Machine Learning for Data Science (CS4786)
Lecture 26

Differential Privacy and Machine Learning

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2017fa/

ML Requires Data

ML Requires Data

- By definition, ML is the task of automatically learning from examples or instances (Data)

ML Requires Data

- By definition, ML is the task of automatically learning from examples or instances (Data)

- Often Data used can have privacy concerns:

ML Requires Data

- By definition, ML is the task of automatically learning from examples or instances (Data)
- Often Data used can have privacy concerns:
    - Medical records of patients (Eg. learn how much smoking affects chances of getting cancer)

ML Requires Data

- By definition, ML is the task of automatically learning from examples or instances (Data)
- Often Data used can have privacy concerns:
  - Medical records of patients (Eg. learn how much smoking affects chances of getting cancer)
  - User search logs (Eg. learning personalized query retrieval for searches)

ML Requires Data

- By definition, ML is the task of automatically learning from examples or instances (Data)
- Often Data used can have privacy concerns:
  - Medical records of patients (Eg. learn how much smoking affects chances of getting cancer)
  - User search logs (Eg. learning personalized query retrieval for searches)
  - Genetic information (Eg. to learn genetic predispositions)

AOL Data Release

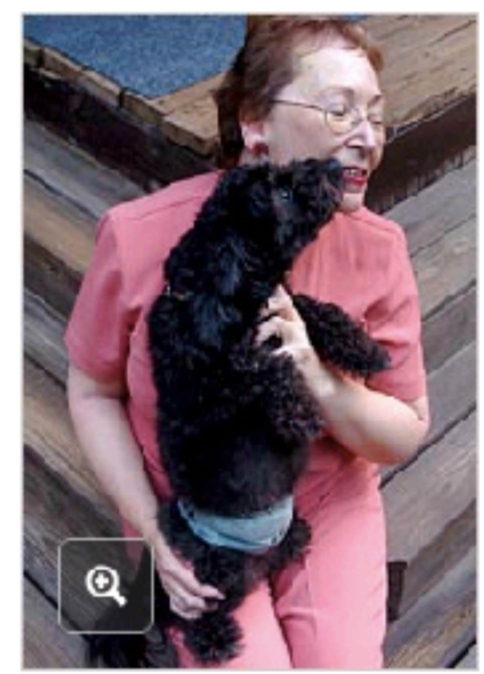
TECHNOLOGY

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 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 everything."



Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.

Erik S. Lesser for The New York Times

AOL Data Release

TECHNOLOGY

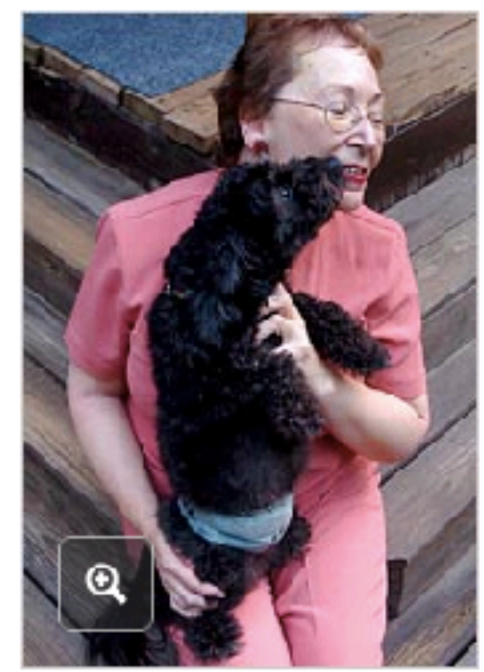
A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

- AOL released internet query

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 everything."



Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.
Erik S. Lesser for The New York Times

RYAN SINGEL SECURITY 03.12.10 02:48 PM

NETFLIX CANCELS RECOMMENDATION CONTEST AFTER PRIVACY LAWSUIT

Netflix is canceling its second $1 million Netflix Prize to settle a legal challenge that it breached customer privacy as part of the first contest's race for a better movie-

Netflix Challenge [NS'08]

Netflix Challenge [NS'08]

NETFLIX

Netflix Challenge [NS'08]

NETFLIX

Movies

User 1		5		1	
User 2	1			1	1
User 3		4			

- Given ratings by users for some movies

Netflix Challenge [NS'08]

Movies

User 1	?	5	?	1	?
User 2	1	?	?	1	1
User 3	?	4	?	?	?

