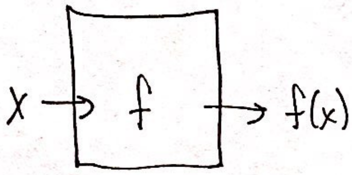
Learning Parities with queries

Goldreich - Levin
Kushilevits - Mansour



Given f, θ

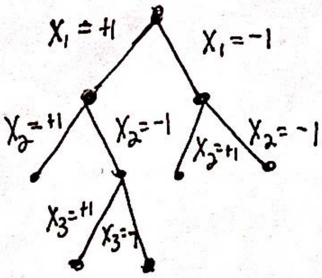
1) Output all coeffs S s.t. $|\hat{f}(S)| \geq \theta$ output all close linear fctns
2) All output coeffs S satisfy $|\hat{f}(S)| \geq \frac{\theta}{2}$ no junk

(Probably) Can't do it with random examples

Can we do it with queries to f? Yes!

Last time: warmup only one big $\hat{f}(S)$. Algorithm: find S bit-by-bit.

General Case / Main idea: "Exhaustive search with good pruning"



idea: find method, using sampling, to estimate total energy in subtree, & don't go down paths with low energy.

How do we prune?

Define quantity:

Fix $0 \leq K \leq n$ current "level" of search

$S_1 \subseteq [K]$ current "node" of search

where are $\underset{\substack{\uparrow \\ \text{suffix of } S}}{S_2}$, $\underset{\substack{\uparrow \\ \text{prefix of } T}}{T_1}$? we'll see them in the analysis

main point:
index 1 → prefix
2 → suffix

note 2^K such fctns

$$f_{K,S_1} : \{\pm 1\}^{n-K} \to \mathbb{R}$$

$$\text{s.t. } f_{K,S_1}(x) = \underbrace{\sum_{T_2 \subseteq \{K+1,\dots n\}} \hat{f}(S_1 \cup T_2)\,\chi_{T_2}(x)}$$

will call it with $y = x_{K+1} \cdots x_n$

note: all Fourier coeffs $\hat{f}(S)$ s.t. S aggrees with S_1 on $[K]$

Sanity checks:

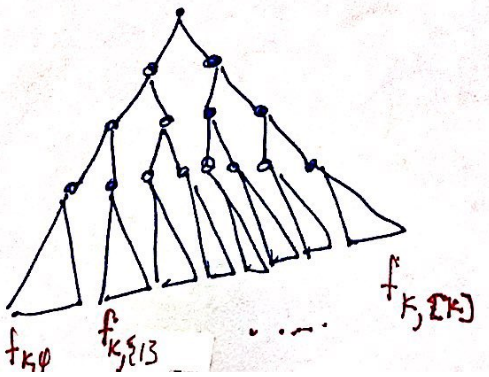
1) $K=0$

$$f_{0,\varphi}(x) = \sum_{\substack{T_2 \subseteq [n] \\ \uparrow \\ \text{since } K=0}} \underset{\substack{\uparrow \\ \text{since } S_1 = \varphi}}{\hat{f}(T_2)}\,\chi_{T_2}(x) = f(x)$$

2) $K=n$

$$f_{n,S_1}(x) = \underset{\substack{\uparrow \\ \text{sum only over empty set}}}{\hat{f}(S_1)} \nwarrow \uparrow_{\text{since } T_2 = \varphi}$$

Back to picture



partition Fourier coeffs of S into 2^K subsets

Plan: Go down paths with $E[f_{k_s}^2(x)] \geq \theta^2$:

1. Can we compute it?
2. Does it bring us to the right leaves?
   - do we get all heavy leaves?
   - do we get junk? (light leaves)
3. How many paths do we take? (do we need lots of dead ends?) (i.e. is runtime good?)

Not too many paths! (Answer to 3)

Lemma "not too many" lemma (not just at end but also at any stage in algorithm)

f Boolean

(1) $\le \frac{1}{\theta^2}$ s's satisfy $|\hat{f}(s)| \ge \theta$

(2) $\forall\ 0 \le k \le n$, $\le \frac{1}{\theta^2}$ fctns f_{k,s_1} have $E_x[f^2_{k,s_1}] \ge \theta^2$

Pf.

