Differential Privacy

Machine Learning Meetup

Manuel Haußmann

February 9, 2017

Table of Contents

Outline for section 1

In machine learning the quality of the ingredients, the quality of the data provided, has a massive impact on the intelligence that is produced.
– Neil Lawrence[1]

- Data are our resource $\rightarrow$ the more we have the better(?)

[1] http://www.theguardian.com/media-network/2015/dec/14/openai-benefit-humanity-data-sharing-elon-musk-peter-thiel

In machine learning the quality of the ingredients, the quality of the data provided, has a massive impact on the intelligence that is produced.
– Neil Lawrence[1]

- Data are our resource $\rightarrow$ the more we have the better(?)
- What about privacy?

[1] http://www.theguardian.com/media-network/2015/dec/14/openai-benefit-humanity-data-sharing-elon-musk-peter-thiel

In machine learning the quality of the ingredients, the quality of the data provided, has a massive impact on the intelligence that is produced.
– Neil Lawrence[1]

- Data are our resource $\rightarrow$ the more we have the better(?)
- What about privacy? We want fair trade ingredients

What if we just anonymize the data?

[1] http://www.theguardian.com/media-network/2015/dec/14/openai-benefit-humanity-data-sharing-elon-musk-peter-thiel

Anonymization gone wrong

- Medical Records
  (Sweeney, 1997), (Sweeney, Abu, Winn, 2013)...

[2] See e.g. https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/

Anonymization gone wrong

- Medical Records
  (Sweeney, 1997), (Sweeney, Abu, Winn, 2013)...
- AOL Search Data
  3 Months worth of search data released

[2] See e.g. https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/

Anonymization gone wrong

- Medical Records
  (Sweeney, 1997), (Sweeney, Abu, Winn, 2013)...
- AOL Search Data
  3 Months worth of search data released
- Netflix Challenge

[2] See e.g. https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/

Anonymization gone wrong

- Medical Records
  (Sweeney, 1997), (Sweeney, Abu, Winn, 2013)...
- AOL Search Data
  3 Months worth of search data released
- Netflix Challenge
- New York Taxi Data[2]

[2] See e.g. https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/

Anonymization gone wrong

- Medical Records
  (Sweeney, 1997), (Sweeney, Abu, Winn, 2013)...
- AOL Search Data
  3 Months worth of search data released
- Netflix Challenge
- New York Taxi Data[2]

$\Rightarrow$ Linkage Attacks

[2] See e.g. https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/

Anonymization gone wrong

- Medical Records
  (Sweeney, 1997), (Sweeney, Abu, Winn, 2013)...
- AOL Search Data
  3 Months worth of search data released
- Netflix Challenge
- New York Taxi Data[2]

⇒ Linkage Attacks ⇒ Data cannot be fully Anonymized and Remain Useful

[2] See e.g. https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/

Problems

- What if we just anonymize the data?

Problems

- What if we just anonymize the data?
- How about we only allow aggregate over large groups of individuals?

Problems

- What if we just anonymize the data?
- How about we only allow aggregate over large groups of individuals?
- How about we place a guy in the middle who checks the queries?

Problems

- What if we just anonymize the data?
- How about we only allow aggregate over large groups of individuals?
- How about we place a guy in the middle who checks the queries?
- How about we just release summary statistics?

Problems

- What if we just anonymize the data?
- How about we only allow aggregate over large groups of individuals?
- How about we place a guy in the middle who checks the queries?
- How about we just release summary statistics?
- Then we just release ”ordinary” facts?

Problems

- What if we just anonymize the data?
- How about we only allow aggregate over large groups of individuals?
- How about we place a guy in the middle who checks the queries?
- How about we just release summary statistics?
- Then we just release ”ordinary” facts?
- Well, as long as most people are protected, who cares about ”a few”?

Let's focus on what we want

- My data should have no impact on the released results

Let's focus on what we want

- My data should have no impact on the released results
- An attacker, shouldn't be able to learn anything new about me

Let's focus on what we want

- My data should have no impact on the released results
- An attacker, shouldn't be able to learn anything new about me
- Demand by Tore Dalenius in 1977: Anything that can be learned about a respondent from the statistical database, should be learnable without access to the database.

From having access to a study, Alice should not be able to figure out whether Bob participated or not

Let's focus on what we want

- My data should have no impact on the released results
- An attacker, shouldn't be able to learn anything new about me
- Demand by Tore Dalenius in 1977: Anything that can be learned about a respondent from the statistical database, should be learnable without access to the database.

