VC theory

To simplify the math we will work with we replace binary classifiers with sets. Hypothesis class will become a set system.

$$h \longleftrightarrow h^{-1}(1) = \{x \in X : h(x) = 1\}$$

- Let X be a set.
- The set of all subsets is denoted $\mathcal{P}(X)$. It is also called <u>power set</u> of X. (Hence $\mathcal{P}$.)

- Let $H \subseteq \mathcal{P}(X)$ be a set system (i.e. hypothesis class represented as sets)

- Let $S \subseteq X$.

- $\Pi_H(S) = \{S \cap h : h \in H\}$ is set of <u>behaviors</u> of H on S

- Note that $\Pi_H(S) \subseteq \mathcal{P}(S)$

- Note that $|\Pi_H(S)| \leq 2^{|S|}$

(if S is finite)

- $$\Pi_H(m) = \max_{\substack{S \subseteq X \\ |S|=m}} |\Pi_H(S)|$$

is the maximum possible number of behaviors on set of size m.

- If $S \subseteq S'$ then

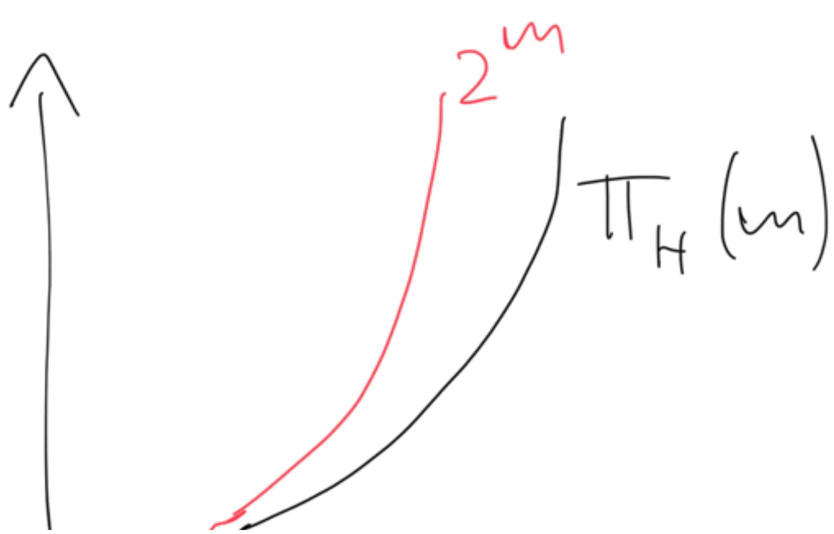
$$|\Pi_H(S)| \leq |\Pi_H(S')|$$

therefore

$\Pi_H(m)$ is a non-decreasing function of m.

- We say H <u>shatters</u> S iff $\Pi_H(S) = \mathcal{P}(S)$.

- <u>Vapnik-Chervonenkis dimension</u> of H is the size of the largest shattered set.
It is denoted $VC(H)$.



- If H does not shatter any set $VC(H) = -1$.

- If H shatters sets of arbitrarily large size, then $VC(H) = +\infty$.

- Example

Let $X = \mathbb{R}^2$ and H be the set of all closed halfspaces.

$$H = \left\{ \{(x_1, x_2) \in \mathbb{R}^2 : ax_1 + bx_2 \le c\} : \begin{array}{l} a, b, c \in \mathbb{R} \\ a^2 + b^2 = 1 \end{array} \right\}$$

$VC(H) = 3$

- All 8 possible behaviors are possible.

- If $|S| = 4$ the H does NOT shatter S.

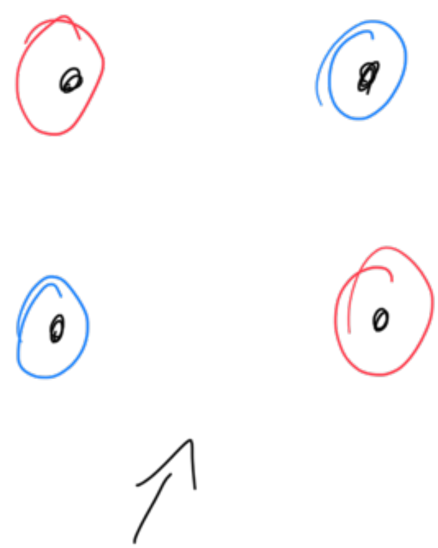
<u>Case 1:</u>

This behavior is impossible.