(1) Parseval's: $1 = E_x[f^2(x)] = \sum_s \hat{f}(s)^2$

So, if $> \frac{1}{\theta^2}$ s's satisfy $|\hat{f}(s)| \ge \theta$

then $\sum_s \hat{f}(s)^2 > \frac{1}{\theta^2} \cdot \theta^2 > 1$ $\rightarrow\leftarrow$

(2) For given k:

Claim: $\forall k,\ s_1 \le k$

$$E_x[f_{k,s_1}(x)^2] = \sum_{T_2 \subseteq \{k+1,\dots,n\}} \hat{f}(s_1 \cup T_2)^2$$

Pf of Claim

$$E_x\Big[\Big(\sum_{T_2} \hat{f}(s_1 \cup T_2)\chi_{T_2}(x)\Big)^2\Big] = E_x\Big[\sum_{T_2, T_2' \subseteq \{k+1\dots n\}} \hat{f}(s_1 \cup T_2)\,\hat{f}(s_1 \cup T_2')\,\chi_{T_2}(x)\,\chi_{T_2'}(x)\Big]$$

$$= \sum_{T_2, T_2'} \hat{f}(s_1 \cup T_2)\,\hat{f}(s_1 \cup T_2)'\ \underbrace{E[\chi_{T_2}(x)\,\chi_{T_2'}(x)]}_{\substack{=1 \text{ if } T_2 = T_2' \\ =0 \text{ o.w.}}}$$

$$= \sum_{T_2} \hat{f}(s_1 \cup T_2)^2$$

∎

Using claim:

$$1 = \sum_{S} \hat{f}(S)^2 = \sum_{S_1 \subseteq [k]} \sum_{T_2 \subseteq \{k+1..n\}} \hat{f}(S_1 \cup T_2)^2$$

$$= \sum_{S_1} E_x[f^2_{k,S_1}(x)] \quad \leftarrow \text{claim}$$

so $\le \frac{1}{\theta^2}$ S_1's can have $E_x[f^2_{k,S_1}(x)] > \theta^2$ ■

Does it bring us to good leaves? (Answer to 2)

Fact: "not missing out" ⇒ find all big Fourier coeffs

for any S_1, if $\exists T_2$ s.t. $|\hat{f}(S_1 \cup T_2)| > \theta$

then $E_x[f^2_{k,S_1}(x)] = \sum_{T_2} \hat{f}(S_1 \cup T_2)^2 \ge \theta^2$

$\Rightarrow$ $E_x[f^2_{k,S_1}(x)]$ is a good measure to use to find heavy Fourier coeffs.

so we find good leaves, but do we output bad leaves too?

no can always estimate F.C. of each leaf found to make sure we aren't outputting junk

Can we estimate $f_{K,S_1}(X)$? (answer to 1)

Bad idea: estimte each $\hat{f}(S_1 \cup T_2)$ $\forall T_2$ (too much time too many T_2)

Better idea: use following lemma (+ queries)

"$f_{K,S_1}(x)$ estimation" Lemma

$f : \{\pm1\}^n \to \{\pm1\}$

$0 \le K \le n$

$S_1 \subseteq [K]$

For $X \in \{\pm1\}^{n-k}$

$$f_{K,S_1}(x) = \underset{y \in \{\pm1\}^k}{E}[f(\underset{\substack{\uparrow \\ \text{concatenation}}}{yx}) \chi_{S_1}(y)]$$

} use this to estimate $f_{K,S_1}(x)$

Pf

$$f(yx) = \sum_T \hat{f}(T) \chi_T(yx)$$

$T = T_1 \cup T_2$ $\quad T_1 \subseteq [K]$, $T_2 \subseteq [K+1, \dots, n]$

so $\chi_T(yx) = \chi_{T_1}(y) \chi_{T_2}(x)$

$$E_{y \in \{\pm1\}^k}[f(yx)\chi_{S_1}(y)] = E_y[\sum_{T_1}\sum_{T_2} \hat{f}(T_1 \cup T_2) \chi_{T_1}(y) \chi_{T_2}(x) \chi_{S_1}(y)]$$

$$= \sum_{T_1}\sum_{T_2} \hat{f}(T_1 \cup T_2) \chi_{T_2}(x) \underbrace{E_y[\chi_{T_1}(y)\chi_{S_1}(y)]}_{\substack{= 0 \text{ if } T_1 \neq S_1 \\ 1 \text{ if } T_1 = S_1}}$$