From having access to a study, Alice should not be able to figure out whether Bob participated or not

$\Rightarrow$ Randomization is the key

An Example: Randomized Response

Participating in a Study on whether or not you did X last week you are told to use the following procedure for answering:

1. Flip a coin
2. If **tails**, respond with the truth
3. Else, flip a second coin
    1. If tails: Respond "Yes"
    2. If head: Respond "No"

Expected Number of "Yes" answers: $0.25 \cdot (1-p) + 0.75 \cdot p = 0.25 + p/2$

An Example: Randomized Response

Outline for section 2

Differential Privacy (DP)

Definition: A randomized mechanism $\mathcal{M}$ is called *ε-differentially private*, if for all $S \subseteq \text{Range}(\mathcal{M})$ and for all neighboring databases $\mathcal{D}_1, \mathcal{D}_2$:

$$P(\mathcal{M}(D_1) \in S) \leq \exp(\varepsilon) P(\mathcal{M}(D_2) \in S)$$

Differential Privacy (DP)

Definition: A randomized mechanism $\mathcal{M}$ is called *(ε, δ)-differentially private*, if for all $S \subseteq \text{Range}(\mathcal{A})$ and for all neighboring databases $\mathcal{D}_1, \mathcal{D}_2$:

$$P(\mathcal{M}(\mathcal{D}_1) \in S) \leq \exp(\varepsilon) P(\mathcal{M}(\mathcal{D}_2) \in S) + \delta$$

Differential Privacy (DP)

Definition: A randomized mechanism $\mathcal{M}$ is called *(ε, δ)-differentially private*, if for all $S \subseteq \text{Range}(\mathcal{A})$ and for all neighboring databases $\mathcal{D}_1, \mathcal{D}_2$:

$$P(\mathcal{M}(\mathcal{D}_1) \in S) \leq \exp(\varepsilon) P(\mathcal{M}(\mathcal{D}_2) \in S) + \delta$$

Note: $\exp(\varepsilon) \approx 1 + \varepsilon$

Differential Privacy (DP)

Definition: A randomized mechanism $\mathcal{M}$ is called *(ε, δ)-differentially private*, if for all $S \subseteq \text{Range}(\mathcal{A})$ and for all neighboring databases $\mathcal{D}_1, \mathcal{D}_2$:

$$P(\mathcal{M}(\mathcal{D}_1) \in S) \leq \exp(\varepsilon) P(\mathcal{M}(\mathcal{D}_2) \in S) + \delta$$

Note: $\exp(\varepsilon) \approx 1 + \varepsilon$

Probability that privacy loss does not exceed ε is at most $1 - \delta$

Neighboring Database?

Generally two different interpretations:

- $\mathcal{D}_1$ can be obtained from $\mathcal{D}_2$ by adding or removing one entry (*unbounded DP*)
- $\mathcal{D}_1$ can be obtained from $\mathcal{D}_2$ by changing one entry (*bounded DP*)

Example: Mean Salary in a company: We don't want to hide the fact that Bob works there, only how much he earns

Properties of DP

Post-Processing Let f be some arbitrary randomized mapping and $\mathcal{M}$ be (ε, δ)-DP, then $f \circ \mathcal{M}$ is (ε, δ)-DP.

Properties of DP

Post-Processing Let f be some arbitrary randomized mapping and $\mathcal{M}$ be (ε, δ)-DP, then $f \circ \mathcal{M}$ is (ε, δ)-DP.

Group privacy Any $(\varepsilon, 0)$-DP mechanism is $(k\varepsilon, 0)$-DP for groups of size k

Properties of DP

Post-Processing Let f be some arbitrary randomized mapping and $\mathcal{M}$ be (ε, δ)-DP, then $f \circ \mathcal{M}$ is (ε, δ)-DP.

Group privacy Any $(\varepsilon, 0)$-DP mechanism is $(k\varepsilon, 0)$-DP for groups of size k

Composition Composition of k DP mechanisms, where $\mathcal{M}_i$ is $(\varepsilon_i, \delta_i)$-DP is $(\sum_i \varepsilon_i, \sum_i \delta_i)$-DP

Properties of DP

Post-Processing Let f be some arbitrary randomized mapping and $\mathcal{M}$ be (ε, δ)-DP, then $f \circ \mathcal{M}$ is (ε, δ)-DP.

Group privacy Any $(\varepsilon, 0)$-DP mechanism is $(k\varepsilon, 0)$-DP for groups of size k

Composition Composition of k DP mechanisms, where $\mathcal{M}_i$ is $(\varepsilon_i, \delta_i)$-DP is $(\sum_i \varepsilon_i, \sum_i \delta_i)$-DP

Note: Composition and Group privacy are not the same! *(We can get stronger results for Composition)*

Theorem: For all $\varepsilon, \delta, \delta' \geq 0$, the class of (ε, δ)-DP mechanisms, satisfies $(\varepsilon', k\delta + \delta')$-DP under k-fold adaptive composition for

$$\varepsilon' = \sqrt{2k \ln(1/\delta')}\varepsilon + k\varepsilon(e^{\varepsilon} - 1)$$

Properties of DP

- Protection against linkage attacks (are in general unaffected by auxiliary information)

Properties of DP

- Protection against linkage attacks (are in general unaffected by auxiliary information)
- Independent of an adversaries computational power

Properties of DP

- Protection against linkage attacks (are in general unaffected by auxiliary information)
- Independent of an adversaries computational power
- Quantification of Privacy Loss

An Example: Randomized Response

Claim: The Randomized Response scheme from earlier is $(\ln 3, 0)$ differentially private.

