LAST TIME:

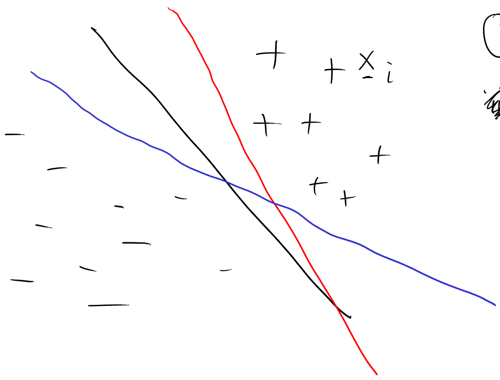
Theorem (Occam's Razor) [Informal Version]:

- If there is a consistent learner L for C using H. Then provided we draw
$$m \geq \frac{1}{\varepsilon}\left(\log|H_n| + \log\frac{1}{\delta}\right)$$
examples from $EX(c, D)$ we obtain a PAC-learning algorithm using L.

What if C or H is infinite?

C is class of axis-aligned rectangles (this is uncountably infinite).

Halfspaces / Linear Separators / Linear Threshold Functions (LTFs)



(i) Can we find a consistent learner?

if y_i is 1, then $\underline{w} \cdot \underline{x}_i + w_0 \geq 0$

y_i is 0 then $\underline{w} \cdot \underline{x}_i + w_0 \leq -\delta$

Feasible solution to such a linear program suffices to yield a consistent learner.

For the rest of the discussion today keep n implicit.

Let $S \subseteq X$, S is a finite set. Let C be a concept class over X.

$$\Pi_C(S) = \{ c_S \mid c \in C \}.$$

If $S = \{x_1, x_2, \ldots, x_m\}$

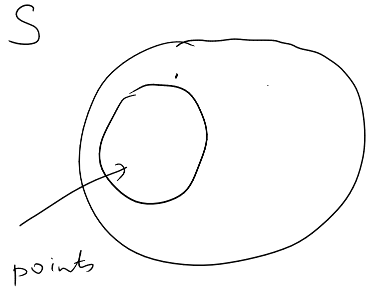
$$\Pi_C(S) = \{ (c(x_1), \ldots, c(x_m)) \mid c \in C \}$$

$$|\Pi_C(S)| \leq 2^m$$

$\Pi_C(S)$ represent all behaviors or dichotomies on S that are induced by C.

Def (Shattering) We say that a finite set S is shattered by a concept class C, if $|\Pi_C(S)| = 2^{|S|}$. Alternatively S is shattered by C if all possible dichotomies of S can be induced by C.

S



all points $x \in S$, s.t. $c(x) = 1$

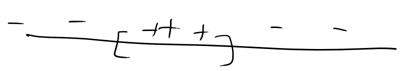
Each $c \in \Pi_C(S)$ can be viewed as a subset of S.

$\Pi_C(S)$ collection of subsets of S.

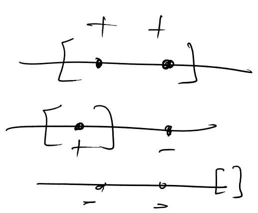
Shattering $\equiv$ $\Pi_C(S)$ is the power set of S.

Def: The Vapnik-Chervonenkis (VC) dimension of C, denoted by $VCD(C)$ or VC-dim (C), is the cardinality d of the largest finite set S shattered by C. (If C shatters arbitrarily large finite sets, we will say $VCD(C) = \infty$.)

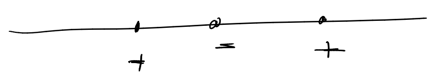
Example: Intervals on $\mathbb{R}$, C_1.



$VCD(C_1)$?

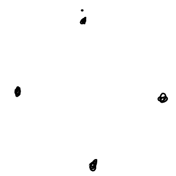


Set of size 2 that is shattered

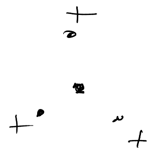


No set of size 3 can be shattered.

- What is the VC dim of RECTANGLES in $\mathbb{R}^2$?



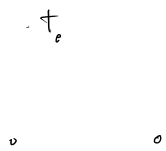
This set is shattered



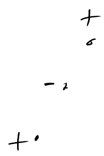
This set is not shattered.

5 points, the There is one point <u>that is not exclusively</u> the leftmost/rightmost/top-most or bottom-most.

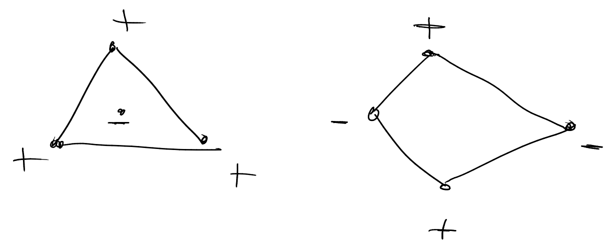
- What is the VC-dim of halfspaces/LTFs in $\mathbb{R}^2$?



This set is shattered



(This set of 3 points is not shattered)



These configurations are not achievable

VC-dim of halspaces in $\mathbb{R}^n$ is $n+1$.

X instance space
C concept class

Growth Function:

<u>Defn</u>: For any natural number m,
define $\Pi_C(m) = \max\{|\underbrace{\Pi_C(S)}_{\text{dichotomies}}| \; |S| = m\}$.

