Last time: - 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
- Start tail bounds

Today:
- "Chernoff bounds" (tail bounds on sums of indep. random variables) + applic. to hypothesis testing
- learning by finding consistent hyp. from a fixed class $\mathcal{H}$ "CHF"
- "Occam's Razor" ("cardinality version")

Questions?

Sol to ps 1 :) don't use AI :)

$\boxed{\text{Check Ed Discussion}}$

Tail Bounds on sums of indep. RV's

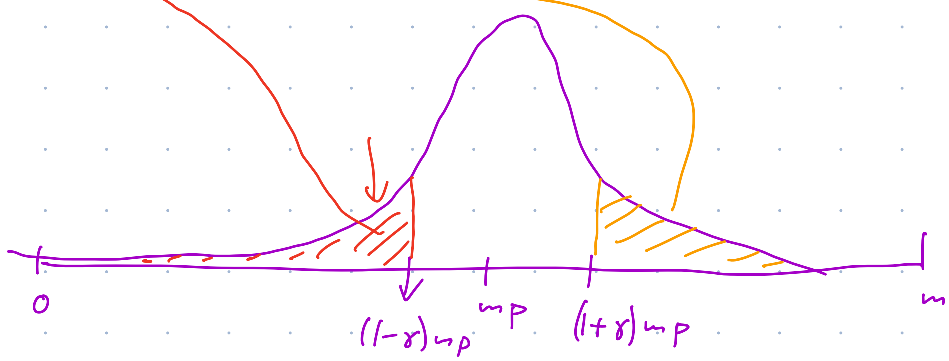
"Chernoff Bound": Let $X_1, \ldots, X_m$ be independent 0/1-valued RV's with $\Pr[X_i = 1] = p \quad \forall i.$
Let $X = X_1 + \cdots + X_m$, so $\mathbb{E}[X] = mp$

$\forall\ 0 \le \gamma \le 1$,

"relative error"

$$\Pr[X \le (1-\gamma)mp] \le \exp\left(-\tfrac{1}{2}\cdot\gamma^2\cdot mp\right), \quad +$$

$$\Pr[X \geq (1+\gamma)mp] \leq \exp\left(-\tfrac{1}{3}\gamma^2 \cdot mp\right).$$



"absolute error" version:

Let $X_1, \dots, X_m$ as above; let $\hat{p} = \frac{X}{m}$ (= empirical estimate of p)

Then for $\varepsilon > 0$

$$\Pr[p - \hat{p} \geq \varepsilon] \leq \exp(-2m\varepsilon^2), \quad \&$$

$$\Pr[\hat{p} - p \geq \varepsilon] \leq \exp(-2m\varepsilon^2).$$

Ex 1: Say team A wins each game w.p. $\frac{1}{2}$, indep. 162 games.

What's $\Pr[A \text{ wins} \leq 47 \text{ games}]$?

Let $X_i = \begin{cases} 1 & \text{win game } i \\ 0 & \text{o/w} \end{cases}$. $m = 162$, $p = \frac{1}{2}$

$$\Pr[X \leq 47] = \Pr[X \leq (1-\gamma)\cdot \overbrace{mp}^{81}] \leq \exp\left(-\tfrac{1}{2}\cdot\gamma^2\cdot mp\right)$$

$(1-\gamma)81 = 47$.

$\gamma \approx 0.42$:

$$\exp\left(-\tfrac{1}{2}\gamma^2 \cdot mp\right) \approx 0.0008$$

Ex: Design survey : goal: estimate frac. of population who support Candidate B. Y/N,

How many people do we need to survey?

want result acc. to $\pm 3\%$ with confidence 95%.