Proof:

$$\frac{P(\text{Response} = \text{"Yes"} | \text{Truth = "Yes"})}{P(\text{Response} = \text{"Yes"} | \text{Truth = "No"})} = \frac{3/4}{1/4} =$$

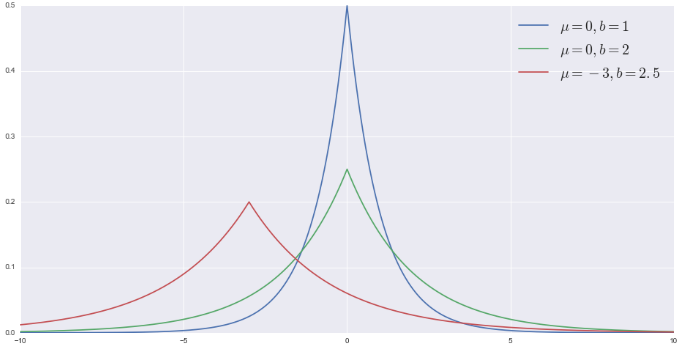
$$\frac{P(\text{Response} = \text{"No"} | \text{Truth = "No"})}{P(\text{Response} = \text{"No"} | \text{Truth = "Yes"})} = 3$$

Laplace Mechanism

Quick Intro to the Laplace Distribution

$$\mathcal{L}ap(x|\mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

$$\mathbb{E}[x] = \mu \qquad \mathrm{var}[x] = 2b^2$$



(Global) Sensitivity of a function

Definition: The ℓ_1 sensitivity of a function f is:

$$\Delta f = \max_{x_1, x_2} |f(x_1) - f(x_2)|_1$$

for two neighboring datasets x_1, x_2

- Measure of how much a single person can influence the outcome.

(Global) Sensitivity of a function

Definition: The ℓ_1 sensitivity of a function f is:

$$\Delta f = \max_{x_1, x_2} |f(x_1) - f(x_2)|_1$$

for two neighboring datasets x_1, x_2

- Measure of how much a single person can influence the outcome.
- Δ for query: "How many Mathematicians?"

(Global) Sensitivity of a function

Definition: The ℓ_1 sensitivity of a function f is:

$$\Delta f = \max_{x_1, x_2} |f(x_1) - f(x_2)|_1$$

for two neighboring datasets x_1, x_2

- Measure of how much a single person can influence the outcome.
- Δ for query: "How many Mathematicians?"
- Δ for query: "How many siblings?"

(Global) Sensitivity of a function

Definition: The ℓ_1 sensitivity of a function f is:

$$\Delta f = \max_{x_1, x_2} |f(x_1) - f(x_2)|_1$$

for two neighboring datasets x_1, x_2

- Measure of how much a single person can influence the outcome.
- Δ for query: "How many Mathematicians?"
- Δ for query: "How many siblings?"
- Δ for query: "Histogram of salary/income?"

Laplace Mechanism

Definition: Given any function f, the *Laplace Mechanism* is defined as:
$\mathcal{M}_L(x, f(\cdot), \varepsilon) = f(x) + (Y_1, \ldots, Y_k)$ where $Y_i \sim \mathcal{L}ap(0, \Delta f/\varepsilon)$ (iid)

Theorem: The Laplace Mechanism preserves $(\varepsilon, 0)$-DP.
Proof: Let x_1, x_2 be two neighboring datasets, then

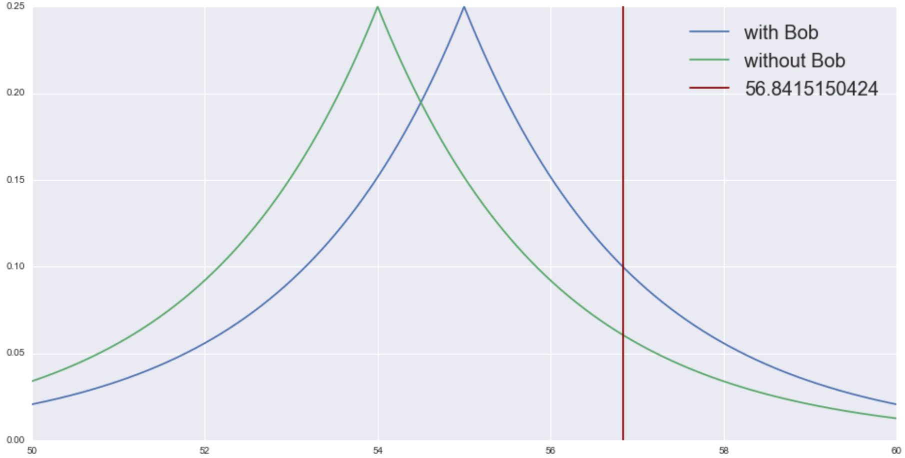
$$\frac{p_{x_1}(z)}{p_{x_2}(z)} = \frac{\exp\left(-\frac{\varepsilon|f(x_1)-z|}{\Delta f}\right)}{\exp\left(-\frac{\varepsilon|f(x_2)-z|}{\Delta f}\right)}$$

