Learning Parity Fctns

PAC Setting:

from which distribution?

Given samples $x, f(x)$

Find χ_s st. χ_s + f agree a lot ← large Fourier Coeffs.

Thought to be hard:

if x from arbitrary distribution then NP-hard
"Maximum likelihood decoding of linear codes"

if x from uniform dist. then ~~still~~ thought to be hard
"hardness of parity with noise"
"hardness of decoding linear codes"
used as hardness assumption eg. in Crypto

if noise random:

"hardness of decoding random linear codes"

"noisy parity"

[A. Blum Kalai Wasserman]: Can solve in $2^{O(n/\log n)}$
used to determine lattice vector of length,
Crytoanalysis
+ other learning problems

What if given query access to f for arbitrary inputs??

Learning Parities with Queries

$$x \to \boxed{f} \to f(x)$$

Given f, θ

1) Output all coeffs S s.t. $|\hat{f}(S)| \geq \theta$ (get all "close" fctns)
2) Only output coeffs S s.t. $|\hat{f}(S)| \geq \frac{\theta}{2}$ (no real junk)

(Using Boolean Parseval's: $\sum \hat{f}^2(S) = 1$ only $O(1/\theta^2)$ such coeffs)

<u>recall</u> $\Pr_x[f(x) = \chi_S(x)] = \frac{1}{2} + \frac{\hat{f}(S)}{2}$

so case 1 $\Rightarrow$ $\Pr_x[f(x) = \chi_S(x)] \geq \frac{1}{2} + \frac{\theta}{2}$

2 $\Rightarrow$ $\leq \frac{1}{2} + \frac{\theta}{4}$

<u>Warmup #0:</u>

poly queries
unbnded time } find <u>all</u> f that agree enough

<u>Warmup #1:</u> (poly queries, poly time) from now on

Suppose f agrees with χ_S everywhere for some S

(i.e. 0-error case)
only one S s.t. $\hat{\chi}_S \neq 0$

Algorithm 1: equation solving for coeffs

Algorithm 2: $\forall\, i \in [n]$ put i in S if $f(1111) \neq f(\underbrace{111\overset{i^{\text{th}}\text{ spot}}{(-1)}111}_{e_i})$

Outputs

Note
if $i \in S$
$\chi_S(u) \cdot \chi_S(u e_i) = -1$

Warmup #2 ($\exists s$ s.t. agreement with $\chi_S \approx 1$ + all other χ_T's is ≈ 0)

Suppose f agrees with χ_S "almost" everywhere ($\leq \frac{1}{\text{poly}(n)}$ - negligible fraction of inputs)
for some S

Note: Can't use previous algorithm since error might be on $(111\cdots1)$

Algorithm:

choose $r \in \{\pm1\}^n$

$\forall i \in [n]$

put i in S if $f(r) \neq f(r \odot e_i)$ — coordinatewise multiplication

Output S

Why? (sketch)
$f(r), f(r \odot e_i)$ agree with $\chi_S(r)\ \chi_S(r \cdot e_i)$ (unif dist) for almost all r

so $\Pr[S \text{ not correct}] \leq 2n \cdot \underbrace{\text{negligible}}_{\text{union bnd}}$

Warmup #3

Suppose f agrees with χ_S on $3/4 + \varepsilon$ for some S
$\varepsilon \geq \frac{1}{\text{poly}(n)}$

(here get better result on # solns than Boolean Parsevals: BP $\Rightarrow \leq 3$
but actually there is only unique soln.)

Algorithm:

Choose $r_1 \cdots r_t \in \{\pm1\}^n$

$\forall i \in [n]$

put i in S if $\underbrace{\text{majority}}_{t \text{ samples}}$ of $f(r_j) \neq f(r_j \odot e_i)$

Output S

(warmup 3 cont)

why?

$$\Pr[\text{"wrong" answer for } r_j \text{ on } i] \quad \text{("wrong"} \equiv (-1)^{1_{i\in S}})$$

$$= \Pr[f(r_j)\cdot f(r_j \oplus e_j)\cdot(-1)^{1_{i\in S}} \neq 1]$$

"right" should be different if $i \in S$, same if $i \notin S$

$$\leq \Pr[f(r_j) \neq \chi_S(r_j)] + \Pr[f(r_j \oplus e_j) \neq \chi_S(r_j \oplus e_j)]$$

uniformly distributed

$$\leq \left(\tfrac{1}{4}-\varepsilon\right) + \left(\tfrac{1}{4}-\varepsilon\right) = \tfrac{1}{2} - 2\varepsilon$$

Union bound on two bad events BUT we are doing union bound on same $f(r_j)$ event over & over & over!!!

