Chapter 7 Vapnik Chervonenkis Classes

Most abstract setup

$(Z, P, \mathcal{F})$ $\quad Z_1, \dots, Z_n$ iid P

$$\hat{f}_n = A(Z^n) \qquad \min P(\hat{f}_n)$$

ERM $\quad \hat{f}_n = \underset{f}{\operatorname{argmin}}\ P_n(f)$

We saw $\quad P(\hat{f}_n) \le L_P^*(\mathcal{F}) + \overbrace{4 E R_n(\mathcal{F}(Z^n))}$

$$+ \sqrt{\frac{2\log(1/\delta)}{n}}$$

with probability $\ge 1-\delta$.

We also saw finite class lemma

$$R_n(A) \le \frac{2L\sqrt{\log N}}{n} \le \frac{2\sqrt{n}\sqrt{\log n^2}}{n} = 2\sqrt{\frac{2\log n}{n}}$$

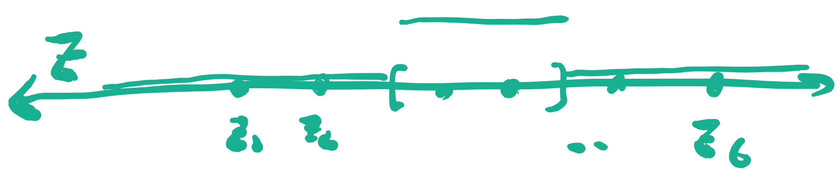
if $A = \{a_1, \dots, a_N\}$, $\quad \|a_i\| \le L \quad 1 \le i \le N.$

For case of intervals - see below

Think about $R_n(\mathcal{F}(Z_n))$ in case $\mathcal{F}$ is a set of binary valued functions.

$f(z_1), \dots, f(z_n) \in \{0,1\}^n$ for all $f \in \mathcal{F}$.

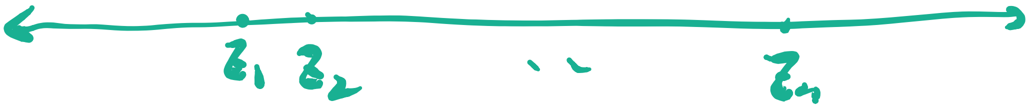
eg. $Z = \mathbb{R}$ f's are indicator functions of intervals



$2^6 = 64$

0 0 1 1 0 0
1 1 1 1 0 0
1 1 1 1 1 1
0 0 0 0 0 0
⋮

00000000
00111000
00111000



$$\overset{n_1}{\longleftrightarrow}\ \overset{n_2}{\longleftrightarrow}\ \overset{n_3}{\longleftrightarrow}$$
$$\|00000\,1111111000\|$$

$n_1+n_2+n_3=n$

$\le n^2$ sequences possible

$$\sqrt{\sum_{i=1}^{n} b_i^2} \le \sqrt{n}$$

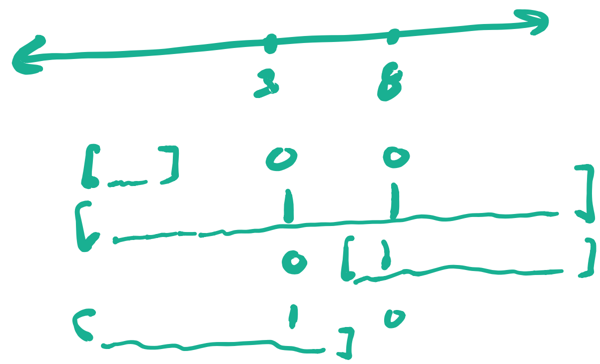
Let $\mathcal{C}$ be a family of subsets of Z.

