Data 102 Spring 2022 Lecture 25

Privacy in Machine Learning

These slides are linked to from the course website, data102.org/sp22

Weekly Outline

- Last lecture: Multi-armed bandits.
  - Application of concentration inequalities to decision making
  - UCB algorithm, Explore then Commit

- This lecture: Privacy and learning.

Announcements

HW 6 is posted, due on April 30

Lab 10 due tomorrow

Discussions as usual

Midterm 2 is graded
- Mean: about 21.5
- Median: about 21

Machine Learning and Privacy

Machine Learning seems to be about general statistics of the distribution, not about any one individual.

If we take two large enough sample sets S and S' from the same distribution, then effectively we should learn the same thing from S or S'.

Machine learning is much more about the distribution D or the sample S as a whole, not so much about a specific $x \in S$. So, we should be able to "preserve the privacy of individuals".

Let's formalize what "privacy" means here.

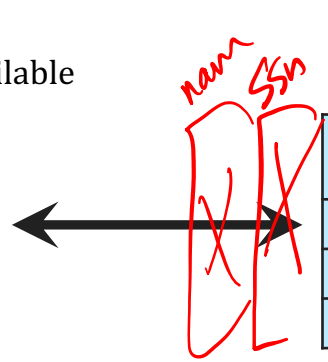
Anonymized Data Sets

Latanya Sweeney

The trouble with “anonymized data” that other easily available data can “re-identify” the data set.

Non-anonymized Publicly available data: Voter Registration

Name	ZIP	DoB	Gender



Anonymized Sensitive Data

Gender	DoB	ZIP	**Entire Medical Record**

At the time GIC released the data, William Weld, then Governor of Massachusetts, assured the public that GIC had protected patient privacy by deleting identifiers. In response, then-graduate student Sweeney started hunting for the Governor’s hospital records in the GIC data. She knew that Governor Weld resided in Cambridge, Massachusetts, a city of 54,000 residents and seven ZIP codes. For twenty dollars, she purchased the complete voter rolls from the city of Cambridge, a database containing, among other things, the name, address, ZIP code, birth date, and sex of every voter. By combining this data with the GIC records, Sweeney found Governor Weld with ease. Only six people in Cambridge shared his birth date, only three of them men, and of them, only he lived in his ZIP code. In a theatrical flourish, Dr. Sweeney sent the Governor’s health records (which included diagnoses and prescriptions) to his office.

Anonymized Data Sets

Latanya Sweeney

The trouble with “anonymized data” that other easily available data can “re-identify” the data set.

Non-anonymized Publicly available data: Voter Registration

Name	ZIP	DoB	Gender

Linkage Attack

Anonymized Sensitive Data

Gender	DoB	ZIP	**Entire Medical Record**

Privacy is not the same as anonymizing the data

k-anonymity

A data release mechanism satisfies the k-anonymity property, if the information for each person that was contained in the release cannot be distinguished from at least $k - 1$ individuals whose information also appeared in the release.

Divide data attributes into “quasi-identifiers” and “sensitive attributes”.

Modify attributes so that there are $\geq k$ rows for each combination of quasi-identifiers that is present.

Can be broken: Repeating each data point k times “meets” the definition. Background knowledge can be harmful.

Example of successful uses: Compromised Credential Checking protocol, to anonymously verify if a password is leaked.

Population level statistics

Only answer queries that are about population as a whole:

November 1

ID	**Midterm 1 Grade**
xx123	Hidden
yy123	
aa000	
zz123	

What's the class average? 72.75

You know your friend aa000 dropped out

November 2

ID	**Midterm 1 Grade**
xx123	Hidden
yy123	
zz123	

What's the class average? 82

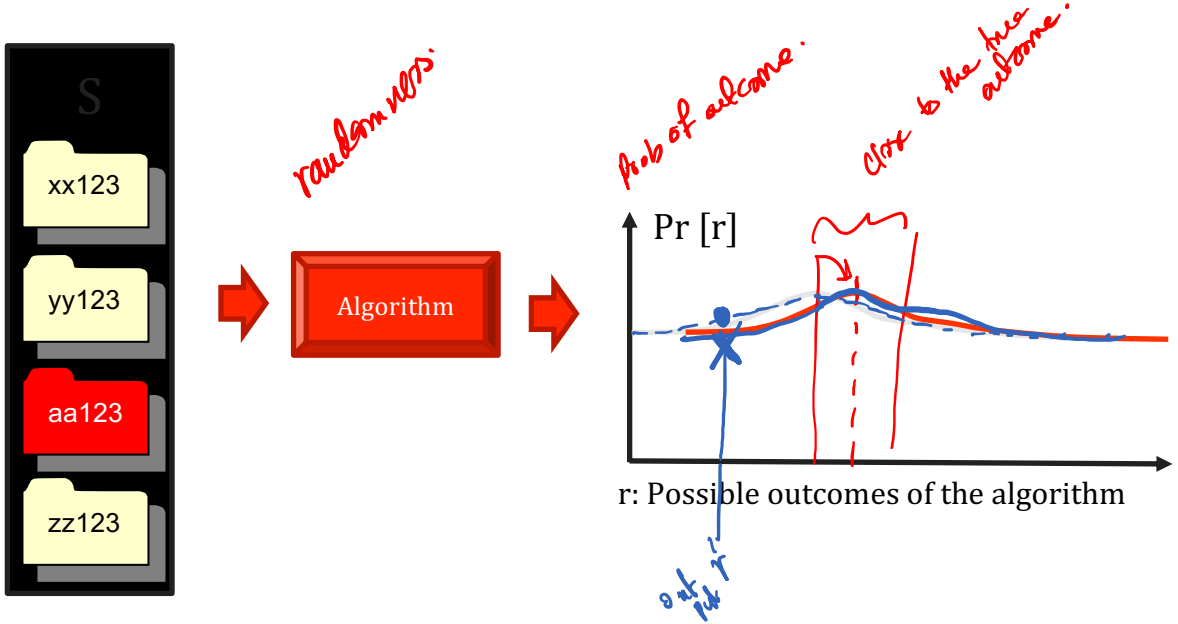
You can figure out aa000's prelim grade $4 \times 72.75 - 3 \times 82 = 45$.

Answering too many queries very accurately reduces privacy.

Privacy while Learning

Privacy is about protecting against inferences using your data.

"An analysis of a dataset S is private if the data analyst knows almost no more about Alice after the analysis than he would have known had he conducted the same analysis on an identical database with Alice's data removed."



Differential Privacy

neighboring data sets.

$S = \{ \cdots (\text{nika}, x) \cdots \}$

$S' = \{ \cdots (\text{nika}, x) \cdots \}$


Cynthia Dwork


Frank McSherry


Kobbi Nissim


Adam Smith

S: The data set, where each person's data is one point $x \in S$.

Differential Privacy

An algorithm $\mathcal{L}$ is α-differentially private if for all pairs of datasets S, S' differing in one user's data, and for all outputs r: $\alpha = 1$, $\alpha = 0.01$

$$\Pr[\mathcal{L}(S) = r] \leq (1 + \alpha) \Pr[\mathcal{L}(S') = r].$$

If the set of potential outcomes is infinite? $[0, 100]$

→ Same condition, this time for any subset A of outcomes

$$\forall \Pr[\mathcal{L}(S) \in A] \leq (1 + \alpha) \Pr[\mathcal{L}(S') \in A]$$

Differential Privacy


Cynthia Dwork


Frank McSherry


Kobbi Nissim


Adam Smith

S: The data set, where each person's data is one point $x \in S$.

