Last time: part 2 of 3-part pf that if VCDIM($\mathcal{C}$)=d, then any CHF for $\mathcal{C}$ using $\mathcal{C}$ as hyp class, run on $m \approx \frac{1}{\varepsilon}(d \log \frac{1}{\varepsilon} + \log 1/\delta)$ examples from $EX(c, \mathcal{D})$, is an (ε, δ)-PAC learner.

Recall: ① · Setup: Ponder VCDIM($\mathcal{C}$). ← a #
We'll consider a function "growth function of $\mathcal{C}$" $\equiv \Pi_{\mathcal{C}}(m)$ } 2 lectures ago
↳ to get more info.

---

② Combinatorics arg: AMAZING THEOREM about how the growth fn behaves, for any $\mathcal{C}$.

last time: $\boxed{\Pi_{\mathcal{C}}(m) \le \left(\frac{em}{d}\right)^d \text{ if } m \ge d}$

---

③ Learning argument: a lot like CHF thm, but using AMAZING THM to control # ways conc in $\mathcal{C}$ can label a data set.

Today: • + finish pf of

• Application: efficient PAC learning of $\mathcal{C}$ = LTFs over $X = \mathbb{R}^n$ or $\{0,1\}^n$

• (maybe) Start unit on boosting confidence + accuracy.

Questions?

Recall: say concept/hyp. $h \in \mathcal{C}$ is bad if

$$\Pr_{x \sim \mathcal{D}}[h(x) \neq c(x)] > \varepsilon, \quad c = \text{target concept in } \mathcal{C}$$

Recall our old CHF thm: (spec. to $\mathcal{H}=\mathcal{C}$)

Thm: Fix any finite c.c. $\mathcal{C}$, any target $c \in \mathcal{C}$, any dist $\mathcal{D}$.
Given m samples from $EX(c,\mathcal{D})$, where

$$m \geq \frac{1}{\varepsilon}\left(\ln|\mathcal{C}| + \ln 1/\delta\right), \qquad \star$$

w.p. $\geq 1-\delta$ all bad $h \in \mathcal{C}$ are inconsistent w/ the m samples.

(key to pf: union bound over all $h \in \mathcal{C}$)

Want analogue for $\mathcal{C}$ infinite.
We'll prove:

Key Thm: Fix any (may be ∞!) c.c. $\mathcal{C}$, any target $c \in \mathcal{C}$, any dist $\mathcal{D}$.
Given m samples from $EX(c,\mathcal{D})$, where

$$m \geq \frac{2}{\varepsilon}\left(\log\left(\Pi_{\mathcal{C}}(2m)\right) + \log^2 1/\delta\right), \quad m \geq 8/\varepsilon \qquad \star$$

w.p. $\geq 1-\delta$ all bad $h \in \mathcal{C}$ are inconsistent w/ the m samples.

(Key to pf: union bound over all poss. labelings $\mathcal{C}$ can induce on a $2m$-elt sample.)

So if m sat. ★, using a CHF on m ex. for $\mathcal{C}$ using $\mathcal{C}$ is an ε,δ-PAC learner.

What's up with

$$\boxed{m \geq \frac{2}{\varepsilon}\left(\ln\left(\Pi_{\mathcal{C}}(2m)\right)+\ln \tfrac{2}{\delta}\right), \quad m \geq 8/\varepsilon \quad ★}$$?

if didn't have Amazing Thm (AT), we'd only know $\Pi_{\mathcal{C}}(2m) \leq 2^{2m}$ + $m > d$. If it were $\frac{1}{2}\cdot 2^{2m}$, could never hold:

says $\boxed{m \geq \frac{2}{\varepsilon}\cdot\left(2m-1+\ln \tfrac{2}{\delta}\right)}$

impossible!

But thanks to AT, have control on $\Pi_{\mathcal{C}}(2m)$: it's $\leq \left(\frac{e\cdot 2m}{d}\right)^d$, so holds for m "not too large".