A set $S=\{z_1,\cdots,z_n\}\subset Z$ is said to be <u>shattered</u> by $\mathcal{C}$ if for any $b\in\{0,1\}^n$ there exists $C\in\mathcal{C}$ so

$$(1_{z_1\in C},\cdots,1_{z_n\in C})=b$$

e.g. $Z = \mathbb{R}$ $\quad \mathcal{C}$ = set of intervals

$S = \{3, 8\}$



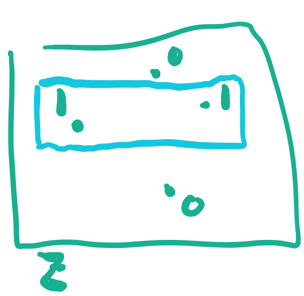
$S = \{3, 8, 11\}$ is not shattered by $\mathcal{C}$ — can't get 101

<u>**Def**</u> The VC dimension of $\mathcal{C}$ is

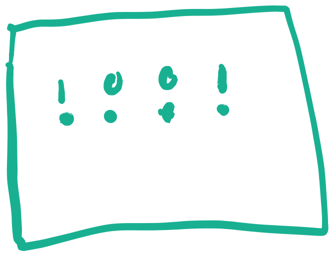
$$\sup\{n : \text{there exists } S = \{z_1, \dots, z_n\} \text{ that is shattered by } \mathcal{C}\}$$

$= V(\mathcal{C})$

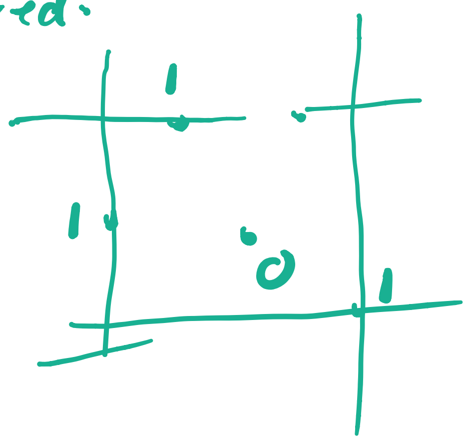
<u>**Examples**</u> Axis aligned interval classifiers



4 point set is shattered



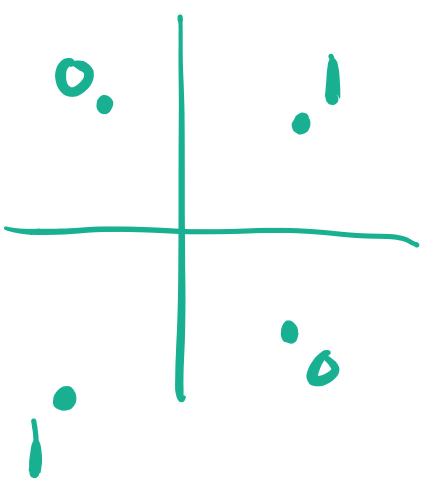
<u>Claim</u> If $|S| = 5$, S cannot be shattered.



$\underline{V(\mathcal{F}) = 4}$

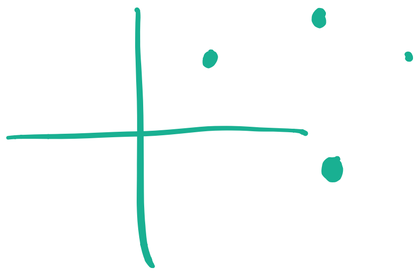
Half space in $\mathbb{R}^2$

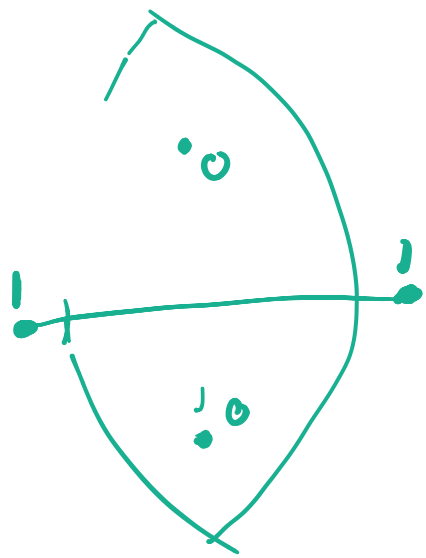
$V(\mathcal{F}) \geq 3$



disks in $\mathbb{R}^2$

$vc(\mathcal{C}) \geq 3$





7.2.5 Dudley class

$$\mathcal{C} = \{\{z : g(z) + h(z) \geq 0\} : g \in \mathcal{G}\}$$

$\mathcal{G}$ = set of linear combinations

$$g(z) = \sum_{i=1}^{m} c_j \psi_j(z)$$

$c_1, \dots, c_m \in \mathbb{R}$

are linearly independent

Then $V(\mathcal{C}) = m$

Proof ($V(C) \geq m$) Let $z_1, \dots, z_m$ be

such that matrix $\longrightarrow$ has full rank.

