Locally private learning without interaction requires separation

Vitaly Feldman

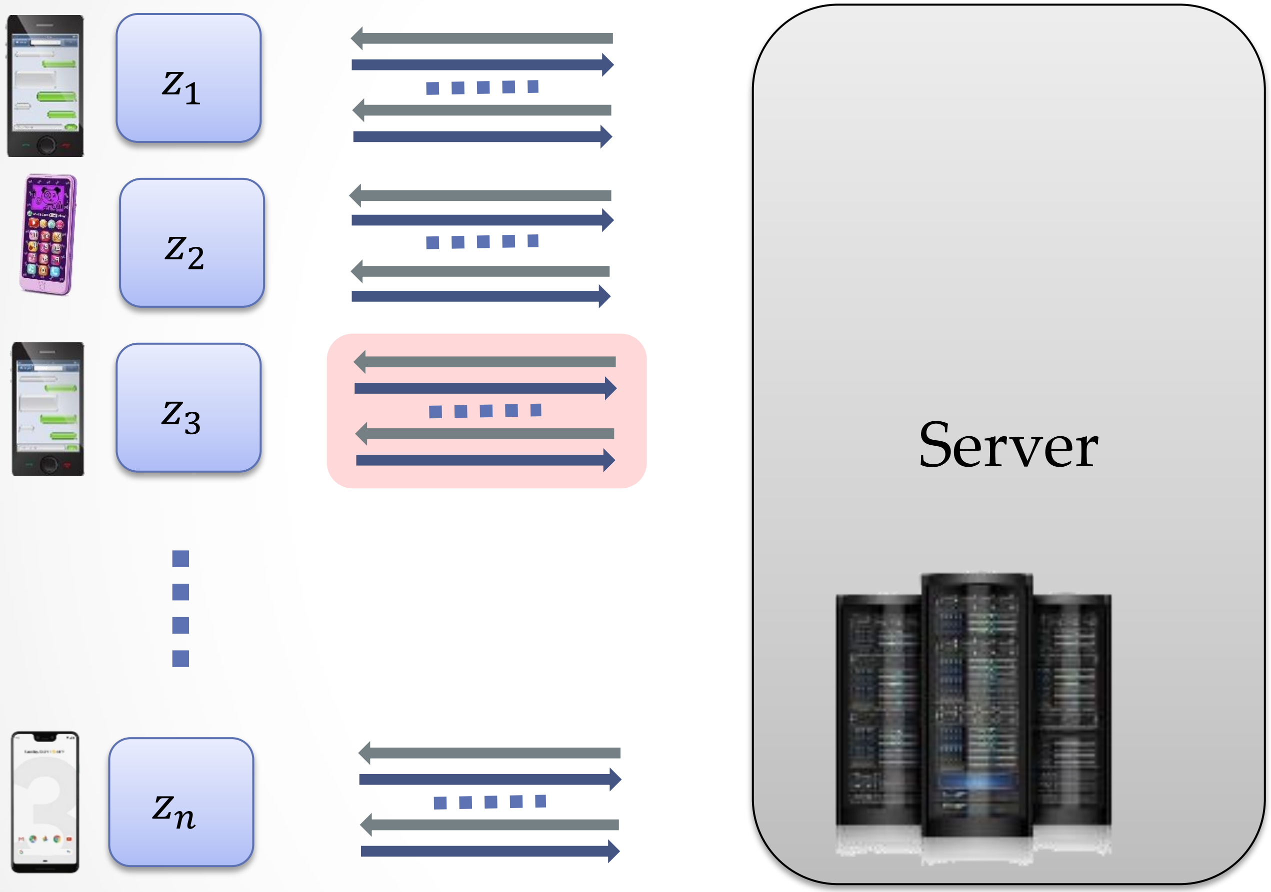
Google Research

with Amit Daniely

Hebrew University

Google

Local Differential Privacy (LDP)