Differential Privacy

An algorithm $\mathcal{L}$ is α-differentially private if for all pairs of datasets S, S' differing in one user's data, and for all outputs r:

$$\Pr[\mathcal{L}(S) = r] \leq (1 + \alpha) \Pr[\mathcal{L}(S') = r].$$

When $\mathcal{L}(\cdot)$ is a learning algorithm, $h = \mathcal{L}(S)$ is a classifier, that can then be applied to any x in the domain X.

Post-processing: If $\mathcal{L}(\cdot)$ is α-differentially private, and f is any function, then $f(\mathcal{L}(\cdot))$ is also ϵ-differentially private.

Differential Privacy: An Example


Cynthia Dwork


Frank McSherry


Kobbi Nissim


Adam Smith

S: The data set, where each person's data is one point $x \in S$.

Differential Privacy

An algorithm $\mathcal{L}$ is α-differentially private if for all pairs of datasets S, S' differing in one user's data, and for all outputs r:

$$\Pr[\mathcal{L}(S) = r] \leq (1 + \alpha) \Pr[\mathcal{L}(S') = r].$$

An example: A hypothetical algorithm for mean estimation that returns 70% determinisitically, regardless of S. Is this differentially private?

Differential Privacy's Promises

- Differential Privacy and Generalization:
  - → If the $h = \mathcal{L}(S)$ doesn't depend heavily on any one sample in x …
  - → The algorithm does not overfit to S.

- Differential privacy promises that $h = \mathcal{L}(S)$ doesn't leak information about whose data was in S.

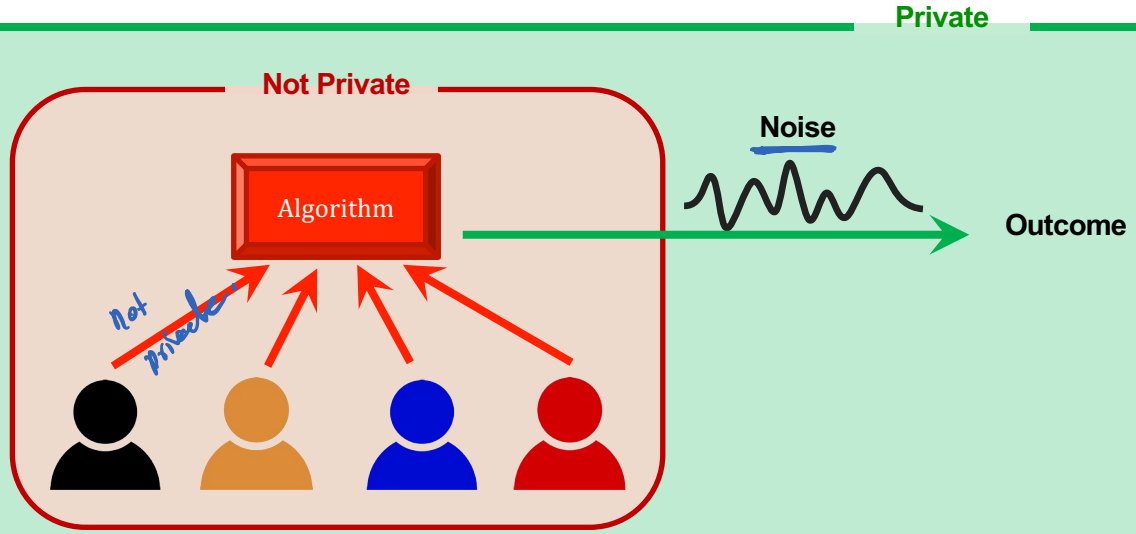
- We can still use differential privacy to find patterns in population:
  →If there is correlations between smoking and lung cancer, we can find it in the data.
  →If x is a smoker $h(x)$ will show high likelihood of getting cancer and can lead to higher health insurance rate for x.
  → **Still private:** This would have happened even if your data wasn't in the medical dataset.

The “Centralized” model of Privacy

Implemented at Census, Facebook/Social Science One

The algorithm sees the data fully, but releases information that is differentially private.

