Lecture 15

Learning parity fits

Without "noise": given samples of \( x \in \mathbb{F}_2^n \), equation solving.
With "noise": find closest parity fit \( \Rightarrow \) find largest Fourier coeff.

NP-hard - find all close parity fits \( \Rightarrow \) find all large enough Fourier coeffs (not necessarily low degree).

(worst case) maximum likelihood decoding of linear codes

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

Thought to be hard

Uniform dist.

Hardness of parity with noise

i.e., given \( x_1, \ldots, x_k \), \( b \) \( \sim \) \( x_i \)'s uniform

\( x_1, \ldots, x_k, b_k \)

find largest Fourier coeff.

Hardness of decoding linear codes

Find large Fourier coeff.

Easter model?

Assume noise is \( \text{random} \)

i.e., flip \( n \) biased coin

+ flip output if coin = \( \text{?} \)

Hardness of decoding random linear codes

Noisy parity problem

Used as assumption in cryptography

Note: A. Blum, Kalai, Wasserman:

Slightly subexponential algorithm exists (for random noise) \( \Rightarrow \) used to determine shortest nonzero vector \( \bullet \) length.
Learning Parities with Queries

Given $f, \theta$

1) Output all coeffs $S$ s.t. $|\hat{f}(s)| \geq \theta$ (get all "close" coeffs)

2) Only output coeffs $S$ s.t. $|\hat{f}(s)| = \frac{\theta}{2}$ (no real junk)

Using Boolean Parseval: $E[\hat{f}^2] = 1$

only $O(1/\theta^2)$ such coeffs

recall

$$Pr_x[f(x) = X_b(x)] = \frac{1}{2} + \frac{\hat{f}(s)}{2}$$

so case 1 $\implies$ $Pr_x[f(x) = X_b(x)] = \frac{1}{2} + \frac{\theta}{2}$

2 $\implies$

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

Warmup #0:

poly queries $\exists$ find all $f$ that agree enough

unbounded time

Warmup #1: (from now on, poly queries, poly time)

Suppose $f$ agrees with $X_s$ everywhere for some $s$

(i.e., 0-error case)

only one $s$ s.t. $X_s \to 0$

Algorithm 1: equation solving for coeffs

Algorithm 2: $i \in [n]$

put $i$ in $S$ if $f(\mathbf{1}i) \neq f(\mathbf{1}i \mathbf{0}e)$

Output $S$
Warmup #3

(\exists s \text{ s.t. } \chi_s \approx 1 + \text{ all other } \chi_s \text{ is } \leq \frac{1}{20})

Suppose \( f \) agrees with \( \chi_s \) "almost" everywhere

for some \( s \)

Note: cant use previous algorithm since error might be on \( \{1111\ldots\} \)

Algorithm:
choose \( r \in \mathbb{Z}_{2^n}^n \)

\[ \forall i \in [n] \]

put \( i \) in \( S \) if \( f(r) \neq f(r \cdot \chi_s) \)

Output \( S \)

Why? (sketch)

\( f(r), f(r \cdot \chi_s) \) agree with \( \chi_s(r), \chi_s(r \cdot \chi_s) \) for

almost all \( r \)

so \( \Pr [S \text{ not correct}] \leq 2n \cdot \text{negligible} \)

union bound

Warmup #4

Suppose \( f \) agrees with \( \chi_s \) on \( \frac{3}{4} + \epsilon \) for some \( s \)

Algorithm:
choose \( r \in \mathbb{Z}_{2^n}^n \)

\[ \forall i \in [n] \]

put \( i \) in \( S \) if \( \text{majority of } f(r) \neq f(r \cdot \chi_s) \)

Output \( S \)

(here get better result than Boolean

PSPACE \( \leq \text{BPP} \))
(warmup \# cont)

why?

