Def: The VC dimension of $\mathcal{H}$ (abbrev. $VC(\mathcal{H})$)

$\exists\, n$ points shatter-able by $\mathcal{H}$

is $\max\{n \in \mathbb{N} : \mathcal{H}$ can shatter n points$\}$.

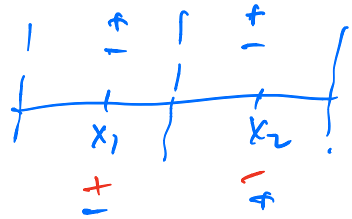
Def: growth function $S(\mathcal{H}, n)$:

$$S(\mathcal{H}, n) = \max_{x_1 \cdots x_n} \left| \Pi_{\mathcal{H}}(\{x_1 \cdots x_n\}) \right|$$

$$\{ (h(x_1) \cdots h(x_n)) : h \in \mathcal{H} \}$$

threshold class

$S(\mathcal{H}, 2) = 3.$



$x_1 = x_2$

$S(\mathcal{H}, n) = 2^n \iff \exists$ dataset of size n, shattered by $\mathcal{H}$.

$$VC(\mathcal{H}) = \max\{ n : S(\mathcal{H}, n) = 2^n \}.$$

Theorem: Suppose $\mathcal{H}$, $VC(\mathcal{H}) = d < \infty$. Then, given a set of n iid examples $(x_1, y_1) \cdots (x_n, y_n)$ drawn from D, then with probability $1-\delta$:

$\log|\mathcal{H}|$

$$\sup_{h \in \mathcal{H}} \left| err(h, S) - err(h, D) \right| \le c \cdot \sqrt{\frac{d \ln \frac{n}{d} + \ln \frac{2}{\delta}}{n}}.$$

$\hat{h}$. $err(\hat{h}, D) \le \min_{h' \in \mathcal{H}} err(h', D) + \tilde{O}\left(\sqrt{\frac{d}{n}}\right)$

$\Rightarrow$ $\mathcal{H}$ is agnostic PAC learnable (by ERM).

$$\tilde{O}(f(n)) = O\left(f(n) \cdot \log\left(\frac{1}{f(n)}\right)\right)$$

Lemma: $|\mathcal{H}| < \infty$.

① $S(\mathcal{H}, n) \overset{?}{\le} |\mathcal{H}|$ ✓

② $VC(\mathcal{H}) \overset{?}{\le} \log|\mathcal{H}|$.

$$\max_{x_1 \ldots x_n} \left|\{(h(x_1) \ldots h(x_n)) : h \in \mathcal{H}\}\right|$$

$$VC(\mathcal{H}) = \max\{n : S(\mathcal{H}, n) = 2^n\}$$

$$\Downarrow ①$$

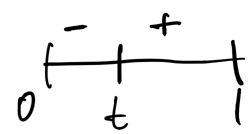
$$2^n \le |\mathcal{H}|$$

$$\Downarrow$$

$$n \le \log|\mathcal{H}|.$$

$$\le \log|\mathcal{H}|.$$

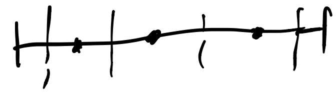
Example:



1. threshold class.

$VC(\mathcal{H}) = 1$.

$S(\mathcal{H}, n) \overset{?}{=} n+1$.



2. interval class in $[0,1]$

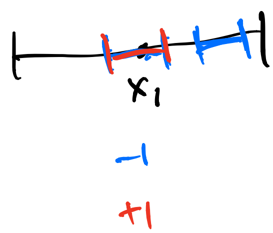
$$\mathcal{H} = \{ h_{a,b}(x) = 2I(a \le x \le b) - 1 : 0 \le a \le b \le 1 \}.$$

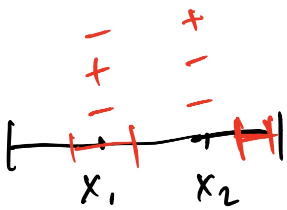
$VC(\mathcal{H}) = 2$

≥ 1

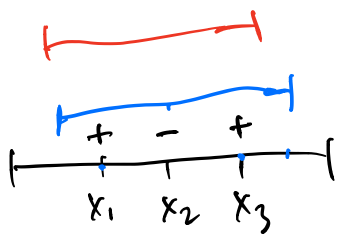
≥ 2.

< 3.





shatterable by $\mathcal{H}$



shatterable by $\mathcal{H}$?

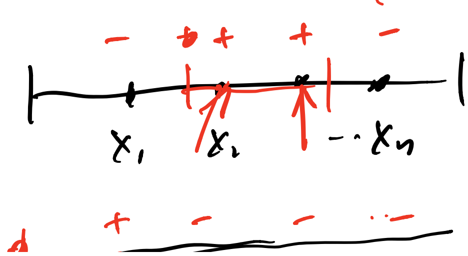
$\forall$ dataset of size 3, they are not shatterable by $\mathcal{H}$.