$\therefore$ get correct answer with prob slightly $> 1/2$

$\therefore$ for i, most r_j are right with prob $> 1 - \delta/n$ — Chernoff: picking $t = \Theta(\log n)$

for all i, most r_j are right with prob $> 1-\delta$

Warmup 4

output all S s.t. f agrees with χ_S on $\geq \frac{1}{2}+\varepsilon$ fraction of inputs

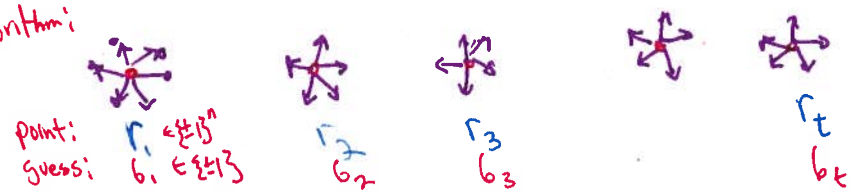
constant

idea 1 Guess answers to $f(r_j)$'s

Since only $O(\log n)$, can run over all possible guesses — saves half the union bound error!!!

Idea 2 Can test candidates & rule out junk

Picture of Algorithm: Pick $r_1 \ldots r_t$ uniformly & compare to all nbrs



Algorithm

- Choose $r_1 \cdots r_t \in \{\pm 1\}^n$ $\quad t = O(\log n)$
- For all possible settings of $\sigma_1 \cdots \sigma_t$

  {"guesses" to values of $\chi_S(r_i)$'s}

  - $\forall i \in [n]$ put i in $S_{\sigma_1 \cdots \sigma_t}$ if

    i.e. by testing if $f(r_j) \neq f(r_j \odot e_j)$ $\Updownarrow$ $\sigma_j \neq f(r_j \odot e_j)$ $\longrightarrow$ majority of $\sigma_j \neq f(r_j \odot e_i)$ (over $j \in [t]$) } generate a candidate for S

  - Sample to see if $\chi_{S_{\sigma_1 \cdots \sigma_t}}$ agrees with f on $\geq \frac{1}{2} + \frac{3}{8}\theta$ inputs

    if yes, output $\chi_{S_{\sigma_1 \cdots \sigma_t}}$ } test candidate & weed out junk

Note: many settings of $\sigma_1 \cdots \sigma_t$ could give good answer since could have lots of linear fctns agreeing with f on enough inputs

Why?

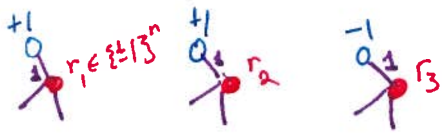
For each S that should be output

consider $\sigma_1 \cdots \sigma_t$ s.t. $\sigma_i = \chi_S(r_i)$

For this setting

(see next page)

Example of what happens with $i=1$ for all guesses of σ_i's:



σ_1	σ_2	σ_3	$f(r_1 \odot 1_{\{1\}}) = +1$	$f(r_2 \odot 1_{\{1\}}) = +1$	$f(r_3 \odot 1_{\{1\}}) = -1$	$1 \in S$?
+	+	+	+ vs. +1	+ vs +	+ vs −	no
+	+	−	+ vs +	+ vs +	− vs −	~~no~~
+	−	+	+ vs +	− vs +	+ vs −	yes
+	−	−	+ vs +	− vs +	− vs −	no
−	+	+	− vs +	+ vs +	+ vs −	~~no~~ yes
−	+	−	− vs +	+ vs +	− vs −	no
−	−	+	− vs +	− vs +	+ vs −	yes
−	−	−	− vs +	− vs +	− vs −	yes

- repeat this for $i = 2, 3, \ldots$
- gives a guess at S $\forall$ settings of σ_i's

For this setting:

$\Pr[\text{wrong answer for } r_j \text{ on } i]$

$= \Pr[\sigma_j \cdot f(r_j \odot e_i) \neq (-1)^{1_{i \in S}}]$

assumption $\Rightarrow \Vert$ $\chi_S(r_j)$ $\cdot$ $\overset{\Vert ?}{\chi_S(r_j \odot e_i) = (-1)^{1 \in S}}$ $\Leftarrow$ always, by def of $1 \in S$

$\leq \Pr[f(r_j \odot e_i) \neq \chi_S(r_j \odot e_i)]$

$\leq \frac{1}{2} - \varepsilon$

Chernoff bnds + $O(\log n)$ r_j's $\Rightarrow$ $\Pr[\text{majority gives wrong answer on } i] \leq 1/2n$

+ union bnd $\Rightarrow$ $\Pr[\text{wrong answer on \underline{any} } i] \leq 1/2$

$\therefore$ S is output with prob $\geq 1/2$

for each S that <u>should not</u> be output:

$\Pr[\text{output } S] \leq \Pr[S \text{ passes testing phase}]$

runtime:

since $t \approx \Theta(\log n)$, need $2^{\Theta(\log n)}$ iterations $\Rightarrow$ poly(n)

Before we start:

Outline

interesting part

- Generate a list $\mathcal{L}$ of candidates for S s.t. $|\hat{f}(S)| \geq \theta$

f agrees with $\chi_S(x)$ on $\geq \frac{1}{2} + \frac{\theta}{2}$ inputs x

– should contain all S s.t. $|\hat{f}(S)| \geq \theta$ } we need to prove this

– hopefully not too large i.e. not too many extra S's } will follow from construction

- Remove bad sets from $\mathcal{L}$ via sampling:

all that is going on here is basic sampling!

$\forall\ S \in \mathcal{L}$, estimate $\hat{f}(S)$ & remove if not $\geq \theta$ – constant

i.e.: f agrees with $\chi_S(x)$ $\geq \frac{1}{2} + \frac{3}{8}\theta$ fraction inputs x

$\Rightarrow$ "never" output coeffs S st. $|\hat{f}(S)| \leq \frac{\theta}{2}$

so we just need to generate $\mathcal{L}$ of reasonable size.

recall $\mathcal{L}$ doesn't need to be bigger than $\theta\left(\frac{1}{\theta^2}\right)$ via Boolean Parseval's

Learning Parity Functions

General Case

Output all S st f agrees with χ_S on $\geq \frac{1}{2} + \varepsilon$ fraction of inputs

$\uparrow$ can be $\frac{1}{poly(n)}$

Show that not too many such S

idea

in earlier warmup, if ε small ($\approx \frac{1}{poly(n)}$)

need more samples for Chernoff to kick in – i.e. if need $poly(n)$ samples

then need $2^{poly(n)}$ guesses!

Fix

Choose many more $r_1 \cdots r_t$ but not independently

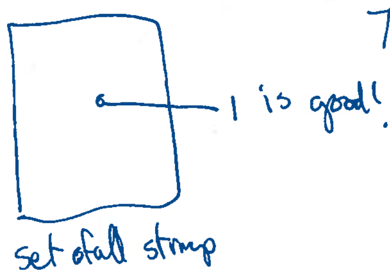
i.e. choose them pairwise independently

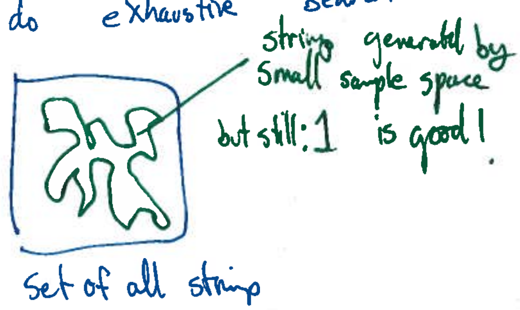
that is – find sample space of poly size

(i.e. $2^{\underbrace{O(\log n)}_{\#\text{p.i. bits needed}}}$)

which behaves in the same way as iid vars.

Then do exhaustive search on sample space!