\[ \Pr \left[ \text{"wrong" answer for } r_i \text{ on } i \right] \]
\[ = \Pr \left[ f(r_i) \neq f(r_i \cdot g) \cdot (-1)^{\text{t-ies} + 1} \right] \]
\[ \uparrow \]
\[ \text{"right" should be different if t-ies same } \]
\[ \leq \Pr \left[ f(r_i) \neq X_s(r_i) \right] + \Pr \left[ f(r_i \cdot g) \neq X_s(r_i \cdot g) \right] \]
\[ \leq \left( \frac{1}{t} - \varepsilon \right) + \left( \frac{1}{t} - \varepsilon \right) = \frac{1}{2} - 2\varepsilon \]
\[ \text{uniformly distributed} \]

Chernoff:

get correct answer with prob \( \geq \frac{1}{2} \) /n

\[ \text{picking } t = \Theta \left( \frac{\log n}{\varepsilon^2} \right) \]

\[ \text{for all } i, \text{ most } r_i \text{ are right with prob } \geq 1 - 8 \]

Warmup 4

output all \( S \) st. \( f \) agrees with \( X_s \) on \( \geq \frac{1}{2} + \varepsilon \) fraction of inputs

\[ \uparrow \]

constant

idea: guess answers to \( f(r_i)'s \)

Since only \( O(\log n) \), can run over all possible guesses
Algorithm

* Choose $r_1, r_k \in \mathbb{Z}_2^{n}$, $t = O(\log n)$

* For all possible settings of $b_1, b_k$
  
  3 guesses to values of $X_S(r_i)$'s

* $A \in [n]$ put $i$ in $S_{b_1, b_k}$ if

  by testing if $A(r_j) \neq f(r_j \oplus e_i) \rightarrow$ majority of $b_j \neq f(r_j \oplus e_i)$
  
  (over $j \in S$)

  Sample to see if $X_{S_{b_1, b_k}}$ agrees
  
  with $f$ on $2 \frac{1}{2} + \frac{3}{8} \theta$ inputs

  if yes, output $X_{S_{b_1, b_k}}$

  test candidate + weed out junk

Note: many settings of $b_1, b_k$ could give good answer since could have lots of linear fits agreeing with $f$ on enough inputs

Why?

for each $S$ that should be output

consider $b_1, b_k$ st. $b_i = X_S(r_i)$

For this setting

(see next page)
For this setting:

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

\[ = P_{\text{wrong answer on } i} \cdot f(r_j, Oe_i) \cdot (H_{i=0}^{i=1} = -1) \]

assumption \( \Rightarrow \)

\[ x_s(r_j) \cdot x_s(r_j, Oe_i) \cdot (H_{i=0}^{i=1} = -1) \]

\[ \leq P_r[ f(r_j, Oe_i) \neq x_s(r_j, Oe_i) ] \]

\[ \leq \frac{1}{2} - \epsilon \]

Chernoff bounds + \( O(\log n) \) 's \( r_j \) 's \( \Rightarrow \) \( P_{\text{wrong answer on } i} \leq \frac{1}{2} \n \)

+ union bound \( \Rightarrow \) \( P_{\text{wrong answer on any } i} \leq \frac{1}{2} \)

\( \therefore S \) is output with prob \( \geq \frac{1}{2} \)

for each \( S \) that should not be output:

\[ P_{\text{output } S} \leq P_{\text{S passed testing phase}} \]
Learning Parity Functions

General Case

Output all $S$ at $f$ agrees with $X_5$ on

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

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

Show that not too many such $S$

Idea

In earlier warmup, if $\varepsilon$ small (\(2^{\frac{1}{\text{poly}(n)}}\))

Need more samples for Chernoff to kick in -- i.e., if need $\text{poly}(n)$ samples

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

Fix

Choose many more $r_1, \ldots, r_k$ but not independently

i.e., choose them pairwise independently

That is -- find sample space of poly size (i.e., $2^{\mathcal{O}(\log n)}$)

Which behaves in the same way as iid vars.

Thus do exhaustive search on sample space!