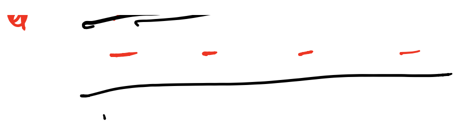
$h_{a,b}(x_1) = +1 \Rightarrow a \le x_1$

$h_{a,b}(x_3) = +1 \Rightarrow b \geq x_3$

$x_2 \in [a,b] \Rightarrow h_{a,b}(x_2) = +1 \neq -1$.

$$S(\mathcal{H}, n) \overset{?}{=} \left[\binom{n}{2} + \binom{n}{1} + \binom{n}{0} \right]$$

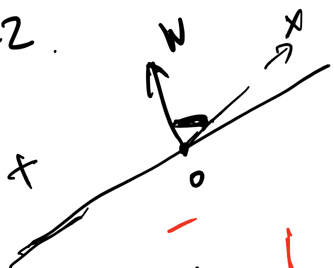




3. Homogeneous linear classifiers in $\mathbb{R}^d$

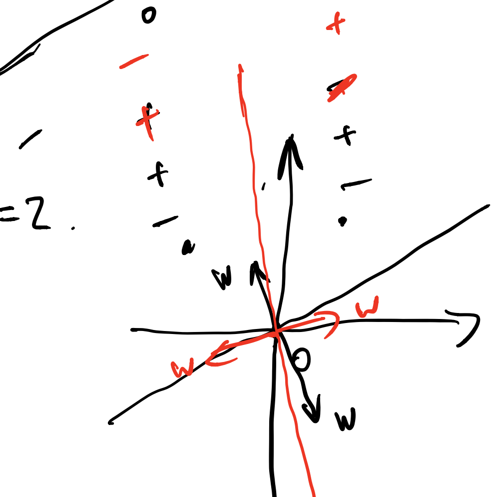
$$\mathcal{H} = \{ h_w(x) = 2I(w \cdot x > 0) - 1 \; : \; w \in \mathbb{R}^d \}$$

d=2.

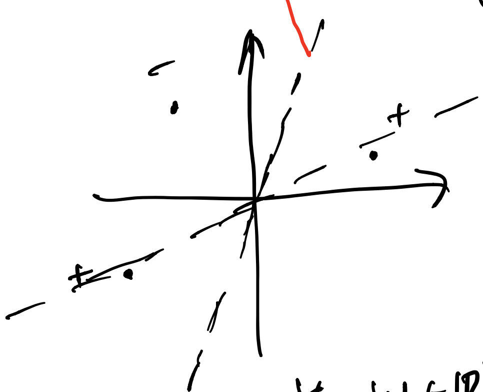


$VC(\mathcal{H}) \overset{?}{=}$

d=2.



$VC(\mathcal{H}) \geq 2$.



① if cannot put the three points on one side

cannot generate

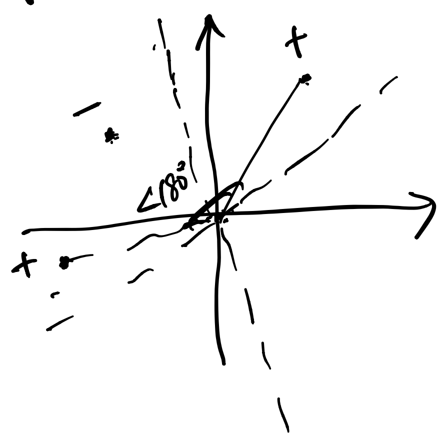
$(+, +, +)$

$\forall\, w \in \mathbb{R}^2$. $w \cdot x_1$ $w \cdot x_2$ $w \cdot x_3$ have different signs

$\rightarrow (+ \; + \; +)$

$\rightarrow (- \; - \; -)$

② If can put all pts on on side



$(+, -, +)$ is not acheivable.

$\forall$ 3 points $\{x_1, x_2, x_3\}$, they are not shatterable by $\mathcal{H}$.

$d=2$

$\Rightarrow VC(\mathcal{H}) = 2$.

For general d. $VC(\mathcal{H}) \overset{?}{=}$

① $\mathcal{H}$ can shatter any linearly independent points.

$x_1 \cdots x_d$ linearly independent

$\Rightarrow$ they are shatterable.

$$\text{sign}\left(\begin{pmatrix} x_1^T \\ \vdots \\ x_d^T \end{pmatrix} \cdot w \right) =$$

full rank.

$$
= \begin{pmatrix} \text{sign}(\langle w, x_1 \rangle) \\ \vdots \\ \text{sign}(\langle w, x_d \rangle) \end{pmatrix}
$$

$$
\underset{1\times(d+1)}{\alpha} \cdot \underset{(d+1)\times d}{X} = 0
$$

X full rank

$\Rightarrow \{X \cdot w,\ w \in \mathbb{R}^d\} = \mathbb{R}^d$

consider any labeling $l = (l_1 \cdots l_d) \in \{\pm 1\}^d$.

