Lecture 15

Learning parity fctns

Without "noise": given samples of $x, f(x) \Rightarrow$ equation solving.
With "noise": find closest parity fctn $\Leftrightarrow$ find largest Fourier coeff
find all close parity fctns $\Leftrightarrow$ find all large enough Fourier coeffs (not necessarily low degree)

NP hard –

(worst case) maximum likelihood decoding of linear codes

i.e. given I/O examples of fctn,
find largest Fourier coeff

Thought to be hard

(uniform dist)

· Hardness of parity with noise

i.e. given
$$\left.\begin{matrix} x_1^1 \cdots x_n^1 & b^1 \\ \vdots & \\ x_1^k \cdots x_n^k & b^k \end{matrix}\right\} x_i\text{'s uniform}$$
find largest Fourier coeff

Hardness of decoding linear codes
Find large Fourier coeffs

Easier model?

Assume noise is $\boxed{\text{random}}$ i.e. flip η biased coin
+ flip output if coin = H

Hardness of decoding random linear codes
Noisy parity problem

Used as assumption in crypto / learning theory!!

Note A. Blum, Kalai, Wasserman:
Slightly subexponential algorithm exists (for random noise) $2^{O(n/\log n)}$ (instead of 2^n) } used to determining shortest lattice vector + length

Learning Parities with Queries

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

Given f, θ

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

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

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

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

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

Warmup #0:

$\left.\begin{array}{l}\text{poly queries} \\ \text{unbnded time}\end{array}\right\}$ find all f ~~that~~ agree enough

Warmup #1: (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 $f(1 \cdots 1) \neq f(\underbrace{1 \cdots 1 \overset{i\text{th spot}}{(-1)} 1 \cdots 1}_{e_i})$

Output S

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

<u>Warmup #2</u> $\left(\exists\, S \text{ s.t. } \underset{\text{agreement with}}{\chi_S \approx 1} \text{ + all other } \chi_T\text{'s is} \approx 0\right)$

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

Note: Can't use previous algorithm since error might be on $(1111\cdots1)$

<u>Algorithm</u>:

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

$\forall\, i \in [n]$

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

$\uparrow$ coordinatewise multiplication

Output S

Why? (sketch)

$f(r), f(r \odot e_i)$ agree with $\chi_S(r)\ \chi_S(r \odot e_i)$ for almost all r

unif dist

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

<u>Warmup #3</u>

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 Parseval's: BP $\Rightarrow \leq 3$ but actually there is only unique soln.

<u>Algorithm</u>:

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?

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

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

"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

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

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

$\therefore$ for i, most r_j are right with prob $> 1 - \delta/n$ — Chernoff: picking $t = \Theta\left(\frac{\log n}{\varepsilon^2}\right)$

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

<u>Warmup ~~4~~</u>

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

constant

<u>idea</u> guess answers to $f(r_j)$'s
Since only $O(\log n)$, can run over all possible guesses

Algorithm

- Choose $r_1 \cdots r_t \in \{\pm 1\}^n$ $\quad t = O(\log n)$
- For all possible settings of $b_1 \cdots b_t$

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

  - $\forall\, i \in [n]$ put i in $S_{b_1 \cdots b_t}$ if

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

  - Sample to see if $\chi_{S_{b_1 \cdots b_t}}$ agrees with f on $\geq \frac{1}{2} + \frac{3}{8}\theta$ inputs

    if yes, output $\chi_{S_{b_1 \cdots b_t}}$ } test candidate & weed out junk

Note: many settings of $b_1 \cdots b_t$ 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 $b_1 \cdots b_t$ s.t. $b_i = \chi_S(r_i)$

For this setting

(see next page)

For this setting:

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

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

assumption $\Rightarrow \Vert$ $\quad \Vert ?$

$$\chi_S(r_j) \cdot \chi_S(r_j \odot e_i) \cdot (-1)^{1 \in S} = -1$$

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

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

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

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

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

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

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

Learning Parity Functions

General Case

Output all S st f agrees with χ_S on

$\geq \frac{1}{2} + \varepsilon$ fraction of inputs

↑ can be $\frac{1}{\text{poly}(n)}$

Show that not too many such S

idea

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

need more samples for Chernoff to

Kick in – i.e. if need poly(n) samples

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

Fix

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

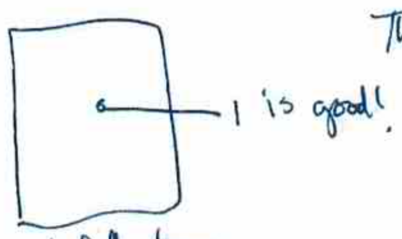
i.e. choose them pairwise independently

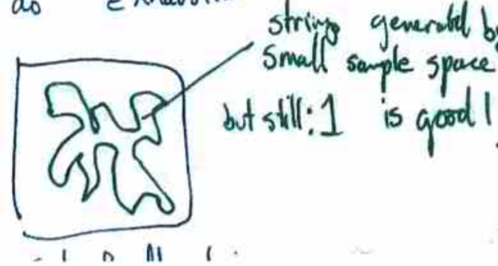
that is – find sample space of poly size

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

which behaves in the same way as iid vars.

Then do exhaustive search on sample space!





Algorithm

- choose $s_1 \ldots s_k \in \{\pm 1\}^n$ $\quad k = \log(t+1)$ $\quad$ # guesses

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

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

- For all possible settings of

$\sigma_1 \ldots \sigma_k \in \{\pm 1\}^k$: $\quad$ { all "guesses" for values of $\chi_S(s_i)$'s }

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

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

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

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

- $\forall i \in [n]$ put i in $S_{\sigma_1 \ldots \sigma_k}$ if majority of $\rho_W \neq f(r_W \oplus e_i)$ $\quad$ ← creates $S_{\sigma_1 \ldots \sigma_k}$

- Test $S_{\sigma_1 \ldots \sigma_k}$ to see if agrees enough with f

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

if yes, output it

Behavior

For $\tilde{s}$ s.t. f agrees with $\chi_{\tilde{s}}$ on $\geq \frac{1}{2}+\varepsilon$ of inputs:

1) if setting of σ_i's agrees with $\chi_{\tilde{s}}$

i.e. $\forall i \quad \sigma_i = \chi_S(s_i)$

then $\forall w$

$$\left.\begin{aligned} p_w &= \prod_{j\in w} \chi_{\tilde{s}}(s_j) && \text{def of } p_w \\ &= \chi_{\tilde{s}}\Big(\bigoplus_{j\in w} s_j\Big) \\ &= \chi_{\tilde{s}}(r_w) && \text{def of } r_w \end{aligned}\right\} \text{so all } p_w\text{'s are consistent with } \sigma$$

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

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

3) $Pr[$ Algorithm generates $\tilde{s}$ when considering $S_{\sigma_1 \cdots \sigma_k}]$:

$Pr[$it get $\tilde{s}$ right on index $i]$

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

Note: if $f(r_w \odot e_i) = \chi_{\tilde{s}}(r_w \odot e_i) \leftarrow$??

$\quad$ & $p_w = \chi_{\tilde{s}}(r_w) \leftarrow$ assumption

then $X_w = 1$

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

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

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

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

$$\Pr\Big[\sum_w X_w < \frac{t}{2}\Big] \leq \frac{\left(\frac{1}{2}\right)^2 - \varepsilon^2}{t\,\varepsilon^2} \leq \frac{1}{t\varepsilon^2} \leq \frac{1}{2n}$$

union bnd: $\Pr[\, S$ 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 $O\left(\frac{1}{\varepsilon^2}\right)$