Last time:
- Any SQ learning alg. automatically yields a PAC alg. that can handle RCN.
  - Many PAC algs can be rephrased as SQ algs... so we get RCN-tolerance!
  - But not all... concept class of Parity Functions
    - is efficiently PAC learnable

Today:
- Unconditional lower bounds on SQ learning:
  some $\mathcal{C}$'s are eff. PAC learnable, but are not eff. SQ learnable.
  (any $\mathcal{C}, \mathcal{D}$ s.t. $\mathcal{C}$ has many functions that are all uncorrelated with each other under $\mathcal{D}$)
  Parity functions: are not eff. SQ learnable
- Start unit on inherent unpredictability: concept classes $\mathcal{C}$ that don't have eff. learning alg's using any eff. evaluable hypothesis class (based on cryptography)

Admin: All lectures now viewable thru Courseworks
(second exam: Thurs Dec 4)

Questions?

Last time: have eff CHF for $\mathcal{C}_{PAR}$ using $\mathcal{C}_{PAR}$;
$|\mathcal{C}_{PAR}| = 2^n$, so running ν on $m = O\left(\frac{1}{\varepsilon}\left(n + \ln\frac{1}{\delta}\right)\right)$ ex
PAC learns $\mathcal{C}_{PAR}$:)

Note... this seems to really use indiv. ex; is there an SQ version? No!

Fix a dist. $\mathcal{D}$ over $\{0,1\}^n$.

We say two concepts $c_1, c_2 : \{0,1\}^n \to \{0,1\}$ are uncorrelated under $\mathcal{D}$ if

$$\Pr_{x \sim \mathcal{D}}[c_1(x) = c_2(x)] = \frac{1}{2} = \Pr_{x \sim \mathcal{D}}[c_1(x) \neq c_2(x)]$$

Fact (lower bd on SQ learnability):

Fix a $\mathcal{C}$, a dist $\mathcal{D}$ over $\{0,1\}^n$.

Suppose $\exists\, c_1, \dots, c_N \in \mathcal{C}$ s.t. $\forall\, 1 \le i < j \le N$,

$c_i \neq c_j$ are uncorr. under $\mathcal{D}$.

Then any SQ alg for $\mathcal{C}$ (under $\mathcal{D}$) that learns even to acc. $\frac{1}{2} + \frac{1}{N^{1/3}}$ must either

1) make $\geq N^{1/3}$ SQ queries, or
2) make an SQ (x, τ) with $\tau \le \frac{1}{N^{1/3}}$.

Super-hazy intuition: • Sps target $c = c_i$ some unknown $i \in [N]$.

Pretend only poss. SQ's (x's) are $x_1, \dots, x_N$ where

$x_j(x, b_{c(x)})$ is " does $b = c_j(x)$ "$_{c(x)}$

$j = i$: $P_x = 1$

$j \neq i$: $P_x = \frac{1}{2}$.

$x, \tau \to$ [STAT $(c, \mathcal{D})$]

Return to $\mathcal{C}_{PAR}$...

Consider $\mathcal{D} = \mathcal{U}$, unif. dist. on $\{0,1\}^n$.

Claim Given any 2 PAR fns $PAR_{S_1}(x) \neq PAR_{S_2}(x)$, they are uncorr. under $\mathcal{D}$.

Pf: Know $S_1 \neq S_2$. So $\exists j$ s.t. $j \in S_1$ but $j \notin S_2$

eg $S_1 = \{1,2,3\}$, $S_2 = \{2,3,4\}$ $(j=1)$.

Break up $\{0,1\}^n$ into 2^{n-1} pairs: each pair's 2 elts disagree only on coord j

$j=1$: $\{00000, 10000\}$ a pair $(n=5)$

$\{00001, 10001\}$ a pair

Fix any pair: PAR_{S_2} is same on the 2 elts of the pair, but PAR_{S_1} is diff. " " " " " " "

$\rightarrow PAR_{S_2} = 0$ on both, $PAR_{S_1} = 0$ on one, 1 on other. ∎

Cor: There's no eff. SQ alg for $\mathcal{C}_{PAR}$.

$(N = 2^n.)$

Conseq. for other concept classes:

- $\mathcal{C}_{DNF}$ = all n-term DNF's over $\{0,1\}^n$.

We saw $|\widehat{PS}|$ any PAR_S with $|S| \le \log n$ can be computed by a $\mathcal{A}$.

