Last time:
- Randomized Weighted Majority alg (RWM)
- Start next unit:

Probably Approximately Correct (PAC) Learning Model

- motivation
- basic definition
- learning $\mathcal{C}$ = intervals over $X = \mathbb{R}$ (start)

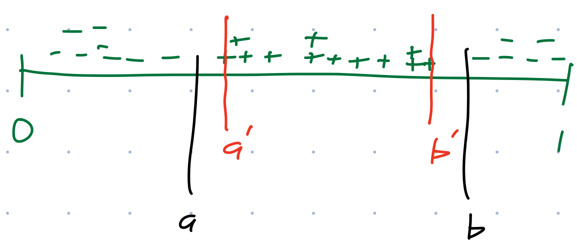
Today:
- Finish PAC learning $\mathcal{C}$ = intervals over $X = \mathbb{R}$
- OLMB $\Rightarrow$ PAC conversion: get all our OLMB results in PAC model :)
- Revisit PAC learning definition:
  - "size" of concepts
  - efficiency of evaluating hyp. h
- "Chernoff bounds" (tail bounds on sums of indep. random variables) & applic. to hypothesis testing

Admin:
- ps1 sol. out tonight
- graded ps1 back soon

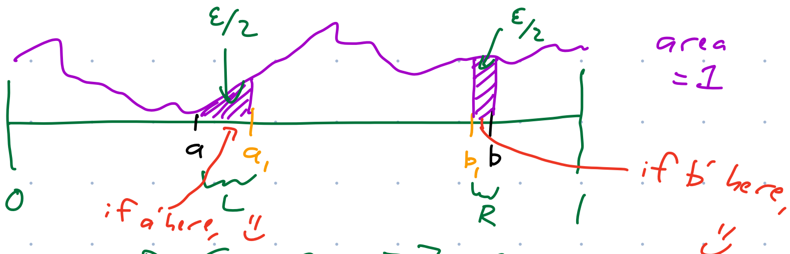
Questions?

Recall our int. learning problem:

Data POV:



"Analysis POV":



Define a_1 = value s.t. $\Pr_{x\sim\mathcal{D}}[x \in \underbrace{[a, a_1]}_{L}] = \varepsilon/2$

b_1 = value s.t. $\Pr_{x\sim\mathcal{D}}[x \in \underbrace{[b_1, b]}_{R}] = \varepsilon/2$.

Obs: if got
- an ex. in $[a, a_1]$, +
- an ex. in $[b_1, b]$,

then error of our hyp. is $\le \varepsilon/2 + \varepsilon/2 = \varepsilon$ ☺.

$\Pr_{x\sim\mathcal{D}}[x \text{ misses } L] = 1 - \varepsilon/2$

So $\Pr[m \text{ indep. ex. } \underline{\text{all}} \text{ miss } L] = (1-\varepsilon/2)^m$

Likewise R $= (1-\varepsilon/2)^m$.

So as long as $\left(1-\frac{\varepsilon}{2}\right)^m \le \frac{\delta}{2}$,

have $\Pr[\text{some ex. hit } L, \text{ and some ex. hit } R] \ge 1 - 2\cdot\frac{\delta}{2} = 1-\delta$.

Recall $1-x \le e^{-x}$, so

$(1-\varepsilon/2)^m \le e^{-\frac{\varepsilon}{2}\cdot m}$; to have this $\le \frac{\delta}{2}$, take $m = \frac{2}{\varepsilon} \cdot \ln(2/\delta)$.

For *indep.* events A, B,

$$\Pr[A \,\&\, B] = \Pr[A] \cdot \Pr[B].$$

For *any* events A, B,

$$\Pr[A \text{ or } B] \leq \Pr[A] + \Pr[B].$$

Note: · δ controls prob. that *sample* is "bad"/atypical;

· ε controls (inevitable) small error.

OLMB to PAC Conversion

Say A is an OLMB alg. w/ mist. bd. M for $\mathcal{C}$.

Can *automatically* get PAC alg for $\mathcal{C}$:

Thm: If , then there's a PAC learner for $\mathcal{C}$ with sample complexity

$$\boxed{m \leq M + \frac{M+1}{\varepsilon} \cdot \ln\left(\frac{M+1}{\delta}\right).}$$

(Comput. efficiency also preserved.)

Pf: Say OLMB alg A is <u>conservative</u> if it only updates its h when makes mistake.

Claim: If $\exists A$, then $\exists$ <u>conservative</u>.

Justif.: The conserv. version, A', simply ignores ex. (doesn't update h) on which no mistake.
This can't make $M+1$ mist. on any seq., or else A would too. $\rightarrow\!\!\leftarrow$. ∎

So, we can assume A is conservative.

A natural PAC alg:

Run A on seq. of indep. ex. from $EX(c, \mathcal{D})$.
Keep track, for each hyp. h that is current hyp, of # examples in a row it got right.
If h gets $\frac{1}{\varepsilon} \cdot \ln\left(\frac{M+1}{\delta}\right)$ ex. right in a row, stop & output that h.

Note: this alg. uses $\leq M + \frac{M+1}{\varepsilon} \cdot \ln\left(\frac{M+1}{\delta}\right)$ ex:

$$\underbrace{\checkmark\ \checkmark\ \checkmark \dots \checkmark}_{\leq \frac{1}{\varepsilon}\cdot\ln\left(\frac{M+1}{\delta}\right)}\ \times\ \checkmark \dots \checkmark\ \times\ \checkmark \dots \checkmark\ \times\ \dots\ \underset{M^{th}\text{ mistake}}{\times}\ \checkmark \dots \checkmark$$

h right

h wrong; new h

$\leq M$ $\times$'s (mistakes

$\leq M+1$ many "runs" of $\leq \frac{1}{\varepsilon}\cdot\ln\left(\frac{M+1}{\delta}\right)$ $\checkmark$'s

Remains to show: $\Pr[\text{output } (1-\varepsilon) \text{ acc. hyp}] \geq 1-\delta$.