$$= \exp\left(\frac{\varepsilon(|f(x_2) - z| - |f(x_1) - z|)}{\Delta f}\right)$$

$$\leq \exp\left(\frac{\varepsilon|f(x_2) - f(x_1)|}{\Delta f}\right) \leq \exp(\varepsilon)$$

Example: Laplace Mechanism

Situation: Study of Drug Usage among Cryptographers.



A quick detour: Laplace vs Gauss

What does the Gaussian Version look like?
Definition: ℓ_2 sensitivity of a function f is

$$\Delta_2 f = \max_{x_1, x_2} ||f(x) - f(y)||_2$$

where x_1, x_2 are neighboring datasets.

A quick detour: Laplace vs Gauss

What does the Gaussian Version look like?
Definition: ℓ_2 sensitivity of a function f is

$$\Delta_2 f = \max_{x_1, x_2} ||f(x) - f(y)||_2$$

where x_1, x_2 are neighboring datasets.

Theorem: Let $\varepsilon \in (0, 1)$ be arbitrary. For $c^2 > 2\ln(1.25/\delta)$, the Gaussian Mechanism with parameter $\sigma \geq c\Delta_2 f/\varepsilon$ is (ε, δ)-DP

What about queries like:

- "Most frequent bachelor degree in this room?"
- "Most frequent eye color?"

Exponential Mechanism

Exponential Mechanism

Definition: The *Exponential Mechanism* $\mathcal{A}_E(x, u, \mathcal{R})$, selects and outputs an element $r \in \mathcal{R}$ with probability proportional to $\exp\left(\frac{\varepsilon u(x,r)}{2\Delta u}\right)$, where u is a suitable utility/scoring function

Theorem: The *Exponential Mechanism* preserves $(\varepsilon, 0)$-differential privacy
Proof: *Analogous to Laplace*

Synthetic Data

So far: We give Alice our function f and she returns a noisy result to use. Can we do this offline on our own?

- There is work on *Synthetic Data* that can be published and freely operated on, but...

Synthetic Data

So far: We give Alice our function f and she returns a noisy result to use. Can we do this offline on our own?

- There is work on *Synthetic Data* that can be published and freely operated on, but...
- ... need to specify which kind of questions will be asked beforehand

Synthetic Data

So far: We give Alice our function f and she returns a noisy result to use.
Can we do this offline on our own?

- There is work on *Synthetic Data* that can be published and freely operated on, but...
- ... need to specify which kind of questions will be asked beforehand
- ... only certain computations can be done with a reasonable accuracy

Synthetic Data

So far: We give Alice our function f and she returns a noisy result to use.
Can we do this offline on our own?

- There is work on *Synthetic Data* that can be published and freely operated on, but...
- ... need to specify which kind of questions will be asked beforehand
- ... only certain computations can be done with a reasonable accuracy
- Maybe intermediate approach? Spend part of your privacy budget on looking at the data and the rest to build a synthetic dataset

Synthetic Data

So far: We give Alice our function f and she returns a noisy result to use. Can we do this offline on our own?

- There is work on *Synthetic Data* that can be published and freely operated on, but...
- ... need to specify which kind of questions will be asked beforehand
- ... only certain computations can be done with a reasonable accuracy
- Maybe intermediate approach? Spend part of your privacy budget on looking at the data and the rest to build a synthetic dataset
- Looks so far like work in progress

SYNTHETIC DATA

A SIMPLE AND PRACTICAL ALGORITHM FOR DIFFERENTIAL PRIVACY
(HARDT, LIGETT, MCSHERRY,NIPS 2012)

For $i = 1, ..., T$

1. Exponential Mechanism: Sample $q_i \in Q$ using EM, parametrized with $\varepsilon/2T$ and score function

$$s_i(D, q) = |q_i(A_{i-1}) - q(D)|$$

2. Laplace Mechanism: Let $m_i = q_i(D) + \mathcal{L}ap(2T/\varepsilon)$
3. Multiplicative Weights:

$$A_i(x) \propto A_{i-1}(x) \exp(q_i(x) \cdot (m_i - q_i(A_{i-1}))/2n)$$

Return $A = \text{avg} A_i$

Synthetic Data

A Simple and Practical Algorithm for Differential Privacy
(Hardt, Ligett, McSherry,NIPS 2012)