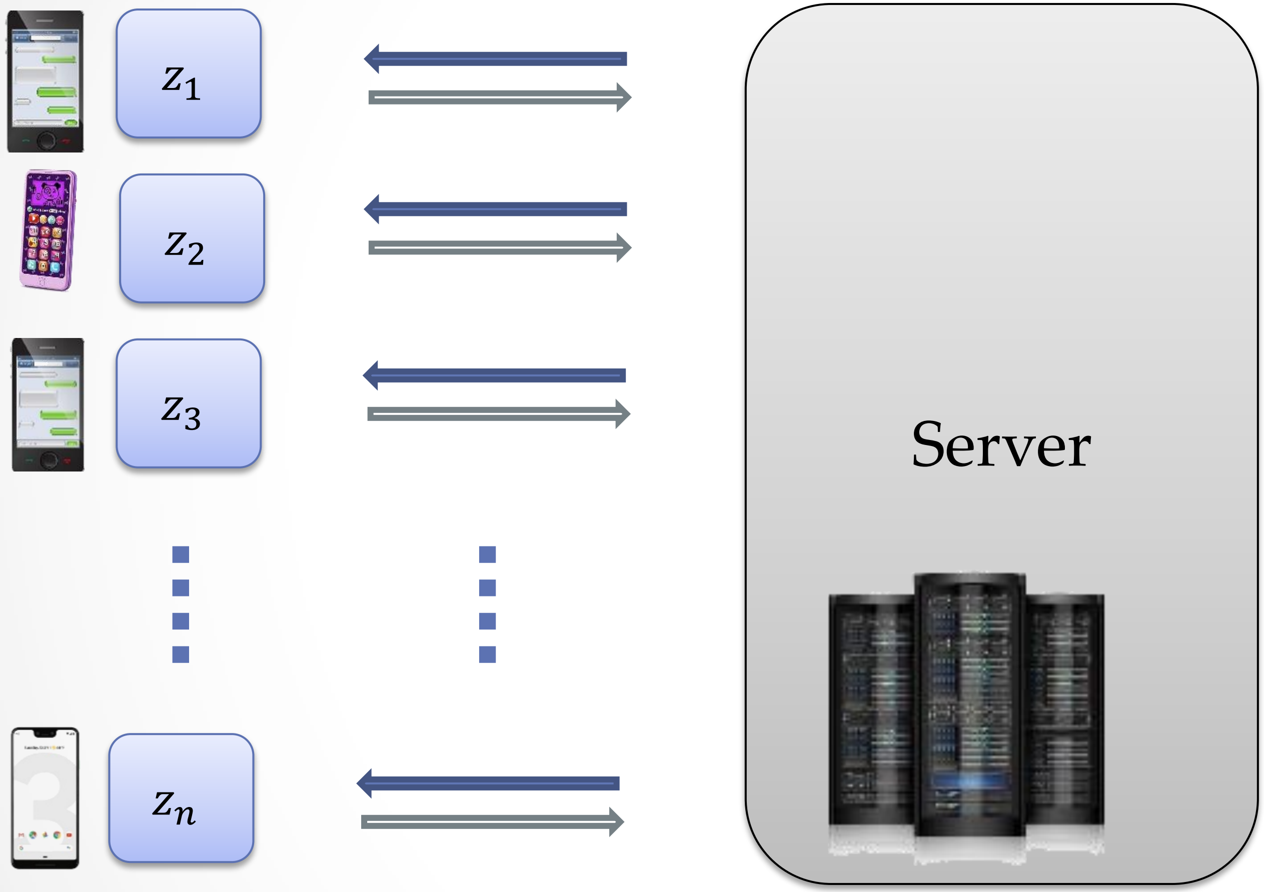


[KLNRS '08]

ϵ-LDP if for every user i, message j is sent using a local $\epsilon_{i,j}$-DP randomizer $A_{i,j}$ and

$$\sum_j \epsilon_{i,j} \leq \epsilon$$

Non-interactive LDP



PAC learning

PAC model [Valiant '84]: Let C be a set of binary classifiers over X
A is a PAC learning algorithm for C if $\forall f \in C$ and distribution D over X, given i.i.d. examples $(x_i, f(x_i))$ for $x_i \sim D$, A outputs h such that w.h.p.

