Monotone Functions

def partial order $\leq$

$$x \leq y \quad \text{iff} \quad \forall i \quad x_i \leq y_i$$

monotone function f

$$x \leq y \Rightarrow f(x) \leq f(y)$$

Learning algorithms for the class of monotone functions?

← here n = dimension, 2^n = domain size

in homework we see $2^{O(\sqrt{n})}$ random samples suffice for uniform distribution

why is this nontrivial?

we said poly samples is easy, the problem is computation time? poly in what?

but need $\text{poly}(\log |\mathcal{C}|)$ samples

↑ all monotone fctns

there are 2^{2^n} fctns, $\geq 2^{2^n/\sqrt{n}}$ monotone fctns

why so many monotone fctns?

consider slice fctns:

f=1 →

f=0 →

$\binom{n}{n/2} \approx 2^{\left(\frac{2^n}{\sqrt{n}}\right)}$ pts in middle row can be set in all possible ways w/o violating monotonicity

$\#\text{fctns} \geq 2^{2^{\left(\frac{2^n}{\sqrt{n}}\right)}}$

a hard distribution: Uniform on middle row

$\Rightarrow$ learning needs $\Omega\left(2^{\left(\frac{2^n}{\sqrt{n}}\right)}\right)$ even with queries in PAC model

Today's question:

what about learning monotone distributions, on uniform distribution, <u>with</u> queries?

here we will get a very slight "win":

slightly better than random guess

All monotone fctns have weak agreement $\left(\frac{1}{2}+\Theta\left(\frac{1}{n}\right)\right)$ with <u>some</u> dictator fctn.

(can get $\frac{1}{2}+\Theta\left(\frac{1}{\sqrt{n}}\right)$ with majority + dictators)

slice fctns have weak agreement with all dictators on uniform dist

<u>Thm</u> $\forall f$ monotone, $\exists\, g \in \{\pm 1, x_1, x_2, \dots, x_n\} = S$

s.t. $\Pr_x[f(x)=g(x)] \geq \frac{1}{2}+\Omega\left(\frac{1}{n}\right)$

Why does this give weak learning algorithm? estimate agreement of f with all members of S + output member with max agreement

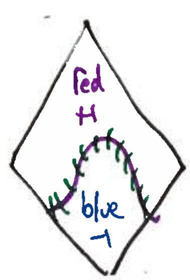
<u>Pf</u>.

<u>Case 1</u> $f(x)$ has weak agreement with $+1$ or -1

<u>Case 2</u> otherwise $\Pr[f(x)=1] \in \left[\frac{1}{4}, \frac{3}{4}\right]$

First a break, before we prove case 2 ...

let's define influence of monotone fctns?



- # nodes $= 2^n$, # edges $= \frac{n \cdot 2^n}{2}$
- each level has $\binom{n}{j}$ weight j nodes
- monotone $\Rightarrow$ no blue above any red
- slicing cube in roughly half cuts many edges + many in same direction
- $\mathrm{Inf}(f) = \frac{\#\text{red-blue edges}}{2^{n-1}}$, $\mathrm{Inf}_i(f) = \frac{\#\text{rb edges in } i^{\text{th}} \text{dir}}{2^{n-1}}$

Another Definition

influence of i^{th} var on $f: \{\pm 1\}^n \to \{\pm 1\}$

$$\mathrm{Inf}_i(f) = \Pr_x[f(x) \neq f(x^{\oplus i})]$$

i^{th} bit of x is flipped

Total Influence

$$\mathrm{Inf}(f) = \sum_{i=1}^{n} \mathrm{Inf}_i(f)$$

NOTE: exactly equal to # red-blue edges

Thm: f monotone $\Rightarrow$ $\mathrm{inf}_i(f) = \hat{f}(\{i\})$

Thm: majority fctn $f(x) \equiv \mathrm{sign}\left(\sum_i x_i\right)$ (n odd) maximizes influence among monotone fctns.

