Learning large-margin halfspaces

- We are given a sample set of \( n \) unit vectors \( x \in \mathbb{R}^d \) labelled with \( y \in \{ \pm 1 \} \).
  \[ S = \{ (x_1, y_1), \ldots, (x_n, y_n) \} \in (\mathbb{R}^d \times \{ \pm 1 \})^n \]

- The samples are assumed to be drawn from a distribution \( D \) with margin \( y \), i.e., there exists a halfspace defined by the unit vector \( w^* \), such that \( y \cdot (w^*, x) \geq y > 0 \) for all \( (x, y) \sim D \).

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

Differential Privacy [DMNS06]
A randomized algorithm \( A \) is \((\epsilon, \delta)\)–differentially private (DP) if for all neighboring datasets \( S, S' \) differing in one point, and for all measurable output sets \( \mathcal{O} \),
\[
\Pr[A(S) \in \mathcal{O}] \leq e^\epsilon \Pr[A(S') \in \mathcal{O}] + \delta.
\]

Our results

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

<table>
<thead>
<tr>
<th>Sample Complexity</th>
<th>Time</th>
<th>Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_{\alpha, \beta, \epsilon, \gamma} )</td>
<td>( \frac{1}{\alpha \beta^2} \cdot \text{polylog} \left( \frac{1}{\alpha \beta \epsilon \gamma} \right) )</td>
<td>( \text{poly} \left( d, \frac{\ln(1/\beta \delta)}{\alpha \gamma} \right) )</td>
</tr>
<tr>
<td>( A_{\alpha, \beta, \epsilon, \gamma} )</td>
<td>( \frac{1}{\alpha \beta^2} \cdot \text{polylog} \left( \frac{1}{\alpha \beta \epsilon \gamma} \right) )</td>
<td>( 2^{O(1/\gamma^2)} \cdot \text{poly} \left( d, \frac{\ln(1/\beta \delta)}{\alpha \gamma} \right) )</td>
</tr>
</tbody>
</table>

- Lower Bound (via a packing argument): Any \((\epsilon, 0)\)–DP algorithm for learning a large-margin halfspace (with constant classification error \( \alpha \)) 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 \( m = \Theta(\ln(\gamma/\alpha))/\beta^2 \). W.h.p., the new sample set still has margin \( 0.9 \gamma \).
- The \((\epsilon, \delta)\)–DP learner \( A_{\alpha, \beta, \epsilon, \gamma} \) 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 \( 1/10 \)–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 \( \alpha \).

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.

Can we design a differentially private \((\alpha, \beta, \gamma)\)– learner whose sample complexity does not depend on the dimension \( d \)?

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.