LIST DECODABLE LEARNING
& MAX-ENTROPY PSEUDODISTNS.

Yet another application of the same method.

Certifiability $\rightarrow$ eff. algorithms

today's example: research direction currently active
~ 2-3 years old.

design algorithms : humongous fraction malicious
corruptions.
"robust machine learning"

<u>new conceptual ideas</u>:

① distribution view to come up with roundings
② use of max-entropy pseudo-distributions.
③ non-standard appearance of certificates..

<u>Problem:</u> List-decoding noisy linear equations

<u>Input:</u> $(x_1, y_1), \ldots, (x_{n'}, y_{n'})$, $\ell_* \in \mathbb{R}^d$ $\boxed{\mathcal{G}}$

$x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$, $\underline{\forall i}$: $y_i = \langle \ell_*, x_i \rangle$

Obs^n: If x_is are linearly indep then can find ℓ_* by Gaussian elimination!

Set $n' = \alpha n$.

$\boxed{0 \leq \alpha < 1/2}$

<u>Corr Input:</u> $\{(x_1, y_1), \ldots, (x_n, y_n)\} = S$

$S \supseteq$ "original uncorrupted set of size αn"

$\boxed{S = G + B}$, $|B| \sim (1-\alpha) \cdot n$.

<u>Goal:</u> ~~Find ℓ_*.~~ → ~~impossible~~

find a <u>list</u> $L = \{\ell_1, \ldots, \ell_k\}$, $|L| = k$

<u>Such that</u> $\ell_* \in L$. for $k \sim 1/\alpha$

<u>Thm:</u> We'll show a $\boxed{n^{O\left(\frac{1}{\alpha^2 \cdot \eta^2}\right)}}$ time algo to calculate a list of size $\sim O(\frac{1}{\alpha})$.

s.t. $\underline{\exists \hat{\ell} \in L}$ $\|\hat{\ell} - \ell_*\|_2 \leq \eta \cdot \|\ell_*\|_2$ whenever G satisfies some <u>niceness</u> condition..

$\boxed{\text{Focus on certificates.}}$

Prover: $\{\ell_1, \dots, \ell_k\} \subseteq \mathbb{R}^d$

$S_1, \dots, S_k$: "Soluble sets"

each $|S_i| = \alpha n$. and check that

$$\langle \ell_i, x_j \rangle = y_j \quad \forall \ (x_j, y_j) \in S_i$$

Question: how large does k need to be for this test to only accept lists that contain ℓ_*?

Answer: $\exp(\Omega(d))$ $\qquad \ell_* = \left(\frac{\pm 1}{\sqrt{d}}\right) \cdot \vec{1}$

Pf: $\underline{g}$: $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, $\dots$, e_d.

$\frac{1}{\sqrt{d}} \qquad \frac{1}{\sqrt{d}} \qquad \frac{1}{\sqrt{d}}$

$\mathcal{B}$:

$$\left(\frac{1}{\sqrt{d}}, \frac{-1}{\sqrt{d}}, \frac{1}{\sqrt{d}}, \frac{-1}{\sqrt{d}}, \ldots \right)$$

In this S: every boolean vector (i.e. coordinates in $\left\{\pm \frac{1}{\sqrt{d}}\right\}^d$) is a solution.

$\rightarrow$ Need $K \geq 2^d$.

$\frac{2}{\alpha}$ copies of e_1 $\quad +\frac{1}{\sqrt{d}}$, $-\frac{1}{\sqrt{d}}$

$\frac{2}{\alpha}$ copies of e_2

In this example $\mathcal{G}$ doesn't matter at all.

No matter what $\mathcal{G}$ is, $\mathcal{B}$ itself has αn-size soluble sets $\forall$ boolean vectors!

We must force the prover to give a soluble set that at least intersects the good part.

Idea: Force the prover to give us a Soluble partition of S.

$$\underset{\ell_1}{\underset{\downarrow}{S_1}} \,\dot\cup\, \overset{\ell_2}{\overset{\nearrow}{S_2}} \,\dot\cup\, \cdots \,\dot\cup\, \underset{\searrow \ell_k}{S_k} = S$$

$(S_i \cap S_j = \emptyset$ whenever $i \neq j)$.

$|S_i| \geq \alpha n.$

$\rightarrow k \leq \frac{1}{\alpha}$, $\{\ell_1, \dots, \ell_k\}$.

Obs^n: $\exists\, i : \ |S_i \cap \mathcal{G}| \geq \frac{1}{k} |\mathcal{G}|$

$= \alpha \cdot |\mathcal{G}|.$

Question: does this mean that $\ell_* \in \{\ell_1 \text{ — } \ell_k\}$?

Let's look at ℓ_i.

