Machine Learning and Differential Privacy

Maria-Florina Balcan

Learning and Privacy

- To do machine learning, we need data.
- What if the data contains sensitive information?
  - medical data, web search query data, salary data, student grade data.
- Even if the (person running the) learning algo can be trusted, perhaps the output of the algorithm reveals sensitive info.
- E.g., using search logs of friends to recommend query completions:

Why are _
Why are my feet so itchy?

Learning and Privacy

- To do machine learning, we need data.
- What if the data contains sensitive information?
- Even if the (person running the) learning algo can be trusted, perhaps the output of the algorithm reveals sensitive info.
- E.g., SVM or perceptron on medical data:
  - Suppose feature j is has-green-hair and the learned w has $w_j \neq 0$.
  - If there is only one person in town with green hair, you know they were in the study.

Learning and Privacy

- To do machine learning, we need data.
- What if the data contains sensitive information?
- Even if the (person running the) learning algo can be trusted, perhaps the output of the algorithm reveals sensitive info.
- An approach to address these problems:

Differential Privacy

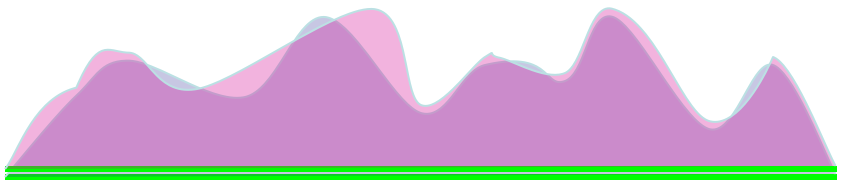
"The Algorithmic Foundations of Differential Privacy". Cynthia Dwork, Aaron Roth. Foundations and Trends in Theoretical Computer Science, NOW Publishers. 2014.

Differential Privacy

E.g., want to release average while preserving privacy.

High level idea:

- What we want is a protocol that has a probability distribution over outputs:

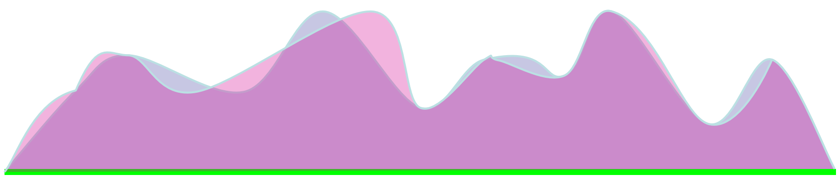


such that if person i changed their input from x_i to any other allowed x_i', the relative probabilities of any output do not change by much.

Differential Privacy

High level idea:

- What we want is a protocol that has a probability distribution over outputs:



such that if person i changed their input from x_i to any other allowed x_i', the relative probabilities of any output do not change by much.

- This would effectively allow that person to pretend their input was any other value they wanted.

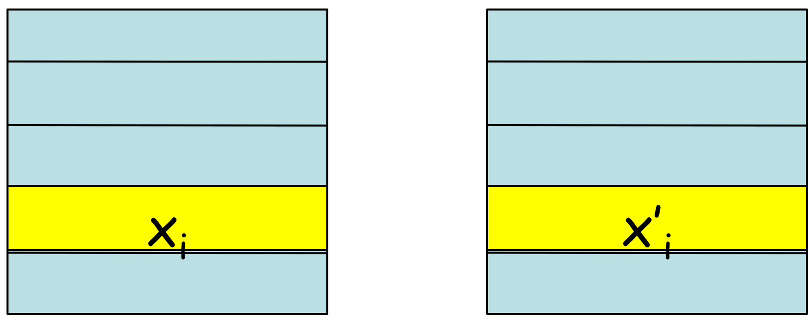
Bayes rule: $$\frac{\Pr(x_i|output)}{\Pr(x_i'|output)} = \frac{\Pr(output|x_i)}{\Pr(output|x_i')} \cdot \frac{\Pr(x_i)}{\Pr(x_i')}$$

(Posterior ≈ Prior)

Differential Privacy: Definition

It's a property of a protocol A which you run on some dataset X producing some output $A(X)$.

