Last time: Handling RCN.

- Revisited an old PAC learning alg (for mon. conj.) that's **not** RCN-tolerant, & adapt it to handle RCN
- Motivated new learning model:

STATISTICAL QUERY (SQ) LEARNING model

- We argued: Any efficient SQ learning algorithm **automatically** yields a efficient PAC learning alg.

Today:

- Efficient SQ learning alg yields efficient PAC learning alg that can tolerate RCN.
  - Many PAC learning algs can be rephrased as SQ algs, so we get RCN-tolerance!
  - But not all...
- **Unconditional** lower bounds on SQ learning: some $\mathcal{C}$'s **are** eff. PAC learnable, but provably **are not** eff. SQ learnable.

Questions?

---

We saw:

Any efficient SQ learning algorithm **automatically** yields a efficient PAC learning alg.

Now:

**Theorem ★**: Let $\mathcal{C}$ be a concept class that's efficiently SQ-learnable.

Then $\mathcal{C}$ is eff. PAC learnable in presence of RCN.

$\forall\ 0 \le \eta < \frac{1}{2}$, time $\mathrm{poly}(\ldots, \frac{1}{1-2\eta})$

Great news! Many PAC algs can be viewed as SQ algorithms, so get RCN-tolerant versions of these:

- Perceptron
- decision lists, conj, disj,
- LTFs

etc...

Pf of ★: key step: simulating a call to $STAT(c, \mathcal{D})$ given $EX^{\eta}(c, \mathcal{D})$.

Let's try.

We're given query $(\chi, \tau)$ + access to $EX^{\eta}(c, \mathcal{D})$.

Must est. $P_\chi := \Pr_{x \sim \mathcal{D}}[\chi(x, c(x)) = 1]$

acc. to $\pm\tau$.

Challenge: noise.

Key idea: break up $X$ into 2 pieces:

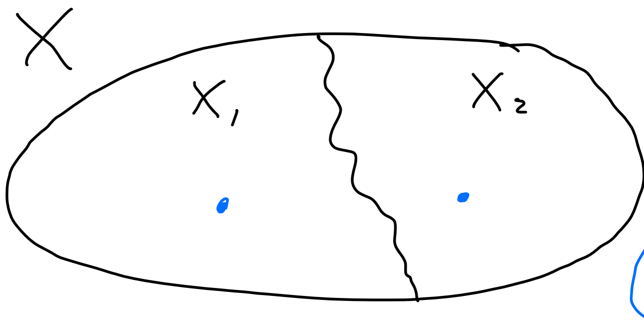
- $X_1 = \{x \in X : \mathcal{X}(x,0) \neq \mathcal{X}(x,1)\}$

  noise matters!

+

- $X_2 = X \setminus X_1 = \{x \in X : \mathcal{X}(x,0) = \mathcal{X}(x,1)\}$

  noise doesn't matter!




Key obs:

given $x$, can check if $x \in X_1$ or $x \in X_2$

$\mathcal{X}(x,0) \stackrel{?}{=} \mathcal{X}(x,1)$

Ex: Suppose $\mathcal{X}(x,b) =$ "$\underline{b=1 \text{ + } x_1 = 0}$".

Have

$$X_2 = \{x \in X : x_1 = 1\}$$
$$X_1 = \{x \in X : x_1 = 0\}$$

---

Setup: Let $p_1 := \Pr_{x \sim \mathcal{D}}[x \in X_1]$, $p_2 = 1 - p_1$

Let $\mathcal{D}_1 = \mathcal{D}$ conditioned on $X_1$

$$\Pr_{x \sim \mathcal{D}_1}[x \in S] = \Pr_{x \sim \mathcal{D}}[x \in S \mid x \in X_1]$$