Say a hyp. h is <u>bad</u> if $er_{\mathcal{D}}(h,c) > \varepsilon$.

We'll argue: $\Pr[\text{our alg. outputs a bad } h] \leq \delta$; then done.

Fix <u>any</u> bad h.

$$\Pr_{x\sim\mathcal{D}}[h(x) \neq c(x)] > \varepsilon, \text{ i.e. } \Pr_{x\sim\mathcal{D}}[h(x) = c(x)] < 1-\varepsilon.$$

So $\Pr[h \text{ gets } k \text{ ex. right in a row}] < \underbrace{(1-\varepsilon)^k}_{< \frac{\delta}{M+1}}$.

For $k = \frac{1}{\varepsilon} \cdot \ln\left(\frac{M+1}{\delta}\right)$,

$(1-x \leq e^{-x})$

Since $\leq M+1$ distinct h's <u>ever</u> generated by the alg A (conservative!), by UB over all <u>bad</u> hyp's A generates ($\leq M+1$ many),

$$\Pr[\underline{\text{any}} \text{ bad } h \text{ is output}] \leq (M+1) \cdot \frac{\delta}{M+1} = \delta. \blacksquare$$

So we have PAC alg's for all our specific $\mathcal{C}$'s from OCMB unit.

<u>Ex</u>: can PAC learn conj. over $\{0,1\}^n$ with

$$m \le n + \frac{n+1}{\varepsilon} \cdot \ln\left(\frac{n+1}{\delta}\right) = O\left(\frac{n}{\varepsilon} \cdot \ln\left(\frac{n}{\delta}\right)\right)$$

examples.

OLMB $\Longrightarrow$ PAC.

PAC $\overset{\text{does not}}{\Longrightarrow}$ OLMB:

$\mathcal{C}$ = intervals
- is PAC learnable
- is not OLMB learnable w/ finite mistake bd.

Refine our notion of PAC learning, by considering
- (1) "size"/"complexity" of target c;
- (2) efficient evaluability of hyp. h.

---

(1): Often natural to have a notion of "size" or "cxity" of concepts $c \in \mathcal{C}$.

Ex: size k of a disj. over $\{0,1\}^n$

Ex: $\mathcal{C}$ = {all DNF formulas over $\{0,1\}^n$}.

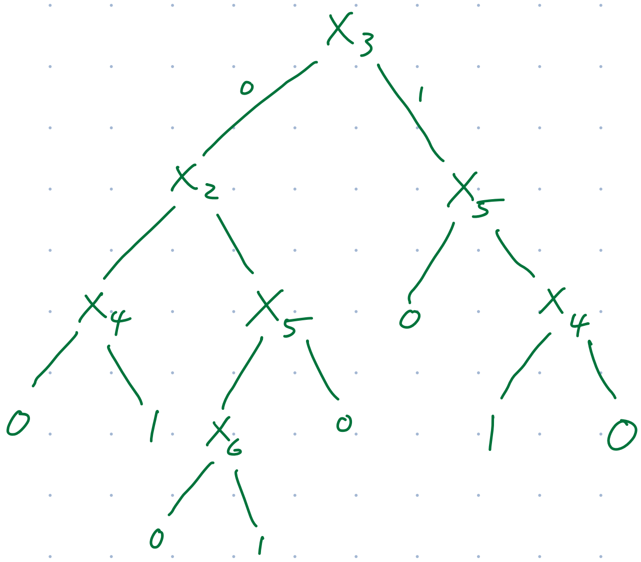
Can define DNFSIZE(f), for any $f: \{0,1\}^n \to \{0,1\}$, as min # terms in any DNF computing f.

(dream)

A DNF learning alg would be allowed to run in time poly$(n, \underline{\text{DNFSIZE}(c)}, \frac{1}{\varepsilon}, \frac{1}{\delta})$.

Other classes also have associated size param's as well:

Ex: a Boolean decision tree is a fn $f: \{0,1\}^n \to \{0,1\}$ like



size = # leaves

depth = depth

---

②: What hyp. classes $\mathcal{H}$ are reasonable?

$\mathcal{C}$ = class of functions

$\mathcal{H}$ = class of representations of functions (i.e. of programs)

$h = x_3 \vee x_4 \vee x_7$

If a learning alg outputs a hyp. h that takes very long to run, not useful...

We say a hyp. class $\mathcal{H}$ over $\{0,1\}^n$ or $\mathbb{R}^n$ is polynomially evaluatable if every $h \in \mathcal{H}$ can be evaluated, on every x, in poly(n) time.

An efficient PAC learner must use a .

Or: view runtime of PAC learner as

- time to construct h, +
- time to evaluate h.

Tail bounds on sums of indep. RV's, & "hypothesis testing".

PAC scenario: have a hyp. h,
have $EX(c,\mathcal{D})$.

Natural q: how good is h? what is $er_{\mathcal{D}}(h,c)$?

Natural answer:

- draw m ex from $EX(c,\mathcal{D})$
  compute $h(x)$ on each.

guess for is $\dfrac{(\text{\# ex. of the } m \text{ that } h \text{ got wrong})}{m}$

How good is this estimate of ?

$\equiv$ to:

Have \$ coin w $Pr\{\text{Heads}\} = p$ unknown

To estimate p: toss coin m times, output $\dfrac{(\#H \text{ you got})}{m}$.

How good is this estimate?

Next time: tools for this.