Algorithm

- choose $u_1, .. u_K \in \{\pm 1\}^n$ $\quad K = \log(t+1)$ $\quad$ # guesses

  $t = \Theta(n/\varepsilon^2)$ $\quad$ # r_i's generated

  $\geq \frac{2n}{\varepsilon^2}$

- For all possible settings of

  $\sigma_1 .. \sigma_K \in \{\pm 1\}^K$: $\quad$ {all "guesses" for values of $\chi_S(u_i)$'s}

  {generate a lot $(2^K \approx n/\varepsilon^2)$ of labelled samples}

  - For every $W \subseteq \{1..K\}$ $\quad W \neq \emptyset$

    set $r_W \leftarrow \bigoplus_{j \in W} u_j$ $\quad$ ← pairwise random bits

    $\rho_W \leftarrow \prod_{j \in W} \sigma_j$ $\quad$ if initial guesses of q_i's "correct" according to χ_S then $\rho_W = \chi_S(r_W)$

  - $\forall i \in [n]$ put i in $S_{\sigma_1 \cdots \sigma_K}$ if $\quad$ ← creates $S_{\sigma_1 \cdots \sigma_K}$

    majority of $\rho_W \neq f(r_W \oplus e_i)$

  - Test $S_{\sigma_1 \cdots \sigma_K}$ to see if agrees enough with f

    ($\geq \frac{1}{2} + \frac{3}{4}\varepsilon$ fraction)

    if yes, output it

Behavior

For s' s.t. f agrees with $\chi_{s'}$ on $\geq \frac{1}{2}+\varepsilon$ of inputs:

1) if setting of σ_i's agrees with χ_s

i.e. $\forall i \quad \sigma_i = \chi_s(u_i)$

then $\forall w \quad p_w = \prod_{j \in w} \chi_{s'}(u_j)$ def of p_w

$= \chi_{s'}\left(\bigoplus_{j \in w} u_j\right)$

$= \chi_{s'}(r_w)$ def of r_w

so all p_w's are consistent with σ

From now on, assume this setting of σ_i's ...

2) r_w's are pairwise independent [in fact, generated via a known construction]

i.e. $\Pr[r_w = b_1 \; \& \; r_{w'} = b_2] = \Pr[r_w = b_1] \cdot \Pr[r_{w'} = b_2]$

also $r_w \odot e_i$'s are p.i.

3) $\Pr[$ Algorithm generate s' when considering $S_{\sigma_1 \dots \sigma_k}]$:

$\Pr[$ it get s' right on index $i]$

$= \Pr[\underbrace{p_w \cdot f(r_w \odot e_i) = (-1)^{1_{i \in s}}}_{\text{indicator } X_w = \begin{cases} 1 & \text{if holds} \\ 0 & \text{o.w.} \end{cases}}]$

Note: if $f(r_w \odot e_i) = \chi_s(r_w \odot e_i) \leftarrow$??

& $p_w = \chi_s(r_w) \leftarrow$ assumption

then $X_w = 1$ so $E[X_w] = \Pr[f(r_w \odot e_i) = \chi_s(r_w \odot e_i)]$ $\Uparrow$ unif dist

$\geq 1/2 + \varepsilon$

$E[X_w] \geq \frac{1}{2} + \varepsilon$ since $r_w \oplus e_i$ uniform dist

Variance $\sigma_w^2 = E[X_w^2] - E[X_w]^2$

$\geq \frac{1}{2} + \varepsilon - (\frac{1}{2} + \varepsilon)^2 = \frac{1}{4} - \varepsilon^2$

$E[\sum_{w \in [K]} X_w] \geq t(\frac{1}{2} + \varepsilon)$

$\underbrace{\Pr[\sum_w X_w < \frac{t}{2}]}_{\leq \Pr\left[\left|\frac{\sum_w X_w}{t} - \frac{1}{2}\right| \geq \varepsilon\right]} \leq \frac{(\frac{1}{2})^2 - \varepsilon^2}{t\ \varepsilon^2} \leq \frac{1}{t\varepsilon^2} \leq \frac{1}{2n}$

union bnd: $\Pr[\$ \text{ not output}] \leq \frac{1}{2}$

Chebyshev:
$X_1 \cdot X_N$ p.i.
$E[X_i] = \mu$
$Var[X_i] = \sigma^2$
$\Pr\left[\left|\frac{\sum X_i}{N} - \mu\right| > \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2 N}$

Also shows:

parity fctns agreeing with f on $\geq \frac{1}{2} + \varepsilon$ is $O(\frac{n}{\varepsilon^2})$