Statistical Learning Theory 3

Tuesday, November 5, 2019 2:34 PM

last 2 lectures:

* Sample Complexity of learning finite $|H|$

$err_P(h^*)=0$ $\quad$ VCD $\leq$ $\quad$ $O\left(\frac{1}{\varepsilon}\left(\ln|H| + \ln(\frac{1}{\delta})\right)\right)$

$err_P(h^*) \neq 0$ $\quad$ VCD $\leq$ $\quad$ $O\left(\frac{1}{\varepsilon^2}\left(\ln(|H|) + \ln(\frac{1}{\delta})\right)\right)$

* Sample Complexity for infinite $|H|$

↳ Growth function : $|H[S]|$: # unique labelings H produces on S.

$$H[m] = \max_{|S|=m} |H[S]|$$

↳ VC Dimension : Max m for which $H[m] = 2^m$.

$O\left(\frac{1}{\varepsilon}\left(\text{VCDim}(H)\ln(\frac{1}{\varepsilon}) + \ln(\frac{1}{\delta})\right)\right)$ $\quad$ $err_P(h^*)=0$

$O\left(\frac{1}{\varepsilon^2}\left(\text{VCDim}(H) + \ln(\frac{1}{\delta})\right)\right)$.

Today: Review VCDim, see some examples.

go from sample complexity ⟶ learnability.

$$S = \{x_1, x_2, \ldots, x_m\} \qquad \{+1,-1\}$$

$$\left\{\left(h(x_1), h(x_2), \ldots, h(x_m)\right)\right\} \quad h \in H$$

$|H[S]| \leq 2^{|S|}$ because $\underbrace{2 \times 2 \times \cdots \times 2}_{m} = 2^m$.

labels for x_1 ... labels for x_2

VCDim is the point where $H[m]$ stops growing exponentially.

$$\underbrace{H[m] \leq O(m^d)}$$

H be any finite hypothesis class. How large can its VCDim be?

$$\underbrace{H[m]}_{\leq |H|} = 2^{m = \geq \log_2(|H|)+1} \implies \overbrace{H[m]}^{|H|} > |H| \quad \times$$

(*) $\text{VCDim}(H) \leq \log_2(H)$. ✓

⊛ $\text{VCDim}(H) \ll \log_2(|H|)$ In many cases.

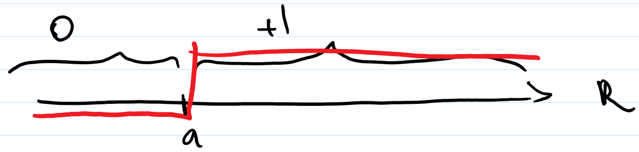
* For inite $|H|$ it's possible to have finite $\text{VCDim}(H)$.

VCDim of 1-Dimensional thresholds (line)

$X = \mathbb{R}$

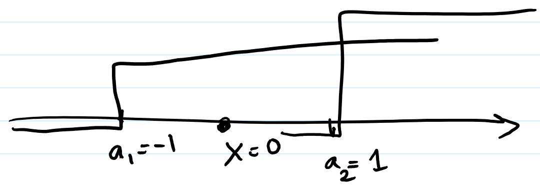
$h_a(x) = \mathbb{1}(x \geq a)$

indicator function.



$\mathcal{H}^{\text{thresholds}} = \{ h_a \mid a \in \mathbb{R} \}$.

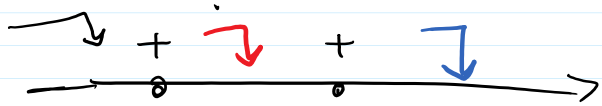
$\text{VCDim}(H^{\text{thresholds}}) = ?$



$h_{a_1}(x) = 1 \qquad h_{a_2}(x) = 0$

$H[S] = 2^{|S|} = 2 \implies \underline{\text{VCDim}(H^{\text{thresholds}}) \geq 1.}$ ①

Can I shatter a set of size 2? I need 4 labelings



$x_1 = 0 \qquad x_2 = 2$

$- \qquad +$

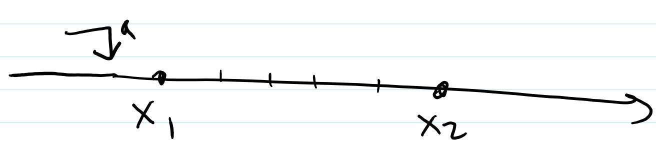
$- \qquad -$

$+ \qquad -$? Impossible

For <u>any</u> set of size 2, it's impossible to shatter it.