If VC dimension of C is d, then $\forall m \le d$, $\Pi_C(m) = 2^m$.

Want to understand behaviour of $\Pi_C(m)$ for $m \ge d$?

<u>Sauer's Lemma</u>: For $m \ge d$, $\Pi_C(m) = O(m^d)$.

<u>Defn</u>: $\Phi_d(m) = \Phi_d(m-1) + \Phi_{d-1}(m-1)$ $\quad$ $\Phi_0(m) = 1 \;\forall m$

$\Phi_d(0) = 1 \;\forall m$.

Easy induction proof:

$$\Phi_d(m) = \sum_{i=0}^{d} \binom{m}{i}. \qquad \left[\binom{m}{i} = \binom{m-1}{i} + \binom{m-1}{i-1} \right]$$

(Exercise).

Lemma: $\Pi_C(m) \le \Phi_d(m)$ where $d = VC\text{-dim}(C)$.

Proof: when $d=0$, $\Pi_C(m) \le \Phi_0(m) = 1$

when $m=0$, $\Pi_C(m) \le \Phi_d(0) = 1$

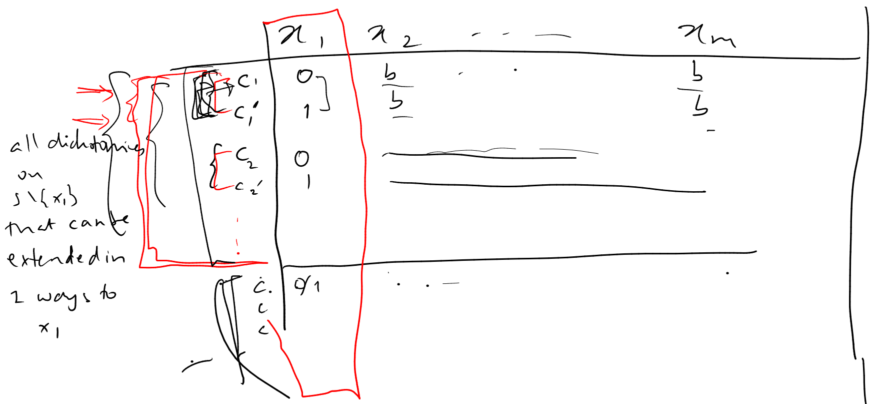
Assume this holds for all $d' \le d$ & $m' \le m$ provided at least one inequality is strict.

Let S be any set of size m.

$S = \{x_1, x_2, \ldots, x_m\}$; Look at x_1

$C_1 = \{c \in \Pi_C(S) \mid c(x_1)=0$ & $\exists c' \in \Pi_C(S)$, s.t $c'(x_1)=1$ & $c(x_i)=c'(x_i) \, \forall i \ne 1\}$

C_1 is a concept class over S (not X)



$$|\Pi_C(S)| = |\Pi_C(S \setminus \{x_1\})| + |\Pi_{C_1}(S \setminus \{x_1\})|$$

$$|\Pi_C(S)| \le \Phi_d(m-1) + \Phi_{d-1}(m-1) = \Phi_d(m)$$

$VC\text{-dim}(C_1) \le d-1$.

Suppose $VC\text{-dim}(C_1) \ge d$, then $\exists T \subseteq S$, st

T is shattered by C_1, $|T| = d$.

- $x_1 \notin T$, x_1 cannot be labelled 1 by any concept in C_1.
- $T \cup \{x_1\}$ is shattered by C.

But $|T \cup \{x_1\}| \ge d+1$ which contradicts $VC\text{-dim}(C) = d$.

Sauer's Lemma : $\Pi_C(m) = O(m^d)$ for $m \geq d$

if $d = \text{VC-dim}(C)$

We proved $\Pi_C(m) \leq \underline{\Phi_d(m)}$

$$\Phi_d(m) = \sum_{i=0}^{d} \binom{m}{i} \cancel{\leq} \frac{m^d}{d!} + \underbrace{\ldots\ldots}_{} \leq \left(\frac{1}{d!}+c\right) m^d \quad \text{for } m \text{ large enough}$$

$$= \binom{m}{0} + \binom{m}{1} + \ldots + \binom{m}{d} \qquad d! \approx \left(\frac{d}{e}\right)^d$$

$$= \left(\frac{m}{d}\right)^d \left(\sum_{i=0}^{d} \left(\frac{d}{m}\right)^d \binom{m}{i} \right)$$

$$\leq \left(\frac{m}{d}\right)^d \left(\sum_{i=0}^{m} \left(\frac{d}{m}\right)^i \binom{m}{i} \right)$$

$$\leq \left(\frac{m}{d}\right)^d \cdot \left(1 + \frac{d}{m}\right)^m \leq \left(\frac{m}{d}\right)^d \cdot e^d = \left(\frac{me}{d}\right)^d = \underline{O(m^d)}$$

$\Pi_C(S)$ $\qquad S \subseteq X$

H_n by $\Pi_C(S)$ in Occam's Razor Theorem.

Replace $\underline{\log |H_n|}$ by $\log |\Pi_C(m)| \approx \log(m^d) \approx \underline{d \cdot \log m}$