- Given ratings by users for some movies
- Predict remaining ratings

Netflix Challenge [NS'08]

NETFLIX

Movies

User 1	?	5	?	1	?
User 2	1	?	?	1	1
User 3	?	4	?	?	?

- Given ratings by users for some movies
- Predict remaining ratings
- Though users were anonymized, some users on dataset were identified

Netflix Challenge [NS'08]

NETFLIX

	Movies				
User 1	?	5	?	1	?
User 2	1	?	?	1	1
User 3	?	4	?	?	?

- Given ratings by users for some movies
- Predict remaining ratings
- Though users were anonymized, some users on dataset were identified
- How?!!!

Netflix Challenge [NS'08]

NETFLIX

	Movies				
User 1	?	5	?	1	?
User 2	1	?	?	1	1
User 3	?	4	?	?	?

Netflix Challenge [NS'08]

NETFLIX

	Movies				
User 1	?	5	?	1	?
User 2	1	?	?	1	1
User 3	?	4	?	?	?

- Some of the users posted reviews (for few movies) on IMDB

Netflix Challenge [NS'08]

NETFLIX

Movies

User 1	?	5	?	1	?
User 2	1	?	?	1	1
User 3	?	4	?	?	?

- Some of the users posted reviews (for few movies) on IMDB
- Only a very small overlap with IMDB was required

Netflix Challenge [NS'08]

Movies

User 1	?	5	?	1	?
User 2	1	?	?	1	1
User 3	?	4	?	?	?

- Some of the users posted reviews (for few movies) on IMDB
- Only a very small overlap with IMDB was required
- You pretty much get the persons viewing record from / Netflix without consent

Privacy Concerns in ML

Privacy Concerns in ML

- Clearly just anonymizing didn’t seem to do the trick here

Privacy Concerns in ML

- Clearly just anonymizing didn’t seem to do the trick here
- Of course in these cases data set was released.

Privacy Concerns in ML

- Clearly just anonymizing didn’t seem to do the trick here
- Of course in these cases data set was released.
- What if we didn’t release data set:

Privacy Concerns in ML

- Clearly just anonymizing didn’t seem to do the trick here
- Of course in these cases data set was released.
- What if we didn’t release data set:
  - Trusted party uses data to learn classifiers or general statistics

Privacy Concerns in ML

- Clearly just anonymizing didn’t seem to do the trick here
- Of course in these cases data set was released.
- What if we didn’t release data set:
  - Trusted party uses data to learn classifiers or general statistics
  - Only releases general statistics?

Privacy Concerns in ML

- Clearly just anonymizing didn't seem to do the trick here
- Of course in these cases data set was released.
- What if we didn't release data set:
  - Trusted party uses data to learn classifiers or general statistics
  - Only releases general statistics?
  - Or classifier learnt from data?

Thought Experiment

Thought Experiment

- Say we release general statistics from a study

Thought Experiment

- Say we release general statistics from a study
- Eg. Smokers Vs Non-smokers (per state or county…)

Thought Experiment

- Say we release general statistics from a study
- Eg. Smokers Vs Non-smokers (per state or county…)
    - We release mean salary in the two groups

Thought Experiment

- Say we release general statistics from a study
- Eg. Smokers Vs Non-smokers (per state or county…)
  - We release mean salary in the two groups
  - Likelihood of Cancer in the two groups

Thought Experiment

- Say we release general statistics from a study
- Eg. Smokers Vs Non-smokers (per state or county…)
  - We release mean salary in the two groups
  - Likelihood of Cancer in the two groups
  - Average number of jobs held by people in the two groups

Thought Experiment

- Say we release general statistics from a study
- Eg. Smokers Vs Non-smokers (per state or county…)
  - We release mean salary in the two groups
  - Likelihood of Cancer in the two groups
  - Average number of jobs held by people in the two groups

What is the problem?

Thought Experiment

- Say we release general statistics from a study
- Eg. Smokers Vs Non-smokers (per state or county…)
  - We release mean salary in the two groups
  - Likelihood of Cancer in the two groups
  - Average number of jobs held by people in the two groups

Thought Experiment

- Say we release general statistics from a study
- Eg. Smokers Vs Non-smokers (per state or county…)
  - We release mean salary in the two groups
  - Likelihood of Cancer in the two groups
  - Average number of jobs held by people in the two groups

Say “Fill Nate” from WA was in the dataset, and is very very rich.

