Admin: • Pick up HW 1 solutions (by door)  
- Reading: 1.1.-1.3 of KV; continue, read all of Chap. 1, 2.  
- OH changes: Rocco OH: WED 9-11.

Do have in-person class Thurs 10/4
Don't ) 1 1 Tues 10/9 (video).

Last time: started PAC model. \(\epsilon, \delta, m = \text{sample complexity} \)  

Today: Online-to-PAC conversion  
- Revisit def. of PAC learning, Markov's ineq.
- Basic tools from prob.: concentration ineq., etc.

OLMB \rightarrow PAC conversion.

Let \( E \) be c.c., suppose we have an OLMB alg. \( A \) for \( E \) w/ m.b. 'M'. Can we get alg for \( E \) in PAC?  

\[ \text{YES} \]

Then: Let \( A \) be OLMB alg for \( E \) w/ m.b. 'M'.  

Then there is a PAC alg. for \( E \) w/ \( m \) sample complexity  

\[ m = M + \frac{M+1}{\epsilon} \cdot \ln \left( \frac{M+1}{\delta} \right). \]

PF: We'll assume (WLOG!) that our OLMB alg \( A \) only changes hyp. when makes mistake.  
The PAC alg: Run \( A \) on seq. of lab. ex. indep. obtained from \( E X(c, \rho) \). If current hyp (h) ever
correctly classifies \( \frac{L \ln \left( \frac{M+1}{\delta} \right)}{\epsilon} \) ex. from EX(0, 0)

in a row, stop and output \( h_i \).

# ex used is \( \leq \frac{1}{\epsilon} \ln \left( \frac{M+1}{\delta} \right) \) \( \leq M \) mistakes.

\[ 
\sqrt{\ldots} \cdot \sqrt{\ldots} \cdot \sqrt{\ldots} \cdot \sqrt{\ldots} \cdot \sqrt{\ldots} \cdot \sqrt{\ldots} \\
\text{length} \leq \frac{1}{\epsilon} \ln \left( \frac{M+1}{\delta} \right) \leq M \text{ mistakes.}
\]

Why is \( A \)'s output hyp. \( \epsilon \)-acc. w.p. \( \geq 1 - \delta \)?

Consider \( i \)-th hyp. \( A \) uses \( h_i \).

If \( \text{err}_x \left( h_i, c \right) \leq \epsilon \): \( C \)

If \( \text{err}_x \left( h_i, c \right) > \epsilon \): \( \Pr \left[ h \text{ gets } \frac{1}{\epsilon} \ln \left( \frac{M+1}{\delta} \right) \text{ ex. right in a row} \right] \)

\[ < (1 - \epsilon)^k = (1 - \epsilon)^{1 - \ln \left( \frac{M+1}{\delta} \right)} = \frac{1}{\epsilon} \ln \left( \frac{M+1}{\delta} \right) \leq \frac{\delta}{M+1} = \frac{\delta}{M+1}.
\]

Since \( A \) only uses \( \leq M+1 \) hyp's \( h_1, \ldots, h_{M+1} \),

\( \forall B \Rightarrow \Pr \left[ \text{any hyp } A \text{ uses has error } > \epsilon \\
\text{ and is output} \right] \leq \left( M+1 \right) \cdot \frac{\delta}{M+1} = \delta. \)
So, $Pr[C A] = \frac{\text{hyp. with error } \leq \epsilon}{1 - \delta}$.

So, have PAC alg for \text{disj}, \text{sparse disj}, \text{conj}, \text{DC}, various LTF types, etc.

\[ \text{Ex: } \text{Disj: Elim. alg: MB } n, \text{ so have PAC alg. s.c. } O\left(\frac{n}{\epsilon} \cdot \log \left(\frac{n}{\delta}\right)\right) \]

\[ \Rightarrow \text{OLMB } \Rightarrow \text{PAC.} \]

\[ \Rightarrow \text{Q: } \text{Does PAC } \Rightarrow \text{OLMB?} \]

\[ \Rightarrow \text{No: see have PAC alg. for } C = \{\text{all intervals of [0,1]}\}, \text{ but know no OLMB alg. w/ finite MB for } C. \]

What about finite domains?

Q: Let $C = \text{c.c. over } \{0,1\}$. For $C = \text{poly}(n)$ time PAC alg for $C \Rightarrow \text{poly}(n)$ time OLMB alg for $C$?