For $i = 1, ..., T$

1. Exponential Mechanism: Sample $q_i \in Q$ using EM, parametrized with $\varepsilon/2T$ and score function

$$s_i(D, q) = |q_i(A_{i-1}) - q(D)|$$

2. Laplace Mechanism: Let $m_i = q_i(D) + \mathcal{L}ap(2T/\varepsilon)$
3. Multiplicative Weights:

$$A_i(x) \propto A_{i-1}(x) \exp(q_i(x) \cdot (m_i - q_i(A_{i-1}))/2n)$$

Return $A = \text{avg} A_i$

Proof of Privacy:

Synthetic Data

A Simple and Practical Algorithm for Differential Privacy (Hardt, Ligett, McSherry,NIPS 2012)

For $i = 1, ..., T$

1. Exponential Mechanism: Sample $q_i \in Q$ using EM, parametrized with $\varepsilon/2T$ and score function

$$s_i(D, q) = |q_i(A_{i-1}) - q(D)|$$

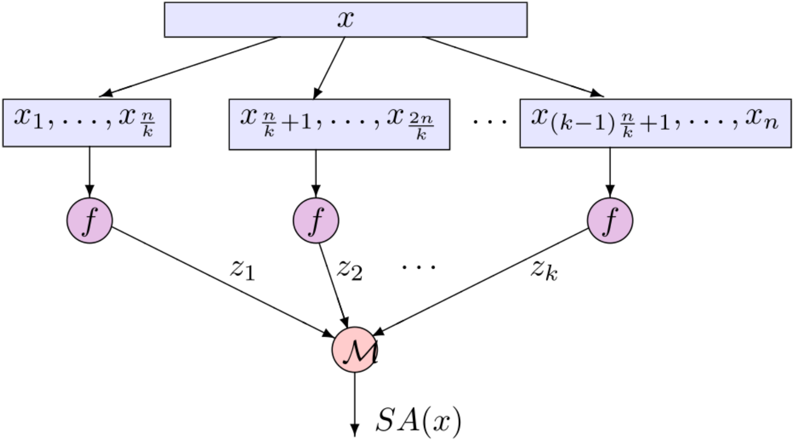
2. Laplace Mechanism: Let $m_i = q_i(D) + \mathcal{L}ap(2T/\varepsilon)$
3. Multiplicative Weights:

$$A_i(x) \propto A_{i-1}(x) \exp(q_i(x) \cdot (m_i - q_i(A_{i-1}))/2n)$$

Return $A = \text{avg} A_i$

Proof of Privacy: $\varepsilon/(2T) + \varepsilon/(2T) = \varepsilon$

Subsample and Aggregate



[Figure7.1 from (Dwork,Roth, 2014)]

What about ε?

Let's say Bob will be in $k = 10000$ $(\varepsilon_0, 0)$-DP databases. Binding his cumulative lifetime privacy loss at $\varepsilon = 1$ with probability $(1 - e^{-32})$ we need $\varepsilon_0 = 1/801$ for each database.

- How about a ε per study?

What about ε?

Let's say Bob will be in $k = 10000$ $(\varepsilon_0, 0)$-DP databases. Binding his cumulative lifetime privacy loss at $\varepsilon = 1$ with probability $(1 - e^{-32})$ we need $\varepsilon_0 = 1/801$ for each database.

- How about a ε per study?
- Or ε per researcher?

What about ε?

Let's say Bob will be in $k = 10000$ $(\varepsilon_0, 0)$-DP databases. Binding his cumulative lifetime privacy loss at $\varepsilon = 1$ with probability $(1 - e^{-32})$ we need $\varepsilon_0 = 1/801$ for each database.

- How about a ε per study?
- Or ε per researcher?
- Allow a total budget of ε for the dataset and bet on innovation to optimize use of this resource.

Results and Extensions

- *What can we learn privately?*, (Kasivisvanathan, et al. 2008)
  ”Therefore, almost anything learnable is learnable privately: specifically, if a concept class is learnable by a (non-private) algorithm with polynomial sample complexity and output size, then it can be learned privately using a polynomial number of samples”
- *Concentrated Differential Privacy*, (Dwork and Rothblum, 2016s)
  Relaxation to (ε, δ), with higher accuracy, while preserving composition results

DP & ML

- Very broad literature: Cryptography & Security, Statistics, Machine Learning, some game theoretic approach etc.
- Many algorithms have a privatized version of them
- DP & ML share a similar goal: Learn information about the distribution of the data, without depending too much/being sensitive on individual data points
- Where to introduce noise?
    - perturb input $\Rightarrow$ similar to our beginning example
    - perturb objective $\Rightarrow$ can be seen as a kind of regularization
    - perturb output $\Rightarrow$ what we have done so far