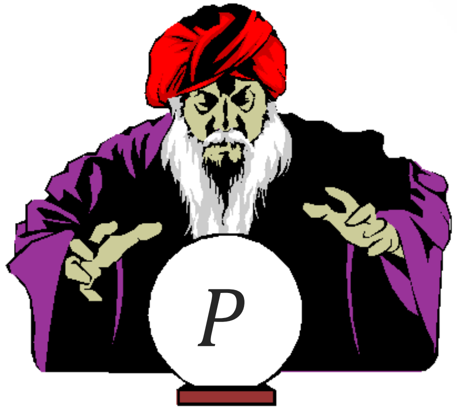
$$\Pr_{x \sim D}[h(x) \neq f(x)] \leq \alpha$$

Distribution-specific learning: D is fixed and known to A

Statistical query model [Kearns '93]

SQ algorithm

$$\xrightarrow{\phi_1}$$
$$\xleftarrow{v_1}$$
$$\xrightarrow{\phi_2}$$
$$\xleftarrow{v_2}$$
$$\vdots$$
$$\xrightarrow{\phi_q}$$
$$\xleftarrow{v_q}$$



SQ oracle

P distribution over Z

$$Z = X \times \{\pm 1\}$$

P is the distribution of $(x, f(x))$ for $x \sim D$

$$\phi_1 : Z \to [0,1] \quad |v_1 - \mathbf{E}_{z\sim P}[\phi_1(z)]| \leq \tau$$

τ is tolerance of the query; $\tau = 1/\sqrt{n}$

[KLNRS '08] Simulation with success prob. $1 - \beta \quad (\epsilon \leq 1)$

- ϵ-LDP with m messages $\Rightarrow$ $O(m)$ queries with $\tau = \Omega\left(\frac{\beta}{m}\right)$
- q queries with tolerance τ $\Rightarrow$ ϵ-LDP with $n = O\left(\frac{q \log(q/\beta)}{(\tau\epsilon)^2}\right)$ samples/messages

Non-interactive if and only if queries are non-adaptive

Known results

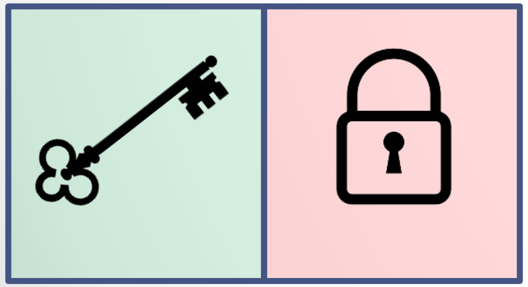
C is SQ-learnable efficiently (non-adaptively) if and only if learnable efficiently with ϵ-LDP (non-interactively)

Examples:

- Yes: halfspaces/linear classifiers [Dunagan,Vempala '04]
- No: parity functions [Kearns '93]
- Yes, non-adaptively: Boolean conjunctions

[KLNRS 08] There exists C that is

1. SQ/LDP-learnable efficiently over the uniform distribution on $\{0,1\}^d$ but
2. requires exponential num. of samples to learn non-interactively by an LDP algorithm

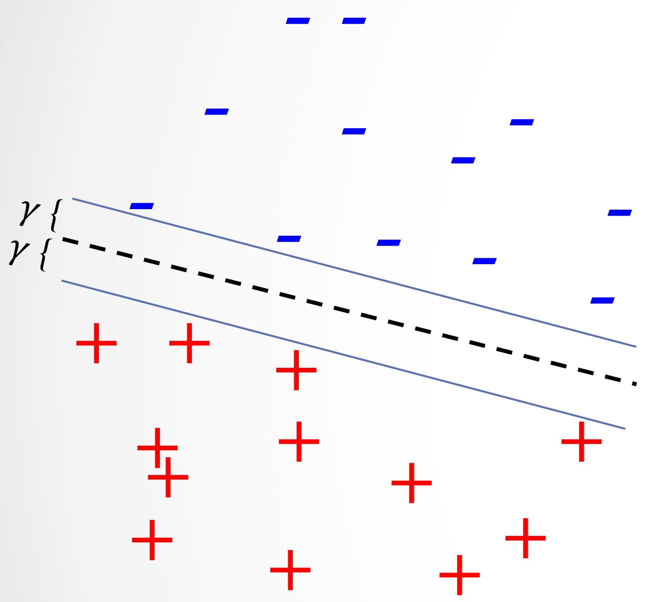


Masked parity

[KLNRS 08]:

Does separation hold for distribution-independent learning?