Any $|S| \geq 2$ $\quad x_1 =$ smallest $\quad x_2 =$ largest

$\underline{h_a(x_1) = +1} \qquad \underline{h_a(x_2) = 0}$



$\downarrow$

$a \leq x_1 \implies a \leq x_2 \implies \underline{h_a(x_2) = +1}$ ✗

$//$

[illegible] x_1 [illegible] x_2 [illegible] a [illegible] x_2 [illegible] X

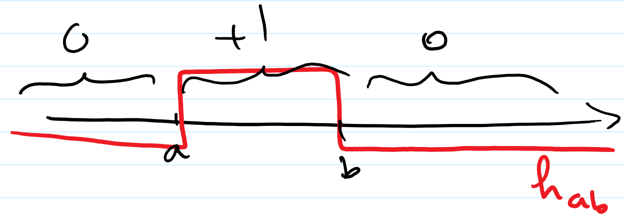
$\hookrightarrow \text{VCDim}(\mathcal{H}^{\text{thresholds}}) < 2$ ②

① + ② $\quad \text{VCDim}(\mathcal{H}^{\text{thresholds}}) = 1.$

VCDim of intervals on a line.

$X = \mathbb{R}$

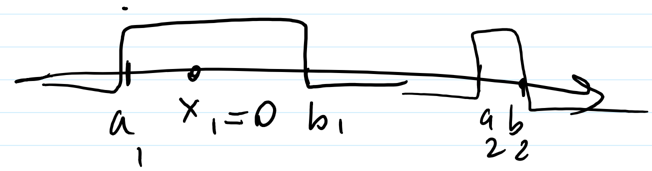
$h_{ab}(x) = \mathbb{1}(a \leq x \leq b)$



$\mathcal{H}^{int} = \{ h_{ab} \mid a, b \in \mathbb{R} \}.$ $\quad \text{VCDim}(\mathcal{H}^{int})$?

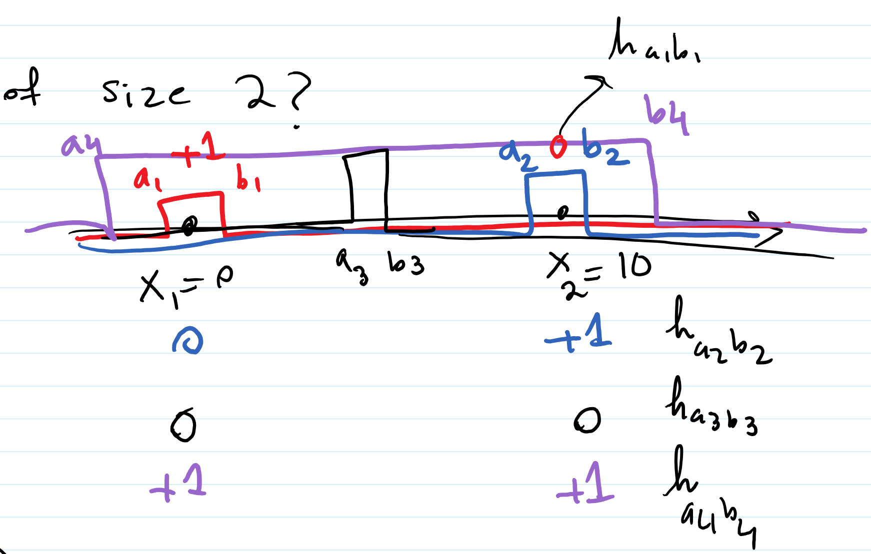
Can I shatter a set of size 1?

