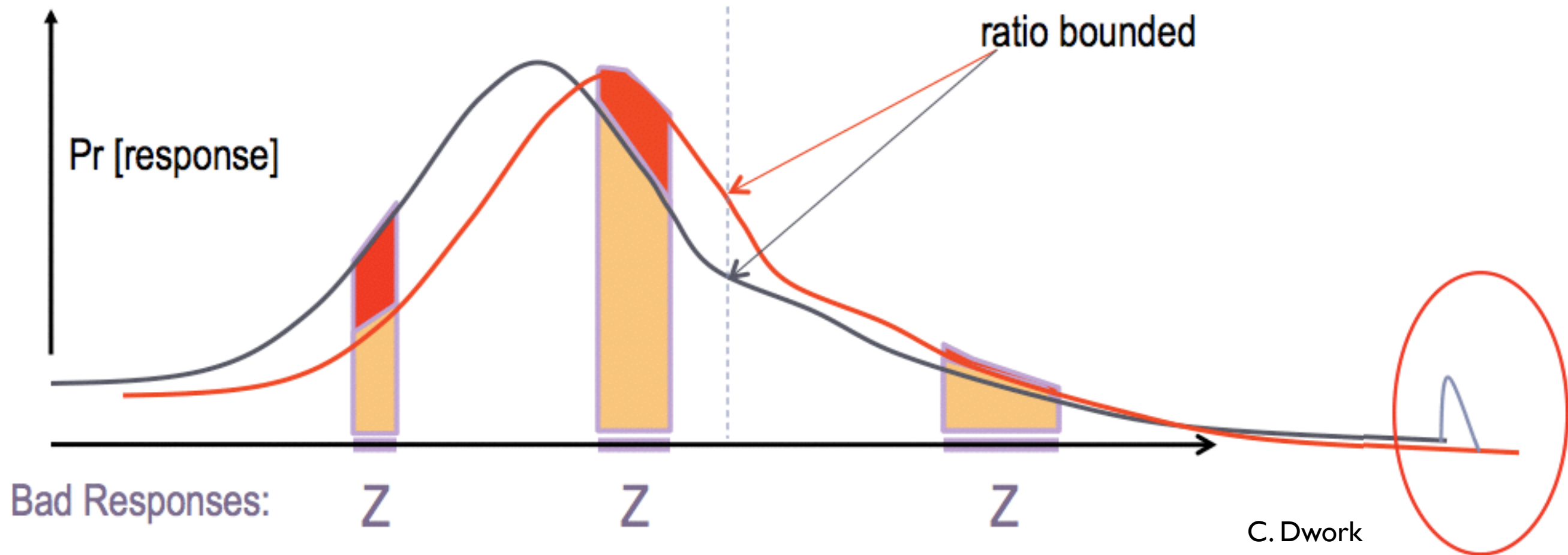
differential privacy

$$\Pr[A(D) \in \mathcal{S}] \leq e^\epsilon \Pr[A(D') \in \mathcal{S}]$$



(ϵ, δ) -differential privacy

$$\Pr[A(D) \in \mathcal{S}] \leq e^\epsilon \Pr[A(D') \in \mathcal{S}] + \delta$$



post-processing

Any subsequent computations on the results of a DP computation maintain the privacy guarantee.

composition

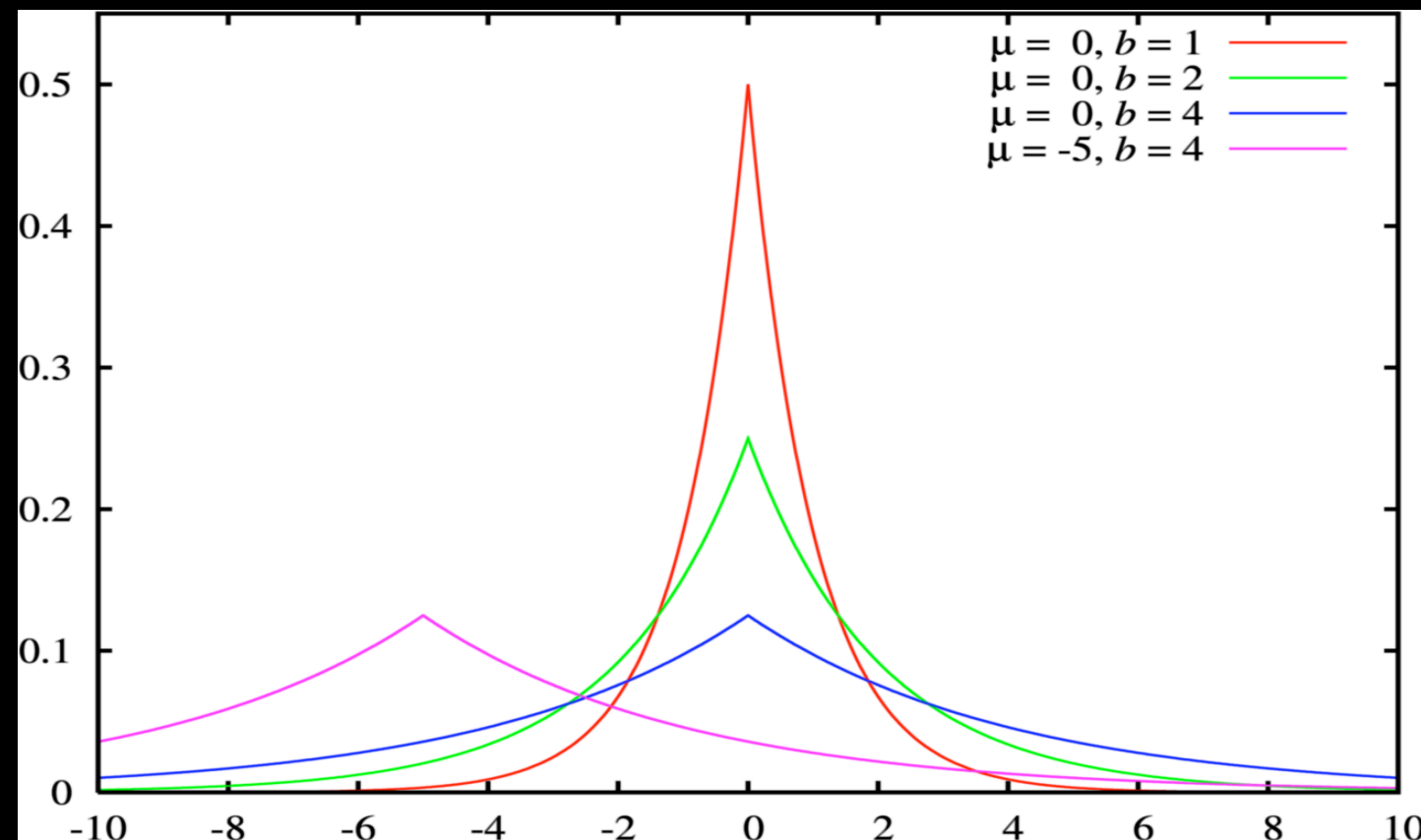


[DworkKenthapadiMcSherryMironovNaor06,DworkLei09]