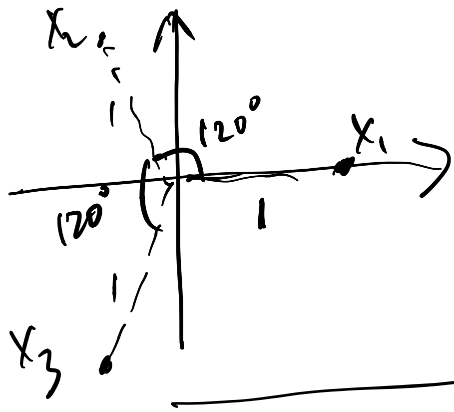
find w. $\text{sign}(X \cdot w) = l$

$w = (X^{-1}) \cdot l$ $\quad \text{sign}(Xw) = \text{sign}(l) = l$.

specifically. can pick $x_1 = e_1 \cdots x_d = e_d$,

they are shattered by $\mathcal{H}$.

② for any $d+1$ points. they are not shatterable by $\mathcal{H}$.



intuition: $\boxed{1x_1 + 1x_2 + 1x_3 = 0}$

$\Rightarrow$ $(+1, +1, +1)$ is not achievable.

if $\exists w$.

$$\left.\begin{array}{l} w \cdot x_1 > 0 \\ w \cdot x_2 > 0 \\ w \cdot x_3 > 0 \end{array}\right] \Rightarrow$$

$$\underbrace{w \cdot (x_1 + x_2 + x_3)}_{0} > 0$$

contradiction.

general case:

for any $d+1$ pts. can find $\alpha_1 \sim \alpha_{d+1}$ not all 0 and there exists i^*. $\alpha_{i^*} > 0$. such that

$$\sum_{i=1}^{d+1} \alpha_i x_i = 0$$

without loss of generality

do define $l_i = \begin{cases} +1, & d_i > 0 \\ -1, & d_i \leq 0 \end{cases}$ $\forall i = 1, \dots, d+1$

If there exists w that achieves $(l_1 \dots l_{d+1})$,

$$\begin{cases} d_i > 0 \Rightarrow l_i = +1 \Rightarrow w \cdot x_i > 0 \\ d_i \leq 0 \Rightarrow l_i = -1 \Rightarrow w \cdot x_i \leq 0 \end{cases}$$

$$\sum_{i=1}^{d+1} d_i \langle w \cdot x_i \rangle \; \cancel{\geq 0} \; > 0$$

$$\underbrace{d_i}_{>0} \underbrace{\langle w \cdot x_i \rangle}_{>0}$$

this contradicts w/ $\sum_{i=1}^{d+1} d_i x_i = 0$.

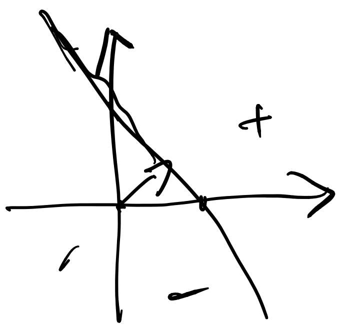
$\Rightarrow$ any $d+1$ points not shatterable by $\mathcal{H}$.

① ② $\Rightarrow$ $VC(\mathcal{H}) = d$.

4. Non-homogeneous linear classifiers in $\mathbb{R}^d$.

$$\mathcal{H} = \{ h_{w,b}(x) = 2\,\mathbb{I}(\underline{w \cdot x + b} > 0) - 1 : w \in \mathbb{R}^d ,\ b \in \mathbb{R} \}$$

$$\begin{pmatrix} w \\ b \end{pmatrix} \quad \begin{pmatrix} x \\ 1 \end{pmatrix}$$



$VC(\mathcal{H}) = d+1$.

(Exercise)

Sauer's Lemma

$\mathcal{H}$ $\quad VC(\mathcal{H}) = d$. $\qquad \underline{|\mathcal{H}| < \infty}$.

$x_1 \cdots x_n$.

$$S(\mathcal{H}, n) \begin{cases} = 2^n, & n \le d \\ \overset{?}{\le} 2^n - 1, & n > d \end{cases}$$

$d=1$

$S(\mathcal{H}, n) = n+1$ (Sauer's Lemma)

$$\le \sum_{i=0}^{d} \binom{n}{i} \quad (*)$$

$\cancel{\#}$ $(*)$ $\begin{cases} \le n^{d+1} & (n \ge 2) \\ \le \left(\frac{e \cdot n}{d}\right)^d & (n \ge d+2) \end{cases}$

Theorem: $\mathcal{H}$. $\underline{VC(\mathcal{H}) = d}$. $S = \{x_1 \dots x_n\}$

then

$$\left| \Pi_{\mathcal{H}}(S) \right| \le \left| \{ T \subseteq S : \mathcal{H} \text{ shatters } T \} \right|$$

$\Downarrow$

$|T| \le d$

$$\overset{?}{\le} \sum_{i=0}^{d} \binom{n}{i}$$

$$\le \{ T \subseteq S : \underline{|T| \le d} \}$$

$|T| = 0$. $\binom{n}{0}$

$= i$ $\binom{n}{i}$