$h_{a_1 b_1}(x) = 1$



$h_{a_2 b_2}(x) = 0$ $\quad \text{VCDim}(\mathcal{H}^{int}) \geq 1.$

Can I shatter a set of size 2?

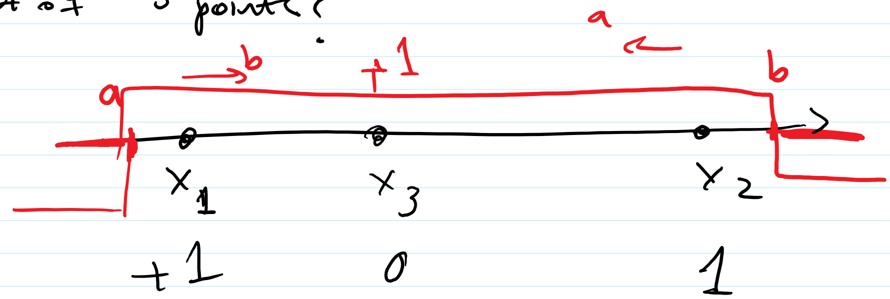


$\text{VCDim}(\mathcal{H}^{int}) \geq 2$ ①

$VCDim(H) \geq 2$. 1

Can I shatter a set of 3 points?

$|S| \geq 3$



x_1 : smallest
x_2 : largest
x_3 : Any other point

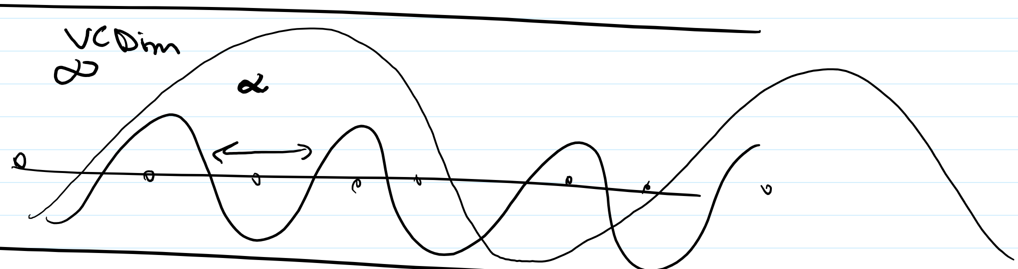
$h_{ab}(x_3) = 1$ X

2

Cannot shatter Any set of 3 points. $VCDim(H^{int}) < 3$.

$VCDim(H^{int}) = 2$

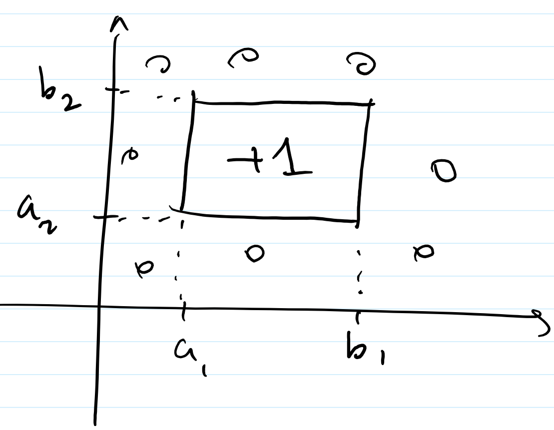
It's possible to have only 1 parameter but VCDim ∞



VCDim Axis-aligned rectangles: 2-D intervals

$X = \mathbb{R}^2$

$h_{a_1 a_2 b_1 b_2}(x) = 1\left(\begin{array}{l} a_1 \leq x_1 \leq b_1 \\ \text{and} \\ a_2 \leq x_2 \leq b_2 \end{array}\right)$



$$H^{int(2)} = \left\{ h_{a_1 a_2 b_1 b_2} \;\middle|\; a_1, a_2, b_1, b_2 \in \mathbb{R} \right\}$$

Can I shatter a set of 4 points?