One point is convex combination of three others

Case 2:

No point is a convex combination of the remaining 3



This behavior is impossible.

The number of behaviors $\Pi_H(m)$ can be upper bounded in terms m and $VC(H)$.

<u>Sauer's lemma</u> (1972)

$$\Pi_H(m) \leq \sum_{i=0}^{VC(H)} \binom{m}{i}$$

Follows from Pajor's lemma

<u>Pajor's lemma</u> (1984)

$$|\Pi_H(S)| \leq |\{S' \subseteq S : H \text{ shatters } S'\}|$$

$|\Pi_H(S)| = |\{ \ldots \}|$

Proof of Pajor's lemma:

Think of $\Pi_H(S)$ as a matrix

$$
\begin{array}{c|cccc}
 & \overbrace{x_1 \quad x_2 \quad \cdots \quad x_m}^{S} \\
\hline
 & 0 & 1 & \cdots & 1 \\
|\Pi_H(S)| & \vdots & & & \\
 & 1 & 0 & \cdots & 0
\end{array}
\qquad m = |S|
$$

- Subset of columns is shattered if any combination of 0/1 is possible

- We need to show that number of rows of the matrix is less that number of shattered subsets of columns

Proof by induction

If A has 1 column

x_1	x_1	x_1	x_1
0 1	0	1	N/A
$\{\emptyset, \{x_1\}\}$	$\{\emptyset\}$	$\{\emptyset\}$	$\emptyset$
$L=2, R=2$	$L=R=1$	$L=R=1$	$L=0, R=0$

Induction step:

- Remove last column x_m.

- We will get some duplicates.
- Move rows that are unique and the first copy of duplicates into matrix B
- Move the second copy of a duplicate row into C
- By construction

$$\#\text{rows}(A) = \#\text{rows}(B) + \#\text{rows}(C)$$

- By induction hypothesis

$\#\text{rows}(B) \leq \text{shattered}(B)$

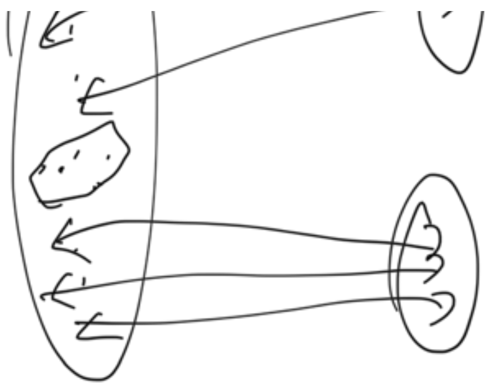
$\#\text{rows}(C) \leq \text{shattered}(C)$

key observation:

$$\text{shattered}(B) + \text{shattered}(C) \leq \text{shattered}(A)$$

- If S' is shattered by B, then A shatters S'.
- If S' is shattered by C, then A shatters $S' \cup \{x_m\}$.

shattered subsets of B



shattered subsets of C

shattered subsets of A

$$
\begin{aligned}
\#\text{rows}(A) &\le \#\text{rows}(B) + \#\text{rows}(C) \\
&\le \text{shatt}(B) + \text{shatt}(C) \\
&\le \text{shatt}(A)
\end{aligned}
$$

$\blacksquare$

Proof of Sauer's lemma

$$\{S' \subseteq S : H \text{ shatters } S'\}$$

$$\cap|$$

$$\{S' \subseteq S : |S'| \leq VC(H)\}$$

$$|\{S' \subseteq S : |S'| \leq VC(H)\}|$$

$$= \sum_{i=0}^{VC(H)} \binom{m}{i} \quad \text{where } m = |S|.$$

- If $i > m$ then $\binom{m}{i} = 0$
- If $m \leq d$ then $\sum_{i=0}^{d} \binom{m}{i} = \sum_{i=0}^{m} \binom{m}{i} = 2^m$
- $\sum_{i=0}^{d} \binom{m}{i}$ is a polynomial in m of degree d.

