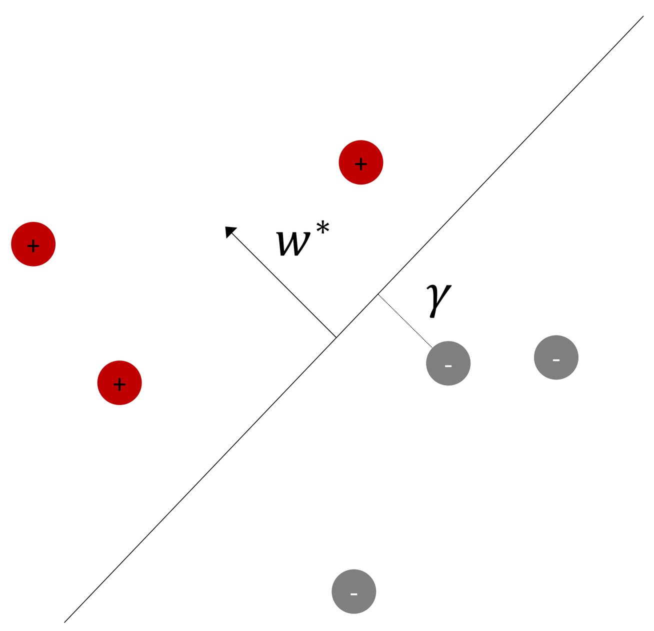
Efficient Private Algorithms for Learning Halfspaces

Huy Lê Nguyễn, Jonathan Ullman, Lydia Zakynthinou

Khoury College of Computer Sciences, Northeastern University

Learning large-margin halfspaces

- We are given a sample set of n unit vectors $x \in \mathbb{R}^d$ labelled with $y \in \{\pm 1\}$.
$$S = \{(x_1, y_1), \ldots, (x_n, y_n)\} \in \left(\mathbb{R}^d \times \{\pm 1\}\right)^n$$
- The samples are assumed to be drawn from a distribution D with **margin γ**, i.e., there exists a halfspace defined by the unit vector w^*, such that $y \cdot \langle w^*, x\rangle \geq \gamma > 0$ for all $(x, y) \sim D$.



- Goal: Design an (α, β, γ) –PAC learner: an algorithm that given a sample set $S \sim D^n$ drawn from any distribution D with margin γ outputs a classifier $\hat{w}$ such that with probability $1 - \beta$, has error at most α, that is,
$$\Pr_{(x,y)\sim D}[y \cdot \langle \hat{w}, x\rangle < 0] \leq \alpha.$$

Differential Privacy [DMNS06]

A randomized algorithm A is (ϵ, δ) – **differentially private** (DP) if for all neighboring datasets S, S' differing in one point, and for all measurable output sets O,
$$\Pr[A(S) \in O] \leq e^{\epsilon} \Pr[A(S') \in O] + \delta.$$

Can we design a differentially private (α, β, γ) – learner whose sample complexity does not depend on the dimension d?

Our results

- We present two differentially private (α, β, γ) –PAC learners that use $\tilde{O}(1/\alpha\epsilon\gamma^2)$ samples:
  - An (ϵ, δ) – DP algorithm that runs in polynomial time with respect to the dimension d and the rest of the parameters $\frac{1}{\alpha}, \frac{1}{\beta}, \frac{1}{\gamma}, \frac{1}{\delta}, \frac{1}{\epsilon}$.
  - An $(\epsilon, 0)$ –DP algorithm that runs in exponential time in $1/\gamma^2$.

	Sample Complexity	Time	Privacy
$A_{\alpha,\beta,\epsilon,\delta,\gamma}$	$\frac{1}{\alpha\epsilon\gamma^2} \cdot \text{polylog}\left(\frac{1}{\alpha\beta\epsilon\delta\gamma}\right)$	$\text{poly}\left(d, \frac{\ln(1/\beta\delta)}{\alpha\epsilon\gamma}\right)$	(ϵ, δ)
$A_{\alpha,\beta,\epsilon,\gamma}$	$\frac{1}{\alpha\epsilon\gamma^2} \cdot \text{polylog}\left(\frac{1}{\alpha\beta\epsilon\gamma}\right)$	$2^{\tilde{O}(1/\gamma^2)} \cdot \text{poly}\left(d, \frac{\ln(1/\beta\delta)}{\alpha\epsilon\gamma}\right)$	$(\epsilon, 0)$

- Lower Bound (via a packing argument): Any $(\epsilon, 0)$ – DP algorithm for learning a large-margin halfspace (with constant classification error α) requires $\Omega(1/\epsilon\gamma^2)$ samples.

Techniques

- Dimensionality Reduction: Pick a random matrix $A \in \mathbb{R}^{m \times d}$ and modify each sample $x \mapsto Ax/\|Ax\|_2$ to be in the reduced space of dimension $m = O(\ln(n/\alpha\beta)/\gamma^2)$. W.h.p., the new sample set still has margin 0.96γ.
  - The (ϵ, δ) – DP learner $A_{\alpha,\beta,\epsilon,\delta,\gamma}$ runs a differentially private ERM algorithm (e.g. the noisy stochastic gradient descent of [BST14]).
  - The $(\epsilon, 0)$ –DP learner $A_{\alpha,\beta,\epsilon,\gamma}$ runs the Exponential Mechanism over a $\gamma/10$ – Net of hypotheses.
- For $n = \tilde{O}(1/\alpha\epsilon\gamma^2)$, both algorithms return a hypothesis with empirical error at most $\alpha/4$, which extends via a generalization bound to true error at most α.

Conclusion

- Yes. There exist differentially private algorithms for learning a large-margin halfspace, with sample complexity $\tilde{O}(1/\alpha\epsilon\gamma^2)$, independent of the dimension d of the data.
- This is comparable to the sample complexity without privacy, which is $O(1/\alpha\gamma^2)$.
- For $(\epsilon, 0)$-DP, we prove that the dependence of the sample complexity on the margin and the privacy parameter is optimal.

We present **differentially private** algorithms for learning a **large-margin** halfspace, with sample complexity that depends only on the margin of the data, and not on the dimension.