Vapnik-Chervonenkis (VC) Classes

ERM: $\hat{f}_n = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} f(z_i)$

$z_1, \dots, z_n \overset{iid}{\sim} P$

Excess risk bound: w.p. $\geq 1 - \delta$,

$$P(\hat{f}_n) - \inf_{f \in \mathcal{F}} P(f) \leq 4\, \mathbb{E} R_n(\mathcal{F}(z^n)) + \sqrt{\frac{2\log(1/\delta)}{n}}$$

where $R_n(\cdot)$ is the Rademacher average,

$$\mathcal{F}(z^n) := \{ (f(z_1), \dots, f(z_n)) : f \in \mathcal{F} \}$$

Binary-valued Hypotheses

: $f : \mathcal{Z} \to \{0,1\}$ $(\{-1,1\})$

$$\mathcal{F}(z^n) = \{ (f(z_1), \dots, f(z_n)) : f \in \mathcal{F} \} \subseteq \underbrace{\{0,1\}^n}_{\text{Boolean } n\text{-cube}}$$

– finite subset of $\mathbb{R}^n$

Finite Class Lemma $\Rightarrow$

$$R_n(\mathcal{F}(z^n)) = \frac{1}{n} \mathbb{E}_\varepsilon \left\{ \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} \varepsilon_i f(z_i) \right| \right\}$$

$$\leq 2 \sqrt{\frac{\log |\mathcal{F}(z^n)|}{n}}$$

$\mathbb{E} R_n(\mathcal{F}(z^n)) \xrightarrow{n \to \infty} 0$ will hold if

$$\frac{1}{n} \sup_{z^n} \log |\mathcal{F}(z^n)| \xrightarrow{n \to \infty} 0$$

$\Rightarrow$ for any P,

$$\mathbb{E}_P R_n(\mathcal{F}(z^n)) \le \sup_{z^n \in \mathcal{Z}^n} R_n(\mathcal{F}(z^n))$$

$$\le 2 \sup_{z^n \in \mathcal{Z}^n} \sqrt{\frac{\log|\mathcal{F}(z^n)|}{n}} \xrightarrow{n\to\infty} 0$$

Vapnik-Chervonenkis combinatorics: only two possibilities

1) $\sup_{z^n \in \mathcal{Z}^n} |\mathcal{F}(z^n)| = 2^n \quad \forall n$

2) $\sup_{z^n \in \mathcal{Z}^n} |\mathcal{F}(z^n)| \le n^d$ for n large enough

$\log(\dots) \le d \log n \qquad \frac{1}{n}\log n \to 0$

Shattering and VC dimension

binary-valued fcns $\longleftrightarrow$ subsets of $\mathcal{Z}$

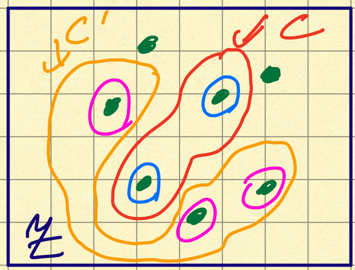
$f: \mathcal{Z} \to \{0,1\} \qquad C_f := \{z \in \mathcal{Z} : f(z) = 1\}$

$\mathcal{C} = \{C : C \text{ a subset of } \mathcal{Z}\}$

S – a finite subset of $\mathcal{Z}$

Def: S is shattered by $\mathcal{C}$ if

$$\forall S' \subseteq S \quad \exists C \in \mathcal{C} \text{ s.t. } S' = S \cap C$$



Shattering $\Leftrightarrow$ ability to assign any binary split to elements of $\mathcal{S}$ using sets in $\mathcal{C}$

Shatter coefficients:

$$\mathbb{S}_n(\mathcal{C}) := \sup_{\substack{\mathcal{S} \subseteq \mathcal{Z} \\ |\mathcal{S}| = n}} \left| \{ \underbrace{C \cap \mathcal{S}}_{\subseteq \mathcal{S}} : C \in \mathcal{C} \} \right|$$

$$0 \leq \mathbb{S}_n(\mathcal{C}) \leq 2^n$$

($\mathbb{S}_n(\mathcal{C}) = 2^n$: $\exists \mathcal{S}$ of size n shattered by $\mathcal{C}$)

Def: $V(\mathcal{C}) := \sup \{ n \in \mathbb{N} : \mathbb{S}_n(\mathcal{C}) = 2^n \}$

- if $V(\mathcal{C}) < \infty$, then we say that $\mathcal{C}$ is a <u>VC class</u> and $V(\mathcal{C})$ is its <u>VC dimension</u>

Prop. if $m > n$ and $\mathbb{S}_m(\mathcal{C}) = 2^m$, then $\mathbb{S}_n(\mathcal{C}) = 2^n$.