- A is *α-differentially private* if for any two neighbor datasets S, S' (differ in just one element $x_i \to x_i'$),



for all outcomes v,

$$e^{-\alpha} \le \Pr(A(S)=v)/\Pr(A(S')=v) \le e^{\alpha}$$

$\approx 1-\alpha$

probability over randomness in A

$\approx 1+\alpha$

Differential Privacy: Definition

It's a property of a protocol A which you run on some dataset X producing some output A(X).

- A is ϵ-*differentially private* if for any two neighbor datasets S, S′ (differ in just one element $x_i \to x_i'$),

View as model of plausible deniability

If your real input is x_i and you'd like to pretend was x_i', somebody looking at the output of A can't tell, since for any outcome v, it was nearly just as likely to come from S as it was to come from S′.

for all outcomes v,

$$e^{-\alpha} \leq \Pr(A(S)=v)/\Pr(A(S')=v) \leq e^{\alpha}$$

$\approx 1 - \alpha$

probability over randomness in A

$\approx 1 + \alpha$

Differential Privacy: Methods

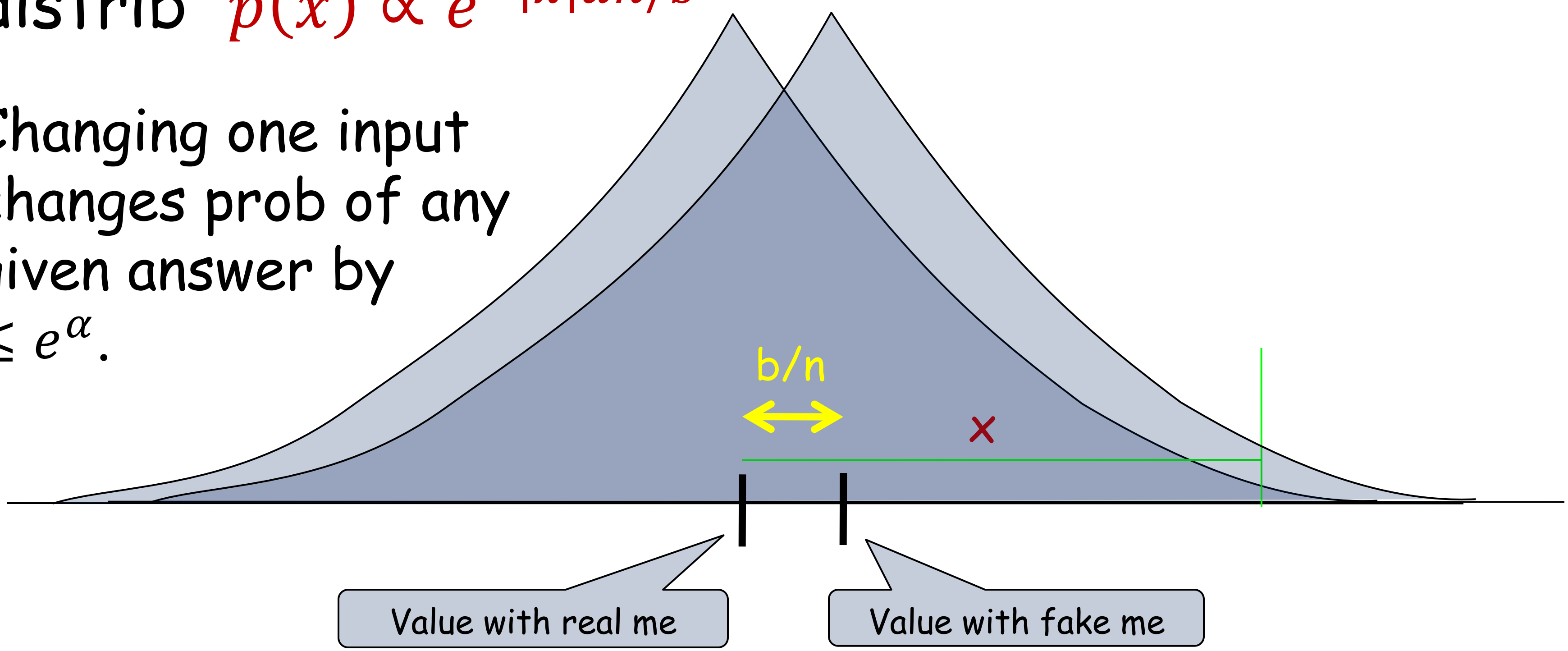
It's a property of a protocol A which you run on some dataset X producing some output A(X).