Need to trust the algorithm.



Privately Releasing Averages (or Sums)

$\frac{1}{2}$ private
2 private

Computing a sum: Add enough noise to obscure participation of a single user in the *aggregated sum*.

"Do you like Pizzas better than Burgers??"

https://tinyurl.com/yjtcp4jj

$(0.43 - 0.51)$

Ensuring (almost) 1-differential privacy: $\alpha = 1$

1. Compute the *exact* answer p.
2. Perturb that answer: $\hat{p} = p + N(0, \sigma^2)$, $\sigma \approx \frac{1}{n}$
3. Release $\hat{p}$

$p \pm$ ~privacy

St. $\frac{1}{22}$ =

0.48 people prefer pizzas to burgers.

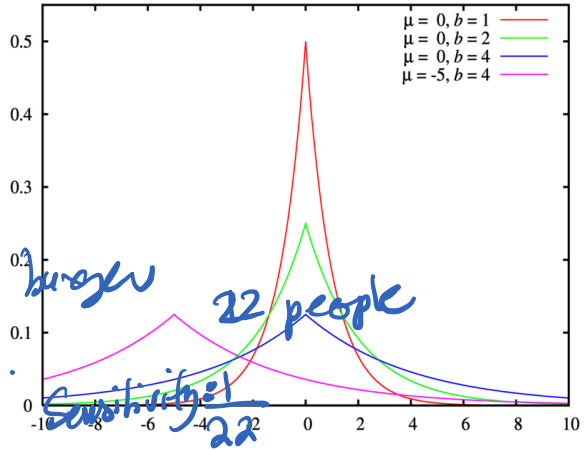
close to true mean p by about $\frac{KC}{22}$

Laplace Mechanism

Laplace Mechanism

Given a query q for data set S:

1. Compute $q(S)$
2. Output $\hat{q} = q(S) + Noise$.



What noise? Previous slide on Gaussian

- Even better: Laplacian noise.

Noise parameter:

1. More noise, more privacy, less accuracy
2. How much? depends on how sensitive $q(S)$ is to an individual.

Sensitivity: Consider "neighboring data sets" S and S'

$$sensitivity = \max_{neighboring\ S,S'} |q(S) - q(S')|$$

Laplace Mechanism

pizzas.

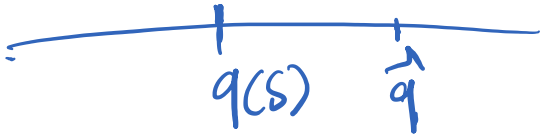
Laplace Mechanism

Given a query q for data set S:

1. Compute $q(S)$.
2. Output $\hat{q} = q(S) + Lap(sensitivity/\epsilon)$.

ϵ-private algorithm

Alg can compute



$\epsilon = 1$

Sensitivity $= \frac{1}{22}$.

$\frac{1}{22} < 0.05$

accurate

Claim: Laplace mechanism is ϵ-differentially private.

Claim: With high probability, Laplace mechanism returns $\hat{q}$ that's within $\frac{sensitivity}{\epsilon}$ of $q(S)$.

→ You will try to prove something similar in the homework.

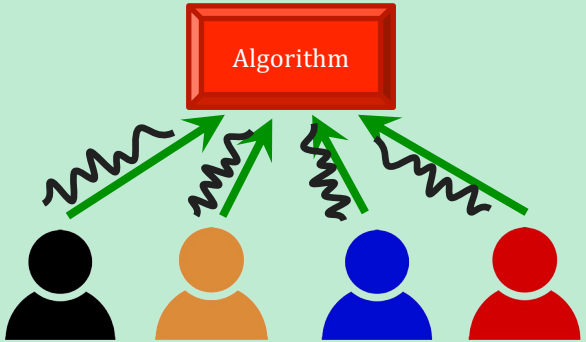
The “Distributed” model of Privacy

Implemented on iOS10, Google Chrome

Privacy protected even from the algorithm collecting the data.

