Admin: *PS5 due today! (last one!)

Rocco OH Fri 9-11

Last time: \( C_{\text{eff}} \) SQ learn. \( \Rightarrow \) \( C_{\text{eff}} \) PAC learn. even \( \cap \) RCN.

Today:
- SQ learn. vs PAC learn.
- """" in presence of RCN
- learning \( \cap \) noise

Last unit: computational hardness of learning. KV chap. 6 (we'll go a diff. route).

\( \Rightarrow "C_{\text{eff}} \) PAC learnable" does not imply
\( "C_{\text{eff}} \) SQ learnable"

\( C = \) all \( 2^n \) for \( PAR_5(x) \), \( \subseteq \{1, \ldots, n\} \).

\( \Rightarrow \) PAC learnable (Gaussian elim. over \( \mathbb{F}_2 \)).

\( \Rightarrow \) This alg doesn't seem comput. \( \cap \) SQ...

\( \Rightarrow C \) not \( \text{eff} \) SQ learnable...
Def: Let $\mathcal{D}$ be a distrib. over $\{0,1\}^n$. Two conc. $c_1, c_2$ are uncorrelated under $\mathcal{D}$ if
\[
Pr(c_1(x) = c_2(x)) = Pr(c_1(x) \neq c_2(x)) = \frac{1}{2}.
\]

FACT: Let $E$ be a conc. over $\{0,1\}^n$, $\mathcal{D}$ a distrib. over $\{0,1\}^n$. Suppose $E$ contains $N$ concepts $c_1, \ldots, c_N$ s.t.
\[\forall 1 \leq i \neq j \leq N, \ c_i \neq c_j \text{ are uncorr. under } \mathcal{D}.
\]
Then any SQ alg for learning $E$ (for distrib $\mathcal{D}$) must either:

i) call STAT $\geq N^{1/3}$ many times, or

ii) call STAT $\leq \frac{1}{N^{1/3}}$ in order to learn to error $\epsilon = 0.49$.

Idea: Concept $c \leftrightarrow$ vector $c$ in bi-dim space.

Uncorr. conc. $c_1, \ldots, c_N \leftrightarrow$ $N$ vect. in $\mathbb{R}^N$.

A SQ $x(\cdot, \cdot)$ : $\leftrightarrow$ a vector in $\mathbb{R}^N$.

partic. value of $P_x$ on $c$: $\leftrightarrow (v, c)$.

SQ learning: Trying to learn $c$ by making "inner product probes": choose set of $v_1, v_2, \ldots, (x_1, x_2, \ldots) + \text{get } (v_1, c) \pm \epsilon$

$(v, c) \pm \epsilon$

e tc.
Existence of $N$ uncorr. concepts: in response to queries $v'_1, v'_2, \ldots \in \mathbb{R}^N$, poss. to get responses

"$(v'_1, c) \approx 0$" $\rightarrow$ ($c$ is in $\mathbb{R}^{N-1}$-dim space)
"$(v'_2, c) \approx 0$" $\rightarrow$ ("" in $\mathbb{R}^{N-2}$-dim space)

e tc.

Even after many queries: still have $\geq 1$ dim. out of luck.

Using FACT:

1) $c = \text{all } 2^N \text{ PAR}_S \text{ fns.}$

Claim: $\mathcal{F} = \text{unif on } S_0, 1^N$. Any 2 distinct

\begin{align*}
\text{PAR}_S, \text{PAR}_T \text{ fns. are uncorrelated.} \\
x_1 \Theta x_2 \Theta x_3 \rightarrow x_2 \Theta x_3 \Theta x_6
\end{align*}

PF: Fix $S \neq T$. Have some $i$ in one but not other: say $x_i \in S$, but $x_i \notin T$.

Fix any outcome of $x_1 \ldots x_n \in S_0, 1^{n-1}$.

$b_2, \ldots, b_n$.

Claim: PAR$_S$ + PAR$_T$ agree on one of $0b_2 \ldots b_n$ $1b_2 \ldots b_n$

+ disagree on other one.

PAR$_S(0b_2 \ldots b_n) \neq \text{PAR}_S(1b_2 \ldots b_n)$, but

PAR$_T(0b_2 \ldots b_n) = \text{PAR}_T(1b_2 \ldots b_n)$.

So any SQ alg for PAR needs $2^\Omega(n)$ time.
Q: Is every $C$ which is \textit{eff} PAC learnable with RCN, also \textit{eff} SQ-learnable? (Maybe yes...?)

Any class $C$ containing $N$ PAR funs:
$C$ can't be SQ learned in time $cN^{1/3}$.

Consider $C = \{ \text{all poly}(n)$-term $\text{ONE}$s $\}$.

Any $(\log n)$-size PAR can be expressed as an $O(1)$-term $\text{ONE}$.

Best runtime to learn $C$? $n^{1/2}$.

$x_1 \oplus x_2 \oplus x_3 = x_1 x_2 x_3 \lor x_1 \overline{x_2} \overline{x_3} \lor \overline{x_1} x_2 x_3 \lor \overline{x_1} \overline{x_2} x_3$.

There are $N = \binom{n}{\log n} = n^{\log n}$ many $(\log n)$-size PAR's...

so any SQ alg for learning $\text{ONE}$ (even under $\Theta = \mathbb{U} = \text{unif}$) must have $n \Omega(\log n)$ runtime.

Same is true for $C = \text{all poly}(n)$-leaf decision trees.
Best runtime known for $n \log n$.

Learning with noise, code:
$Sps E$ is s.t.
- you have learning alg $A$, proper PAC alg for $E$ (no noise)
  - you also have $B$, improper RCN-noise-tolerant PAC alg for $E$.

Can you get best of both worlds? Yes.
Have $EX-c*$.
- Run $B$ to get $h_{\text{acc}}$ hyp. $h$ for $c$.
  \[ \begin{align*}
  \text{Use } h \text{ to relabel ex. from } 0, 1 \text{ now they're noise-free. Use these clean ex. for } A \text{ to get hyp } h' \in E. \end{align*} \]
Last unit: **Hardness of Learning**

**Computational Cryptographic**

"Hardness": neg. results. 3 types:

1) **Info-theoretic.**
   - Mal. noise at rate $n = 0.1$: no PAC alg can learn a dist. c.e. $C$ to acc $\epsilon = 0.15$.
   - No PAC alg for $C = \text{mon disj}$ can use only $\sqrt{\frac{n}{\epsilon}}$ samples.

2) "**Representation-dependent**:" comput. hard to learn 3-term ONE using 3-term ONE as the hyp. representation. **Worst-case.**

3) (NEW!) **Represent.- indep. hardness/**
   "inherent unpredictability".
   "Unless (problem XYZ) has an eff. alg., there exists no PAC alg for $C$ using any efficiently eval. hyp class."

Spzs you prove, with no assump, that

"There is no poly$(n)$-time PAC alg. to learn $C = \{\text{all } n^2\text{-term ONEs}\}$."

\[\]
This implies $P \neq UP$.