Pfs on h.w.

recall H.W.:

Thm f monotone

$$\mathrm{Inf}_i(f) \underset{\text{H.W.}}{=} \hat{f}(\{i\}) \underset{\text{Known}}{=} 2\Pr[f(x) = \underbrace{\chi_{\{i\}}}_{x_i}(x)] - 1$$

Plan:

Show $\mathrm{Inf}_i(f) \geq \Omega\left(\frac{1}{n}\right)$

$$\Rightarrow \Pr[f(x) = x_i] \geq \frac{1}{2} + \frac{\mathrm{Inf}_i(f)}{2}$$

$$\geq \frac{1}{2} + \Omega\left(\frac{1}{n}\right)$$

Will use following tool:

Canonical Path Argument

Plan 1) define canonical path for every red-blue pair of nodes (note such a path must cross at least one red-blue edge)

2) show upper bnd on # of c.p.'s passing through any edge (in particular, any red-blue edge)

3) Conclude lower bnd. on # of red-blue edges

Part 1 of plan:

def. $\forall (x,y)$ s.t. x red & y blue

"Canonical path from x to y" is:

scan bits left to right, flipping where needed

each flip $\rightsquigarrow$ step in path

example

	direction 1	2	3	4
x =	-1	$+1$	$+1$	$+1$
w =	$+1$	$+1$	$+1$	$+1$
z =	$+1$	-1	$+1$	$+1$
y =	$+1$	-1	$+$	-1

$x \to w \to z \to y$

each step is Hamming distance 1

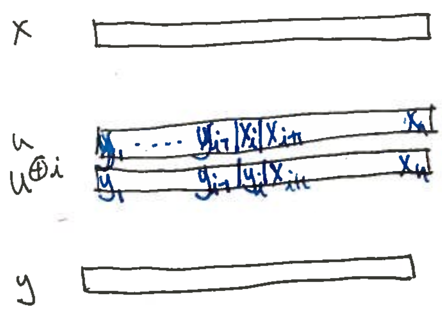
How many red-blue x,y pairs have canonical paths?

recall, $\Pr[f(x)=1] \in [\frac{1}{4}, \frac{3}{4}]$

$$\#\text{ paths} \geq \frac{1}{4}\cdot 2^n \cdot \frac{1}{4}\cdot 2^n = \frac{1}{16}\cdot 2^{2n}$$

Part 2 of plan:

For any (red-blue) edge e, how many x, y pairs can cross it with canonical x-y-path?



$e = (u, u^{\oplus i})$, u: red, $u^{\oplus i}$: blue

edge in i^{th} direction

$\overbrace{y_1 \cdots y_{i-1}}^{\leq 2^{i-1} \text{ settings for prefix of } x}\, x_i\, x_{i+1} \cdots x_n$

$y_1 \cdots y_{i-1}\, y_i\, \underbrace{x_{i+1} \cdots x_n}_{\leq 2^{n-i} \text{ settings for suffix of } y}$

$\therefore \leq 2^n$ total settings of prefix x, suffix y consistent with this edge

Main point: all canonical paths crossing $u, u^{\oplus i}$ agree on $y_1 \cdots y_{i-1}$ + $x_{i+1} \cdots x_n$

Part 3 of plan:

$(\#\text{ red-blue edges})(\max \#\text{ canonical paths that use it}) \geq \#\text{ red-blue canonical paths}$

since each uses ≥ 1 red-blue edge

So

$$\#\text{ red-blue edges} \geq \frac{\overbrace{\tfrac{1}{16}2^{2n}}^{\text{l.b. on \# r-b pairs}}}{\underbrace{2^n}_{\text{u.b. on \# canonical paths crossing any edge}}} = \frac{1}{16}\cdot 2^n$$

so $\exists\, i$ s.t. $\geq \frac{2^n}{16}\cdot\frac{1}{n}$ red-blue edges in direction i

so $\mathrm{Inf}_i(f) \geq \dfrac{\frac{2^n}{16n}}{2^{n-1}} = \dfrac{1}{8n} = \hat{f}(\{i\}) = 2\Pr[f(x) = x_i] - 1$

total # edges in dir i

$\therefore \Pr[f(x) = x_i] \geq \frac{1}{2} + \frac{1}{16 \cdot n}$

Canonical Path argument also used in

- routing
- expansion / conductance of hypercube / other Markov Chains

What good is weak learning?

unclear

here only uniform distribution

if can learn in all distributions,

can do much more

(next result does not apply to monotone function learning...

in particular, this weak notion of learning provably doesn't give anything for stronger learning)

i.e. $\frac{1}{n}$ agreement

i.e. const $> \frac{1}{2}$ agreement