Pf of Key Thm:

Consider doing the following:
Draw $2m$ samples from $EX(c,\mathcal{D})$.

Event A: there exists a bad $h \in \mathcal{C}$ that's consistent w/ S_1. (Goal: show $Pr[A]$ low. $\leq \delta$)

Event B: there is some $h \in \mathcal{C}$ s.t.
- h is consist. w/ S_1, but
- h makes $\geq \frac{\varepsilon}{2} \cdot m$ mist. on S_2

Claim: For $m > 8/\varepsilon$, have $Pr[A] \leq 2 \cdot Pr[B]$.
↳ given this: rest of effort is on showing $Pr[B] \leq \delta/2$.

Pf: $Pr[B] \geq Pr[A \,\&\, B] = Pr[A] \cdot Pr[B|A]$

So: enough to show $Pr[B|A] \geq \frac{1}{2}$.

Since A happens, there's some bad $h^* \in \mathcal{C}$ cons. w/ S_1.

Bad! fresh batch of ex

$Pr[h^*$ makes $\geq \frac{\varepsilon}{2} m$ mist. on $S_2]$: high!

CB: $\mathbb{E}[\#$ mist. of h^* on $S_2] \geq \varepsilon m$

so $\Pr\{h^*$ makes $\overbrace{\text{fewer than half}}^{(1-\gamma),\ \gamma=\frac{1}{2}}$ its exp. # of mist. on m pts $\}$

$\le \boxed{e^{-\varepsilon m \gamma^2/2}} \longrightarrow$ This is $\le \frac{1}{2}$ as long as $m \geq 8/\varepsilon$. ■ (claim)

Rest of pf: show $\Pr\{B\} \le \delta/2$.

B:

Can view B as follows (equiv. to original descrip.):

- Draw set S of $2m$ examples.
- Randomly split S into = sized S_1 + S_2.

B = event that there's some $h \in \mathcal{C}$ • cons. with S_1, +
• wrong on $\geq \frac{\varepsilon}{2} m$ pts in S_2.

Fix a <u>partic.</u> labeling of the $2m$ pts in S (which labels $\geq \frac{\varepsilon}{2} m$ of them wrongly).

Now we'll argue

$$\Pr_{\substack{\text{split of } S \\ \text{into } S_1 + S_2}}\{\text{this labeling puts all wrong pts in } S_2\}$$

very small:

Same as:

2m balls in barrel, at least $k = \frac{\varepsilon}{2} m$ are black, rest white.

Split balls in barrel into m + m randomly.

$\Pr\{\text{all } k \text{ black balls end up in 2nd } m\}$

$$= \binom{m}{k} \Big/ \binom{2m}{k}$$

$$= \frac{m!}{(m-k)!\,k!} \Big/ \frac{(2m)!}{(2m-k)!\,k!}$$

$$= \frac{m(m-1)\ldots(m-k+1)}{2m(2m-1)\ldots(2m-k+1)} \le \frac{1}{2^k}. \qquad k = \frac{\varepsilon m}{2}$$

So each <u>fixed</u> lab. of 2m ex has $\le \frac{1}{2^{\frac{\varepsilon m}{2}}}$ chance of giving us B.

UB over all ($\leq \Pi_e(2m)$!!)

labelings of S: gives

$$\underbrace{Pr[B] \leq \Pi_e(2m) \cdot \frac{1}{2^{\epsilon m/2}}}\ .$$

We're done provided that $\leq \frac{\delta}{2}$:

$$\Pi_e(2m) \cdot \frac{1}{2^{\frac{\epsilon m}{2}}} \leq \frac{\delta}{2} \qquad : \text{iff } (\log)$$

$$\log(\Pi_e(2m)) - \frac{\epsilon m}{2} \leq \log\left(\frac{\delta}{2}\right), \quad \text{i.e.}$$

$$\log(\Pi_e(2m)) + \log 2/\delta \leq \frac{\epsilon m}{2}, \quad \text{i.e.}$$

$$\boxed{m \geq \frac{2}{\epsilon}\left(\log\left(\Pi_e(2m)\right) + \log 2/\delta\right).}$$

End of pf of <u>Key Thm!</u>

Okay, have Key Thm:

$$\boxed{m \geq \frac{2}{\epsilon}\left(\ln(\Pi_e(2m)) + \ln 2/\delta\right), \quad m \geq 8/\epsilon}$$

is enough ex. for a CHF for $\mathcal{C}$ using $\mathcal{C}$ to be a PAC learner.

