Last time: Handling RCN.

- Revisit old PAC alg (for mon. conj.) that's not RCN-tolerant, adapt it to handle RCN
- Motivates new learning model: learning from STATISTICAL QUERIES (SQ)

Today: • 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...
- Unconditional lower bounds on SQ learning: some $\mathcal{C}$'s are eff. PAC learnable, but are not eff. SQ learnable.

Questions?

Thm: Let $\mathcal{C}$ be $\boxed{\text{eff.}}$ SQ learnable.

Then $\mathcal{C}$ is eff. PAC learnable.

Pf sketch: Say the SQ alg. makes M calls to STAT$(c, \mathcal{D})$ oracle, & each is with tol. param. $\tau \geq \tau_0$.

$M \leq \text{poly}(\dots)$, $\frac{1}{\tau_0} \leq \text{poly}(\dots)$, each χ query eff. computable.

The PAC alg. simulates each call to STAT$(c, \mathcal{D})$ as described last time, using δ/M as fail. prob.

UB $\Rightarrow$ $\Pr[\text{any sim. of STAT fails}] \le M \cdot \delta/M = \delta$.

This works: $1-\delta$ prob. of ε-acc. hyp.

τ is eff. by .

Now the interesting part:

Thm: Let $\mathcal{C}$ be eff. SQ learnable.

Then $\mathcal{C}$ is eff. PAC learnable even with RCN.

This is useful, b/c for many eff. PAC learnable $\mathcal{C}$'s, the alg. can be adapted to an SQ alg; we get RCN-tol. versions of the algs!

Pf sketch: We'll sketch how to eff. sim. a call to STAT$(c,\mathcal{D})$, given $\underbrace{\text{access to } EX^{\eta}(c,\mathcal{D})}$.

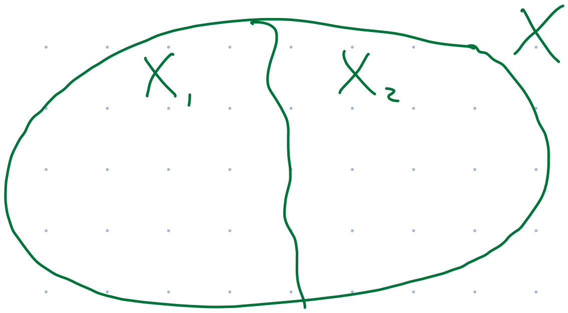
So, we're given a SQ (χ,τ) + .

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

Idea: decompose X (= domain of c) into 2 pieces:

$X_1 = \{x \in X : \chi(x,0) \neq \chi(x,1)\}$ (label always matters; pts where χ is predictably affected by noise)

$X_2 = X \setminus X_1 = \{x \in X : \chi(x,0) = \chi(x,1)\}$ (label never matters; χ is unaffected by noise)



Note: given $x \in X$, can determine whether $x \in X_1$ or X_2:

eval. $\chi(x,0)$, $\chi(x,1)$

<u>Ex:</u> $\chi(x,b) =$ "$b = 1 \,\&\, x_1 = 0$". $\quad (X = \{0,1\}^n)$

For this χ, $X_2 = \{x \in X : x_1 = 1\}$

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

Let $p_1 := \Pr_{x \sim \mathcal{D}}[x \in X_1]$, p_2 similarly.

Let $\mathcal{D}_1$ be $\mathcal{D}$ conditioned on X_1: for any $S \subseteq X_1$

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

<u>Lemma 1:</u> $P_\chi = \Pr_{(x,c(x)) \sim EX(c,\mathcal{D})}[\chi(x,c(x)) = 1]$ is $=$ to

$$p_1 \cdot \frac{\Pr_{(x,b) \sim EX^\eta(c,\mathcal{D}_1)}[\chi(x,b) = 1] - \eta}{1 - 2\eta} + \Pr_{(x,b) \sim EX^\eta(c,\mathcal{D})}[\chi(x,b) = 1 \,\&\, x \in X_2]$$

<u>Lemma 2:</u> It's poss. to eff. estimate P_χ given access to $EX^\eta(c,\mathcal{D})$.

<u>Pf of L1:</u> Have

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

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

$$= \Pr_{(x,b)\sim EX^{\eta}(c,\mathcal{D})}[\mathcal{A}(x,b)=1 \,\&\, x\in X_2]$$

b/c "labels don't matter" on X_2

$$= \underbrace{\Pr_{(x,c(x))\sim EX(c,\mathcal{D})}[x\in X_1]}_{=p_1} \cdot \Pr_{(x,c(x))\sim EX(c,\mathcal{D})}[\mathcal{A}(x,c(x))=1 \mid x\in X_1]$$

$$= \Pr_{(x,c(x))\sim EX(c,\mathcal{D}_1)}[\mathcal{A}(x,c(x))=1]$$

Since $\mathcal{D}_1$ is over X_1, flipping label always changes $\mathcal{A}$: have

$$\Pr_{(x,b)\sim EX^{\eta}(c,\mathcal{D}_1)}[\mathcal{A}(x,b)=1] = \overset{\text{no noise}}{(1-\eta)} \cdot \overset{\text{true pred. value on noiseless ex.} =1}{\Pr_{(x,c(x))\sim EX(c,\mathcal{D}_1)}[\mathcal{A}(x,c(x))=1]}$$

