Last time:

Start unit on sample complexity of PAC learning / VCDIM

✓ • lower bound on #: $\Omega(\mathrm{VCDIM}(\mathcal{C})/\varepsilon)$ $(\delta = 0.1)$

• start upper bound on #: roughly $O(\mathrm{VCDIM}(\mathcal{C})/\varepsilon)$

1) Setup: think more deeply about $\mathrm{VCDIM}(\mathcal{C}) = d$.

$$\Pi_{\mathcal{C}}(S) = \{c \cap S : c \in \mathcal{C}\}$$

$$\Pi_{\mathcal{C}}(m) = \max_{|S|=m} |\Pi_{\mathcal{C}}(S)| \quad \Big\}\text{ "growth function" of } \mathcal{C}.$$

$\mathrm{VCDIM}(\mathcal{C}) = \max m$ s.t. $\Pi_{\mathcal{C}}(m) = 2^m$

Today: Continue (finish?)

2) Amazing Theorem (combinatorics) about $\Pi_{\mathcal{C}}(m)$, for any $\mathcal{C}$.

3) Learning-style argument: like CHF thm: bad hypotheses, union bd but now over labelings (# is bounded by A.T.)

Reminder: Midterm next Tues – be on time

Lecture Videos available

Questions?

sol. to H2 out today

Recall: $\mathrm{VCDIM}(\mathcal{C}) \leq$ max value d s.t. for $m \leq d$
there exists a $S \subseteq X$, $|S| = d$, $2.718\ldots$
with $\binom{m}{d} + \binom{m}{1} + \cdots + \binom{d}{d}$, $\Pi_{\mathcal{C}}(S) = 2^d$, i.e. $\leq \left(\frac{em}{d}\right)^d$

max d! s.t. $\Pi_{\mathcal{C}}(d) = 2^d$.

↳ For $m > d$, can $\Pi_{\mathcal{C}}(m)$ be almost 2^m? NO.

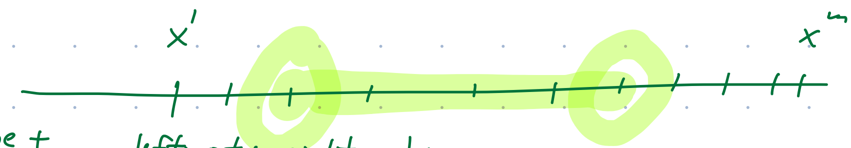
Let's look again at intervals...

$\mathcal{C}$ = intervals, $\Pi_{\mathcal{C}}(1)=2$

$\Pi_{\mathcal{C}}(2)=4$

$\Pi_{\mathcal{C}}(3)=7$

$\Pi_{\mathcal{C}}(m)=?$



all $-$; one $+$; leftmost $+$, rightmost $+$

$$\binom{m}{0}+\binom{m}{1}+\binom{m}{2} = O(m^2)$$

$$\binom{m}{0}=1,\quad \binom{m}{1}=m,\quad \binom{m}{2}=\frac{m(m-1)}{2}$$

much slower than 2^m !

Not a coincidence!

(Sauer–Shelah Thm)

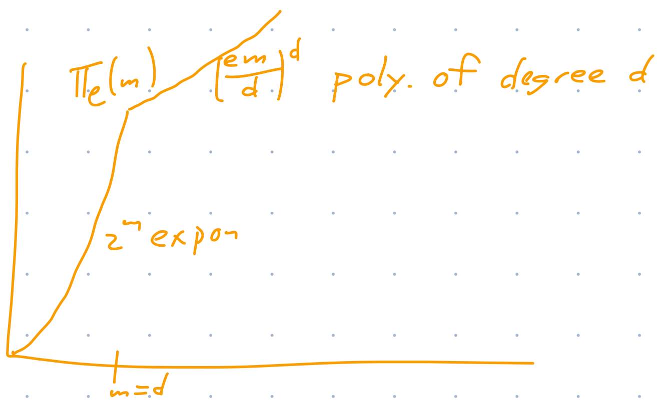
Amazing Theorem: Fix any $\mathcal{C}$. Let $d = \mathrm{VCDIM}(\mathcal{C})$. Have

- $\Pi_{\mathcal{C}}(m) = 2^m$ for $m \le d$
  - $= \binom{m}{0}+\cdots+\binom{m}{d}$ for $m \le d$

(!) • $\Pi_{\mathcal{C}}(m) \le \binom{m}{0}+\binom{m}{1}+\cdots+\binom{m}{d} \le \underbrace{\left(\frac{em}{d}\right)^d}$ for $m > d$

$e = 2.718\ldots$

$\downarrow$ amazing!