- If run multiple DP algorithms, the ϵ s and δ s add up
- Allows simple privacy analysis of complex algorithms
- Holds even if subsequent computations chosen as function of previous results
- More subtle analysis gives even better guarantees

Laplace mechanism

For numeric computations, direct noise addition of a particular form and magnitude preserves differential privacy



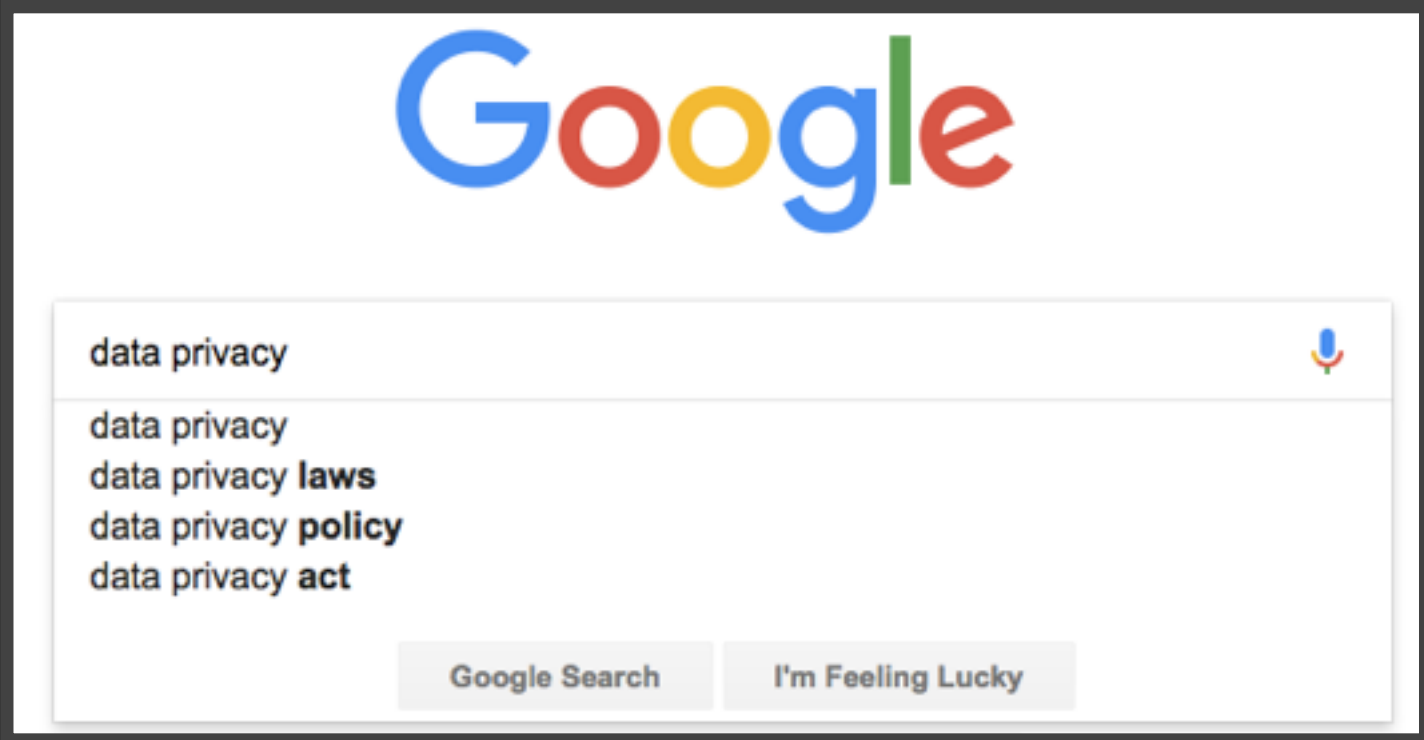
today

- Formalizing privacy: differential privacy

 (private) Empirical risk minimization (ERM)

- Accuracy-First Private ERM

NETFLIX



Google

data privacy

data privacy
data privacy laws
data privacy policy
data privacy act

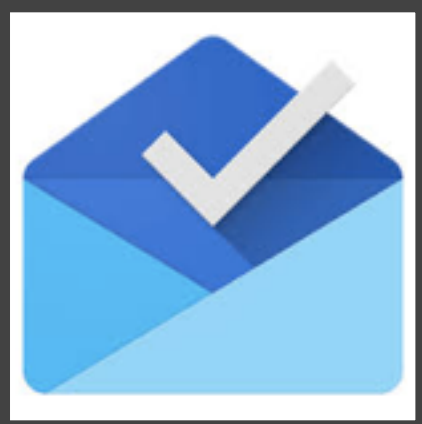
Google Search I'm Feeling Lucky



facebook



amazon



what do companies do with our private data?

- learn to predict what you'll type
- learn to predict what ads you'll click on
- learn to predict what you'll buy
- ...

what do companies do with our private data?

- learn to predict what you'll type
- learn to predict what ads you'll click on
- learn to predict what you'll buy
- ...