Thought Experiment: subtler

- Building classifier and releasing only the classifier
  - “Assume” chain smoking has some correlation with lower income
  - Say we have classifier from two or more counties/hospital, one of them has “Fill Gates”
  - Say we use regression for learning the classifier
  - By looking at weight put on income column of dataset, we can infer if “Fill gates” was part of study and which hospital

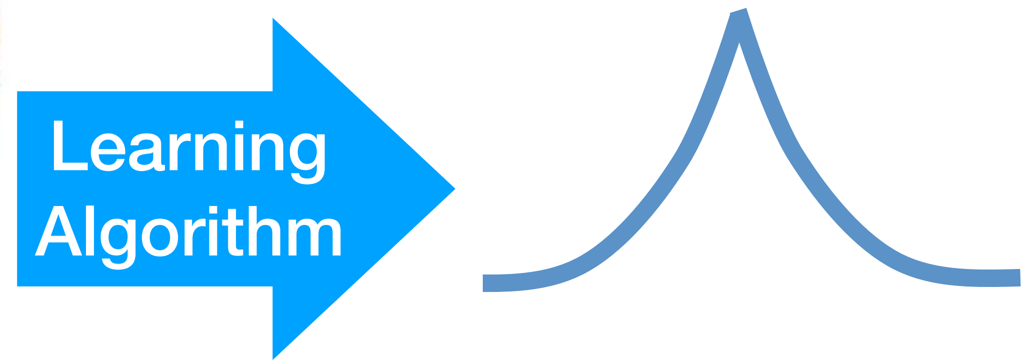
Defining Privacy

Defining Privacy

Dataset + 

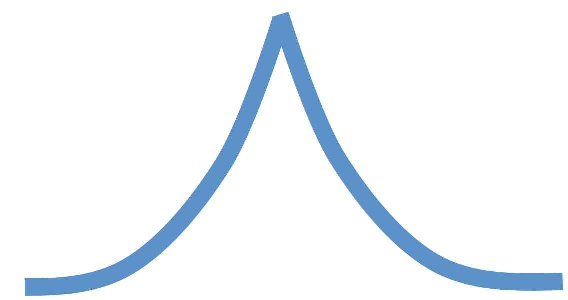
Defining Privacy

Dataset + 

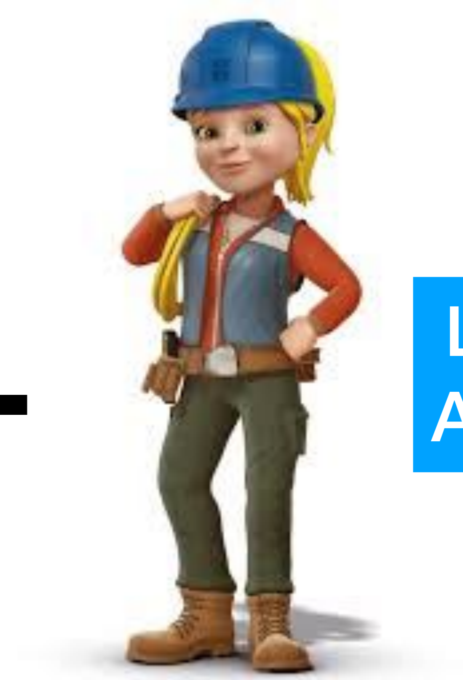
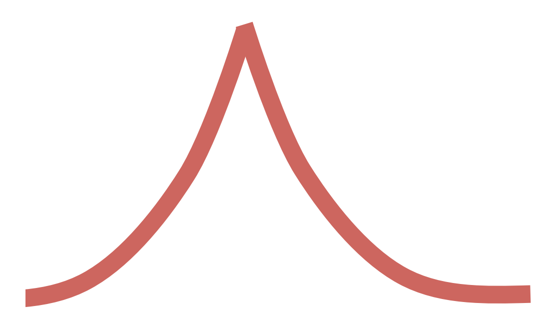