(Interlude start – no mention of $\mathcal{C}$ for a bit)

Def: Given $d, m \geq 0$, define $\Phi_d(m)$ as

- $\Phi_0(m) = 1 \quad \forall m$
- $\Phi_d(0) = 1 \quad \forall d$
- For $d, m \geq 1$, $\boxed{\Phi_d(m) = \Phi_d(m-1) + \Phi_{d-1}(m-1)}$

$\Phi_d(m)$

$d \backslash m$	0	1	2	3	4
4	1	2	4	8	16
3	1	2	4	8	15
2	1	2	4	7	11
1	1	2	3	4	5
0	1=1	1	1	1	1

$\swarrow = +$

Fact: Have $\Phi_d(m) = \sum_{i=0}^{d} \binom{m}{i}$.

Pf: Induc. on $m+d$.

Base case: $d=0$: $\Phi_0(m) = 1 \stackrel{?}{=} \binom{m}{0}$ yes.

$m=0$: $\Phi_d(0) = 1 \stackrel{?}{=} \sum_{i=0}^{d} \binom{0}{i} = \binom{0}{0} = 1$ yes.

General/inductive step:

$$\Phi_d(m) \overset{\text{def}}{=} \underbrace{\Phi_d(m-1)} + \underbrace{\Phi_{d-1}(m-1)}$$

$$\overset{\text{ind.hyp}}{=} \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d-1} \binom{m-1}{i}$$

$$= \sum_{i=0}^{d} \left(\binom{m-1}{i} + \binom{m-1}{i-1} \right) \qquad \text{b/c } \binom{m-1}{-1} = 0$$

$$= \sum_{i=0}^{d} \binom{m}{i}, \qquad \text{true b/c } \boxed{\binom{m-1}{i} + \binom{m-1}{i-1} = \binom{m}{i}:}$$

---

to choose i elts from $\{1, \dots, m\}$

- if include elt m, have $\binom{m-1}{i-1}$ ways to choose rest;
- if don't " " , have $\binom{m-1}{i}$ ways to choose the i elts.

(End of interlude)

Back to $\mathcal{C}$ (any concept class, $d = \text{VCDIM}$)

<u>AT:</u> $\Pi_{\mathcal{C}}(m) \leq \Phi_d(m)$. ☆

<u>Pf:</u> Induc. on $m+d$

Base case: $m = 0$. $\Pi_{\mathcal{C}}(0) = 1$, is indeed $\leq \Phi_d(0) = 1$ ✓

Base case $d = 0$: For $\mathcal{C}$ s.t. $\text{VCDIM}(\mathcal{C}) = 0$, have $\Pi_e(m) = 1$, b/c must have $|\mathcal{C}| = 1$. ✓

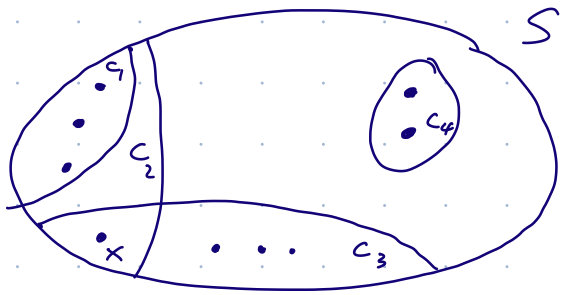
Induc. step: assume ✩ for all m', d' with $m'+d' < m+d$ & $m' \le m$, $d' \le d$.

Fix any $S \subseteq X$ with $|S| = m$:

goal is to bound $|\Pi_e(S)|$ by $\Phi_d(m)$.