(Proof in lecture notes)

Key point: $\mathbb{S}_n(\mathcal{C}) = 2^n$ iff

$\exists \mathcal{S} = \{z_1, \dots, z_n\} \subseteq \mathcal{Z}$ s.t. $\forall b = (b_1, \dots, b_n) \in \{0,1\}^n$

$\exists C^b \in \mathcal{C}$ s.t. $b_i = \mathbb{1}_{\{z_i \in C^b\}}$ $\forall i \in [n]$

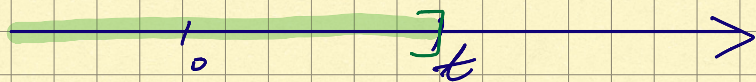
Examples of VC classes

Key idea: to show that $V(\mathcal{C}) = d$,

(i) exhibit at least one $\mathcal{S}$ of size d shattered by $\mathcal{C}$

(ii) prove that no set of size $d+1$ can be shattered by $\mathcal{C}$.

① Semi-infinite intervals $\mathcal{C} = \{(-\infty, t] : t \in \mathbb{R}\}$



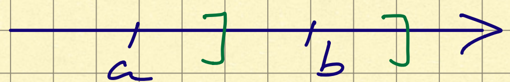
$V(\mathcal{C}) = 1$

(i) $\mathcal{S} = \{a\}$

$\{a\} \cap (-\infty, t_1] = \emptyset$

$\{a\} \cap (-\infty, t_2] = \{a\}$

(ii) $\mathcal{S} = \{a, b\}$, $a < b$



$\mathcal{S}'$: $\emptyset$ ✓

$\{a\}$ ✓

$\{a, b\}$ ✓

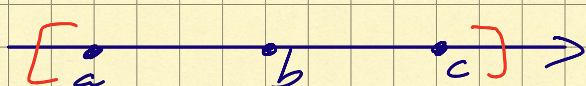
$\{b\}$ ✗

2) Closed intervals $\mathcal{C} = \{[a, b] : -\infty < a \leq b < \infty\}$

$V(\mathcal{C}) = 2$

(i) $\mathcal{S} = \{a, b\}$ $\qquad \mathcal{S}'$: $\emptyset$, $\{a\}$, $\{b\}$, $\{a, b\}$

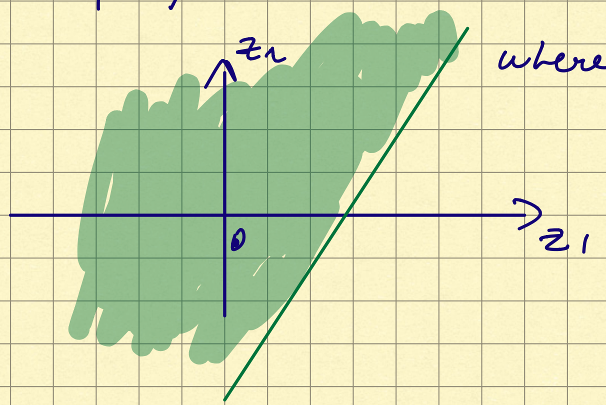
(ii) $\mathcal{S} = \{a, b, c\}$ $\qquad a < b < c$



$\mathcal{S}' = \{a, c\}$ ✗

3) Closed half-planes (in $\mathbb{R}^2$)

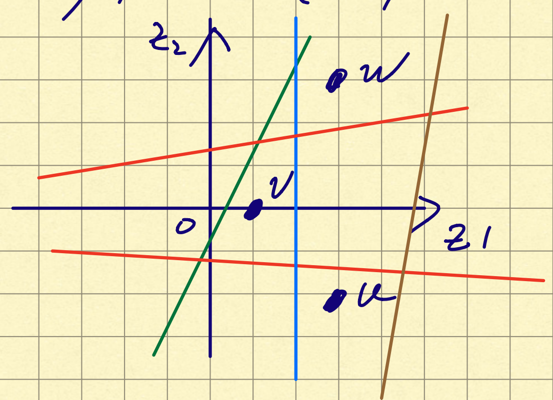
half-plane in $\mathbb{R}^2$: $\{(z_1, z_2) : w_1 z_1 + w_2 z_2 + b \geq 0\}$