Outline for section 3

Conclusion

- Still a lot of open questions (how to choose ε, how to get rid of the intermediary curator, better compositions for reducing privacy leakage, popular implementations ...)
- But a very fast growing field (given that term and definition stem from 2006.)
- Differential Privacy looks like a very promising way to conduct privacy preserving ML
- See, *No Free Lunch in Data Privacy* (Kifer and Machanvajjhala, 2011) for a critical discussion of DP
- Data Trusts?[3]

[3] http://inverseprobability.com/2016/05/29/data-trusts

Main Sources

- *"A Firm Foundation for Private Data Analysis", (Dwork, 2011)*
- *"Algorithmic Foundations of Differential Privacy"*, by Cynthia Dwork and Aaron Roth
- *"Differential Privacy and Learning: The Tools, The Results, and The Frontier"*, NIPS Tutorial, 2014 by Katrina Ligett

Outline for section 4

Privacy Preserving Logistic Regression

Chaudhuri, Monteleoni, NIPS 2008

Simple approach using that the sensitivity of logistic regression is $2/n\lambda$

1. Compute w^* by the usual regularized logistic regression on $(x_1, y_1), \ldots, (x_n, y_n)$
2. pick noise vector $\eta \sim \mathcal{L}ap(2/(n\lambda\varepsilon))$
3. Return $w^* + \eta$

PRIVACY PRESERVING LOGISTIC REGRESSION

CHAUDHURI, MONTELEONI, NIPS 2008

More sophisticated

1. Pick $b \sim \mathcal{L}ap(1/\varepsilon)$
2. Given $(x_1, y_y), \ldots, (x_n, y_n)$ and regularizer λ, compute

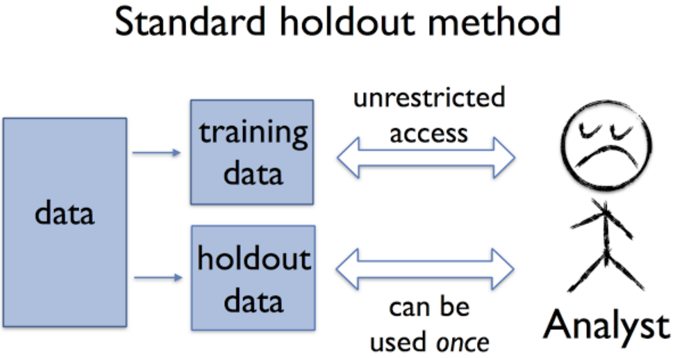
$$w^* = \arg\min_w \frac{1}{2}\lambda w^\top w + \frac{b^\top w}{n} + \tfrac{1}{n}\sum_i \log(1 + \exp(-y_i w^\top x_i))$$

3. Return w^*

Reusable Holdout

Generalization in Adaptive Data Analysis and Holdout Reuse (Dwork et al., NIPS 2015)

Ideal Situation[4]:

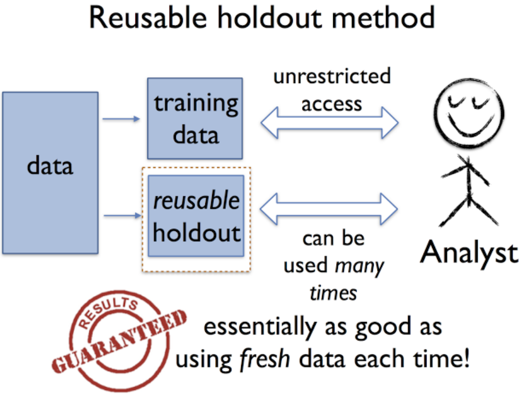


[4] Image due to Moritz Hardt via http://googleresearch.blogspot.de/2015/08/the-reusable-holdout-preserving.html

Reusable Holdout

Generalization in Adaptive Data Analysis and Holdout Reuse (Dwork et al., NIPS 2015)

Suggested Solution[5]:



[5] Image due to Moritz Hardt via http://googleresearch.blogspot.de/2015/08/the-reusable-holdout-preserving.html

Reusable Holdout

Generalization in Adaptive Data Analysis and Holdout Reuse (Dwork et al., NIPS 2015)

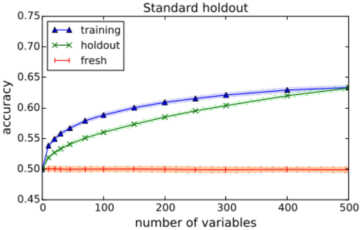
Input: Training set S_t, Holdout set S_h, threshold T, noise rate σ, Budget B

sample $\gamma \sim \mathcal{L}ap(2\sigma)$; $\hat{T} \leftarrow T + \gamma$ and for each query ϕ:

1. if $B < 1$ return $\emptyset$
2. else
   1. sample $\eta \sim \mathcal{L}ap(4\sigma)$
   2. if $|\mathcal{E}_{S_h}[\phi] - \mathcal{E}_{S_t}[\phi]| > \hat{T} + \eta$
      - sample $\xi \sim \mathcal{L}ap(\sigma), \gamma \sim \mathcal{L}ap(2\sigma)$
      - $B \leftarrow B - 1$, $\hat{T} \leftarrow T + \gamma$
      - output $\mathcal{E}_{S_h}[\phi] + \xi$
   3. else output $\mathcal{E}_{S_t}[\phi]$