- Can we achieve it?
- Sure, just have A(X) always output 0.
- This is perfectly private, but also completely useless.
- Can we achieve it while still providing useful information?

Laplace Mechanism

Say have n inputs in range [0,b]. Want to release average while preserving privacy.

- Changing one input can affect average by ≤ b/n.
- Idea: take answer and add noise from Laplace distrib $p(x) \propto e^{-|x|\alpha n/b}$
- Changing one input changes prob of any given answer by $\le e^{\alpha}$.



Laplace Mechanism

Say have n inputs in range [0,b]. Want to release average while preserving privacy.

- Changing one input can affect average by ≤ b/n.
- Idea: : compute the true answer and add noise from Laplace distrib $p(x) \propto e^{-|x|\epsilon n/b}$
- Amount of noise added will be ≈ $\pm b/(n\epsilon)$.
- To get an overall error of $\pm\,\gamma$, you need a sample size $n = \frac{b}{\gamma\alpha}$.
- If you want to ask k queries, the privacy loss adds, so to have ϵ-differential privacy *overall*, you need $n = \frac{kb}{\gamma\alpha}$.

Laplace Mechanism

Good features:

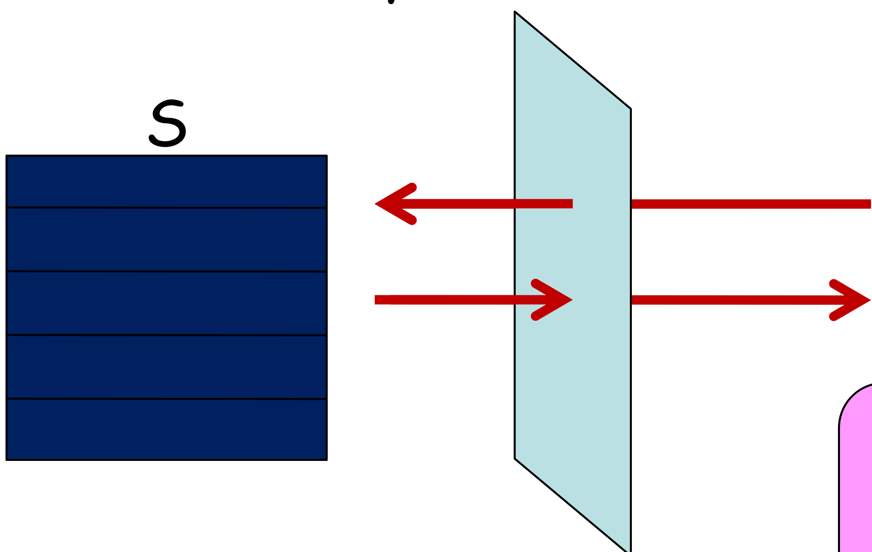
- Can run algorithms that just need to use approximate statistics (since just adding small amounts of noise to them).
- E.g., "approximately how much would this split in my decision tree reduce entropy?"

More generally

- Anything learnable via "Statistical Queries" is learnable differentially privately.

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

- Statistical Query Model [Kearns93] :



$q(x,l)$

$\Pr_D[q(x,f(x))=1] \pm \gamma$.

- What is the error rate of my current rule?
- What is the correlation of x_1 with f when x_2=0? …

- Many algorithms (including ID3, Perceptron, SVM, PCA) can be re-written to interface via such statistical estimates.

Laplace Mechanism

Problems:

- If you ask many questions, need large dataset to be able to can give accurate and private answers to all of them. (privacy losses accumulate over questions asked).
- Also, differential privacy may not be appropriate if multiple examples correspond to same individual (e.g., search queries, restaurant reviews).

More generally

Problems:

- The more interconnected our data is (A and B are friends because of person C) the trickier it becomes to reason about privacy.
- Lots of current work on definitions and algorithms.