String generated by small sample space but still: 1 is good!
Algorithm

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

- \( t = \Theta(n/\epsilon^2) \) \# r's generated

- \[ a = \frac{2n}{\epsilon^2} \]

- For all possible settings of
  \( b_1, \ldots, b_k \in \{ \pm 1 \}^k \)  
  \( z \) all "guesses" for values of
  \[ X_s(s_i) s_j \] \( \text{generate a lot } (2^k = n/\epsilon^2) \text{ of labelled samples} \)

- For every \( w \subseteq \{1, \ldots, K\} \quad W \neq \emptyset \)
  set \( r_w = \bigoplus_{j \in w} s_j \) \quad \( \leftarrow \) pairwise random bits

  \[ p_w = \bigoplus_{j \in w} b_j \quad \text{if } \text{"correct" guess of } \text{"}\]

  \[ \text{initial guess of } s_i \text{" according to } X_s \]

- \( \forall i \in [n] \) put \( i \) in \( S_{b_1, \ldots, b_k} \) if
  \( \text{majority of } p_w + f(r_w \Theta s_i) \)

- Test \( S_{b_1, \ldots, b_k} \) to see if agrees enough with \( f \)

  \( \text{if yes, output } \) \[ \frac{1}{2} + \frac{3}{4} \epsilon \] for }
Behavior

For \( S \) s.t. \( f \) agrees with \( \chi_S \) on \( \frac{1}{2} + \varepsilon \) of inputs:

1) if setting of \( \delta_i \)'s agrees with \( \chi_S \)
   i.e. \( \forall i \quad \delta_i : \chi_S (s_i) \)
   then \( \forall w \quad p_w = \prod_j \chi_S (s_j) \) def of \( p_w \)
   \[ = \chi_S (\Theta_{j \in w} s_j) \]
   \[ = \chi_S (r_w) \] def of \( r_w \)

From now on, assume this setting of \( \delta_i \)'s...

2) \( r_w \)'s are pairwise independent \([in fact, generated via a known construction] \)
   i.e. \( \Pr [r_w = b_1 \land r_w = b_2] = \Pr [r_w = b_1]. \Pr [r_w = b_2] \)
   also \( r_w \circ e_i \)'s are p.i.

3) \( \Pr [ \text{Algorithm generates } S \text{ when considering } S_{b_1, \ldots, b_k} ] : \)
   \[ \Pr [ \text{it gets } S \text{ right on index } i ] \]
   \[ = \Pr [ p_w \cdot f (r_w \circ e_i) \cdot (-1)^{\sum_{j \neq i} e_j} = 1 ] \]
   indicator \( \chi_w = 1 \) if holds

Note: if \( f (r_w \circ e_i) = \chi_S (r_w \circ e_i) \) \( \leq \ldots \)

then \( \chi_w = 1 \)
\[ E[X_w] = \frac{1}{a} + \varepsilon \]

since \( r \) is \( \text{uniform dist} \)

\[ \text{Variance } \sigma_w^2 = E[E[X_w]] - E[X_w]^2 \]
\[ \geq \frac{1}{2} + \varepsilon - \left( \frac{1}{a} + \varepsilon \right)^2 = \frac{1}{a} - \varepsilon^2 \]

\[ E\left[ \sum_{w \in [k]} X_w \right] \geq t \left( \frac{1}{a} + \varepsilon \right) \]

\[ \Pr \left[ \sum_{w} X_w < \frac{t}{2} \right] \leq \frac{\left( \frac{1}{a} - \varepsilon^2 \right)}{t \varepsilon^2} \leq \frac{1}{t \varepsilon^2} \leq \frac{1}{2n} \]

union bound \( \Pr [ \text{not output} ] \leq \frac{1}{2} \)

Also shows:

\[ \text{parity test agrees with } \frac{1}{a} + \varepsilon \text{ is } O \left( \frac{1}{\varepsilon^2} \right) \]