$$+ \overset{\text{noise}}{\eta} \cdot \overset{\text{true pred. value on noiseless ex} = 0}{\Pr_{(x,c(x))\sim EX(c,\mathcal{D}_1)}[\mathcal{A}(x,c(x))=0]}$$

$$= (1-\eta)\cdot \Pr_{(x,c(x))\sim EX(c,\mathcal{D}_1)}[\mathcal{A}(x,c(x))=1] + \eta\left(1 - \Pr_{(x,c(x))\sim EX(c,\mathcal{D}_1)}[\mathcal{A}(x,c(x))=1]\right)$$

$$= \eta + (1-2\eta)\cdot \Pr_{(x,c(x))\sim EX(c,\mathcal{D}_1)}[\mathcal{A}(x,c(x))=1]$$

Solve for , & get

$$\underset{(x,c(x))\sim EX(c,\vartheta)}{\Pr}[\chi(x,c(x))=1 \mid x\in X_1] = \frac{\underset{(x,b)\sim EX^{\eta}(c,\vartheta_1)}{\Pr}[\chi(x,b)=1]-\eta}{1-2\eta}$$

Pf sketch for L2:

Argue that we can estimate

$$\underbrace{p_1\cdot\frac{\underset{(x,b)\sim EX^{\eta}(c,\vartheta_1)}{\Pr}[\chi(x,b)=1]-\eta}{1-2\eta}}+\underbrace{\underset{(x,b)\sim EX^{\eta}(c,\vartheta)}{\Pr}[\chi(x,b)=1 \;\&\; x\in X_2]}$$

given $EX^{\eta}(c,\vartheta)$.

$\underbrace{\quad}$: Can est. p_1: repeatedly draw $(x,b)\sim EX^{\eta}(c,\vartheta)$ & eval. $\chi(x,0)$ & $\chi(x,1)$

Recall $p_1=\underset{x\sim\vartheta}{\Pr}[x\in X_1]$, & $\vartheta_1=\vartheta/X_1$.

Note that $\frac{\underset{(x,b)\sim EX^{\eta}(c,\vartheta_1)}{\Pr}[\chi(x,b)=1]-\eta}{1-2\eta}$ is $\le \frac{1-\eta}{1-2\eta} < \frac{1}{1-2\eta}$.

So if p_1 very small (say, $\le \kappa\cdot(1-2\eta)$), then $\underbrace{\quad}$ is $<\kappa$, & don't need to try to est. the quantity.

O/w, p_1 not too small, & given $EX^\eta(c,\vartheta)$, with $\frac{1}{p_1}$ slowdown, can simul. draws from $EX^\eta(c,\vartheta_1)$, & hence est.

: Can est. $\Pr_{(x,b)\sim EX^\eta(c,\vartheta)}[\chi(x,b)=1 \;\&\; x\in X_2]$:

draw $(x,b)\sim$, eval. $\chi(x,0)$, $\chi(x,1)$

check =

& that $\chi(x,b)=1$.

So can simulate $STAT(c,\vartheta)$ given $EX^\eta(c,\vartheta)$.

(Lemma 2).

Note:

If we worked out # ex. required for all est. to be acc. enough, we'd see need est.

to have error $< (1-2\eta)$, hence need at least $\left(\frac{1}{1-2\eta}\right)^2$ rand ex from $EX^\eta(c,\vartheta)$.

- Details on est. accuracy: 5.4.2 of book

- We pretended we know n. Can avoid: run alg. with guesses
$n=0$
$n=10$
$n=20$
$n=30$
$\vdots$
some guess almost right

Natural question:

Is every $\mathcal{C}$ that is efficiently PAC learnable, also eff. SQ learnable?

NO.

$X = \{0,1\}^n$

Define $\mathcal{C}_{PAR}$ = all 2^n parity functions $S \subseteq [n]$

$PAR_S : \{0,1\}^n \to \{0,1\}$

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

e.g. $PAR_{\{1,3,4,5\}}(x) = x_1 + x_3 + x_4 + x_5 \mod 2$.

Claim: There is a $\mathrm{poly}(n, \frac{1}{\varepsilon}, \log\frac{1}{\delta})$-time PAC learner for $\mathcal{C}_{PAR}$.

<u>Pf sketch:</u> Solve linear equations mod 2.

For $i \in [n]$, let $a_i = \begin{cases} 0 & \text{if } x_i \text{ not in target } PAR_S \ (i \notin S) \\ 1 & \text{if } x_i \text{ is in } PAR_S \ (i \in S) \end{cases}$

Each ex. gives a lin. eq. on a_i's:

e.g. $n=5$

$$(\underbrace{11001}_{x};\ \underbrace{\overset{\text{pos ex}}{1}}_{c(x)}) \rightsquigarrow a_1 + a_2 + a_5 = 1$$

$$(11000;\ 0) \rightsquigarrow a_1 + a_2 = 0$$

Eff.

Solving syst. of lin. eq. (obtained from the examples)

$\rightsquigarrow$

eff CHF for $\mathcal{C}_{PAR}$ using $\mathcal{C}_{PAR}$, so can PAC learn.

Next time: no $\overset{\text{eff}}{\text{SQ}}$ alg. for $\mathcal{C}_{PAR}$.