Distribution of outcome

Defining Privacy

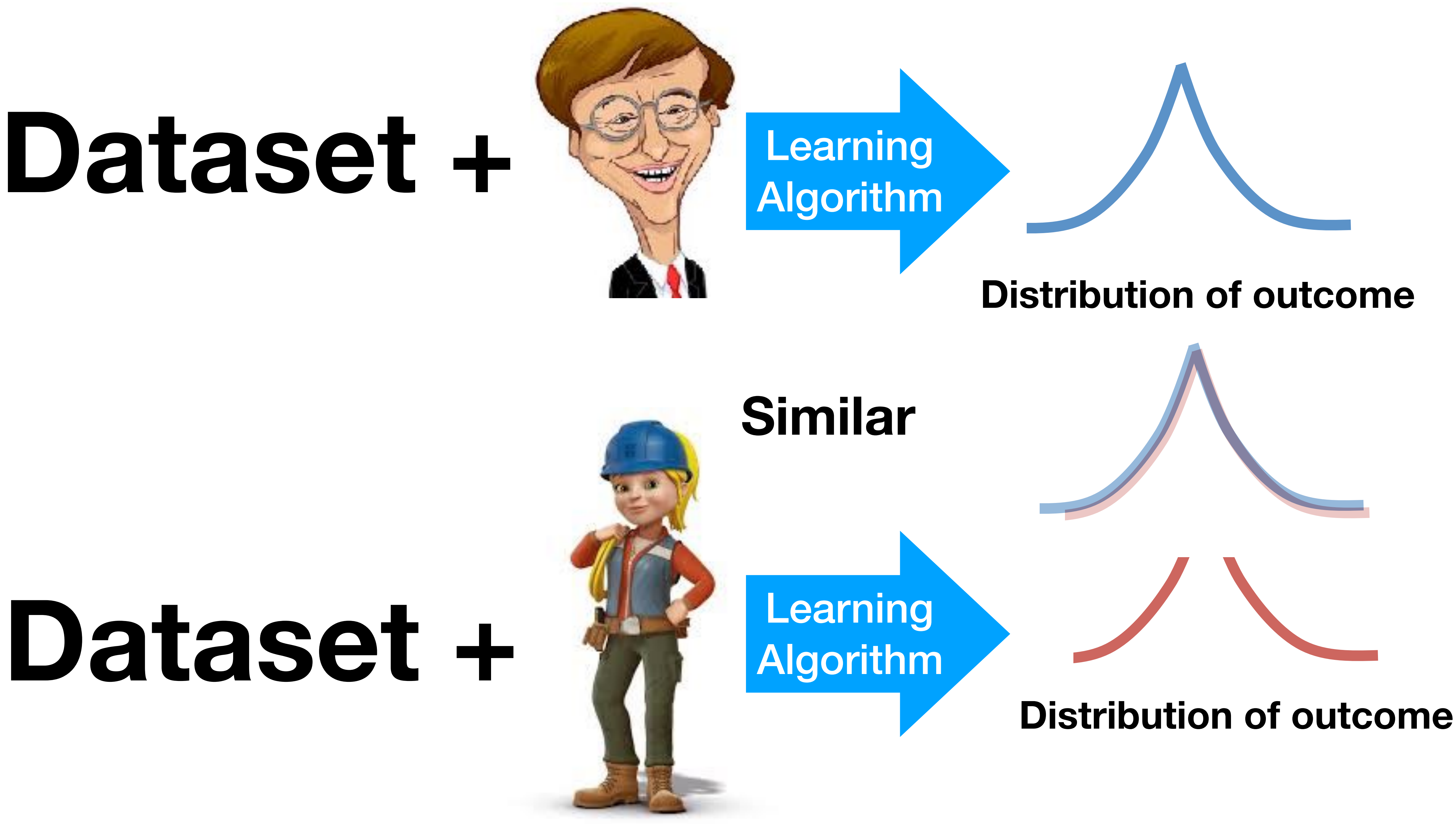
Dataset +  Learning Algorithm 

Distribution of outcome

Dataset +  Learning Algorithm 

Distribution of outcome

Defining Privacy



Differential Privacy

Differential Privacy

- A deterministic algorithm cannot preserve privacy

Differential Privacy

- A deterministic algorithm cannot preserve privacy
- Say $S = (Data_1, \ldots, Data_n)$ is the data provided to learning algorithm (be it clustering, supervised learning etc).

Differential Privacy

- A deterministic algorithm cannot preserve privacy
- Say S = ($\text{Data}_1,\ldots,\text{Data}_n$) is the data provided to learning algorithm (be it clustering, supervised learning etc).
- Say (randomized) learning algorithm A takes this training data and returns solution as A(S)

Differential Privacy

- A deterministic algorithm cannot preserve privacy

- Say S = ($Data_1$,…,$Data_n$) is the data provided to learning algorithm (be it clustering, supervised learning etc).