Under suitable assumption from crypto: NO

Idea: View $\{0,1\}^{n}$ as $\{1,2,\ldots,2^n\}$.

Via crypto, can come up with "one-way flashlight" function $c$ on $\{1,2,\ldots,2^n\}$. 

\[ \text{Final words:} \]
If given \( c(i) \): easy to compute \( c(i) \) for \( i \leq j \) but computationally hard to compute \( c(k) \) for \( k > j \).

**OLMB model:** adv. gives 1, 2, ..., always hard to compute \( c(i) \) on next pt.

**PAC model:** Draw 1000 examples:

\[ \rightarrow \text{have high prob. that some ex in these draws is in top 1\% (0.1\%) of } \mathbf{Y}, \text{ no matter what } \mathbf{Y} \text{ is.} \]

Back to PAC def: two issues.

\#1: \textit{"size" of concept } \( c \).

- Some cc's: \( c \) natural to assign a \textit{size}, \( \text{size}(c) \), to each \( c \in C \).
  - \( \text{size}(c) \) measures "how } c \text{ is } c".

Basically, \# bits needed to describe \( c \).

**Ex:** \( C = \text{ all ONF formulas over } \{0,1\}^n \).

Any \( f: \{0,1\}^n \rightarrow \{0,1\} \) can be written as a ONF.

Some \( f \)'s: can \( \Rightarrow \) ONF with
few terms. \( f(x) = x, x_2, x_3, \overline{x_1} x_4 x_5 x_6 \)

2-term ONF.

Other hideous \( f \) needs \( 10^{100} \) terms.

We view size \( (\cdot) \) as size of smallest representation of the function in \( C \).

\( C = \text{ONFs} \)

\( f(x) = x, x_2, \overline{x_3} x_4 \) size = 2

2-term ONF.

\[
\begin{align*}
&x_1 x_2 x_7 \lor x_1 x_2 \overline{x_7} \lor x_3 x_4 x_8 x_9 \lor x_3 x_4 \overline{x_8} \overline{x_9} \\
&\phantom{=} \lor x_3 x_4 x_8 \overline{x_9} \lor x_3 x_4 \overline{x_8} x_9
\end{align*}
\]

6-term ONF

Ex: \( C = \text{all decision trees over } 0, 1^n \)

DT size of a function = size (\#leaves) of smallest DT computing the fn.
Refined notion of efficient PAC learning: if \( C \) has size assoc. \( \leq \) it, \( (X=\mathbb{R}^n, 50/15) \) + target fn has size \( c = 5 \), 
efficient PAC alg: should run in time 
\[ \text{poly}(\frac{n}{\varepsilon}, \frac{1}{\delta}, 5). \] (isn't necessary) 
Alg gets all these as inputs.

For some \( C \)'s, not really interesting to consider size.

\( C = \text{conj: 05; size } \leq n \)

Issue #2: PAC learning \( C \not\subset \text{hyp. class } \mathcal{H} \).

\( \mathcal{H} \): a class of programs. If PAC alg \( A \) outputs a \( h \in \mathcal{H} \) where program \( h \) is very inefficient: useless.

Say a hyp. class \( \mathcal{H} \) is polynomially evaluable \( (X=\mathbb{R}^n \text{ or } \{0,1\}^n) \) if every \( h \in \mathcal{H} \) is a program which, on any input \( x \in X \), runs in \( \text{poly}(n) \) time.

Right notion of eff. PAC learning:
must use polynomial eval. hyp. class \( \mathcal{H} \).

(If you don't require this, easy but useless PAC alg. for \( \approx \) any \( E \): "offloads all comput. onto \( h \).")

---

Set stage: (Read: Appendix of KV book)

(Chernoff, Hoeffding, Markov)

Motiv: Sps PAC universe.

Have \( h : X \to \{0,1\} \), \( \text{Have } \mathbb{E}(c, \Theta) \).

? how good is \( h \)?

Get \( (x', c(x')) \cdots (x^{100}, c(x^{100})) \)

Eval. \( h(x') \)

\[
\tilde{P} = \frac{\# \mathcal{H}}{m}
\]

Next time:

\[
\hat{p} = \frac{\# \mathcal{H}}{m}
\]
let us discuss acc., conf. of estim. p.