---

$P_\chi = A\cdot B + C.$

**Lemma 1:** Have

$$P_\chi = \overset{A}{\overset{\shortparallel}{p_1}} \cdot \overbrace{\frac{\Pr_{(x,b)\sim EX^\eta(c,\mathcal{D}_1)}[\chi(x,b)=1] - \eta}{1-2\eta}}^{B} + \underbrace{\Pr_{(x,b)\sim EX^\eta(c,\mathcal{D})}[\chi(x,b)=1 \wedge x\in X_2]}_{C}.$$

**Lemma 2:** Possible to est $A\cdot B + C$ given $EX^\eta(c,\mathcal{D})$.

---

**Pf of L1:**

$$P_\chi = \Pr_{x\sim\mathcal{D}}[\chi(x,c(x))=1]$$

$$= \Pr_{x\sim\mathcal{D}}[\chi(x,c(x))=1 \wedge x\in X_1] + \Pr_{x\sim\mathcal{D}}[\chi(x,c(x))=1 \wedge x\in X_2]$$

$$= \underbrace{\Pr_{x\sim\mathcal{D}}[x\in X_1]}_{=p_1} \cdot \Pr_{x\sim\mathcal{D}}[\chi(x,c(x))=1 / x\in X_1] + \Pr_{x\sim\mathcal{D}}[\chi(x,c(x))=1 \wedge x\in X_2]$$

= b/c for $x\in X_2$, noise doesn't matter

$$= \overset{A}{\overset{\shortparallel}{p_1}} \cdot \Pr_{x\sim\mathcal{D}}[\chi(x,c(x))=1 / x\in X_1] + \underbrace{\Pr_{(x,b)\sim EX^\eta(c,\mathcal{D})}[\chi(x,b)=1 \wedge x\in X_2]}_{C}$$

Need to show: $\Pr_{x\sim\mathcal{D}}[\chi(x,c(x))=1 \mid x\in X_1] = B$.

$$\Pr_{x\sim\mathcal{D}}[\chi(x,c(x))=1 \mid x\in X_1] = \Pr_{x\sim\mathcal{D}_1}[\chi(x,c(x))=1]$$

Have (b/c in $X_1$ / under $\mathcal{D}_1$, noise always matters)

$$\begin{aligned}
\Pr_{(x,b)\sim EX^{\eta}(c,\mathcal{D}_1)}[\chi(x,b)=1] &= \overset{\text{no noise}}{(1-\eta)}\cdot\Pr_{x\sim\mathcal{D}_1}[\chi(x,c(x))=1] \\
&\quad + \overset{\text{noise}}{\eta}\cdot\Pr_{x\sim\mathcal{D}_1}[\chi(x,c(x))=0] \\
&= (1-\eta)\cdot\Pr_{x\sim\mathcal{D}_1}[\chi(x,c(x))=1] \\
&\quad + \eta\left(1-\Pr_{x\sim\mathcal{D}_1}[\chi(x,c(x))=1]\right) \\
&= \eta + (1-2\eta)\cdot\Pr_{x\sim\mathcal{D}_1}[\chi(x,c(x))=1].
\end{aligned}$$

Solve :

$$\Pr_{x\sim\mathcal{D}_1}[\chi(x,c(x))=1] = \underbrace{\frac{\Pr_{(x,b)\sim EX^{\eta}(c,\mathcal{D}_1)}[\chi(x,b)=1]-\eta}{1-2\eta}}_{=\,B.}$$

---

Pf of L2:

Possible to est $A \cdot B + C$ given $EX^{\eta}(c, \mathcal{D})$.

$$A = p_1 = \Pr_{x \sim \mathcal{D}}[x \in X_1]$$

H/T

- Call $EX^{\eta}(c, \mathcal{D})$, get $(x, b)$, $x \sim \mathcal{D}$

- Call $\mathcal{X}(x, 0)$ & $\mathcal{X}(x, 1)$. $\neq$ : $x \in X_1$ :)

Use CB: $\frac{\log 1/\delta}{\tau^2}$ coin tosses give $\pm\tau$-acc est w.p. $1-\delta$.