Margin Complexity



Margin complexity of C over X - $\mathbf{MC}(C)$:
smallest M such that exists an embedding $\Psi: X \to \mathbf{B}_d(1)$ under which every $f \in C$ is linearly separable with margin $\gamma \geq \frac{1}{M}$

Positive examples $\{\, \Psi(x) \mid f(x) = +1\}$
Negative examples $\{\, \Psi(x) \mid f(x) = -1\}$

Lower bound

Thm: Let C be a negation-closed set of classifiers.
Any non-interactive 1-LPD algorithm that learns C
with error $\alpha < 1/2$ and success probability $\Omega(1)$ needs

$$n = \Omega\left(\mathbf{MC}(C)^{2/3}\right)$$

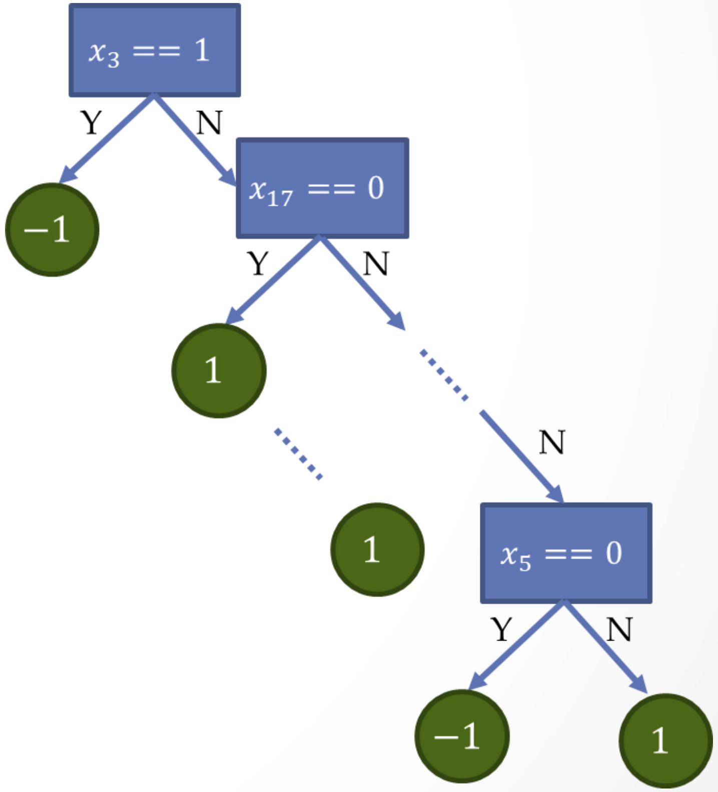
Corollaries:

- Decision lists over $\{0,1\}^d$: $n = 2^{\Omega(d^{1/3})}$
  [Buhrman,Vereshchagin,de Wolf '07]

(Interactively) learnable with $n = \text{poly}\left(\frac{d}{\alpha\epsilon}\right)$ [Kearns '93]

- Linear classifiers over $\{0,1\}^d$: $n = 2^{\Omega(d)}$
  [Goldmann,Hastad,Razborov '92; Sherstov '07]

(Interactively) learnable with $n = \text{poly}\left(\frac{d}{\alpha\epsilon}\right)$ [Dunagan,Vempala '04]



Upper bound

Thm: For any C and distribution D there exists a non-adaptive ϵ-LPD algorithm that learns C over D with error α and success probability $1-\beta$ using

$$n = \mathrm{poly}\left(\mathbf{MC}(C) \cdot \frac{\log(1/\beta)}{\alpha\epsilon}\right)$$

Instead of fixed D

- access to public unlabeled samples from D
- (interactive) LDP access to unlabeled samples from D

Lower bound holds against the hybrid model

Lower bound technique

Thm: Let C be a negation-closed set of classifiers.
If exists a non-adaptive SQ algorithm that uses q queries of tolerance $1/q$ to learn C with error $\alpha < 1/2$ and success probability $\Omega(1)$ then

$$\mathbf{MC}(C) = O\left(q^{3/2}\right)$$

Correlation dimension of C - $\mathrm{CSQdim}(C)$ [F. '08] : smallest t for which exist t functions $h_1, \ldots, h_t: X \to [-1,1]$ such that for every $f \in C$ and distribution D exists i such that

$$\left|\mathop{\mathbf{E}}_{x\sim D}[f(x)h_i(x)]\right| \geq \frac{1}{t}$$

Thm: [F. '08; Kallweit,Simon '11]:

$$\mathbf{MC}(C) \leq \mathrm{CSQdim}(C)^{3/2}$$

Proof

If exists a non-adaptive SQ algorithm A that uses q queries of tolerance $1/q$ to learn C with error $\alpha < 1/2$ then

$$\mathrm{CSQdim}(C) \le q$$

Let $\phi_1, \ldots, \phi_q : X \times \{\pm 1\} \to [0,1]$ be the (non-adaptive) queries of A