- Say (randomized) learning algorithm A takes this training data and returns solution as A(S)

- Algorithm A is (ϵ,δ)- differentially private if for all samples S and S' that only differ by one data point and any set C

$$P(A(S) \in C) \leq e^{\epsilon} P(A(S') \in C) + \delta$$

Differential Privacy

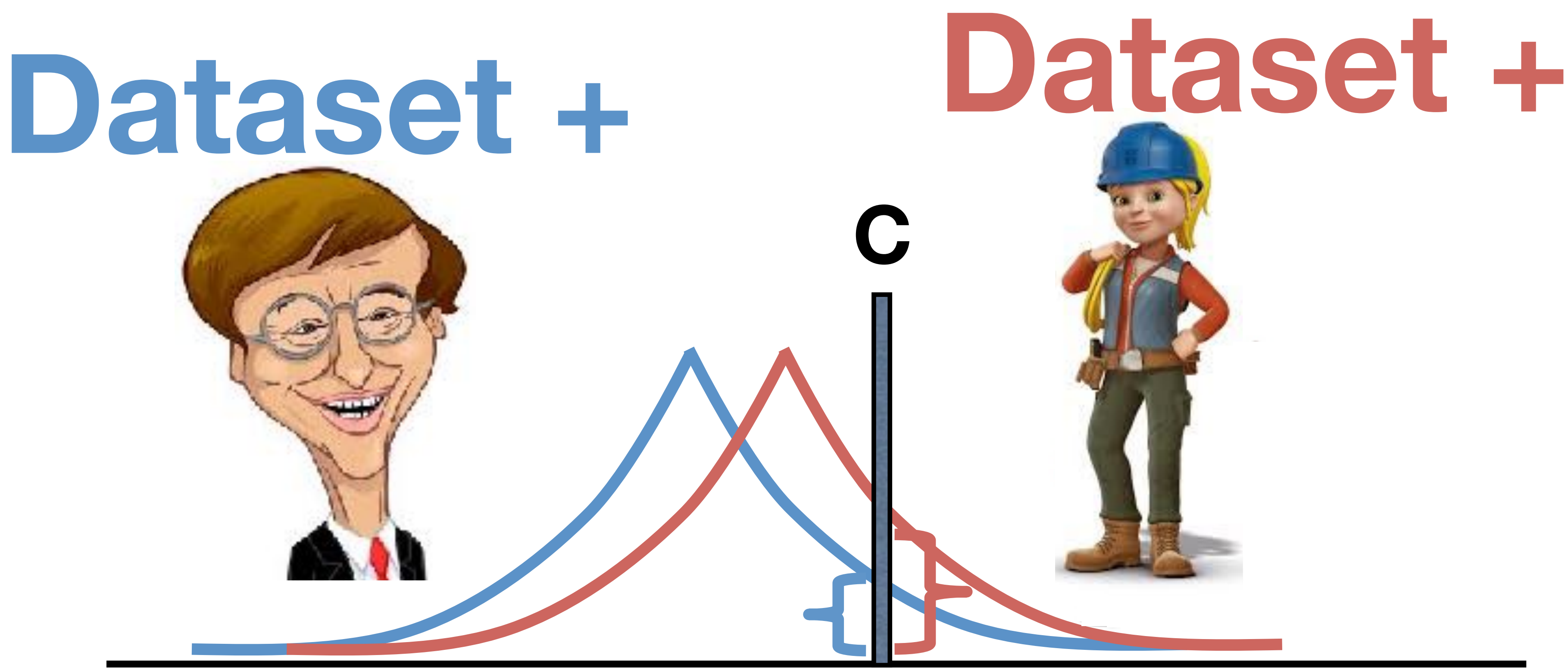
- A deterministic algorithm cannot preserve privacy
- Say S = ($Data_1$,…,$Data_n$) is the data provided to learning algorithm (be it clustering, supervised learning etc).
- Say (randomized) learning algorithm A takes this training data and returns solution as A(S)
- Algorithm A is (ϵ,δ)- differentially private if for all samples S and S' that only differ by one data point and any set C