- Never hold private data; no breach or subpoena risk.
- Good for when the data could be legal risk or embarrassing.

Private

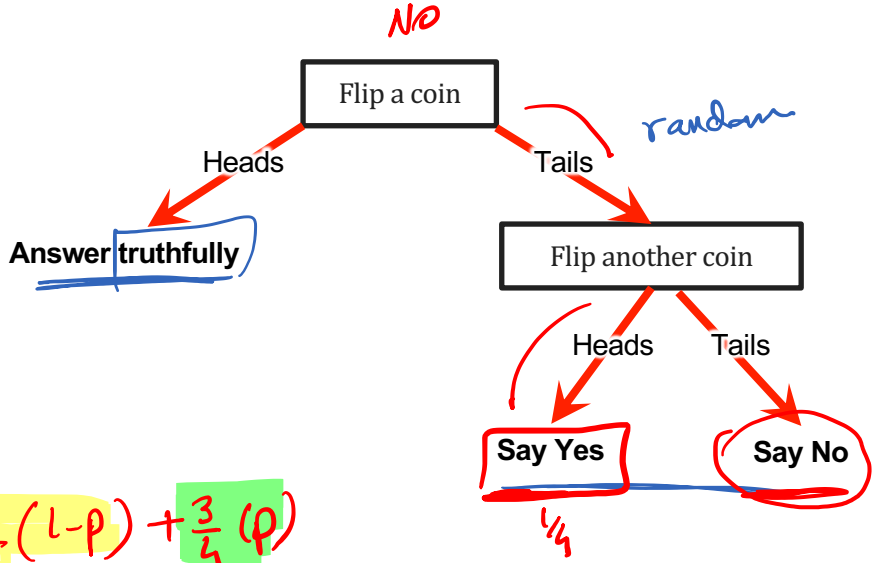


Randomized Response (RR)

Computing a sum: Each person adds noise to their response.

"Have you ever drunk so much alcohol that you threw up?"

Ensuring 2-differential privacy:



How do we compute the actual average?

$\hat{p}$ = fraction of RR=Yes ⟹ Alg compute.

p = " " " truth=Yes → Not directly

Relationship between $p, \hat{p}$: In expectation

Truth = NO (1-p) → RR = Yes = 1/4; RR = NO = 3/4

Truth = Yes → RR = NO = 1/4; RR = Yes = 3/4

$$\hat{p} = \frac{1}{4}(1-p) + \frac{3}{4}(p)$$

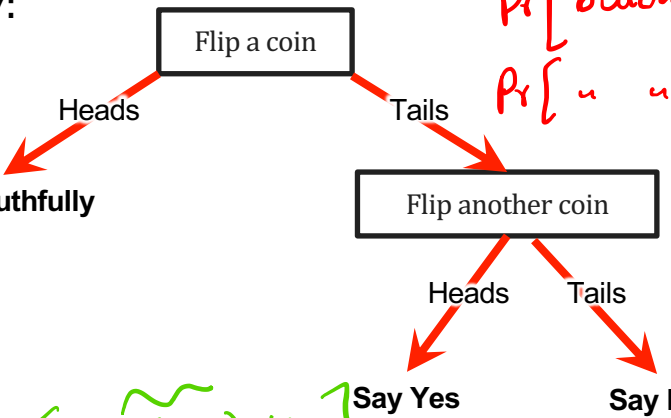
$$p = 2\hat{p} - \frac{1}{2}$$

Randomized Response

Computing a sum: Each person adds noise to their response.

"Have you ever drunk so much alcohol that you threw up?"