Extrapolate from lots of data a rule that maps individual behavior into a specific outcome

Empirical Risk Minimization (ERM)

- Setting: there are true labels for points, drawn from an underlying distribution
- Learner has access to a training set
- Pick a hypothesis (function mapping points to labels) from among a given set, to minimize mistakes on the training set

prior work: private ERM

- output and objective perturbation [Chaudhuri Monteleoni 2008, Chaudhuri Monteleoni Sarwate 2011, Kifer Smith Thakurta 2012, Rubinfeld Bartlett Huang Taft 2009]
- covariance perturbation [Smith Upadhyay Thakurta 2017]
- exponential mechanism [Bassily Smith Thakurta 2014, McSherry Talwar 2007]
- stochastic gradient descent [Bassily Smith Thakurta 2014, Duchi Jordan Wainwright 2013, Jain Kothari Thakurta 2012, Song Chaudhuri Sarwate 2013, Williams McSherry 2010]

Accuracy-First Private ERM: flip the theorem

- Flip existing “utility theorem” for existing private ERM algorithm to solve for the smallest epsilon (and other parameters) consistent with accuracy requirement
- Run existing algorithm with resulting epsilon
- Only prior theoretically sound approach

Accuracy-First Private ERM: flip the theorem

- Problem: utility theorems are worst-case (algorithms are often providing much better accuracy/privacy than they promise)
- Sloppy constants
- Specific dataset may allow for much better privacy/utility tradeoff

Accuracy-First Private ERM: search for the best possible epsilon

- Idea: try values of epsilon until find one that satisfies accuracy constraint
- Problems:
 - Search is data-dependent, so pays for every attempt
 - Not a priori clear how to bound privacy loss with usual notion
 - search could run a long time (forever?)
 - selected privacy parameter is function of the data

today

- Formalizing privacy: differential privacy
- (private) Empirical risk minimization (ERM)

 Accuracy-First Private ERM

Accuracy-First Private ERM: this paper

- A principled version of “epsilon search”
- Give a meta-method for this search applicable to several classes of private learning algorithms

High-level approach

- Initially, compute very private hypothesis
- Degrade the privacy guarantee by doubling until the accuracy guarantee is met

High-level approach

- Initially, compute very private hypothesis
- Degrade the privacy guarantee by doubling until the accuracy guarantee is met
 - To not pay extra, use correlated noise across rounds, so can “subtract” noise from previous round to get next

High-level approach

- Initially, compute very private hypothesis
- Degrade the privacy guarantee by doubling until the accuracy guarantee is met
 - To not pay extra, use correlated noise across rounds, so can “subtract” noise from previous round to get next
 - New algorithm to minimize costs of checking whether accuracy guarantee is met

High-level approach

- Initially, compute very private hypothesis
- Degrade the privacy guarantee by doubling until the accuracy guarantee is met
 - To not pay extra, use correlated noise across rounds, so can “subtract” noise from previous round to get next
 - New algorithm to minimize costs of checking whether accuracy guarantee is met
 - Pay only privacy cost of final hypothesis (earlier attempts are free) + checking

Ex post privacy

Ex post privacy

Doesn't satisfy a priori ϵ -DP for any fixed ϵ , but if terminates after k rounds, seems to satisfy bounded "ex post" privacy loss

Ex post privacy

Doesn't satisfy a priori ϵ -DP for any fixed ϵ , but if terminates after k rounds, seems to satisfy bounded "ex post" privacy loss

Ex post privacy

Doesn't satisfy a priori ϵ -DP for any fixed ϵ , but if terminates after k rounds, seems to satisfy bounded "ex post" privacy loss

c.f. privacy odometers [Rogers Roth Ullman Vadhan 2016]

Ex post privacy

Definition. The **ex-post privacy loss** of a randomized algorithm $A : X^* \rightarrow O$ on outcome o is the maximum over pairs of neighboring data sets D, D' of

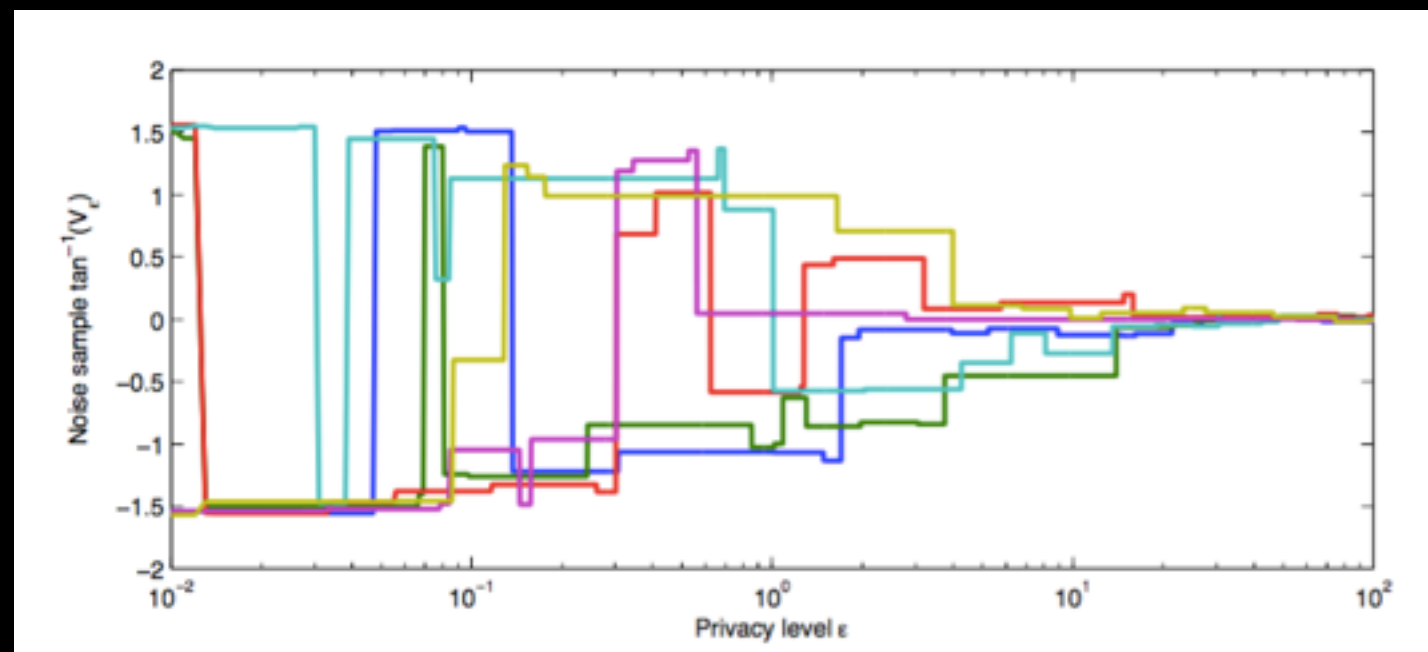
$$\log(\Pr[A(D) = o] / \Pr[A(D') = o])$$

Definition. Consider function $E:O \rightarrow (\mathbf{R}_{\geq 0} \cup \{\infty\})$ on the outcome of algorithm $A : X^* \rightarrow O$. Given outcome $o = A(D)$ we say A satisfies $E(o)$ **ex-post differential privacy** if for all $o \in O$, $\text{Loss}(o) \leq E(o)$.

