## Last time:

Start *generic* bounds: Algorithms + lower bounds that apply to *any* (finite) concept class. [Give up on comput. eff...]

pos results!
- Halving algorithm: $\log_2 |\mathcal{C}|$ mistake bound
- Randomized halving alg.: $\mathbb{E}[\#\text{mistakes}] \le \ln|\mathcal{C}| + O(1)$ (under "oblivious adversary" assumption)

neg. result
- Start discussion of VC dimension of $\mathcal{C}$

## Today:

- VCDIM($\mathcal{C}$) as lower bound on best poss. mistake bound of any OLMB alg. for $\mathcal{C}$
- Predicting from Expert Advice
  - Weighted Majority (WM) alg
  - Randomized WM alg (RWM)

rebranded noise-tolerant HA, RHA

**Reminder:** PS 1 due today, PS 2 out

Questions?

---

## Recall shattering + VCDIM($\mathcal{C}$)   VC($\mathcal{C}$)

**Def:** Fix $\mathcal{C}$ over domain $X$. Let $S \subseteq X$.

$S$ is shattered by $\mathcal{C}$ means:

| (subset POV): | (labeling POV) |
|---|---|
| $\forall\, U \subseteq S\ \exists$ some $c \in \mathcal{C}$ (subset of $X$) s.t. $c \cap S = U$. | For every labeling of pts in $S$ by 0/1, some $c \in \mathcal{C}$ induces that labeling. |

VCDIM($\mathcal{C}$) = size of largest $S \subseteq X$ that's shattered by $\mathcal{C}$.

---

**Equiv:** VC($\mathcal{C}$) = smallest $d$ s.t. no $(d+1)$-elt. set $S \subseteq X$ is shattered.

To show $VC(\mathcal{C}) = d$: must $(VC(\mathcal{C}) \geq d)$

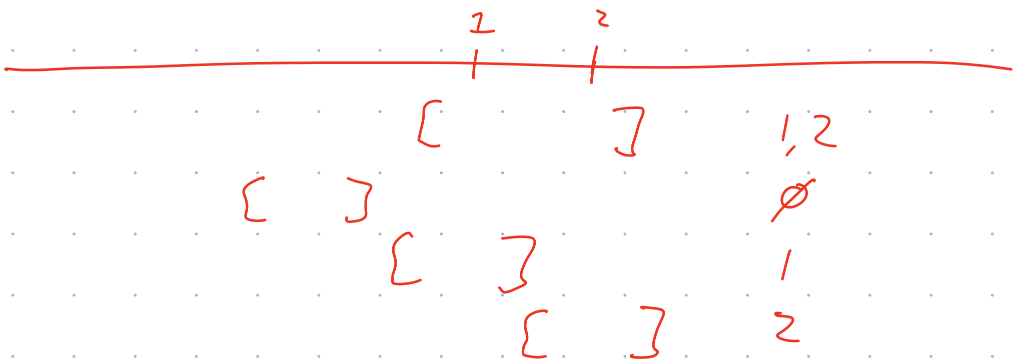
- exhibit an $S \subseteq X$, $|S| = d$, that $\mathcal{C}$ shatters;
- $(VC(\mathcal{C}) \leq d)$: argue <u>no</u> $S \subseteq X$ of size <u>$d+1$</u> is shattered.

If $\forall d \ \exists \ |S| = d$ shattered: we say $VC(\mathcal{C}) = \infty$.

