Differential Privacy

for Machine Learning

Dai

February 11, 2016

Outline

What is Privacy?

- "Publishing Data Won't harm you"
- Two ways to give privacy:
  - Erasure
  - Random Noise (Plausible Deniability)

Erasure Example: Netflix Prize

- Netflix Challenge provides anonymized user profiles for analytics
  - Profiles include Movie ratings, review dates
  - 1mil+ ratings, 480k users, 18k movies
- 96% of released profiles can be identified with at most 8 rating dates by matching to IMDb

Randomization Example: Randomized Responses

- For collecting sensitive data:
  1. Flip a coin
  2. If Tails, record truthfully
  3. If Heads, flip again and record "yes" if Heads, "no" if Tails
- Provides "Plausible Deniability"

The impossibility of Perfect Privacy

- Linkage Attacks: Correlated datasets link erased values
- Re-sampling: Correlated queries average noisy values
- Both methods provide *some* privacy, but how much? (Not enough)

Definition (Differential Privacy, Informally)

"Opting into a Differentially Private database won't harm you"

The Promise of Differential Privacy

- ✓ Limit the total *information*, of any kind, provided by a query
- ✓ Conceptually: Small changes to the database lead to small changes in expected output
- × NOT guaranteed: Public Linkage

Example (Public Linkage)

Alice smokes (publicly). A study is published linking smoking to cancer. Alice is not in the study, but her insurance goes up.

The Promise of Differential Privacy

- ✓ Limit the total *information*, of any kind, provided by a query
- ✓ Conceptually: Small changes to the database lead to small changes in expected output
- × NOT guaranteed: Public Linkage

Example (Public Linkage)

Alice smokes (publicly). A study is published linking smoking to cancer. Alice is not in the study, but her insurance goes up.

Participating in a study can be *no worse* than not participating

Differential Privacy

Given: Noisy query Q, Information ϵ, Probability δ

Definition (Q is $(\epsilon, \delta)-$Differentially Private)

$\forall$ Neighbor Datasets: $||x - x'||_1 \leq 1$
$\forall$ Selection Criteria: $S \subseteq Range(Q)$

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

Equivalently:

$$\left| \log \frac{P(Q(x) = a)}{P(Q(x') = a)} \right| \leq \epsilon \quad \text{with probability at least } 1 - \delta$$

Redux: Randomized Responses

Recall:

1. Flip a coin
2. Answer truthfully only if tails
3. Else flip again and answer Heads=T, Tails=F

Proposition: This mechanism has Has $(\ln 3, 0)-$ DP (1.6 bits)

Proof.

The max difference is between

$$P_{T|T} = P(\text{Response=T}|\text{Actual=T}) = 3/4$$

$$P_{T|F} = P(\text{Response=T}|\text{Actual=F}) = 1/4$$

so

$$log\frac{P_{T|T}}{P_{T|F}} = \log\frac{3/4}{1/4} = \log 3$$

□

General Masking: Laplace Noise

For deterministic q, add Laplace Noise independently to each component

Definition (L1-sensitivity)

$$\Delta_q = \max_{x,x'} ||q(x) - q(x')||_1$$

Definition (Laplace / "Symmetric Exponential" Distribution)

$$L^x_{\Delta q/\epsilon}(z) = \frac{\epsilon}{2\Delta_q} e^{\frac{-\epsilon}{\Delta_q}|q(x)-z|}$$

Not Too Wrong

For query q returning a $k-$vector, and $1 \geq d > 0$

$$P\left(\left|\left|q(x) - L^{x}_{\Delta q/\epsilon}\right|\right|_{\infty} \geq \log\left(\frac{k}{d}\right)\frac{\Delta_q}{\epsilon}\right) \leq d$$

Example

Count 10000 interesting google searches

- Each person can affect only one bin, so $\Delta_f = 1$
- For $\epsilon = 1$, with $d = 95\%$ probability, no count will be off by more than $\ln \frac{10000}{0.05} \approx 12.2$

✓ Independent of total population size!

× But bad for outlier analysis...

Useful $(\epsilon, 0)-$DP Algorithms

- Folds: Max, Count, Mean, etc.
- $K-$means
- $K-$medians
- Vertex Cover
- Empirical risk minimization
- More discovered frequently!

Compositions

- Proving DP for an algorithm is tedious

Compositions

- Proving DP for an algorithm is tedious
- Fortunately, DP algorithms compose cleanly!
- Nice theoretical structure guides good implementation

Query Chaining

Post-Processing (Functor Instance)

If Q is $(\epsilon, \delta)-$DP then
$\forall f : Range(Q) \to Range(Q)$, $f \circ Q$ is also $(\epsilon, \delta)-$DP

Post-Processing Immunity.

$$\log \frac{P(g(Q(x)) \in S) - \delta}{P(g(Q(x')) \in S)} = \log \frac{P(Q(x) \in g^{-1}(S)) - \delta}{P(Q(x') \in g^{-1}(S))} \leq \epsilon$$

□

Successive queries are additive (Applicative Instance)

If Q is $(\epsilon, \delta)-$DP and Q' is $(\epsilon', \delta')-$DP
then $Q \circ Q'$ is $(\epsilon' + \epsilon, \delta' + \delta)-$ DP