Ensuring 2-differential privacy:

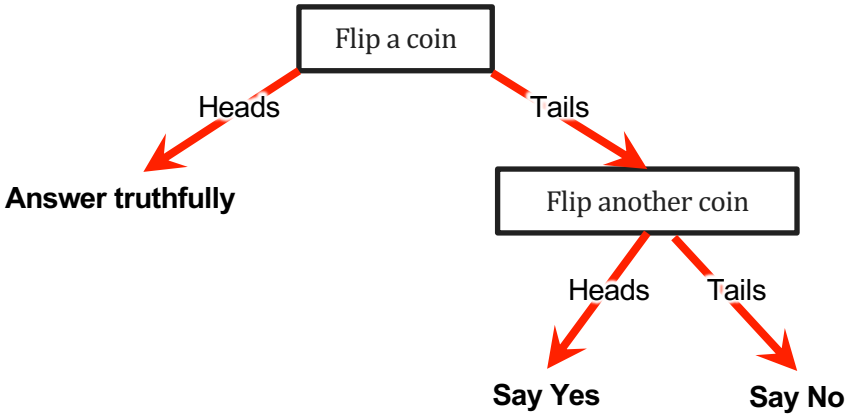


Answer: $p = 2\hat{p} - 0.5$. Where, $\hat{p}$: fraction of people whose response was Yes.

The standard deviation is about $\sigma \approx \frac{1}{2\sqrt{n}}$.

https://tinyurl.com/mwtu7mx7

Randomized Response is 2-DP



Comparison between the two

Distributed setting: randomized response

- Error of $\pm O\left(\frac{1}{\alpha\sqrt{n}}\right)$, for $\alpha = 2$ and $n = 100$, error is $\approx \pm 0.1$.
- But very private. Everybody has *plausible deniability.*
- Needs more data: Facebooks and Googles can afford it.

Centralized model: Laplace Mechanism:

- Error of $\pm O\left(\frac{1}{\alpha n}\right)$, for $\alpha = 1$ and $n = 100$, error is $\approx \pm 0.01$.
- But not that private!
- Needs less data: Smaller stakeholders can also afford it.

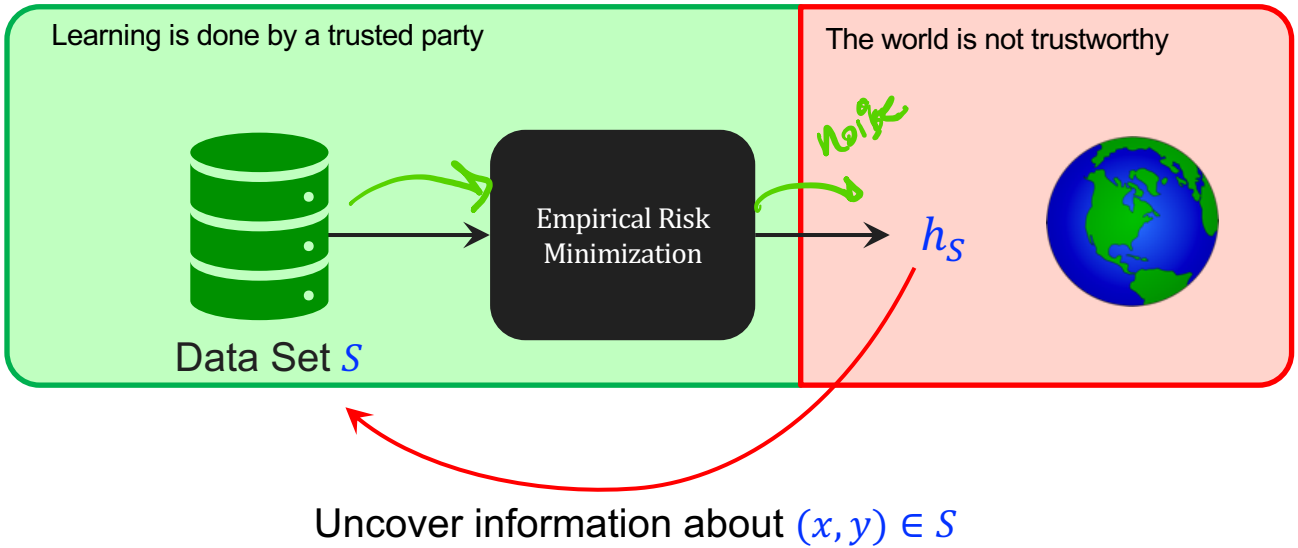
Machine Learning and Privacy

Recall that our goal is given a class H, find h such that

- $err_P(h) \leq \epsilon$ if we are in the realizable setting, or
- $err_P(h) \leq \min_{h^* \in H} err_P(h^*) + \epsilon$ in the general case.