Additive bd: $\varepsilon = 0.03$, want $\exp(-2m\varepsilon^2) = 2.5\%$

so $e^{-0.0018m} = 0.025$

$$\Downarrow \quad m = \frac{1}{0.0018} \cdot \ln(40)$$

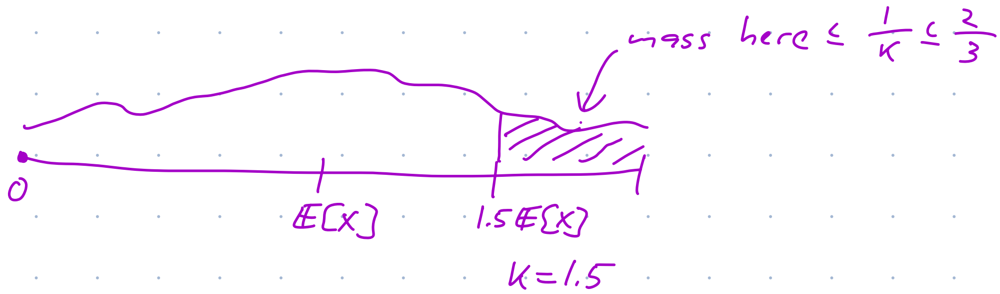
:) Strong! Expon convergence.

:(Strong assumption: X = sum of independent RV's.

What can we say for less structured X's?

"Markov's inequality": Let X be non-negative RV. Then for $k > 1$,

$$\Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}.$$



<u>Ex:</u> Let X = # children in random US household.

Say $\mathbb{E}[X] = 1.85$.

Then $\Pr[X \geq 10]$ must be $\leq \frac{1}{5}$

(o/w avg would be ≥ 2!).

$\quad\quad\quad\quad\quad\quad\quad\quad C \quad\quad\quad\quad H$

<u>Learning by Finding Consistent Hyp's:</u>

New way to PAC learn: find a CH from some a priori fixed $\mathcal{H}$.

<u>Thm:</u> Fix a $\mathcal{C}$ + an $\mathcal{H}$. Fix any $c \in \mathcal{C}$.
Fix any dist $\mathcal{D}$.

Let $(\underset{\text{in } X}{x^1}, \underset{0/1}{c(x^1)}), \ldots, (x^m, c(x^m))$ be i.i.d. draws from $EX(c,\mathcal{D})$,

where $m \geq \frac{1}{\varepsilon}\left(\ln|\mathcal{H}| + \ln\frac{1}{\delta}\right)$.

Say an $h: X \to \{0,1\}$ is <u>bad</u> if $err_{\mathcal{D}}(h,c) > \varepsilon$.

Then

$$\Pr\{\underline{\text{any}} \text{ bad } h \in \mathcal{H} \text{ is consistent w/ } \} \le \delta.$$

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

Have

$$\Pr\{h \text{ cons. with } m \text{ lab. ex. from } EX(c, \mathcal{D})\} < (1-\varepsilon)^m.$$

$\mathcal{H}$ has $\le |\mathcal{H}|$ bad h's; so by union bd.,

$$\Pr\{\underline{\text{any}} \text{ bad } h \in \mathcal{H} \text{ is consistent w/ } m \text{ lab. ex. from } EX(c, \mathcal{D})\} \le |\mathcal{H}| \cdot (1-\varepsilon)^m.$$

This is $\le \delta$, b/c $m = \frac{1}{\varepsilon} \cdot \left(\ln |\mathcal{H}| + \ln 1/\delta\right)$.

$$|\mathcal{H}| \cdot (1-\varepsilon)^m = |\mathcal{H}| \cdot \left(\left(1-\varepsilon\right)^{\frac{1}{\varepsilon}}\right)^{\ln(\mathcal{H}) + \ln(1/\delta)}$$

$$\le |\mathcal{H}| \cdot e^{-\left(\ln\left(\frac{|\mathcal{H}|}{\delta}\right)\right)}$$

$$= |\mathcal{H}| \cdot \frac{\delta}{|\mathcal{H}|}. \quad \blacksquare$$

Interpreting / using it:

Fix $\mathcal{C}, \mathcal{H}$.

<u>Def</u>: A <u>consistent hyp. finder for $\mathcal{C}$ using $\mathcal{H}$</u> (CHF) is an alg B: $\forall m$, if B is given

m ex. $\left\{ (x^1, c(x^1)), \; \vdots \; , (x^m, c(x^m)) \right\}$ all labeled acc. to some $c \in \mathcal{C}$,

$$\longrightarrow \boxed{B} \longrightarrow h$$

then B outputs some $h \in \mathcal{H}$ s.t. $h(x^i) = c(x^i) \; \forall i \in [m]$

Means: if have CHF for $\mathcal{C}$ using $\mathcal{H}$, can use it as (ε, δ) PAC learner for $\mathcal{C}$ using $\mathcal{H}$: run it on $m \geq \frac{1}{\varepsilon} \cdot (\ln |\mathcal{H}| + \ln 1/\delta)$ ex from $EX(c, \mathcal{D})$.