There are $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\log n}$ many $S \subseteq [n]$, $|S| \le \log n$

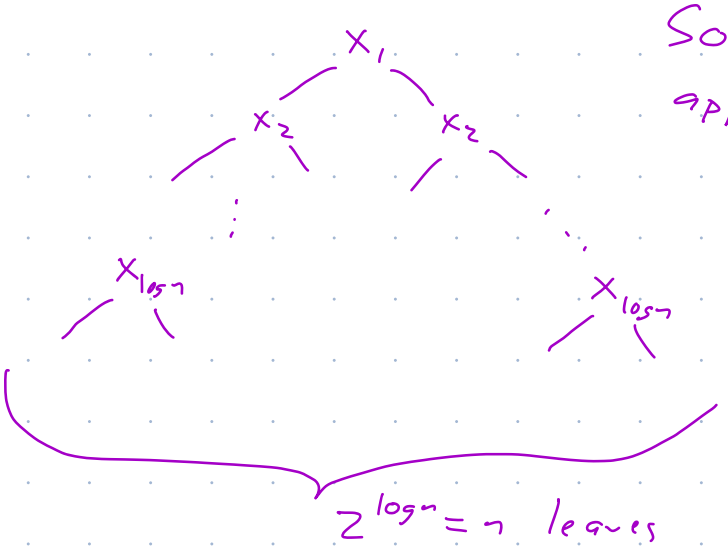
$\geq \left(\frac{n}{\log n}\right)^{\log n} = n^{\Omega(\log n)}$

$\binom{x}{y} \geq \left(\frac{x}{y}\right)^y$

$\Rightarrow$ So $\mathcal{C}_{DNF}$ has $N \geq n^{\Omega(\log n)}$ many uncorr. concepts vis-a-vis $\mathcal{A}$.

So best poss. SQ alg to learn DNFs must run in $n^{\Omega(\log n)}$ time.

Similarly, $\mathcal{C}_{DT}$ = decision trees with $\le n$ leaves: any $(\log n)$-junta also expr. as



So same l.b. applies.

Noise and Hyp. Choice

Sps have
- alg A: eff. proper PAC learner for $\mathcal{C}$ (in orig. noiseless model).
- alg B: eff PAC learner for $\mathcal{C}$ even with RCN (but B <u>not</u> proper).

Can always get best of both worlds: proper + handling RCN:
- run B (on $EX^m(c,\mathcal{D})$) with very small error param. It gives a super-accurate hyp h
- use h to relabel fresh batch of ex. drawn from $EX^m(c,\mathcal{D})$, use those to run A, + it's fine b/c h "got rid of the noise".

COMPUTATIONAL REPRESENTATION-INDEP. HARDNESS of LEARNING (HoL)

3 types of HoL results:

1) "information-theoretic" hardness: hardness irrespective of learner's runtime; based on insuff. data

- No PAC learner can learn mon conj using $\sqrt{\frac{n}{\varepsilon}}$ examples.

- C_{ALL} := all 2^{2^n} Bool fns $c: \{0,1\}^n \to \{0,1\}$.
  VCDIM $= 2^n$; so no poly(n)-time learner.

Other 2 types: computational.
Learning would be possible, given the examples, with enough runtime; but no poly(n)-time learner exists.

2) Representation-dependent HoL: hardness for learning using a partic. hypoth. representation.

(Ex: our 3-term DNF learning hardness result.)

3) (now): Representation-independent HoL:

"Unless (comput. problem XYZ) is efficiently solvable, no eff. alg. can PAC learn C using any poly-time evaluable $\mathcal{H}$."

→ functions in C are inherently unpredictable

---

2), 3) - type results require assumptions; any pf that, e.g., "C_{DNF} n-term DNFs are not

"PAC learnable in poly(n) time", immediately implies $P \neq NP$.

Sketch:

NP = comput. problems for which it's poss. to efficiently verify a sol. that's given to you

↳ Graph 3-COL:

P = comput. problems for which it's poss. to efficiently find a sol.

If $P = NP$, the NP problem of verifying that a given cand. n-term DNF hyp. is consistent w/ a given input data set... is in P. So there would be an eff CHF for C_{DNF} using C_{DNF}, + C_{DNF} would be eff. PAC learnable.

Fact of the matter:

- for 2) -type HoL, we can prove them using "worst-case" hardness assumptions.

↓

$P \neq NP$: no eff. alg solves worst-case instances.

- for 3) -type HoL, we can only prove these under stronger "average-case" hardness assump.

↳ There is an input distr. P of problem instances s.t. any poly-time alg. fails w.h.p. on random instances drawn from P.

(↓ with high prob.)

e.g. "There is no poly(n)-time alg. which, given a # $N = p \cdot q$ where p, q are uniform random n-bit primes, successfully factors N with probability $\geq \frac{1}{100}$."