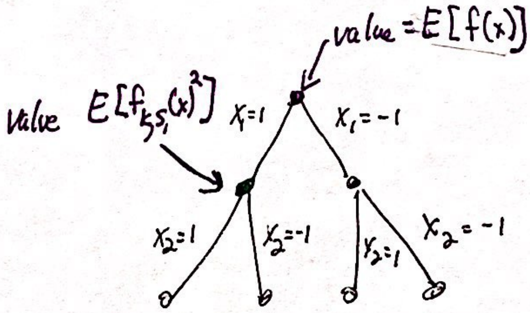
$$= \sum_{T_2 \subseteq [K+1, \dots, n]} \hat{f}(S_1 \cup T_2) \chi_{T_2}(x)$$

$$= f_{K,S_1}(x)$$

∎

Chernoff + sample $\Rightarrow$ Can get γ-additive estimate with prob $\ge 1-\delta$ with $O(\frac{1}{\gamma^2} \log \frac{1}{\delta}$ queries)

Algorithm:



Ideal KM algorithm: (given k, S_1)

```
If $k=n$
    output $S_1$
Else
    if $E_x[f^2_{k,S_1\cup\{k+1\}}(x)] \geq \theta^2/2$
        recurse on $(k+1, S_1 \cup \{k+1\})$   (else, kill this subtree)
    if $E_x[f^2_{k,S_1}(x)] \geq \theta^2/2$
        recurse on $(k+1, S_1)$   (else, kill this subtree)
```

Minor Problem - can only **estimate** $E_x[f^2_{k,S}(x)]$:

```
repeat m times & output average:
    pick random $x$
    estimate $f_{k,S}(x)$ as before
    $y_x \leftarrow$ square of
```

Chernoff $\Rightarrow$ good additive estimate so change alg threshhold to $\theta^2/2$ in order to make sure we get big coefts

(won't output junk since test before output, still won't go down too many paths.)

Thm $f : \{\pm 1\}^n \to \{\pm 1\}$

$\forall \theta > 0$, KM-Algorithm outputs set $S' = \{S_1 \dots S_\ell\}$ s.t. $\ell = O(\frac{1}{\theta^2})$

s.t. with prob $\geq 1 - \delta$

- $\forall\ S_i \in S'$, $|\hat{f}(S)| \geq \frac{\theta}{2}$ (not too many)
- $\forall\ S_i \notin S'$, $|\hat{f}(S)| \leq \theta$ (not too few)

+ query/time complexity is $(\text{poly}(n, \frac{1}{\theta}), \log \frac{1}{\delta})$

Pf

(1) if $\hat{f}(S) < \frac{\theta}{2}$ then $\underset{f_{n,S}}{\underset{\shortparallel}{\hat{f}(S)^2}} \leq \frac{\theta^2}{4}$ + S won't be output by algorithm

(2) if $\hat{f}(S) > \theta$, then $\forall k$, if S_1 + S agree on $[1..k]$

"not too few" fact $\Rightarrow E_x[f^2_{k,S_1}(x)] \geq \theta^2$

(3) total # nodes explored at level k is $\leq \frac{1}{\theta^2}$

∴ total # nodes visited $\leq \frac{n}{\theta^2}$

■

What are the values of the $\hat{f}(S)$'s?

KM.9

Cool Applications: • Decision trees of size $\underbrace{\le t}$

previously we needed a depth bound

- all fctns of small L_1 norm

$$L_1(f) = \sum_S |\hat{f}(S)|$$

by setting $\theta \leftarrow \frac{\varepsilon}{L_1(f)}$

in time $\text{poly}(n, L_1(f), 1/\varepsilon)$

if don't know $L_1(f)$

assume 1, 2, 4, 8

test hypothesis at each round & continue if it is not good