Ex: $\mathcal{C}$ = class of mon. disj.
$\mathcal{H} =$ 〃

Recall our OLMB alg for : elim.
This is a CHF for $\mathcal{C}$ using $\mathcal{H}$.

So thm says: run this on $m = \frac{1}{\varepsilon} \cdot \ln\left(\frac{|\mathcal{H}|}{\delta}\right)$ ex, it's a PAC learner.

Last time: our PAC sample cxity from OLMB $\Rightarrow$ PAC conv. was

$$M + \frac{M+1}{\varepsilon} \cdot \ln\left(\frac{M+1}{\delta}\right) \qquad M = n$$

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

Now: $m = \frac{1}{\varepsilon} \cdot \ln\left(\frac{|\mathcal{H}|}{\delta}\right)$, $|\mathcal{H}| = 2^n$

↳ so $m = O\left(\frac{n}{\varepsilon} \cdot \ln\left(1/\delta\right)\right)$.

Emphasize: • if have CHF, get PAC learner

big "if"

• Does not mean building a lookup table to agree with n examples, is a succ. PAC learner: no fixed $\mathcal{H}$.

Refinement: "Occam's Razor"

"Entities should not be multiplied unnecessarily"

— 1320 William of Ockham

↳ there are few short explan: unlikely any of them which are not actually good, will do well.

One view on CHF thm:

$h \in \mathcal{H}$ is a "short explanation" of the data.

$$\left.\begin{array}{c} (x^1, c(x^1)) \\ \vdots \\ (x^m, c(x^m)) \end{array}\right\} :$$

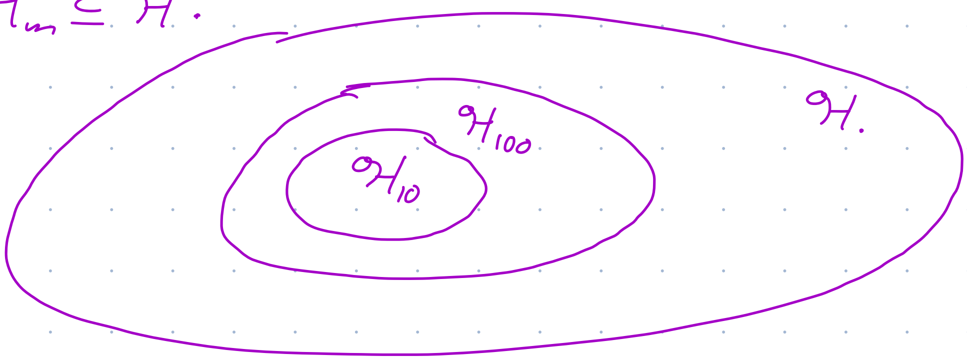
• m bits of data to "explain" (labels).

• Any $h \in \mathcal{H}$ can be encoded/described with $\log|\mathcal{H}|$ bits of info.

So if $m \gg \log|\mathcal{H}|$: h is a "short explation" of the data.

Prev thm: For all m, CHF gives $h \in \mathcal{H}$ consistent.

Maybe: for small m, can find a consistent $h \in \mathcal{H}_m \subseteq \mathcal{H}$.



If so, can take advantage.

Another version of prev. thm:

"Occam's razor, cardinality version": Fix m.
Suppose there's a $\mathcal{H}_m \subseteq \mathcal{H}$ of hyp's, where

$$m \geq \frac{1}{\varepsilon}\left(\ln|\mathcal{H}_m| + \ln 1/\delta\right)$$

$|\mathcal{H}_m| \leq \delta \cdot e^{\varepsilon m}$

st. given any set S of m ex. lab. by some $c \in \mathcal{C}$, there's an $h \in \mathcal{H}_m$ consistent.
Let L be an alg. which, given such an S of size m,

outputs a cons. $h \in \mathcal{H}_m$.
Then L run on m ex from $EX(c, \mathcal{D})$
is an (ε, δ) PAC learner.

Pf: same as before!