$$P(A(S) \in C) \leq e^{\epsilon} P(A(S') \in C) + \delta$$

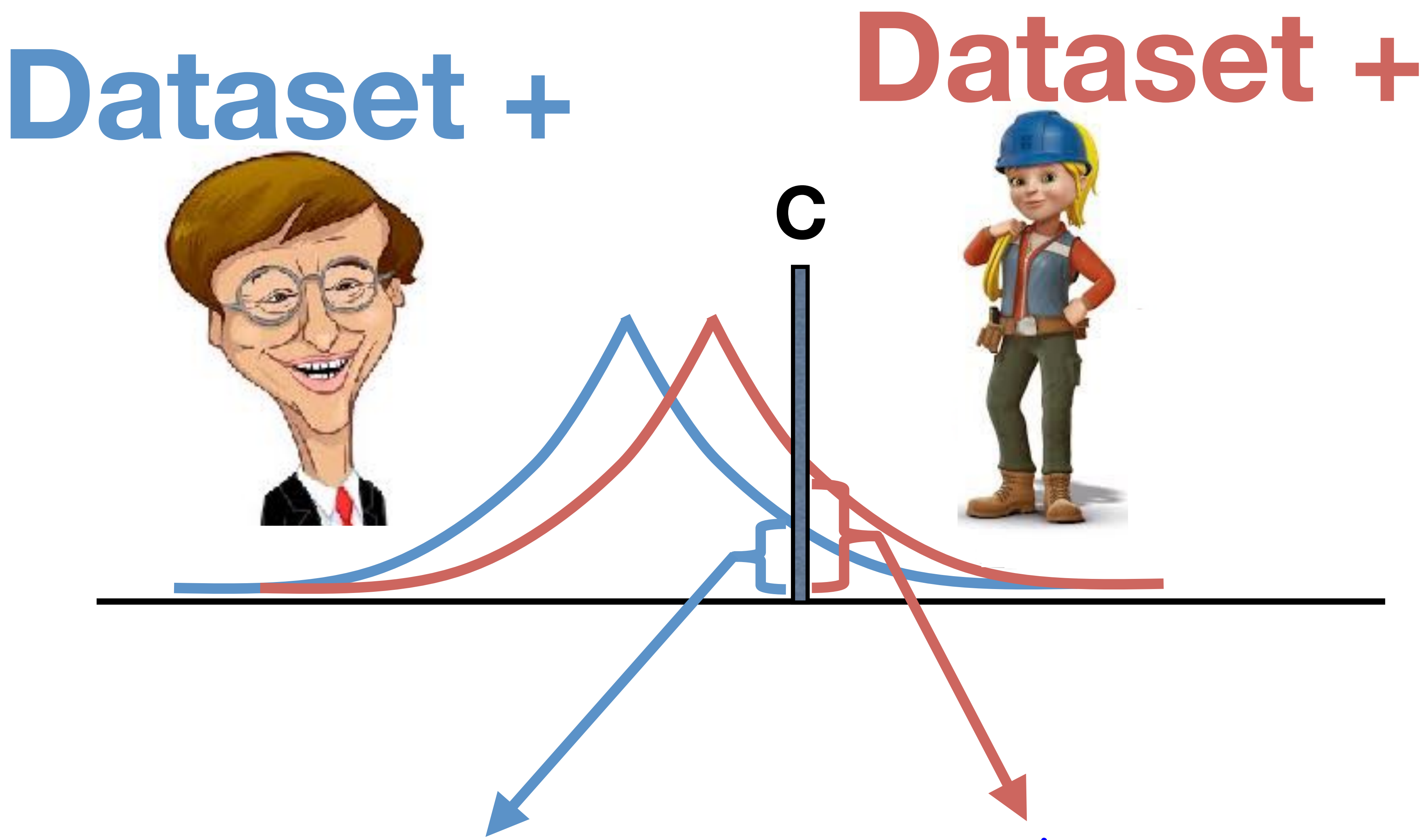
- δ=0 is called pure differential privacy