We did this by using Empirical Risk Minimization on a sample set S.

Selection of h^* that is almost a minimizer but does it privately



Does this actually happen?!

Learning models leak training data [Fredrickson, Jha, Ristenpart. ’15]

Apply the learned model to some made up data and reconstruct some of training data.

Reconstructed Image ✓

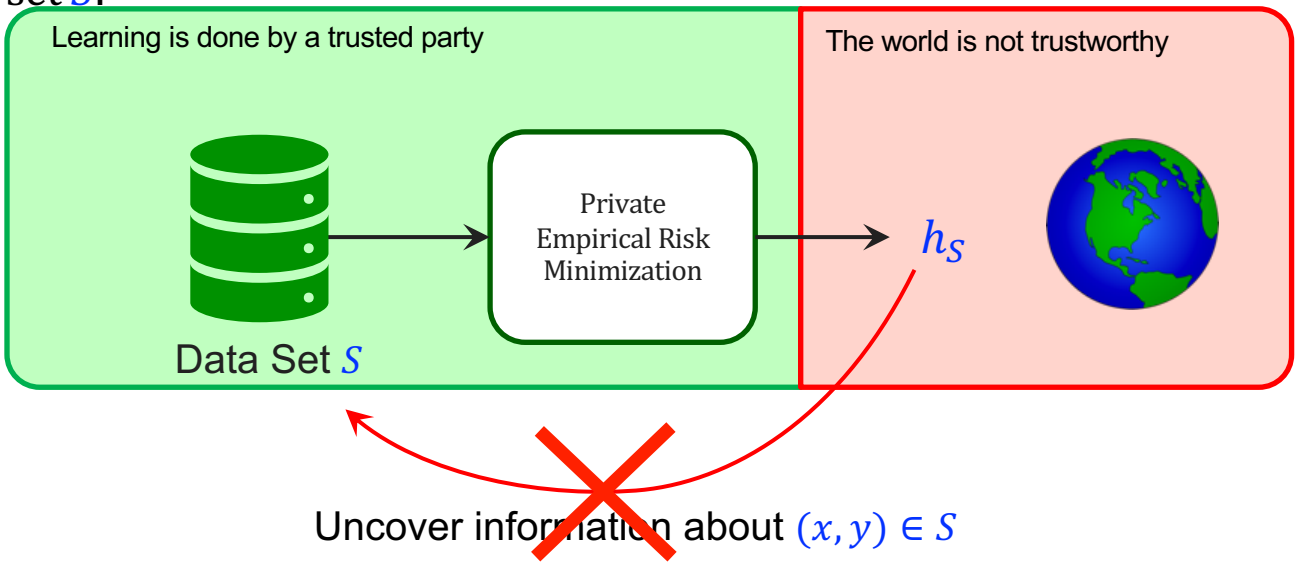
Real image ✓

Machine Learning and Privacy

Recall that our goal is given a class H, find h such that

- $err_P(h) \leq \epsilon$ if we are in the realizable setting, or
- $err_P(h) \leq \min_{h^* \in H} err_P(h^*) + \epsilon$ in the general case.

We did this by using Empirical Risk Minimization (ERM) on a sample set S.