In the subset $S_i \cap \mathcal{G}$

both ℓ_i & ℓ_* label all equations

Correctly

What happens if $l_i \neq l_*$.

It means that $(x_j, y_j) \in S_i \cap G$.

$$\langle l_i, x_j \rangle = \langle l_*, x_j \rangle$$

$$\Rightarrow \langle \underline{l_i - l_*}, x_j \rangle = 0.$$

$\downarrow$

Conclusion: there's an α frac of G such that there's a non-zero vector orthogonal to all x_j in this α-frac.

Def. (Anti-concentration):

A $\mathbb{R}^d$-valued random var Y (mean 0) is said to be δ-anti-concentrated if

$$\underline{\forall v} \in \mathbb{R}^d, \underline{v \neq 0}, \Pr[\langle Y, v \rangle = 0] < \delta.$$

A subset $T \subseteq \mathbb{R}^d$ is said to be δ-anti-conc^n if Uniform dist^n on T is δ-anti-concentrated.

Obs^n: If $\mathcal{Y}$ is δ- anti-concentrated for $\delta = d$, then in the above discussion

$l_i = l_*$.

$\exists$ list with a soluble partition as certificate always contains l_*.

Anti-conc^n is well-studied & interesting.

(a) Y is $N(0, \mathbb{I})$, Y is: 0 - anti-concentrated

$Y = \{y_1 \text{———} y_n\} \sim N(0, \mathbb{I})$, $n \geq \boxed{\frac{100 d}{\delta}}$.

then Y is also - δ-anti-concentrated. (w.p. ≥ 0.99).

Hoeffding + Union bound argument.

Carbery-Wright: Polynomials of gaussians —

:

(b) Uniform dist^n on $\{0,1\}^n$ is not

... anti-concentrated (but is $\frac{1}{2}$-ρ anti-concent-

$\frac{1}{2}$-anti-c...

$\forall \beta > 0$)

$v = \boxed{e_i}$: $\Pr[\langle x, e_i \rangle = x_i = 0] = \frac{1}{2}$.

More generally, $U \sim \{0, 1, \ldots, q\}^n$

is not $\frac{1}{q}$-anti-concentrated..

<u>Littlewood-Offord</u> : $v : |v_i| \leq \left(\sqrt{\sum_i v_i^2}\right)\beta$

$\downarrow$

Erdős $\Pr[\langle x, v \rangle = 0] \leq O(\beta)$.

$\downarrow$

<u>Prop</u>: there exists a $\mathcal{G}$ where x_is are distributed uniformly on $\{0,1\}^n$ s.t. List-decodable noisy linear equation is impossible to solve on $\mathcal{G}$...

<u>Saw</u>: Soluble sets do not work

Soluble partitions are enough $\forall$ $\mathcal{G}$

that are α-anti-concentrated.

... need not exist.

but soluble partition,

Let's generalize soluble partition to get a certificate for good solutions that exists $\forall$ α-anti-concentrated $\mathcal{G}$.

Why did soluble partition work?

① "<u>Coverage</u>": every equation in the input is part of some soluble set.

② "<u>largeness</u>": each soluble set is of size $\geq \alpha n$.

<u>Idea</u>: View soluble partition as a distn μ over subsets of $[n]$ of size $\geq \alpha n$.

$$S_1 \uplus S_2 \uplus \ldots \uplus S_K = S$$

$\rightarrow$ uniform distn on $\{S_1 — S_K\}$

& each subset should be soluble.

how to phrase coverage in this language?

$$\Pr_{S_i \sim \mu}\left[(x_j, y_j) \in S_i \right] = \frac{1}{K} = \alpha$$

Equivalently, μ is a dist^n on $W \in \{0,1\}^n$

$\sum w_i = \alpha n$, $\forall i: \mathbb{E}_{\mu} w_i = \alpha$

(1)

Def: (Maximal Coverage distribution):

μ on W satisfy (1) is said to be "maximal coverage" if $\sum_i \left(\mathbb{E}_{\mu} w_i\right)^2$ is minimized $\approx \|\mathbb{E} w\|_2^2$

over all μ supported on soluble sets.

$\mathbb{E}_{\mu} \sum_i w_i = \alpha n$,

under no other constraints $\sum_i \left(\mathbb{E}_{\mu} w_i\right)^2$ is minimized when

$\mathbb{E}$ ...

$$\boxed{\mathbb{E}_{\mu} w_i = \alpha}$$

<u>Claim</u>: Suppose μ is a maximal coverage dist^n over soluble subsets of $[n)$ of size αn. <u>Then</u>,

$$\underline{\mathbb{E} \sum_{i \in \mathcal{G}} w_i} = \underline{\mathbb{E}_{S \sim \mu} |S \cap \mathcal{G}|} \geq \underline{\alpha \cdot |\mathcal{G}|}.$$

<u>Proof</u>: Idea:

$$wt(\mathcal{G}) = \frac{1}{\alpha n} \cdot \sum_{i \in \mathcal{G}} \mathbb{E} w_i$$

" expected frac of w that intersects with $\mathcal{G}$ "

Our goal: $wt(\mathcal{G}) \geq \alpha$.