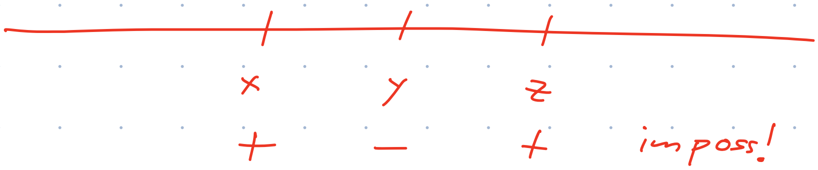
---

<u>Ex 1:</u> $X = \mathbb{R}$, $\mathcal{C}$ = all intervals $[a, b]$

- $VC(\mathcal{C}) \geq 2$: say $S = \{1, 2\}$

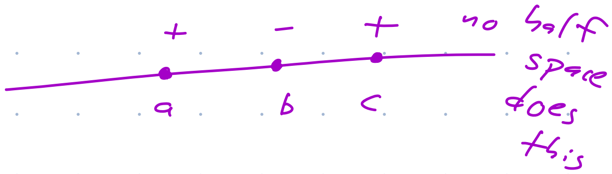



- $VC(\mathcal{C}) \leq 2$: take any 3 pts $x < y < z$

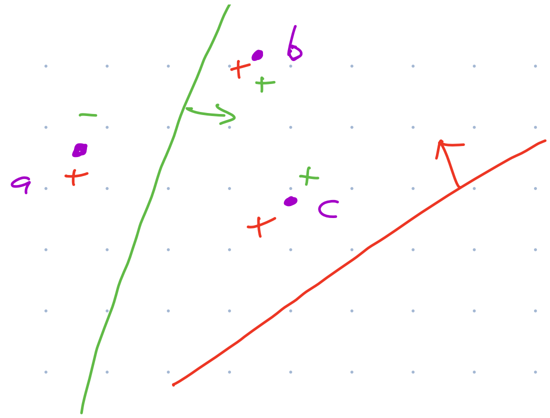



---

<u>Ex#2:</u> $X = \mathbb{R}^2$, $\mathcal{C}$ = halfspaces (LTFs).

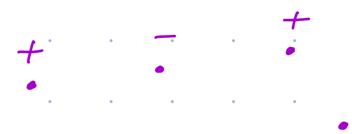



- $VC(\mathcal{C}) \geq 3$: take $a, b, c$ any 3 non-colinear pts



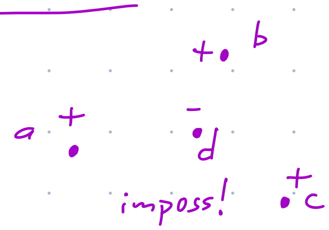

all 8 lab. possible by halfspaces

- To show $VC(\mathcal{C}) = 3$: must show $\leq 3$, i.e. show for <u>any</u> 4 pts, $\exists$ some labelling that LTFs can't induce.
  - <u>Case 1:</u> the 4 pts include 3 colinear pts



impossible ✓

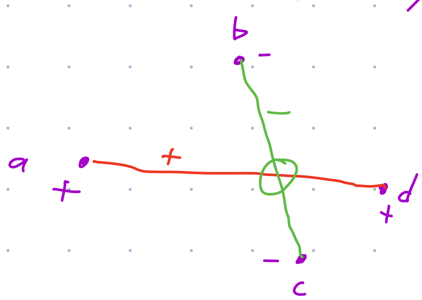
  - <u>Case 2:</u> no 3 of the 4 pts are colinear.




imposs!

<u>Case 2a:</u> 4th pt d is inside $\Delta$. ✓

<u>Case 2b:</u> ✓ d outside $_a\overset{b}{\Delta}_c$, i.e. 4pts form quadrilat.




imposs!

So indeed $vc(\mathcal{C}) = 3$.

**Fact:** for $X = \mathbb{R}^n$, $\mathcal{C}$ = all halfspaces over $\mathbb{R}^n$,

$VC(\mathcal{C}) = n+1$.

---

**Ex #3:** $X = \{0,1\}^n$, $\mathcal{C}$ = mon conj.

- $VC(\mathcal{C}) \geq n$: here's a shattered set:

  $n = 5$

  01111
  10111
  11011
  11101
  11110

  $x_1 \wedge x_5$

- $VC(\mathcal{C}) \leq n$:

  For a set of size $k$ (k ex.) to be shatt. by $\mathcal{C}$, must have $|\mathcal{C}| \geq 2^k$ ☺

  $|\mathcal{C}| = 2^n$, so $VC(\mathcal{C}) \leq n$.

---

Conn. between $VC(\mathcal{C})$ + OLMB:

**Fact:** If $VC(\mathcal{C}) = d$, then any OLMB alg must have worst-case MB $\geq d$.

**Pf:** Say $S = \{x^1, \ldots, x^d\}$ is shatt. by $\mathcal{C}$.
Consider $x^1, \ldots, x^d$ as first $d$ trials.

No matter what $h^1(x^1)$
$h^2(x^2)$
$\vdots$

$h^d(x^d)$ are,

$\exists\, c \in \mathcal{C}$ s.t. all are wrong. ■

---

Obliv. adv. ? Still get $\mathbb{E}[\#\text{mist.}] \geq \frac{d}{2}$ for any alg.