<u>Lemma</u>: For any $m, d \geq 0$

$$\sum_{i=0}^{d} \binom{m}{i} \leq 1 + m^d$$

Proof:

Left-hand side is number of subsets of $\{1, 2, \ldots, m\}$ of size $i = 0, 1, \ldots, d$.

Encode any subset of size $i \neq 0$, $i \leq d$, as a d-tuple

$$\{x_1, \ldots, x_i\} \rightarrow (x_1, x_2, \ldots, x_i, \underbrace{x_i, \ldots, x_i}_{\substack{d-i \\ \text{extra} \\ \text{copies}}})$$

This mapping is clearly one-to-one. Therefore

$$\sum_{i=1}^{d} \binom{m}{i} \leq m^d$$

Since $\binom{m}{0} = 1$ we are done ∎

Lemma: If $m \geq d \geq 1$ then

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

Proof:

Note that $\frac{m}{d} \geq 1$.

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

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

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

$$= (m/d)^d \left(1+\frac{d}{m}\right)^m$$

$$\leq (m/d)^d \left(e^{\frac{d}{m}}\right)^m$$

$$= \left(\frac{me}{d}\right)^d$$

<u>Example (intervals):</u>

$X = \mathbb{R}$

$H = \{[a,b] : a,b \in \mathbb{R},\ a \leq b\}$

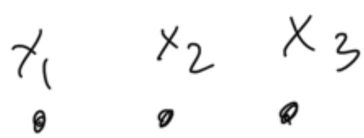
$VC(H) = 2$

$\{x_1\}$

$\{x_2\}$

$\{x_1, x_2\}$

$\emptyset$



$\{x_1, x_3\}$

is impossible

Number of behaviors on sample of size m

- Empty set
- m singleton sets
- $\binom{m}{2}$ intervals with two distinct end-points

$$\#\text{behaviors} = 1 + m + \binom{m}{2}$$

$$= \binom{m}{0} + \binom{m}{1} + \binom{m}{2}$$

Halfspaces

$X = \mathbb{R}^d$

- Homogenous closed half space

$$a_1 x_1 + a_2 x_2 + \cdots + a_d x_d \geq 0$$

$(a_1, a_2, \dots, a_d) \neq (0, 0, \dots, 0)$

- Non-homogenous closed halspace

$$a_1 x_1 + a_2 x_2 + \dots + a_d x_d \geq b$$

$$(a_1, a_2, \dots, a_d) \neq (0, 0, \dots, 0), \; b \in \mathbb{R}$$

Class of halfspaces

- $H_d = \{ \{ (x_1, x_2 \dots x_d) \in \mathbb{R}^d :$
$a_1 x_1 + a_2 x_2 + \dots + a_d x_d \geq b \}$
$(a_1, a_2, \dots, a_d) \neq (0, 0, \dots 0),$
$a_1, a_2, \dots, a_d, b \in \mathbb{R} \}$

- Class of homogeneous

halfspaces is defined analogously.

<u>Lemma</u>

$$VC(H_d) \geq d+1$$

<u>Proof:</u>

If S are vertices of any simplex, then H_d shatters S. Largest simplex has $d+1$ vertices. For example