Correlated noise: key idea

[Koufogiannis Han Pappas 2017]

- Algorithm: continuous random walk starting at private data v , s.t. marginal distribution at each point in time is Laplace centered at v , with variance increasing over time.
- More private points can be derived from less private ones
- Reverse process



Checking algorithm: key ideas

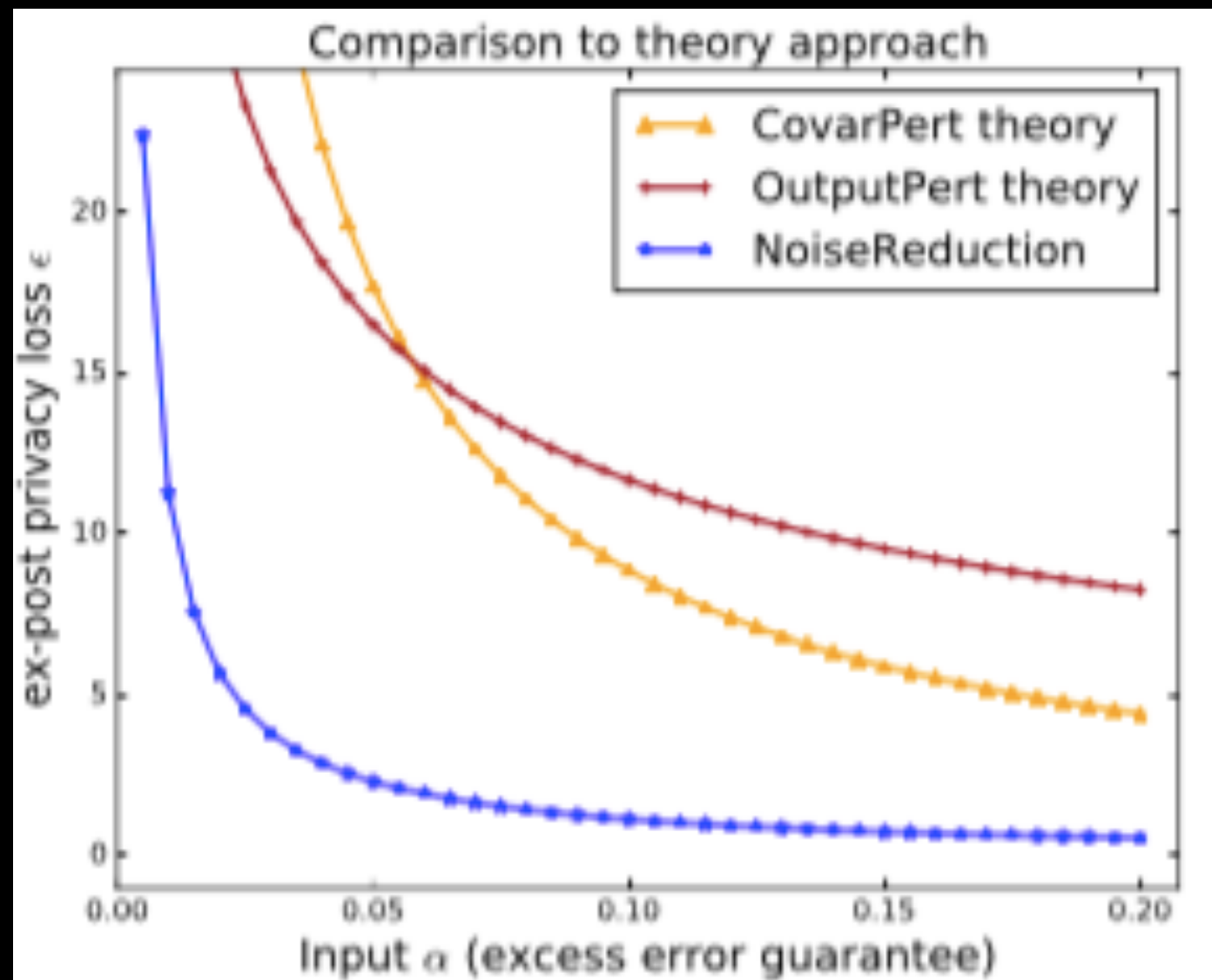
- Existing algorithm AboveThreshold takes dataset and sequence of (adaptively chosen) queries, and privately outputs first query to exceed a given threshold. Pays much less than the composition of the queries.
- For us, “queries” depend on the data, so naively would need to publish (and pay for) them all