✓0 : no points labels 1.

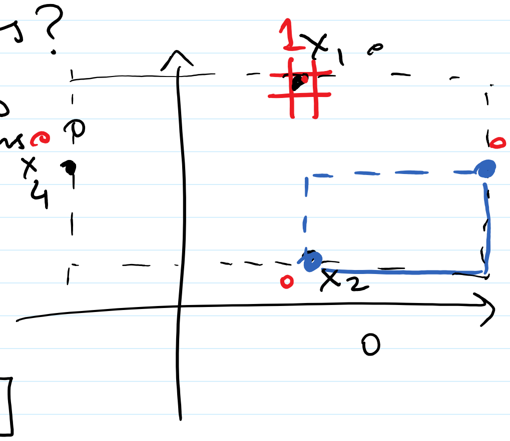
1 : Singletons: If I want to $x_1 = 1$ and others 0

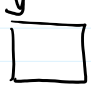
$a_1 b_1 = x_{11}$ $\quad$ $a_2 b_2 = x_{12}$

2 = Doubles

3 = 3 point labeled 1

✓4. 4 points labelled 1





↳ finish this at home

for any $S' \subseteq S$

S' is labeled 1

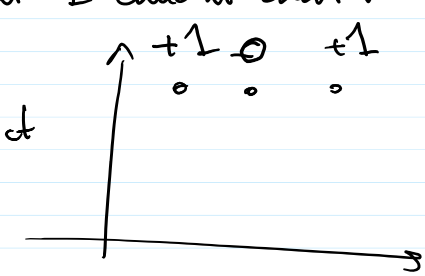
$S \setminus S'$ is labeled 0

Idea: This set is designed so that the bounding box of S' doesn't have any points from $S \setminus S'$.

there is a rectangle that only captures S'.

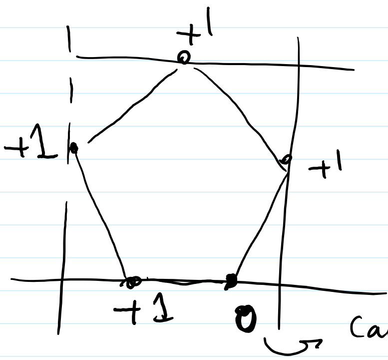
$VCDim(H^{int2}) \geq 4$. 1

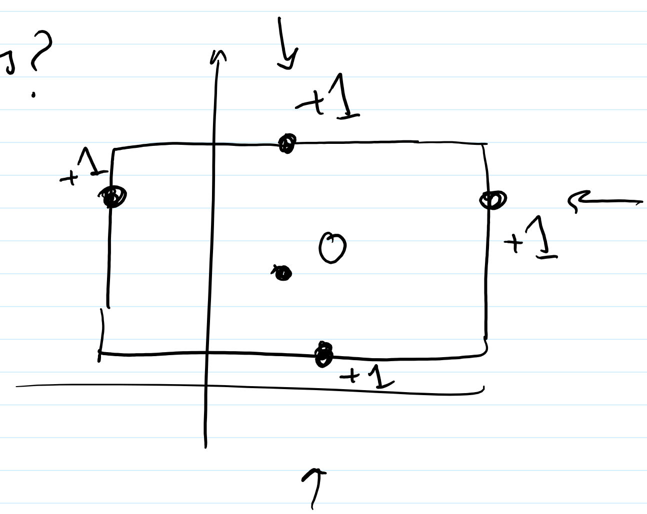
there exists a set of 3 points that I couldn't shatter



↳ but this does not contradict $VCD(H) \geq 3$.

Can I shatter a set of 5 points?





↳ Cannot produce 0.

int(2)

$\hookrightarrow \mathrm{VCDim}(\mathcal{H} \quad) < 5$ (2)

(1) + (2) $\mathrm{VCDim}(H^{\mathrm{int}(2)}) = 4$ ✓