- $S = \{(0,\dots,0), (1,0,\dots 0)$
  $(0,1,0\dots 0), \dots, (0,\dots,0,1)\}$

- $|S| = d+1$
- $S = \{\bar{0}, e_1, e_2, \dots, e_d\}$

$S' \subseteq S$

- If $\bar{0} \notin S'$

$$\sum_{e_i \in S'} x_i - \sum_{e_i \notin S'} x_i \geq 1$$

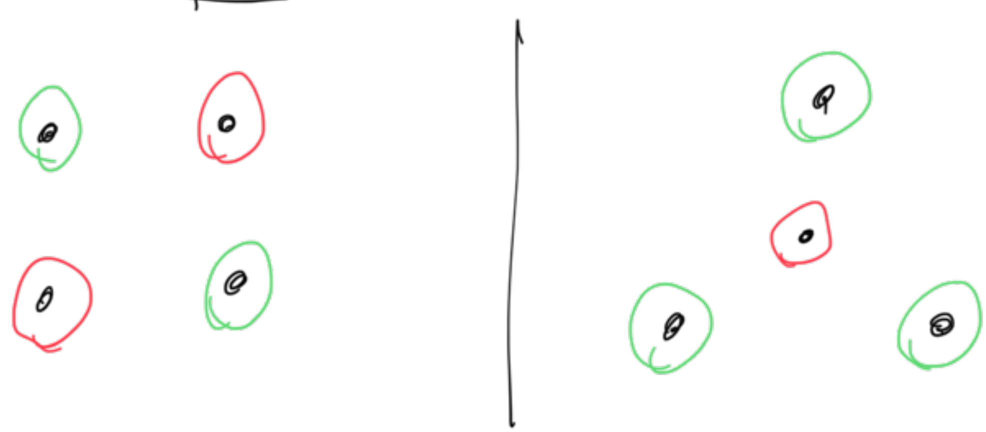
- If $\bar{0} \in S'$

$$\sum_{e_i \in S'} x_i - \sum_{e_i \notin S'} x_i \geq 0$$

Radon's theorem:

If $S \subseteq \mathbb{R}^d$ is a set of $d+2$ points, it can be partitioned into two disjoint sets such that their convex hulls intersect.

<u>Example:</u> in $\mathbb{R}^2$



<u>Proof:</u>

Let $S = \{x_1, x_2, \ldots, x_{d+2}\}$

Consider the system of equations

$$\sum_{i=1}^{d+2} a_i = 0$$ // 1 equation

$$\sum_{i=1}^{d+2} a_i x_i = \bar{0}$$ // d equations one for each coordinate

The unknowns are $a_1, a_2, \ldots, a_{d+2}$!

There must exist a non-zero solution. In particular there is at least one positive and at least one negative a_i

$$I = \{ x_i \in S : a_i > 0 \}$$

$$J = \{x_i \in S : a_i \leq 0\}$$

- Let $A = \sum_{x_i \in I} a_i$
- Clearly $A > 0$
- $\underbrace{\sum_{x_i \in I} a_i}_{A} + \underbrace{\sum_{x_i \in J} a_i}_{-A} = 0$
- We find point p in the intersection of convex hulls of I and J.

$\sum a_i \quad \sum -a_i$

$$p = \sum_{x_i \in I} \underbrace{\frac{a_i}{A}}_{>0} x_i = \sum_{x_i \in J} \underbrace{\frac{-a_i}{A}}_{\geq 0} x_i$$

Note that $\displaystyle\sum_{x_i \in I} \frac{a_i}{A} = 1$

and likewise $\displaystyle\sum_{x_i \in J} -\frac{a_i}{A} = 1$

- So p is a convex combination of points in I.
- Also p is a convex combination of points in J.
- So the convex hulls intersect.

Lemma:

$$VC(H_d) \leq d+1$$

Proof:

Let $S \subseteq \mathbb{R}^d$ be of size $d+2$. We show that H_d does NOT shatter S.

According to Radon's theorem we can split S into two sets, convex hulls of which intersect.

- $S = I \cup J$ $\quad I \cap J = \emptyset$

$p \in conv(I) \cap conv(J)$

- Suppose $h \in H_d$ is

such that

$$h \cap S = I$$

$$h^c \cap S = J$$

since h, h^c are convex

- $h \supseteq \mathrm{conv}(I)$
- $h^c \supseteq \mathrm{conv}(J)$

$p \in h \cap h^c$ which is a a contradiction since h, h^c are disjoint

---

Theorem

$$VC(H_d) = d+1$$

For homogenous halfspaces