Applicable algorithms

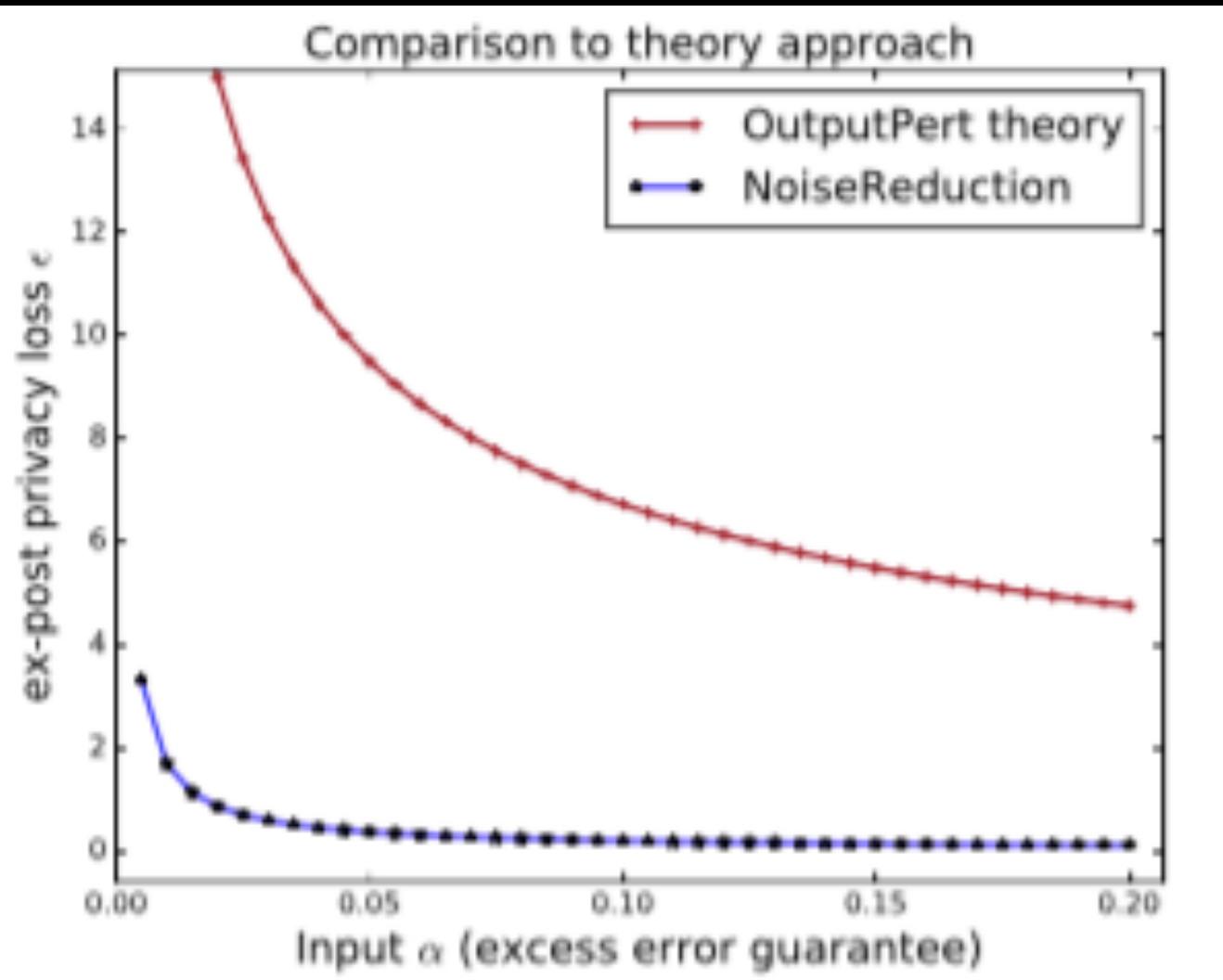
- Our approach applies to any ERM technique that can be described as a post-processing of a Laplace mechanism, e.g.,
- output perturbation (add Laplace noise to result)
- covariance perturbation (perturb covariance matrix of data, then optimize using noisy data)

Empirical results: summary

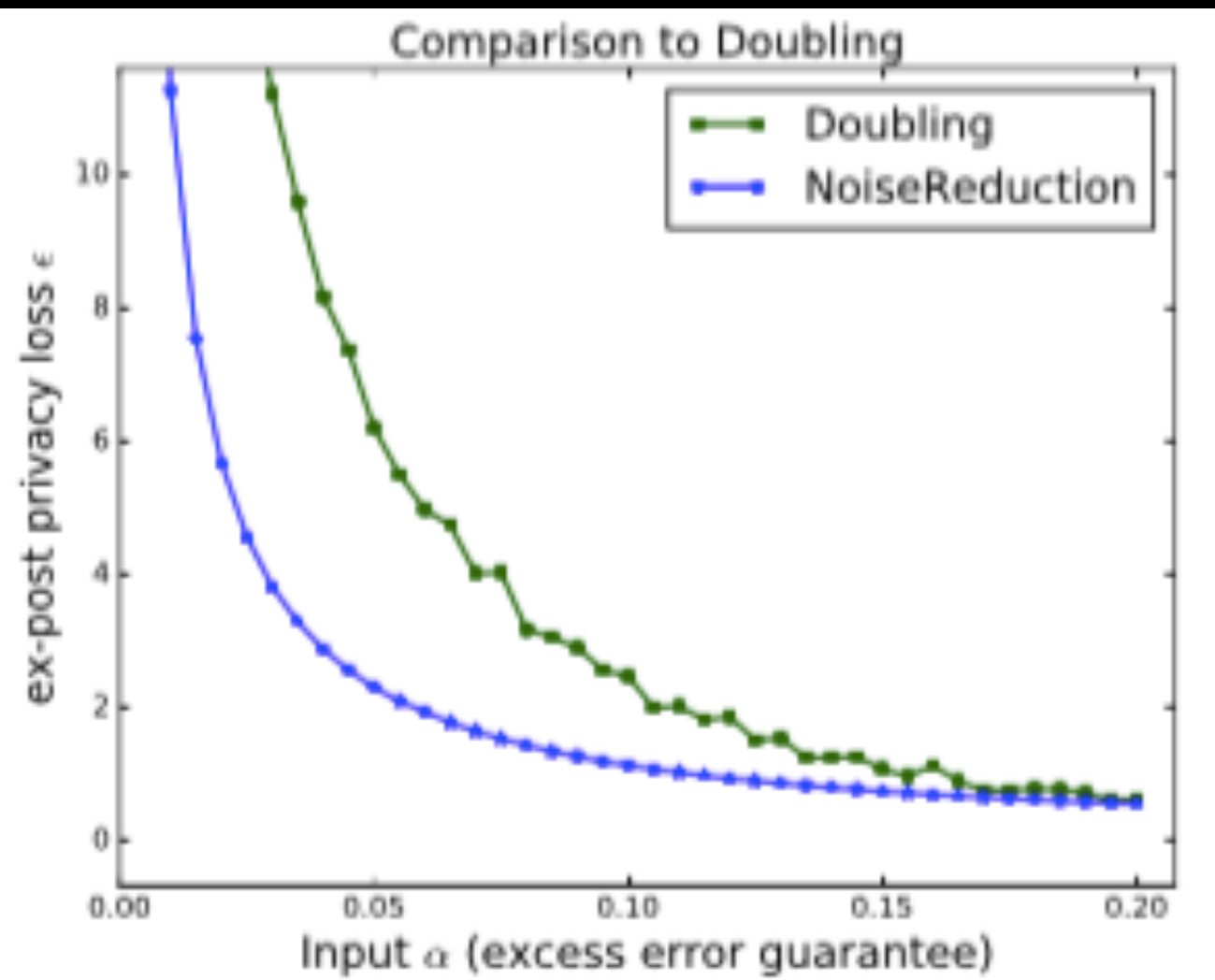
- Our approach massively outperforms inverting the theory curve
- Also improves on a baseline “epsilon-doubling” approach