<u>We'll show</u>: if $wt(\mathcal{G}) < \alpha$, then there's another dist^n μ' s.t. $\|\mathbb{E}_{\mu'} w\|_2^2 < \|\mathbb{E}_{\mu} w\|_2^2$

Construct μ': for some $\lambda \in (0,1)$.

Consider $\mu' = \underline{(1-\lambda)\mu + \lambda \cdot \mu_*}$. $\rightarrow$

$\boxed{\mu_*: \text{ output } \mathcal{G} \text{ w.p. } 1}$

w.p. $(1-\lambda)$ output a sample from μ.
w.p. λ. output a sample from μ_*.

Lemma:
Suppose prover gives you a maximal coverage distn on αn-size soluble sets.

Let $\underbrace{K = \frac{20}{\alpha - \delta}}$, Let $\mathcal{G}$ be δ-anti-concn for $\delta < \alpha$.

then let $L = \{\ell_1, \ldots, \ell_K\}$ be produced by sampling K iid soluble sets from μ. Then $\ell_* \in L$ w.p. ≥ 0.99.

Pf: Let S_1 —— S_K be the soluble sets sampled.

$$\mathbb{E}\, |S_i \cap \mathcal{G}| \geq \alpha n \quad \forall i$$

then: $\rightarrow \Pr\left[\frac{|S_i \cap \mathcal{G}|}{|\mathcal{G}|} \geq \frac{\alpha+\delta}{2}\right] \geq \frac{\alpha-\delta}{2}$.

$\downarrow$ $\qquad > \delta$

$\mathbb{E}\frac{|S_i \cap \mathcal{G}|}{|\mathcal{G}|} < \underbrace{\frac{\alpha-\delta}{2}}\cdot 1 + \left(1-\frac{\alpha-\delta}{2}\right)\cdot\frac{\alpha+\delta}{2} < \alpha$

$$\Pr\left[\exists i : |S_i \cap \mathcal{G}| \geq \frac{\alpha+\delta}{2}|\mathcal{G}|\right]$$

$S_1 \dots S_k$

$$\geq 1 - \left(1-\frac{\alpha-\delta}{2}\right)^{\frac{20}{\alpha-\delta}}$$

$$\sim 1 - e^{-10}$$

$$\geq \underline{0.99}$$

Say S_i is such a set:

l_i & l_* both satisfy all eqⁿs in $S_i \cap \mathcal{G}$

and since $\mathcal{G}$ is δ-anti-concⁿ

$\Rightarrow \quad l_i = l_*$. $\square$

<u>Conclusion</u>: We've ~~a~~ found a cert that allows us to get a good list of size $\sim O(\frac{1}{\alpha})$

whenever $\mathcal{G}$ is $\frac{\alpha}{2}$-anti-concⁿ.

Certifiability to algorithm (Really sketchy part)

<u>Natural algo:</u> Find a distⁿ μ on $\{0,1\}^n$ → max-coverage

s.t. μ's support satisfies:

$$\underbrace{w,\ell}\quad \mathcal{A}_{w,\ell}: \left\{ \begin{array}{l} w_i^2 = w_i \quad \forall i \\ \sum_i w_i = \alpha n \\ \forall i:\ w_i (y_i - \langle x_i, \ell \rangle) = 0 \\ \|\ell\|_2^2 \le 1 \end{array} \right.$$

if $w_i = 1$ then $y_i = \langle x_i, \ell \rangle$

→ Can "round" μ by "sampling $\sim O(\frac{1}{\alpha})$" times.

<u>Eff. version:</u>

Find a p.d. μ of deg t that minimizes $\|\tilde{\mathbb{E}}_\mu w\|_2^2$ & satisfies: $\mathcal{A}_{w,\ell}$

→ convex function of $\tilde{\mathbb{E}} w$

Q: Why is this efficiently computable? ($n^{O(t)}$ time) ✓

1) Analysis: 2) Rounding:

Our analysis earlier relied on anti-concentration of $\mathcal{G}$.

In order for us to do the same type of analysis within SoS $\rightarrow$ we need to phrase anti-concentration as a poly inequality on $\mathcal{G}$.