Ex seq: $x^1, \ldots, x^d$

Adv, once & for all at start, picks unif rand one of the $2^d$ lab's of , chooses as target a c giving that labeling.

So learner's task on 1st d ex $\approx$ predicting d fair coin tosses ;

---

<u>Note:</u> means our OLMB alg (elim.) for mon conj. is optimal !

---

# Predicting from Expert Advice:

## Weighted Majority & Randomized WM

Consider following setting: go to track with group of friends.

Race 1
Race 2
⋮

Need 0/1 pred. on each race.

Your goal: combine friends' pred's for each race, + make $$.

- No absolute guarantee: if all friends do badly, you will to

- Turns out: you can ensure you'll do almost as well as most successful friend, without knowing in advance who it is!

---

Weighted Maj Alg: ← with parameter $0 \le \beta < 1$

- pool of $N$ "experts",
- sequence of trials.

Each expert makes 0/1 prediction at each trial.

Alg: • we maintain weight $w_i$ for expert $i$; initially $w_i = 1$ $\forall i \in [N]$.

- At each trial, expert $i$ predicts $z_i \in \{0,1\}$

- Let $q_0 = \sum_{i: z_i = 0} w_i$ , $q_1 = \sum_{i: z_i = 1} w_i$ .

Predict $\begin{cases} 0 & \text{if } q_0 \ge q_1 \\ 1 & \text{if } q_1 > q_0 \end{cases}$ .

- Given outcome of trial: for each $i \in [N]$ s.t. $z_i$ was wrong, set $w_i \leftarrow \beta \cdot w_i$.

---

$\beta = 0$: exactly HA.
Analogy:

Recall HA terrible if noise. WM is still good even if no expert is perfect!

---

expert $i$ $\longleftrightarrow$ $i^{th}$ concept in $\mathcal{C}$ ($c_i$)

pred. $z_i$ of expert $i$ on trial $t$ $\longleftrightarrow$ output of $c_i$ on $t^{th}$ example

---

Thm: For any seq. of trials, if best expert in pool makes $m$ mistakes, then WMA makes at most

$$\boxed{\frac{\log N + m \log \frac{1}{\beta}}{\log\left(\frac{2}{1+\beta}\right)}}$$

mistakes.

Ex: $\beta = \frac{1}{2}$ : $2.41(m + \ln N)$.

$\beta = \frac{3}{4}$ : $\approx 2.2m + 5.2 \ln N$

$\varepsilon \to 0$, $\beta = 1 - \varepsilon$ : $\approx 2m + \frac{2}{\varepsilon} \cdot \ln N$

---

Pf: Let $W = w_1 + \dots + w_n$. (Initially $W = N$.)

At a partic. trial, $W = q_0 + q_1$.

At each mistake: at least half of $W$ "predicted wrong" & is mult. by $\beta$.

So if $W =$ tot wt before mistake,

after the mistake, new tot wt is

$$\leq \frac{1}{2}W + \frac{1}{2}W \cdot \beta = \left(\frac{1+\beta}{2}\right)W.$$

So if weighted maj alg makes $M$ mistakes,

have $W \leq N\left(\frac{1+\beta}{2}\right)^M.$

LB on $W$: $W \geq$ wt of best single expert

& this always is $\geq \beta^m$.

So have $\beta^m \leq W \leq N\left(\frac{1+\beta}{2}\right)^M$ $\blacksquare$

Take log:

$$m \cdot \log \beta \leq \log N + M \cdot \log\left(\frac{1+\beta}{2}\right)$$

Negate:

$$m \log \frac{1}{\beta} \geq -\log N + M \log\left(\frac{2}{1+\beta}\right)$$

divide

$$\frac{m \log \frac{1}{\beta} + \log N}{\log\left(\frac{2}{1+\beta}\right)} \geq M$$

---

Next time RWM

Start Probably Approximately Correct (PAC) Learning