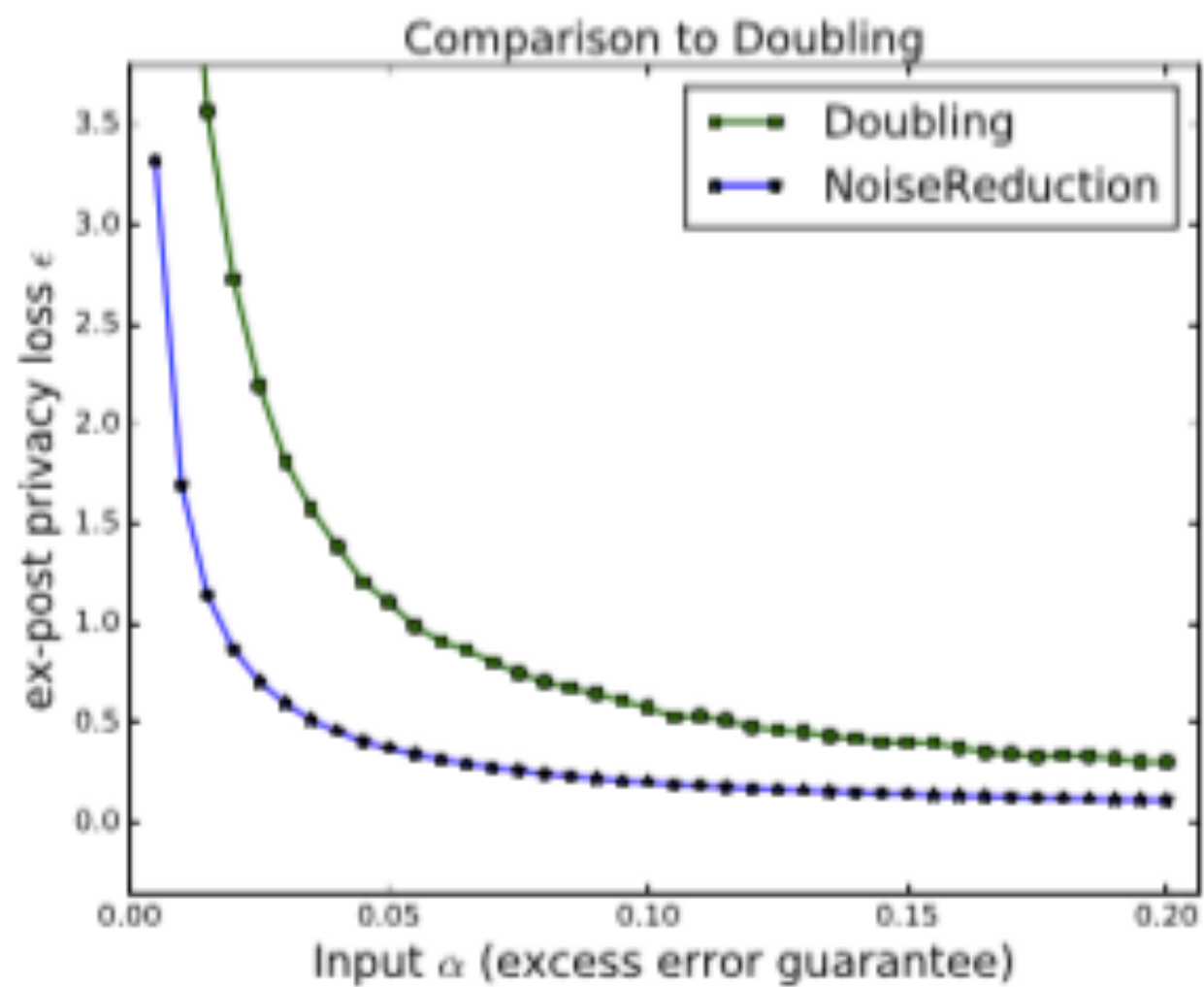
**(a) Linear (ridge) regression,
vs theory approach.**



(b) Regularized logistic regression, vs theory approach.



**(c) Linear (ridge) regression,
vs DOUBLINGMETHOD.**



**(d) Regularized logistic regression,
vs DOUBLINGMETHOD.**

Future directions

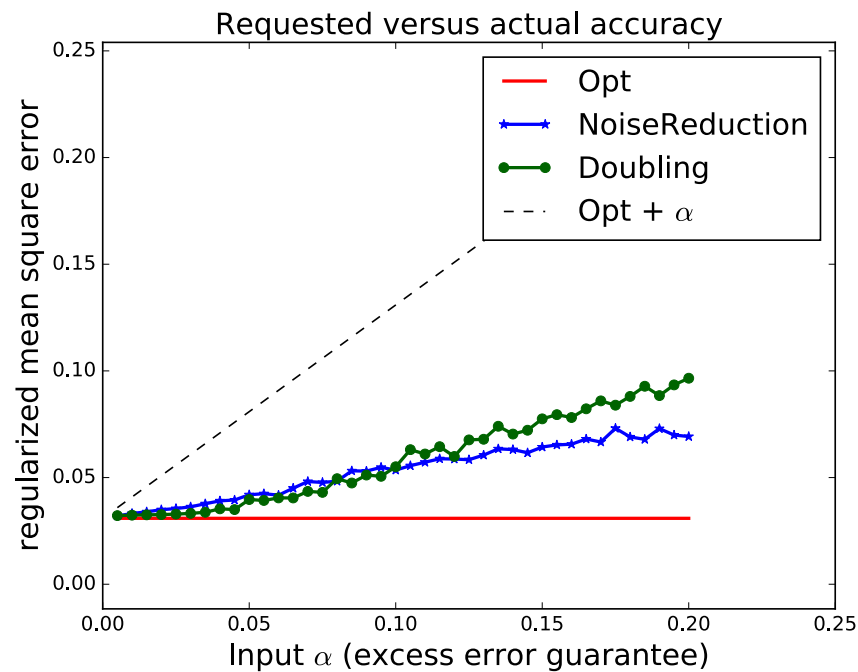
- Empirically, privacy loss from “testing” hypotheses significantly larger than from “generating” them. Loose analysis? (e.g., currently using a theoretical bound on the maximum norm of any hypothesis to compute the sensitivity of queries)
- InteractiveAboveThreshold for (ϵ, δ) -differential privacy