where w_1 or $w_2 \neq 0$, $b \in \mathbb{R}$

$V(\mathcal{C}) = 3$

(i) $S = \{u, v, w\}$ s.t. u, v, w are not collinear

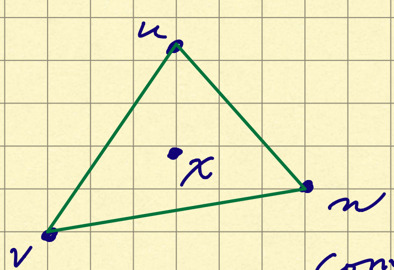


(ii) $S = \{u, v, w, x\}$

two possibilities: a) one pt., say x, is in $\mathrm{conv}(u, v, w)$

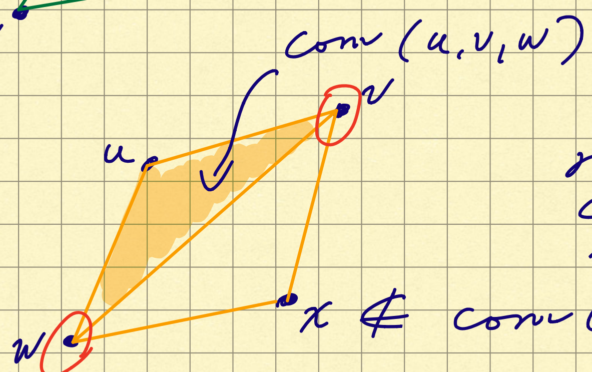
b) points in S are <u>affinely independent</u>

a)



no half-plane can contain u, v, w but not x

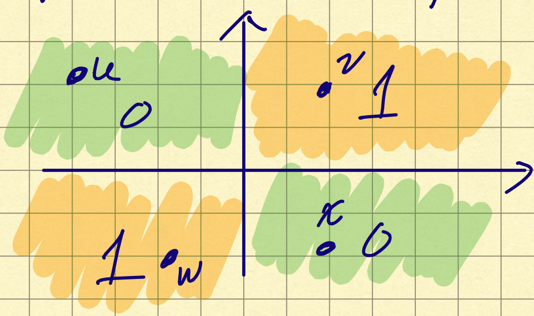
b)



no half-plane can contain just v, w but not u, x

$x \notin \mathrm{conv}(u, v, w)$

XOR counterexample (Minsky-Papert)



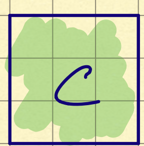
$\mathcal{C}$: half-spaces in $\mathbb{R}^d$

$\{ z \in \mathbb{R}^d : \langle w, z \rangle + b \geq 0 \}$ $\quad w \in \mathbb{R}^d \setminus \{0\}$
$b \in \mathbb{R}$

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

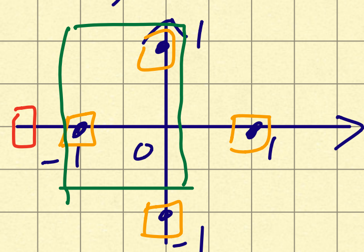
4) <u>axis-aligned rectangles</u>

$\mathcal{C} = \{ [a_1, b_1] \times [a_2, b_2] : a_1 \leq b_1, \; a_2 \leq b_2 \}$



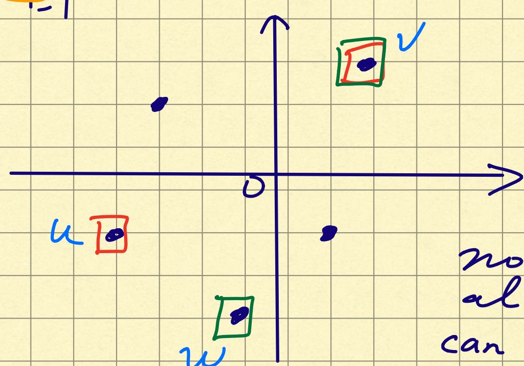
$V(\mathcal{C}) = 4$

(i)



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

(ii) $|S| = 5$



no axis-aligned rect. can contain u, v, w but not the rest

$\mathcal{C}$: axis-aligned hyperrectangles in $\mathbb{R}^d$

$V(\mathcal{C}) = 2d$.

<u>Next lecture:</u>

1) VC dimension of "finite-dim." fcn classes
2) growth of shatter coeffs (Sauer-Shelah lemma)