Decompose

$$\phi(x,y) = \underbrace{\frac{\phi(x,1) + \phi(x,-1)}{2}}_{g} + \underbrace{\frac{\phi(x,1) - \phi(x,-1)}{2}}_{h} \cdot y$$

$$\mathop{\mathbf{E}}_{x\sim D}[\phi_i(x, f(x))] = \mathop{\mathbf{E}}_{x\sim D}[g_i(x)] + \mathop{\mathbf{E}}_{x\sim D}[f(x) h_i(x)]$$

If $\left|\mathop{\mathbf{E}}_{x\sim D}[f(x)h_i(x)]\right| \le \frac{1}{q}$ then $\mathop{\mathbf{E}}_{x\sim D}[\phi_i(x, f(x))] \approx \mathop{\mathbf{E}}_{x\sim D}[\phi_i(x, -f(x))]$

If this holds for all $i \in [q]$, then the algorithm cannot distinguish between f and $-f$

Cannot achieve error $< 1/2$

Upper bound

Thm: For any C and distribution D there exists a non-adaptive ϵ-LPD algorithm that learns C over D with error $\alpha < 1/2$ and success probability $1-\beta$ using

$$n = \mathrm{poly}\left(\mathbf{MC}(C) \cdot \frac{\log(1/\beta)}{\alpha\epsilon}\right)$$

Margin complexity of C over X - $\mathbf{MC}(C)$:
smallest M such that exists an embedding $\Psi: X \to \mathbf{B}_d(1)$ under which every $f \in C$ is linearly separable with margin $\gamma \geq \frac{1}{M}$

Thm [Arriaga,Vempala '99; Ben-David,Eiron,Simon '02]:
For every every $f \in C$, random projection into $\mathbf{B}_d(1)$ for $d = O(\mathbf{MC}(C)^2 \log(1/\beta))$ ensures that with prob. $1-\beta$, $1-\beta$ fraction of points are linearly separable with margin $\gamma \geq \frac{1}{2\,\mathbf{MC}(C)}$

Algorithm

Perceptron: if $\mathrm{sign}(\langle w_t, x\rangle) \neq y$ then update $w_{t+1} \leftarrow w_t + yx$

Expected update: $\mathop{\mathbf{E}}_{(x,y)\sim P}[yx \mid \mathrm{sign}(\langle w_t, x\rangle) \neq y]$

$$\Vert$$

$$\mathop{\mathbf{E}}_{(x,y)\sim P}[yx \cdot \mathbb{1}(\mathrm{sign}(\langle w_t, x\rangle) \neq y)] \,/\, \boxed{\mathop{\mathbf{Pr}}_{(x,y)\sim P}[\mathrm{sign}(\langle w_t, x\rangle) \neq y]} \quad \text{scalar} \geq \alpha$$

$$\Vert$$

$$\mathop{\mathbf{E}}_{(x,y)\sim P}\left[x \cdot \frac{y - \mathrm{sign}(\langle w_t, x\rangle)}{2}\right] = \frac{\underbrace{\mathop{\mathbf{E}}_{(x,y)\sim P}[xy]}_{\text{non-adaptive}} + \underbrace{\mathop{\mathbf{E}}_{(x,y)\sim P}[x\,\mathrm{sign}(\langle w_t, x\rangle)]}_{\text{independent of the label}}}{2}$$

Estimate the mean vector with ℓ_2 error

- LDP [Duchi,Jordan,Wainright '13]
- SQs [F.,Guzman,Vempala '15]

Conclusions

- New approach to lower bounds for non-interactive LDP
  - Reduction to margin-complexity lower bounds
- Lower bounds for classical learning problems
- Same results for communication constrained protocols
  - Also equivalent to SQ
- Interaction is necessary for learning
- Open:
  - Distribution-independent learning in $\mathrm{poly}(\mathbf{MC}(C))$
  - Lower bounds against $2+$ round protocols
  - Stochastic convex optimization

https://arxiv.org/abs/1809.09165