Theorem 3.2. *The instantiation of $\text{CovNR}(D, \{\varepsilon_1, \dots, \varepsilon_T\}, \alpha, \gamma)$ outputs a hypothesis θ_p that with probability $1 - \gamma$ satisfies $L(\theta_p) - L(\theta^*) \leq \alpha$. Moreover, it is \mathcal{E} -ex-post differentially private, where the privacy loss function $\mathcal{E}: (([T] \cup \{\perp\}) \times \mathbb{R}^p) \rightarrow (\mathbb{R}_{\geq 0} \cup \{\infty\})$ is defined as $\mathcal{E}((k, \cdot)) = \varepsilon_0 + \varepsilon_k$ for any $k \neq \perp$, $\mathcal{E}((\perp, \cdot)) = \infty$, and $\varepsilon_0 = \frac{16(\sqrt{1/\lambda}+1)^2 \log(2T/\gamma)}{n\alpha}$ is the privacy loss incurred by IAT.*

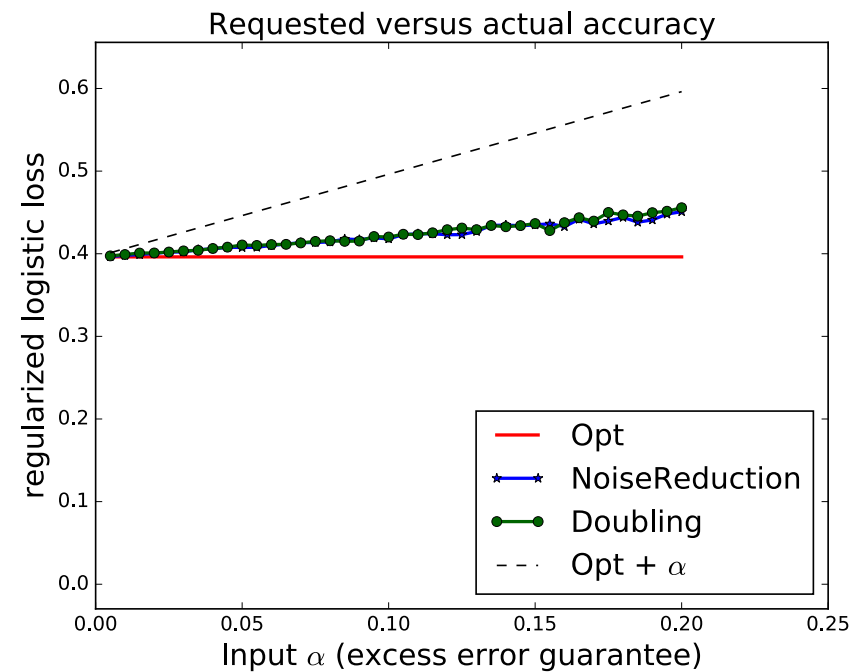
Theorem 3.4. *The instantiation of $\text{OUTPUTNR}(D, \varepsilon_0, \{\varepsilon_1, \dots, \varepsilon_T\}, \alpha, \gamma)$ is \mathcal{E} -ex-post differentially private and outputs a hypothesis θ_p that with probability $1 - \gamma$ satisfies $L(\theta_p) - L(\theta^*) \leq \alpha$, where the privacy loss function $\mathcal{E}: (([T] \cup \{\perp\}) \times \mathbb{R}^p) \rightarrow (\mathbb{R}_{\geq 0} \cup \{\infty\})$ is defined as $\mathcal{E}((k, \cdot)) = \varepsilon_0 + \varepsilon_k$ for any $k \neq \perp$, $\mathcal{E}((\perp, \cdot)) = \infty$, and $\varepsilon_0 \leq \frac{32 \log(2T/\gamma) \sqrt{2 \log 2/\lambda}}{n\alpha}$ is the privacy loss incurred by IAT.*

Theorem A.1. *For any sequence of 1-sensitive queries f_1, \dots, f_T such $\text{InteractiveAboveThreshold}$ is (α, β) -accurate for*

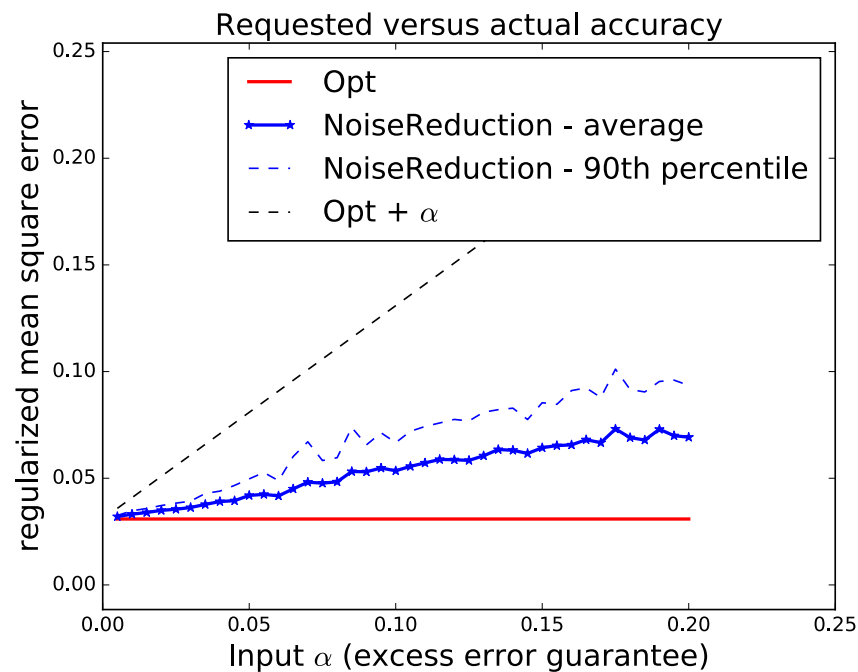
$$\alpha = \frac{8\Delta(\log(T) + \log(2/\gamma))}{\varepsilon}.$$