Reusable Holdout

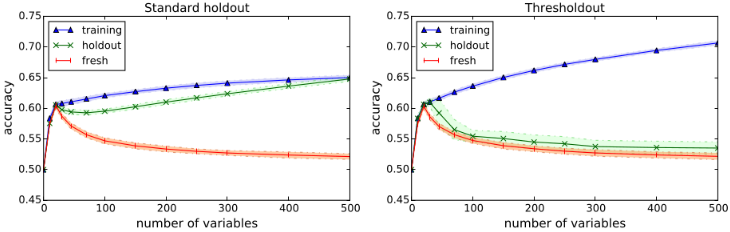
Generalization in Adaptive Data Analysis and Holdout Reuse (Dwork et al., NIPS 2015)



[(Dwork et al., 2015)]

Reusable Holdout

Generalization in Adaptive Data Analysis and Holdout Reuse (Dwork et al., NIPS 2015)



[(Dwork et al., 2015)]

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

Bayes as we know and love him

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

- See posterior as EM with utility $u(X, \theta) = \log P(X, \theta)$

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

Bayes as we know and love him

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

- See posterior as EM with utility $u(X, \theta) = \log P(X, \theta)$
- Draw θ from

$$f(\theta; X, \varepsilon) \propto \exp\left(\frac{\varepsilon \log P(\theta, X)}{2\Delta \log P(\theta, X)}\right)$$

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

Bayes as we know and love him

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

- See posterior as EM with utility $u(X, \theta) = \log P(X, \theta)$
- Draw θ from

$$f(\theta; X, \varepsilon) \propto \exp\left(\frac{\varepsilon \log P(\theta, X)}{2\Delta \log P(\theta, X)}\right)$$

- Sensitivity:

$$\begin{aligned}\Delta \log P(X, \theta) &= \max \left|\log P(\theta, X^{(1)}) - \log P(\theta, X^{(2)})\right| \\ &= \max_{x, x', \theta} \left|\log P(x'|\theta) - \log P(x|\theta)\right|\end{aligned}$$

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

- **Theorem:** If $\log P(X, \theta) \leq C$, releasing one sample from the posterior distribution $P(\theta|X)$ with any prior is $2C$-DP

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

- **Theorem:** If $\log P(X, \theta) \leq C$, releasing one sample from the posterior distribution $P(\theta|X)$ with any prior is $2C$-DP
- Can rewrite f as Boltzman distribution

$$f(\theta; X, \varepsilon) \propto \exp\left(\frac{\varepsilon \log P(\theta, X)}{2\Delta \log P(\theta, X)}\right)$$
$$\propto \exp\left(\frac{-E(\theta)}{T}\right)$$

with $E(\theta) = -u(X, \theta) = -\log P(\theta, X)$, $T = \frac{2\Delta u(X,\theta)}{\varepsilon}$

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

Note:

- $\varepsilon = 0$ corresponds to sampling from uniform distribution $\Rightarrow$ perfect privacy
- $\varepsilon = 2\Delta \log P(\theta, X)$ gives us samples from the posterior
- $\varepsilon \to \infty$ sample most likely θ (cap it at '=')
- For privacy budget $\varepsilon' \geq 2q\Delta \log P(\theta, X)$ with $q \in \mathbb{N}$, can draw q posterior samples within our budget

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

Note:

- $\varepsilon = 0$ corresponds to sampling from uniform distribution $\Rightarrow$ perfect privacy
- $\varepsilon = 2\Delta \log P(\theta, X)$ gives us samples from the posterior
- $\varepsilon \to \infty$ sample most likely θ (cap it at '=')
- For privacy budget $\varepsilon' \geq 2q\Delta \log P(\theta, X)$ with $q \in \mathbb{N}$, can draw q posterior samples within our budget

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

What can we say when working with the exponential family?

$$\text{Exp Family:} \quad P(x|\theta) = h(x)g(\theta)\exp(\theta^\top S(x))$$

$$\text{Conj Prior:} \quad P(\theta|\chi,\alpha) = f(\chi,\alpha)g(\theta)^\alpha \exp(\alpha\theta^\top \chi)$$

$$\text{Posterior:} \quad P(\theta|X,\chi,\alpha) \propto g(\theta)^{N+\alpha}\exp\left(\theta^\top\left(\sum_i S(x_i) + \alpha\chi\right)\right)$$

with a sensitivity of
$\Delta \log P(\theta, X) = \sup|\theta^\top(S(x') - S(x)) + \log h(x') - \log h(x)|$

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

But:

- Data interacts only through the sufficient statistic $S(X) = \sum_i S(x_i)$.