Weren't we promised

$$m \approx \frac{1}{\varepsilon}\left(d \log \frac{1}{\varepsilon} + \log 1/\delta\right)$$

is enough ex. for a CHF for $\mathcal{C}$ using $\mathcal{C}$ to be a PAC learner?

We get $\not P$ using Amazing Thm.

AT: $\Pi_{\mathcal{C}}(2m) \le \left(\frac{2em}{d}\right)^d$ for $m \geq d$,

so $\log \Pi_{\mathcal{C}}(2m) \le d \log\left(\frac{2em}{d}\right)$.

So suff to have m satisfy

$$\boxed{m \geq \frac{2}{\varepsilon}\left(d \log\left(\frac{2em}{d}\right) + \ln 2/\delta\right), \quad m \geq 8/\varepsilon}$$

Holds provided

const like 8 or so

$$\boxed{m \geq K\left(\frac{d}{\varepsilon} \cdot \log \frac{1}{\varepsilon} + \frac{1}{\varepsilon} \log \frac{1}{\delta}\right).}$$

ignore δ part, constants, this is $\approx$

$$m \geq \frac{2}{\varepsilon}\left(d \log\left(\frac{m}{d}\right)\right), \quad \text{i.e.}$$

$$\underbrace{\frac{m}{d}}_{B} \geq \underbrace{\boxed{\frac{2}{\varepsilon}}}_{A} \log\left(\frac{m}{d}\right)^{B}$$

$$B \geq A \log B$$

$$\frac{B}{\log B} \geq A \quad : \quad B = 2A\log A \text{ is enough.}$$

$$\boxed{\frac{m}{d} = \frac{2}{\varepsilon}\log\frac{2}{\varepsilon} \quad \text{is enough.}}$$

End of upper bd on sample cxity that suffices for PAC learning!

Ex / Application:

Consider $\mathcal{C}$ = all LTFs over $\mathbb{R}^n$.

Fact 1: $\text{VCDIM}(\mathcal{C}) = n+1$.

(recall: $n=2$, we argued VCDIM $= 3$).

Fact 2: There is a computationally efficient

CHF for $\mathcal{C}$ = (all LTFs) using $\mathcal{C}$.

"polynomial-time linear programming"

Say $\underline{w_1 x_1 + \dots + w_n x_n \geq \theta}$ is an unknown LTF.

- Pos. ex. $((\overbrace{3, 1, -5, \dots, 6}^{\text{pt in } \mathbb{R}^n}), +)$: means

$$\boxed{3w_1 + w_2 - 5w_3 \dots + 6w_n \geq \theta}$$

- Neg ex. $((4, -2, 7, \dots, 3.3), -)$ means

$$\boxed{4w_1 - 2w_2 + 7w_3 \dots + 3.3w_n < \theta}$$

"linear programming feasibility" (LP): given a coll. of linear inequalities, is there a sol. sat. all of them? if so, find a sol.

Have Poly-time algs for this! :)

This is a CHF for $\mathcal{C}$ using $\mathcal{C}$, & it's comput. efficient.

Summarizing, here's an eff PAC learning alg for $\mathcal{C}$ = LTFs over $\mathbb{R}^n$:

- draw $m \approx \frac{1}{\varepsilon}\left(n \cdot \log \frac{1}{\varepsilon} + \ln \frac{1}{\delta}\right)$ ex;
- run poly-time alg for LP to get a consistent $w_1, \dots, w_n, \theta$;

output hyp $\text{sign}(w \cdot x - \theta)$.

Next time: unit on <u>boosting</u>.