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., 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 $w_j \neq 0$.
  - If there is only one person in town with green hair, you know they were in the study.

A preliminary story

- A classic result from theoretical crypto:
  - Say you want to figure out the average numeric grade of people in the room, without revealing anything about your own grade other than what is inherent in the answer.
  - Turns out you can actually do this. In fact, any function at all, "secure multiparty computation".
  - It's really cool. Want to try?
  - Anyone have to go to the bathroom?
  - What happens if we do it again?

Differential privacy "lets you go to the bathroom in peace"
**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:

$$
\frac{Pr(x_i \text{ output})}{Pr(x'_i \text{ output})} = \frac{Pr(\text{output} | x_i)}{Pr(\text{output} | x'_i)} \cdot \frac{Pr(x_i)}{Pr(x'_i)}
$$

(Posterior $\approx$ Prior)

Differential Privacy: Definition

- $A$ is $\epsilon$-differentially private if for any two neighbor datasets $S, S'$ (differ in just one element $x_i \rightarrow x'_i$),

  View as model of plausible deniability

  (pretend after the fact that my input was really $x'_i$)

$$
\epsilon \leq \frac{Pr(A(S)=v)}{Pr(A(S')=v)} \leq e^\epsilon
$$

Differential Privacy: Methods

- 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 $\leq b/n$.
- Idea: take answer and add noise from Laplace distribution $p(x) \propto e^{-|x|/\epsilon}/b$.
- Changing one input changes probability of any given answer by $\leq e^\epsilon$.

$$
\epsilon \leq \frac{Pr(A(S)=v)}{Pr(A(S')=v)} \leq e^\epsilon
$$

(Laplace Mechanism)

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

- Changing one input can affect average by $\leq b/n$.
- Idea: take answer and add noise from Laplace distribution $p(x) \propto e^{-|x|/\epsilon}/b$.
- Amount of noise added will be $\approx \pm b/(n\epsilon)$.
- To get an overall error of $\pm \gamma$, you need a sample size $n = \frac{b}{\gamma \epsilon}$.
- Get a utility/privacy/database-size tradeoff.
- If want to estimate mean of a distribution up to $\gamma$ and the database is an iid sample, then for $\gamma < \epsilon$ you can get privacy “for free”.
Laplace mechanism more generally

- E.g., \( f \) = standard deviation of income
- E.g., \( f \) = result of some fancy computation.

\[
\text{Global Sensitivity of } f: \quad GS_f = \max_{x, x'} |f(x) - f(x')|
\]
- Just add noise \( \text{Lap}(GS_f/\epsilon) \).

What can we do with this?

- Interface to ask questions
- Run learning algorithms by breaking down interaction into series of queries with noisy answers.
- \textbf{But}, each answer leaks some privacy:
  - If \( k \) questions and want total privacy loss of \( \epsilon \), better answer each with \( \epsilon/k \).

Can run SQ algorithms

- Anything learnable via Statistical Queries is learnable differentially privately using Laplace mechanism.

Statistical query model:

- Many algorithms can be re-written to interface via such statistical estimates.

Privately learnable = SQ-learnable?

- \textbf{[KLNS08]:} Actually, anything learnable is learnable in principle with DP.
  - Exponential mechanism for general classes
    - Assign score to each \( f \in \mathcal{C} \), exponentially decaying in its suboptimality.
    - Choose from this dist by \( q(x) \leq \gamma \).
  - Efficient algorithm for \( \mathcal{C} = \text{parity functions} \).
  - Interesting since not known to be efficiently learnable with noise, and provably not SQ-learnable.
  - SQ-learnable = learnable with local privacy, where no centralized database at all.

Local Sensitivity

- Consider \( f \) = median income
  - On some databases, \( f \) could be “very” sensitive. E.g., 3 people at salary=a, 3 people at salary=b, and you.
  - But on many databases, it’s not.
  - If \( f \) is not very sensitive on the actual input \( X \), does that mean we don’t need to add much noise?

\[
\text{Local Sensitivity} \quad LS_f = \max_{x \in X} |f(x) - f(x')|
\]
Local Sensitivity

- Consider $f = \text{median income}$.
- If $f$ is not very sensitive on the actual input $X$, does that mean we don't need to add much noise?
- Be careful: what if sensitivity itself is sensitive?

Smooth Sensitivity

- \cite{NRS07} prove can instead use (roughly) the following smooth bound instead:
  \[
  \max_Y \{ LS(Y)e^{-\Delta(X,Y)} \}
  \]

Smooth Sensitivity

- In principle, could apply sensitivity idea to any learning algorithm (say) that you'd like to run on your data.
- But might be hard to figure out.

Objective perturbation \cite{CMS08}

- Idea: add noise to the objective function used by the learning algorithm.
- Natural for algorithms like SVMs that have regularization term.
- \cite{CMS} show how to do this, if use a smooth loss function. Also show nice experimental results.

So far: learning as goal, privacy as constraint

Now: learning as tool for achieving stronger privacy
Idea: back to SQ's from class of small VC dim

• Fix a class $Q$ of statistical (i.e., counting/n) queries you care about (e.g., all 2-d marginals).
• VC-dimension bounds: whp a random subsample of size $O(\text{VCdim}(Q)/\alpha^2)$, will approximate all $q \in Q$ up to $\pm \alpha$.
• If $n \gg \text{VCdim}(Q)/\alpha^2$, this offers at least $(\epsilon, \delta)$ privacy. Maybe can invert?

With probability $\frac{1}{1-\epsilon}$, nothing is revealed about you, with prob $\epsilon$, everything is revealed about you. We want: with prob 1, very little is revealed about you.

[BLR08] building on [KLNRS08]: Use this with the "exponential mechanism": Explicit distrib over sets of size $m = O(\text{VCdim}(Q)/\alpha^2)$

Pr($S'$) $\propto e^{-\delta S' (\text{penalty}(S'))}$ penalty($S'$) $=$ maxgap$_S(Q)$

• Solve for $n$ s.t. bad $S'$ (penalty > $\alpha$) have prob $\ll 1/2^m$.
• $\epsilon \approx m = O(\text{VCdim}(Q)/\alpha^2)$

Get $n = O(d \cdot \text{VCdim}(Q)/(\alpha \cdot \epsilon))$ sufficient to whp output good sanitized db.