Differential Privacy

Differential Privacy

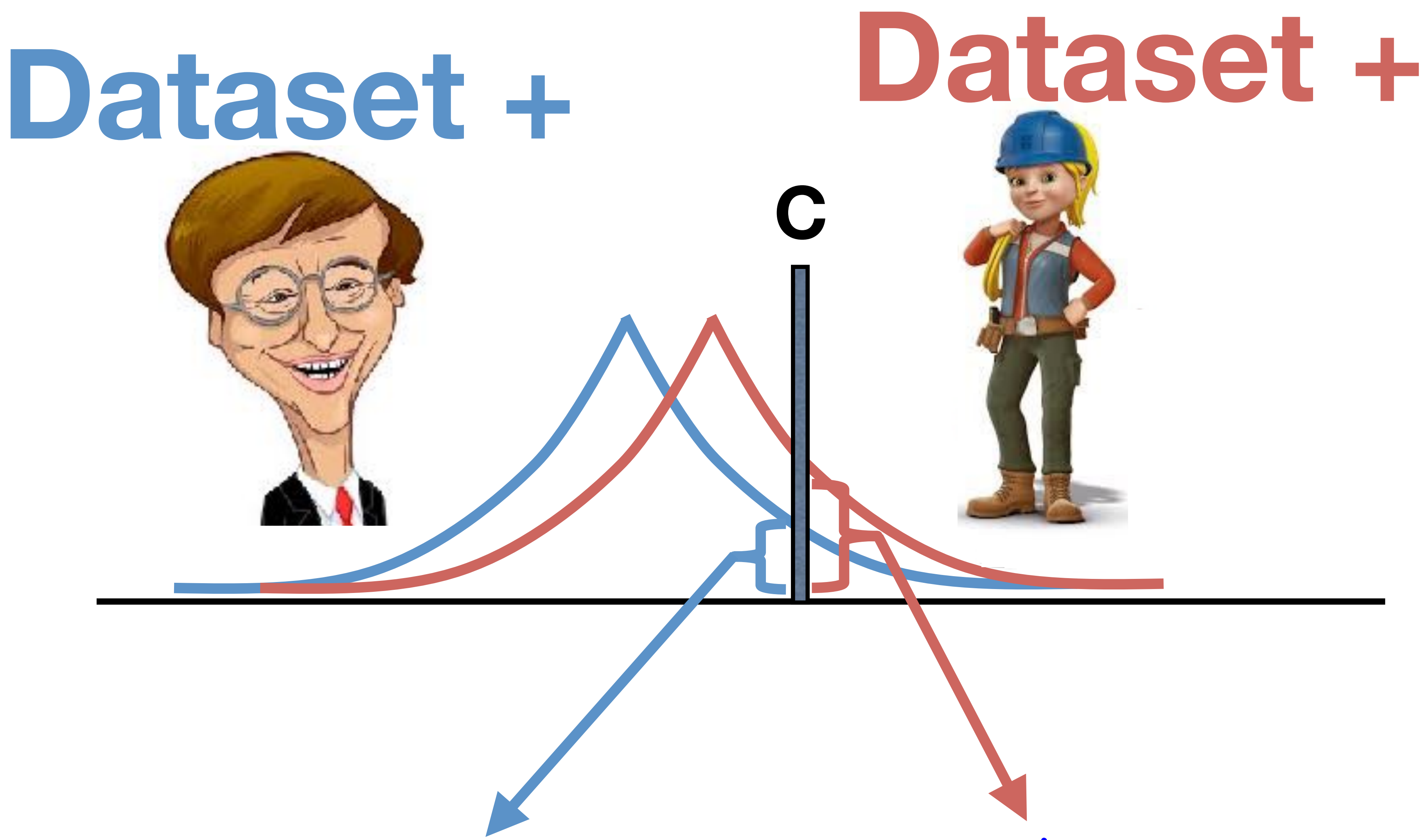


Differential Privacy



$$P(A(S) \in C) \leq e^{\epsilon} P(A(S') \in C)$$

Differential Privacy



$$P(A(S) \in C) \leq \underset{\substack{\wr\wr \\ \textbf{1 (for small } \epsilon\textbf{)}}}{e^{\epsilon}} P(A(S') \in C)$$

Differential Privacy

Differential Privacy

- Any deterministic algorithm either has to produce constant outcome (making it useless)

Differential Privacy

- Any deterministic algorithm either has to produce constant outcome (making it useless)
- If it doesn't define C to be the singleton set of outcome under say S.

Differential Privacy

- Any deterministic algorithm either has to produce constant outcome (making it useless)
- If it doesn't define C to be the singleton set of outcome under say S.
- Then probability of this set under S' is 0

Differential Privacy

- Any deterministic algorithm either has to produce constant outcome (making it useless)
- If it doesn't define C to be the singleton set of outcome under say S.
- Then probability of this set under S' is 0
- But under S probability is 1

Differential Privacy

- Any deterministic algorithm either has to produce constant outcome (making it useless)
- If it doesn't define C to be the singleton set of outcome under say S.
- Then probability of this set under S' is 0
- But under S probability is 1
- Hence cannot be differentially private

Obtaining Differential Privacy

- Typical mechanism: Add noise to outcome or inside algorithm
- More privacy we want the more noise we add

Back to Example I

- First lets begin with the example of releasing mean incomes (smokers Vs non-smokers)
- Say incomes $I_1,\ldots,I_n$ are the income of subjects in the sample
- First compute mean $M = \frac{1}{n}\sum_{t=1}^{n} I_t$
- Add noise to it M + 2 max_income Laplace(0,1)/ ϵ

Why it works?