Private ERM

Differential Privacy ERM

An algorithm $\mathcal{L}$ that returns $h \in H$ is α-differentially private if for all pairs of datasets S, S' differing in one data point, and for every $h \in H$:

$$\Pr[\mathcal{L}(S) = h] \leq (1 + \alpha) \Pr[\mathcal{L}(S') = h].$$

Bad solution 1: ERM

→ Not random and $\mathcal{L}(S) = \text{argmin}_{h \in H}\, err_S(h)$ can deterministically change. Not differentially private.

Bad solution 2: Ignore S and fix an h let $\mathcal{L}(S) = h$.

→ Differentially private, but no learning is being done (ignores S).

What we need:

→ Choose a hypothesis $\mathcal{L}(S)$ such that $err_S\big(\mathcal{L}(S)\big)$ is close to optimal.

→ Allow randomness in the choice $\mathcal{L}(S)$.

Private ERM : Exponential Mechanism

- Let $m = |S|$.
- For all $h \in H$ compute $\mathrm{err}_S(h)$.
- Output $h \in H$ with probability

$$p(h) = \exp\left(-\frac{\alpha}{2m}\,\mathrm{err}_S(h)\right) \Big/ \sum_{h \in H} \exp\left(-\frac{\alpha}{2m}\,\mathrm{err}_S(h)\right)$$

The nice things about the exponential mechanism:

- It is α-differentially private.
- With probability 0.99, it returns $h_S \in H$ such that

$$\mathrm{err}_S(h_S) \leq \underbrace{\min_{h \in H} \mathrm{err}_S(h)}_{\leq \min_{h^* \in H} \mathrm{err}_P(h^*) + \epsilon} + \underbrace{\frac{\log(|H|)}{\alpha\, m}}_{\leq \epsilon}$$

Sample complexity bounds

If $m \geq \Omega\left(\frac{1}{\epsilon^2}\log(|H|)\right)$

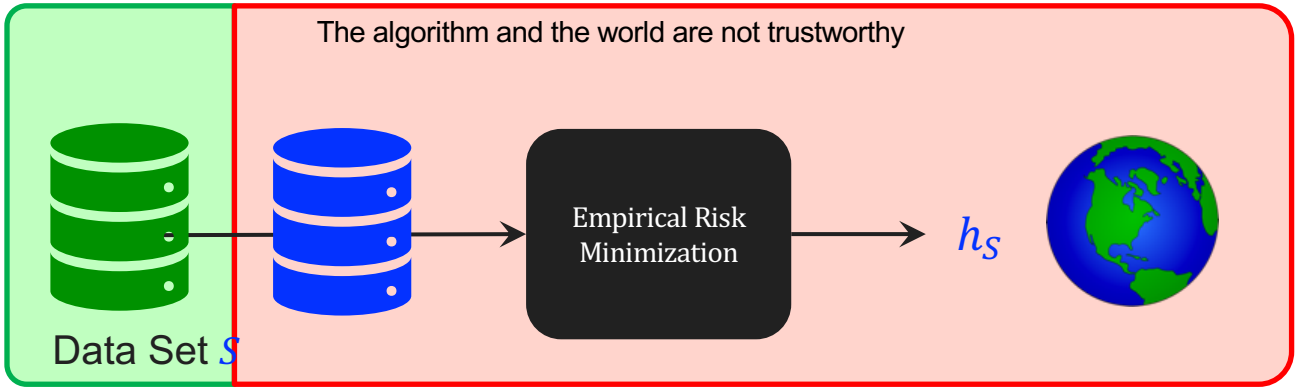
If $m \geq \Omega\left(\frac{1}{\alpha\epsilon}\log(|H|)\right)$

Data Release: Machine Learning and Privacy

Recall that our goal is given a class H, find h such that

- $err_P(h) \leq \epsilon$ if we are in the realizable setting, or
- $err_P(h) \leq \min_{h^* \in H} err_P(h^*) + \epsilon$ in the general case.

We did this by using Empirical Risk Minimization on a sample set S.



There are ways to create a synthetic data set S' from S while preserving differential privacy