Consider a "special" $x \in S$:

("twins" c_1 & c_2: differ only in x)



Have $\underline{|\Pi_e(S \setminus \{x\})| \le \Phi_d(m-1)}$ by induc. assump.

This undercounts $\Pi_e(S)$ exactly b/c of "twins", i.e. pairs of distinct elts of $\Pi_e(S)$ differing only by x.

Let
$$\mathcal{C}' = \{c \in \Pi_e(S) : x \notin c, \text{ and } c \cup \{x\} \in \Pi_e(S)\}$$
(sets like c_1).

Have $$|\Pi_e(S)| = |\mathcal{C}'| + |\Pi_e(S \setminus \{x\})|.$$

But, have $\mathcal{C}' = \Pi_{e'}(S \setminus \{x\})$. So, have

$$|\Pi_{\mathcal{C}}(S)| = \underbrace{|\Pi_{\mathcal{C}'}(S \setminus \{x\})|}_{\text{if knew this } \le \Phi_{d-1}(m-1),} + \underbrace{|\Pi_{\mathcal{C}}(S \setminus \{x\})|}_{\le \Phi_d(m-1) \text{ induc.}}$$

then would $= \Phi_d(m)$ (def of $\Phi_d(m)$) + done!

To show: $|\Pi_{\mathcal{C}'}(\underbrace{S \setminus \{x\}}_{\text{size } m-1, \text{ so}})| \le \Phi_{d-1}(m-1)$; holds by i.h. provided $\mathrm{VCDIM}(\mathcal{C}') \le d-1$.

Suppose $S' \subseteq S \setminus \{x\}$ is shatt. by $\underline{\mathcal{C}'}$.

Then b/c of twins, $\underline{S' \cup \{x\}}$ is shatt. by $\mathcal{C}$.

Largest set shatt. by $\mathcal{C}$ is size d: so
" " " " $\mathcal{C}'$ is size $\le d-1$.

So $\mathrm{VCDIM}(\mathcal{C}') \le d-1$, + done. ■

End of AT proof!

---

3) Learning argument

Recall: we say hyp h^{ϵ} is bad if $err_{\mathcal{D}}(h,c) > \varepsilon$,
i.e. $\Pr_{x \sim \mathcal{D}}[h(x) \ne c(x)] > \varepsilon$.

Recall our old CHF thm, rephrased:

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

$$m \geq \frac{1}{\varepsilon}\left(\log|\mathcal{C}| + \log\frac{1}{\delta}\right),$$

Key to pf: union bd over bad $h \in \mathcal{C}$.

w.p. $\geq 1-\delta$, all bad $h \in \mathcal{C}$ are inconsistent with data set

Now: analogue for any $\mathcal{C}$, even infinite. We'll prove:

New Thm Fix any ~~finite~~ $\mathcal{C}$, any dist $\mathcal{D}$, any target concept $c \in \mathcal{C}$.
Given m data set examples drawn from $EX(c, \mathcal{D})$, where

$\circledast$
$$m \geq \frac{2}{\varepsilon}\left(\log\left(\Pi_{\mathcal{C}}(2m)\right) + \log\frac{2}{\delta}\right),$$

Key to pf: union bd over possible labelings of a $2m$-elt set.

w.p. $\geq 1-\delta$, all bad $h \in \mathcal{C}$ are inconsistent with data set

$\circledast$ is great, b/c of AT: AT says

$$\log\left(\Pi_{\mathcal{C}}(2m)\right) \leq \log\left(\left(\frac{2em}{d}\right)^d\right), \text{ so}$$

$\circledast$ becomes
$$\boxed{m \geq \frac{2}{\varepsilon}\left(d\log\left(\frac{2em}{d}\right) + \log\frac{2}{\delta}\right)}$$

this is achieved once $m \gtrsim \frac{1}{\varepsilon}\left(d + \log\frac{1}{\delta}\right)$

Next time: IOU

$$\binom{m}{0}+\binom{m}{1}+\cdots+\binom{m}{d} \le \left(\frac{\overset{2.718\ldots}{\overset{\parallel}{e}}\,m}{d}\right)^{d}$$

+ proof of New Thm