$$\begin{pmatrix} \psi_1(z_1) & \cdots & \psi_m(z_1) \\ \psi_1(z_2) & & \vdots \\ \vdots & & \\ \psi_1(z_m) & & \psi_m(z_m) \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

Then for any $b_1, \dots b_m \in \{1, -1\}$

$$g(z_i) + h(z_i) = b_i \qquad 1 \leq i \leq m$$

$$\sum_{j=1}^{m} c_j \psi_j(z_i) + h(z_i) = b_i$$

$$\begin{pmatrix} -h(z_1) + b_1 \\ \vdots \\ -h(z_m) + b_m \end{pmatrix} = \begin{pmatrix} \quad & & \quad \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

Examples Closed half spaces in $\mathbb{R}^d$

$$\left\{ \underline{\langle w, z \rangle - b} \geq 0 : \begin{array}{l} w \in \mathbb{R}^d \\ b \in \mathbb{R} \end{array} \right\}$$

$$\psi_1(z) = z_1, \cdots \quad \psi_d(z) = z_d \quad \psi_{d+1}(z) = 1$$

$$h \equiv 0$$

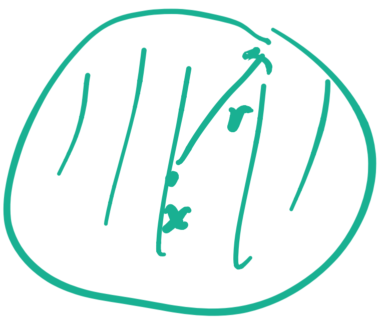
$$w_1\psi_1(z) + \cdots + w_d\psi_d(z) - b\,\psi_{d+1}(z) \geq 0$$

$$\therefore V(\mathcal{C}) = d+1$$

Balls in $\mathbb{R}^d$

$$C = \{ z \in \mathbb{R}^d : \underline{\|z - x\|^2 \le r^2} \}$$

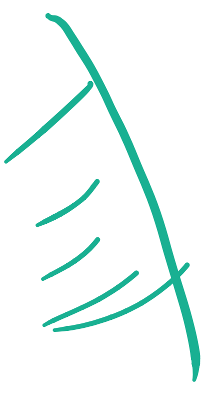
$$V(\mathcal{C}) = d+1$$



$$\sum_{i=1}^{d} (z_i - x_i)^2 - r^2$$

$$\underbrace{\sum_{i=1}^{d} z_i^2}_{h(z)} - \sum_{i=1}^{d} 2 z_i x_i + \underbrace{\sum_{i=1}^{d} x_i^2 - r^2}_{b}$$

$$\psi_i(z) = z_i \quad 1 \le i \le d \qquad \psi_{d+1}(z) = 1$$



Definition Given Z, $\mathcal{C}$ — set of subsets of Z

The n^{th} shatter coefficient

$$S_n(\mathcal{C}) = \text{maximum \# of binary sequences that we can get using } n \text{ points } z_1, \cdots, z_n \in Z.$$

$$= \max_{S=\{z_1, \cdots, z_n\}} |\{S \cap C : C \in \mathcal{C}\}|$$

$$S_n(\mathcal{C}) = 2^n \iff V(\mathcal{C}) \geq n$$

Lemma Sauer-Shelah Lemma

Let $\mathcal{F}$ be a class of binary valued functions with $V(\mathcal{F}) = d$. Then $S_n(\mathcal{F}) \leq \binom{n}{\leq d}$

where $\binom{n}{\leq d} \triangleq \sum_{i=0}^{d} \binom{n}{i}$

and $\binom{n}{\leq d} \leq (n+1)^d$.