So, can eff est. $p_1 \overset{\cdot}{=} A$

We want est. of $A \cdot B$, not just $A$.

$$B \leq \frac{1}{1-2\eta}.$$

If $p_1 < \gamma \cdot (1-2\eta)$, then $A \cdot B < \gamma$, &

$0$ is a $\pm\gamma$-acc. est. of $A \cdot B$.

So done unless our est of $p_1$ is $\geq \gamma \cdot (1-2\eta)$

B involves $\Pr_{(x,b)\sim EX^{\eta}(c,\theta_1)}[\chi(x,b)=1]$ ; have access to $EX^{\eta}(c,\theta)$.

attack by drawing $(x,b)$ from $\underline{EX^{\eta}(c,\theta)}$,

check if $\underline{x \in X_1}$ ; if so, got $(x,b) \sim EX^{\eta}(c,\theta_1)$ ☺

happens w. prob. $p_1$

Since $p_1 > \gamma \cdot (1-2\eta)$, slowdown of this filtering is $\leq \frac{1}{\gamma \cdot (1-2\eta)}$.

So can get acc. est. of B eff, & can est. $A \cdot B$.

---

Last bit: estimate

$$C = \Pr_{(x,b)\sim EX^{\eta}(c,\theta)}[\chi(x,b)=1 \text{ \& } x \in X_2]$$

Draw $(x,b) \sim EX^{\eta}(c,\theta)$,

compute $\chi(x,0)$ & $\chi(x,1)$ to determine whether $x \in X_2$.

So use CB: $\approx \frac{\log 1/\delta}{\tau^2}$ samples enough

for $\pm\tau$-acc. est.

---

So, SQ learning is great — gives us RCN PAC "for free".

---

Q: Is every $\mathcal{C}$ that's eff. PAC learnable also eff SQ-learnable?

A: **NO** Parities.

---

Recall : $X = \{0,1\}^n$.

For $S \subseteq [n]$, "parity fn on $S$"

$$PAR_S(x) = \sum_{i \in S} x_i \mod 2$$

$\mathcal{C}_{PAR} = \{\text{all } 2^n \ PAR_S \text{ functions}\}$ $\quad S \subseteq [n]$

$|\mathcal{C}_{PAR}| = 2^n$

Saw (HW): There's an eff. ($\mathrm{poly}(n, \frac{1}{\varepsilon}, \log \frac{1}{\delta})$ time) PAC learner for PAR.

Solving systems of equations mod 2.

Let $a_1, \dots, a_n$ are 0/1 vars

$a_i = 1 \iff i \in S$ $\qquad S =$ target $PAR_S$.

$a_i = 0 \iff i \notin S$

Each ex: a lin eq. in $a_1, \dots, a_n$ mod 2.

$n = 5$

$(\underbrace{1,1,1,0,0}_{x};\ \overset{pos}{1})$ means $a_1 + a_2 + a_3 = 1$

$(00110;\ 0)$ $\qquad a_3 + a_4 = 0$

Can combine / solve: $a_1 + a_2 + a_4 = 1$ mod 2

Solve system of eq's to find consistent $a_1, \dots, a_n$

assignment sat. all eq's: gives

a consistent PAR hypothesis.

So have comput. eff. CHF,

$|\mathcal{C}| = 2^n$

$$m = \frac{1}{\varepsilon}\left(\ln 2^n + \ln 1/\delta\right) \text{ ex,}$$

run our : PAC learns PARITY.

---

Q: how to do an SQ version?

Feels like it very much "looks at individual examples...."

---

<u>Provably</u> impossible for <u>any</u> SQ alg. to eff. learn $\mathcal{C}$ = all parities.

---

Next time: - discuss "uncorrelated" functions.

- if $\mathcal{C}$ contains many uncorrelated functions, no SQ algs can learn $\mathcal{C}$ efficiently.

- New unit: computational/cryptographic hardness of learning.