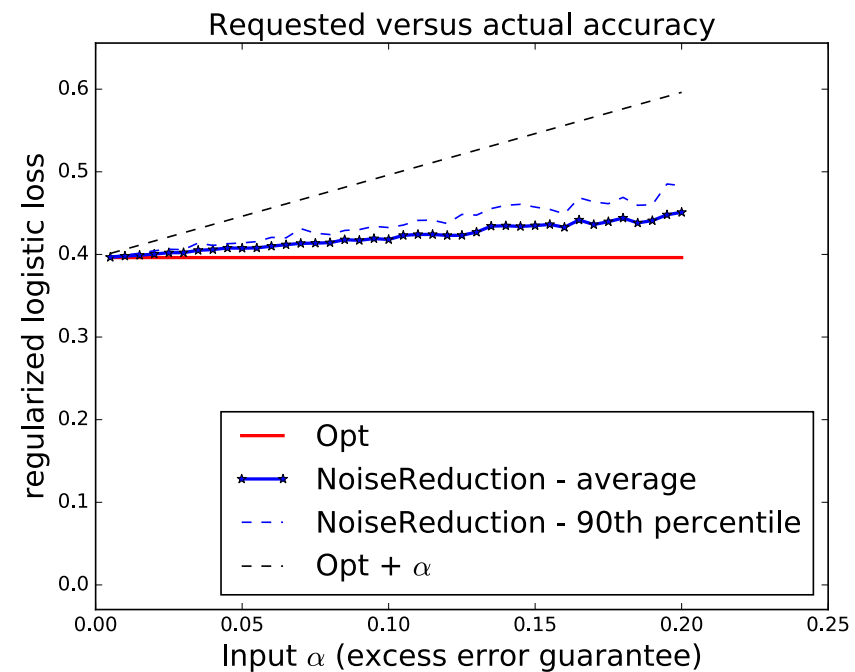
(a) **Linear (ridge) regression.**



(b) **Regularized logistic regression.**

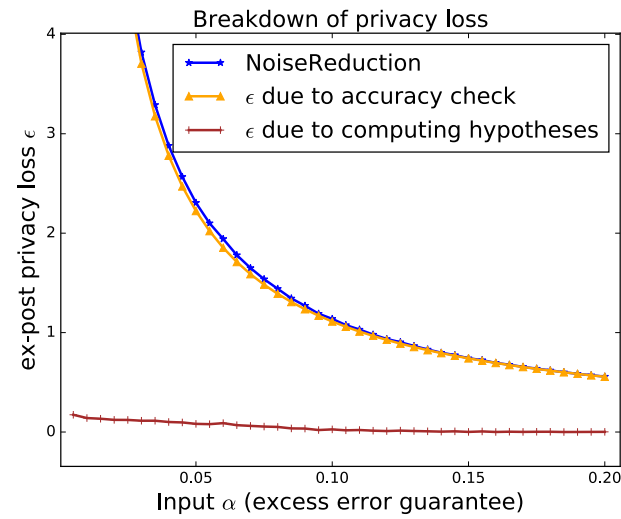


(c) **Linear (ridge) regression.**

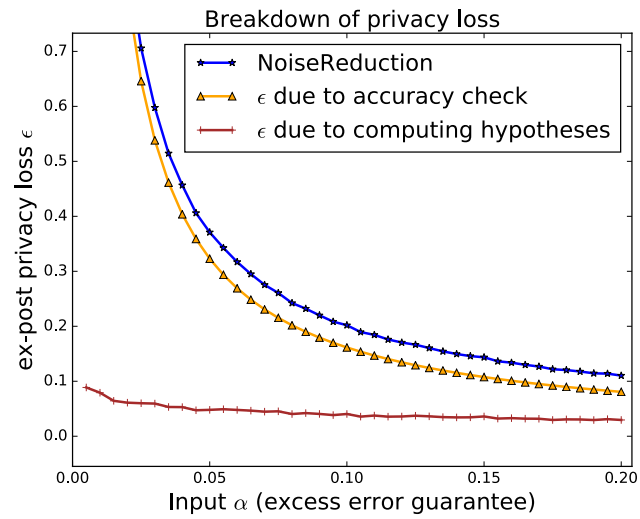


(d) **Regularized logistic regression.**

Figure 2: **Empirical accuracies.** The dashed line shows the requested accuracy level, while the others plot the actual accuracy achieved. Due most likely due to a pessimistic analysis and the need to set a small testing threshold, accuracies are significantly better than requested for both methods.

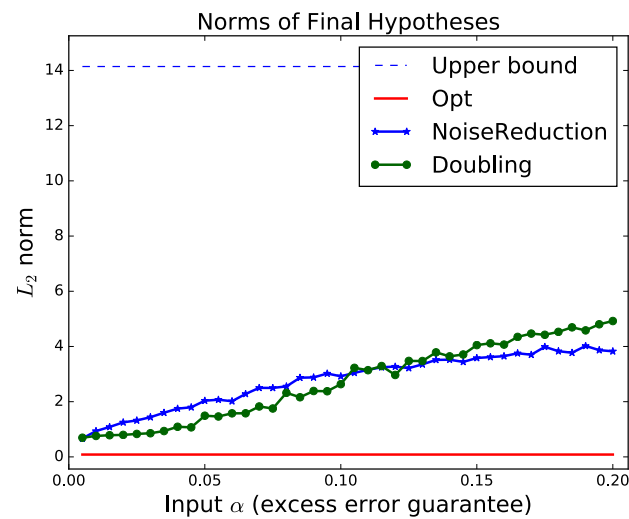


(a) **Linear (ridge) regression.**

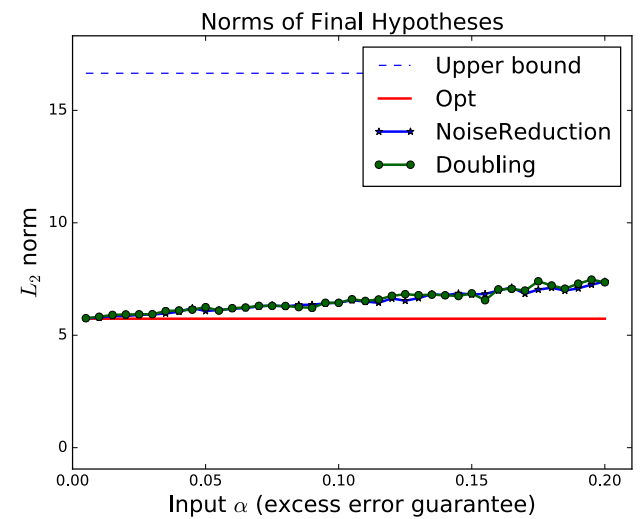


(b) **Regularized logistic regression.**

Figure 3: **Privacy breakdowns.** Shows the amount of empirical privacy loss due to computing the hypotheses themselves and the losses due to testing their accuracies.



(a) **Linear (ridge) regression.**



(b) **Regularized logistic regression.**

Figure 4: **L_2 norms of final hypotheses.** Shows the average L_2 norm of the output $\hat{\theta}$ for each method, versus the theoretical maximum of $1/\sqrt{\lambda}$ in the case of ridge regression and $\sqrt{2\log(2)}/\lambda$ in the case of regularized logistic regression.