Why it works?

- Take any arbitrary (possibly deterministic) function f(S).

Why it works?

- Take any arbitrary (possibly deterministic) function f(S).
- Say $B = \max_{S,S'} |A(S) - A(S')|$ where S and S' differ on one data point

Why it works?

- Take any arbitrary (possibly deterministic) function f(S).
- Say $B = \max_{S,S'} |A(S) - A(S')|$ where S and S' differ on one data point
- A(S) = f(S) + B Laplace(0,1)/ϵ

Why it works?

- Take any arbitrary (possibly deterministic) function f(S).
- Say $B = \max_{S,S'} |A(S) - A(S')|$ where S and S' differ on one data point
- A(S) = f(S) + B Laplace(0,1)/ϵ
- A is (ϵ,0)-differentially private

Why it works?

$$\frac{P(A(S) = x)}{P(A(S') = x)}$$

Why it works?

$$\frac{P(A(S) = x)}{P(A(S') = x)} = \frac{e^{-\epsilon|f(S)-x|/B}}{e^{-\epsilon|f(S')-x|/B}}$$

Why it works?

$$\frac{P(A(S) = x)}{P(A(S') = x)} = \frac{e^{-\epsilon|f(S)-x|/B}}{e^{-\epsilon|f(S')-x|/B}}$$

$$= e^{\epsilon\frac{|f(S')-x|-|f(S)-x|}{B}}$$

Why it works?

$$\frac{P(A(S)=x)}{P(A(S')=x)} = \frac{e^{-\epsilon|f(S)-x|/B}}{e^{-\epsilon|f(S')-x|/B}}$$

$$= e^{\epsilon\frac{|f(S')-x|-|f(S)-x|}{B}}$$

$$\leq e^{\epsilon\frac{|f(S')-f(S)|}{B}}$$

Why it works?

$$\frac{P(A(S)=x)}{P(A(S')=x)} = \frac{e^{-\epsilon|f(S)-x|/B}}{e^{-\epsilon|f(S')-x|/B}}$$

$$= e^{\epsilon\frac{|f(S')-x|-|f(S)-x|}{B}}$$

$$\leq e^{\epsilon\frac{|f(S')-f(S)|}{B}}$$

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

Back to Example II

Back to Example II

- For the classification/regression problem we can of course use the Laplace mechanism

Back to Example II

- For the classification/regression problem we can of course use the Laplace mechanism

- Can we do better?

Back to Example II

Back to Example II

- Yes! For instance for linear classifiers.

Back to Example II

- Yes! For instance for linear classifiers.
- Say we use SVM or logistic regression as follows:

$$f(S) = \mathrm{argmin}_{\mathbf{w}} \frac{1}{n} \sum_{t=1}^{n} \ell(\mathbf{w}^\top x_t, y_t) + \lambda \|\mathbf{w}\|^2$$

Back to Example II

- Yes! For instance for linear classifiers.

- Say we use SVM or logistic regression as follows:

$$f(S) = \mathrm{argmin}_{\mathbf{w}} \frac{1}{n} \sum_{t=1}^{n} \ell(\mathbf{w}^\top x_t, y_t) + \lambda \|\mathbf{w}\|^2$$

- It can be shown that if ||x||’s < 1:

$$\|f(S) - f(S')\| \leq \frac{1}{\lambda n}$$

Back to Example II

- Yes! For instance for linear classifiers.

- Say we use SVM or logistic regression as follows:

$$f(S) = \operatorname{argmin}_{\mathbf{w}} \frac{1}{n} \sum_{t=1}^{n} \ell(\mathbf{w}^\top x_t, y_t) + \lambda \|\mathbf{w}\|^2$$

- It can be shown that if $\|x\|$'s < 1:

$$\|f(S) - f(S')\| \leq \frac{1}{\lambda n}$$

- Add vector version of noise to f, only scale now is of order $O(1/\epsilon\lambda n)$

Differential Privacy in ML

- Differential private versions of PCA, clustering algorithms, deep learning etc. have been explored
- Nice properties of Differential Privacy
    - post processing is ok
    - compostability lemma
- Recently Differential Privacy was used as tool to allow statistically safe reuse of data