Pre-Processing

Definition (Group Privacy)

For databases differing in more entries at once: If Q is $(\epsilon, \delta)-$DP then:
For groups $||x - x'||_1 \leq k$, Q is $(k\epsilon, ke^{(k-1)\epsilon}\delta)-$*Group Private*

Definition ($k-$continuous)

$f : A \to B$ is $k-$continuous if $||f(x) - f(x')|| \leq k||x - x'||$

Pre-Processing (Contravariant Instance)

In particular, if Q is $(\epsilon, \delta)-$DP, and f is $k-$continuous, then

$$Q \circ f \text{ is } (k\epsilon, ke^{(k-1)\epsilon}\delta) - \text{DP}$$

Pre-Processing

Definition (Group Privacy)

For databases differing in more entries at once: If Q is $(\epsilon, \delta)-$DP then:
For groups $||x - x'||_1 \leq k$, Q is $(k\epsilon, ke^{(k-1)\epsilon}\delta)-$*Group Private*

Definition ($k-$continuous)

$f : A \to B$ is $k-$continuous if $||f(x) - f(x')|| \leq k||x - x'||$

Pre-Processing (Contravariant Instance)

In particular, if Q is $(\epsilon, \delta)-$DP, and f is $k-$continuous, then

$$Q \circ f \text{ is } (k\epsilon, ke^{(k-1)\epsilon}\delta) - \text{DP}$$

Only feasible for $(\epsilon, 0)-$DP queries!

Adaptive Chaining (Monad Instance)

- In reality, adversaries can choose queries based on past results
- Instead, ask what queries they *would make*, if the answer turned out a certain way
  - i.e. "Continuation Passing" form
  - Sew the queries together at the server end by correlating the noise

Naive adaptive composition of k $(\epsilon, \delta)-$DP queries

$$(\sqrt{2k \ln(1/\delta')}\epsilon + k\epsilon(e^{\epsilon} - 1), k\delta + \delta')\text{-DP}$$

But clever methods allow exponential (in DB size) chaining, by "preparing" queries

Adaptive Chaining (Monad Instance)

- In reality, adversaries can choose queries based on past results
- Instead, ask what queries they *would make*, if the answer turned out a certain way
  - i.e. "Continuation Passing" form
  - Sew the queries together at the server end by correlating the noise

Naive adaptive composition of k $(\epsilon, \delta)-$DP queries

$$(\sqrt{2k \ln(1/\delta')}\epsilon + k\epsilon(e^{\epsilon} - 1), k\delta + \delta')\text{-DP}$$

But clever methods allow exponential (in DB size) chaining, by "preparing" queries

Still under active research

The Problem of "P-Hacking"

- Over-fitting is a serious problem in machine learning
- Over-fitting is a serious Practical problem in biology

The Problem of "P-Hacking"

- Over-fitting is a serious problem in machine learning
- Over-fitting is a serious Practical problem in biology
- Typical confidence statistics assume data and analysis are uncorrelated (almost never true!)
- Conservative analysis requires partitioning precious test data

Privacy $\approx$ Stability

- "Stable" ML processes resistant to overfitting item
- "Differential Privacy" looks a lot like "Stability"
  - DP limits training-specific entropy leaks

"any adaptive analysis that is carried out in a differentially private manner must lead to a conclusion that generalizes to the underlying distribution"

"Thwarting" Adaptive Analysis

- Many Stable ML algorithms exist, but DP gives us *composition*
- Adaptive Privacy can be reused for Adaptive Analysis

Transfer Theorem

Definition (Informal)

The *Sample Complexity* of an analysis is the number of i.i.d. samples needed to "safely" perform the analysis

Major Result

For r rounds of adaptive analysis, totaling m queries in X, accurate to tolerance t, has Sample Complexity the minimum of

- $O\left(\frac{(\log m^{3/2})\sqrt{\log |X|}}{t^{7/2}}\right)$ (slow)
- $O\left(\frac{\log m \log |X|}{t^4}\right)$ (slow)
- $O\left(\frac{\sqrt{m}(\log m)^{3/2}}{t^{5/2}}\right)$ (fast)
- $O\left(\frac{r \log m}{t^2}\right)$ (fast)

For Further Reading


C. Dwork, A. Roth
The Algorithmic Foundations of Differential Privacy.
Foundations and Trends in Theoretical Computer Science, 2014.


C. Dwork et. al.
Preserving Statistical Validity in Adaptive Data Analysis.
STOC, 2015.


M. Gaboardi, A. Haeberlen, J. Hsu, A. Narayan, B. Pierce
Linear Dependent Types for Differential Privacy.
POPL, 2013

APPENDIX

Accounting For Utility

- Consider an auction:
  - Bidders confidentially provide their maximum bid
  - The final sale price is made public
  - Can't just add any noise, or it won't sell
- More generally, a map $u(DB, QueryResult) \rightarrow Utility$
  - When $u(x, z) = ||q(x) - z||$, recover the usual case.

$$\ln \frac{\exp(\epsilon u(x, r)/\Delta_u)}{\exp(\epsilon u(y, r)/\Delta_u} = \epsilon[u(x, r) - u(y, r)]/\Delta_u \leq \epsilon$$