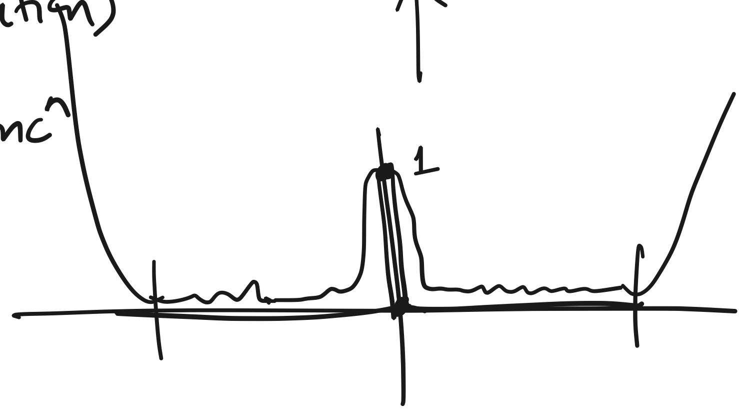
$\mathcal{G}$: δ-anticonc^n

$$= \forall v \in \mathbb{R}^d, \ \Pr[\langle x, v\rangle = 0] < \delta$$

$$= \forall v \in \mathbb{R}^d, \ \mathbb{E}[\underbrace{\mathbb{1}(\langle x, v\rangle = 0)}] < \delta$$

Def: (Certifiable Anti-concentration)

$\mathcal{G}$ is t-certifiably δ-anti-conc^n if there's polynomial p on $\mathbb{R}$



s.t. 1) $p^2(0) = 1$

$$\vdash_t^v \left\{ \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} p^2(\langle x_i, v\rangle) < \delta \right\}$$

Obs^n: If $\mathcal{G}$ is t-certifiably δ-anti-conc^n then $\mathcal{G}$ is also δ-anti-conc^n.

$$\Pr_{x_i \sim \mathcal{G}}[\langle x_i, v\rangle = 0] = \mathbb{E}_{x_i \sim \mathcal{G}}\Big[\underbrace{\mathbb{1}(\langle x_i, v\rangle = 0)}_{\leq\, p^2(\langle x_i, v\rangle)}\Big]$$

$$< \mathbb{E}_{x_i \sim \mathcal{G}}\left[p^2(\langle x_i, v\rangle)\right]$$

$$< \delta$$

Fact: Let $\mathcal{G}$ be a large enough iid sample from $N(0, \mathbb{I})$. Then $\mathcal{G}$ is $O(\frac{1}{\delta^2})$-certifiably δ-anti-conc^n.

is there a better upper bound?

* do the analysis with "certifiable anticonc^n"
* do it in a way that allows a rounding algorithm

Lemma: Suppose $\mathcal{G}$ is t-certifiably

δ-anti-concentrated. Then,

$$\mathcal{A}_{w,\ell} \vdash^{w,\ell}_{\cancel{t}} \left\{ \underbrace{\frac{1}{|\mathcal{G}|} \cdot \sum_{i \in \mathcal{G}} w_i \|\ell - \ell_*\|_2^2}_{} \le \frac{\alpha^2 \cdot \eta^2}{4} \right\} \quad *$$

Imagine that $\underline{\frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} w_i = 1}$

We have no guarantees on $\mathcal{B}$.

Let μ be a p.d. $\mathcal{A}_{w,\ell}$ of deg $\ge t$.

then, $\frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} \tilde{\mathbb{E}}_\mu \underbrace{w_i^2 \cdot \|\ell - \ell_*\|_2^2} \le \alpha^2 \cdot \eta^2/4$

$$\frac{1}{|\mathcal{G}|} \cdot \sum_{i \in \mathcal{G}} \tilde{\mathbb{E}}_{\underline{\mu}} \|\underline{w_i \ell - w_i \ell_*}\|_2^2 \le \_\_\_$$

Cauchy-Schwarz:

$$\frac{1}{|\mathcal{G}|} \cdot \sum_{i \in \mathcal{G}} \| \tilde{\mathbb{E}}\, w_i \ell - \underbrace{\tilde{\mathbb{E}}\, w_i}_{} \cdot \ell_* \|_2^2 \le \frac{\alpha^2 \cdot \eta^2}{4}$$

$$\frac{1}{|\mathcal{G}|} \cdot \sum_{i \in \mathcal{G}} \tilde{\mathbb{E}} w_i \cdot \left\| \frac{\tilde{\mathbb{E}} w_i l}{\tilde{\mathbb{E}} w_i} - l_* \right\|_2^2 \leq \frac{\alpha^2 \cdot \eta^2}{4}$$

Rounding: ① pick $O\left(\frac{1}{\alpha}\right)$ indices $i_1, i_2, \ldots, i_r$

② Output $\frac{\tilde{\mathbb{E}} w_{i_j} \cdot l}{\tilde{\mathbb{E}} w_{i_j}}$

With 99% prob: one of the above vectors is close to l_*.