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

But:

- Data interacts only through the sufficient statistic $S(X) = \sum_i S(x_i)$.
- Use Laplace mechanism to get privacy instead:

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

But:

- Data interacts only through the sufficient statistic $S(X) = \sum_i S(x_i)$.
- Use Laplace mechanism to get privacy instead:
  - $\hat{S}(X) = \text{proj}(S(X) + (Y_1, ..., Y_N)$

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

But:

- Data interacts only through the sufficient statistic $S(X) = \sum_i S(x_i)$.
- Use Laplace mechanism to get privacy instead:
  - $\hat{S}(X) = \text{proj}(S(X) + (Y_1, ..., Y_N)$
  - $Y_j \sim \mathcal{L}ap(\Delta S(X)/\varepsilon)$

On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

But:

- Data interacts only through the sufficient statistic $S(X) = \sum_i S(x_i)$.
- Use Laplace mechanism to get privacy instead:
  - $\hat{S}(X) = \text{proj}(S(X) + (Y_1, ..., Y_N)$
  - $Y_j \sim \mathcal{L}ap(\Delta S(X)/\varepsilon)$
- where $\Delta S(X) = \sup_{x,x'} ||S(x') - S(x)||_1$

On the Theory and Practice of Privacy-Preserving BDA

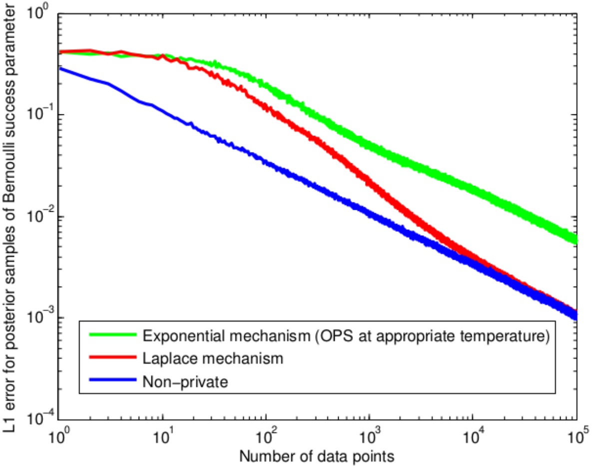
(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)

But:

- Data interacts only through the sufficient statistic $S(X) = \sum_i S(x_i)$.
- Use Laplace mechanism to get privacy instead:
  - $\hat{S}(X) = \text{proj}(S(X) + (Y_1, ..., Y_N)$
  - $Y_j \sim \mathcal{L}ap(\Delta S(X)/\varepsilon)$
- where $\Delta S(X) = \sup_{x,x'} ||S(x') - S(x)||_1$
- Example: beta posterior has $S(x) = [x, 1-x]$ giving us a sensitivity of 2

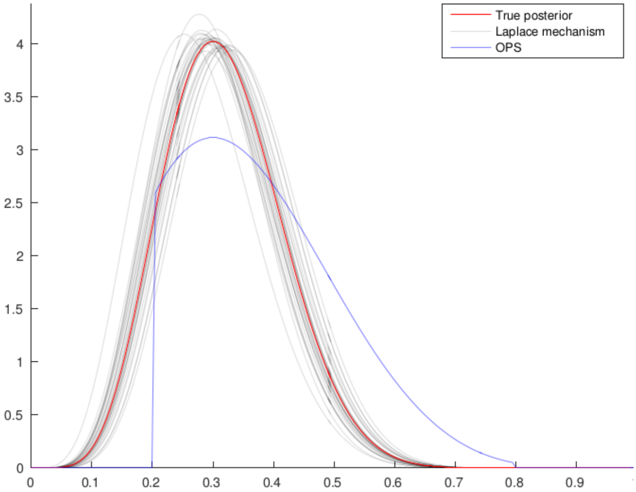
On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)



On the Theory and Practice of Privacy-Preserving BDA

(Foulds, Geumlek, Welling, Chaudhuri, UAI 2016)



Conclusion

- Still a lot of open questions (how to choose ε, how to get rid of the intermediary curator, better compositions for reducing privacy leakage, popular implementations ...)
- But a very fast growing field (given that term and definition stem from 2006.)
- Differential Privacy looks like a very promising way to conduct privacy preserving ML
- See, *No Free Lunch in Data Privacy* (Kifer and Machanvajjhala, 2011) for a critical discussion of DP
- Data Trusts?[6]

[6] http://inverseprobability.com/2016/05/29/data-trusts

Main Sources

- *"A Firm Foundation for Private Data Analysis", (Dwork, 2011)*
- *"Algorithmic Foundations of Differential Privacy"*, by Cynthia Dwork and Aaron Roth
- *"Differential Privacy and Learning: The Tools, The Results, and The Frontier"*, NIPS Tutorial, 2014 by Katrina Ligett