COL774
Machine Learning
NOV 10, 2021

Last Class:

Learning Theory

$m, \underline{k}, \delta, \gamma$:- finite-sized hypothesis class.

↳ V.C-Dim ($\mathcal{H}$)

$\mathcal{H}$:- infinite sized

$\mathcal{H}$:- set of all linear classifiers

$\Pr(\forall h \in \mathcal{H} \, |\varepsilon(h) - \hat{\varepsilon}(h)| \le \gamma) \ge (1-\delta)$

Uniform convergence bound

↳ Vapnik - Chervonenkis Dimension.

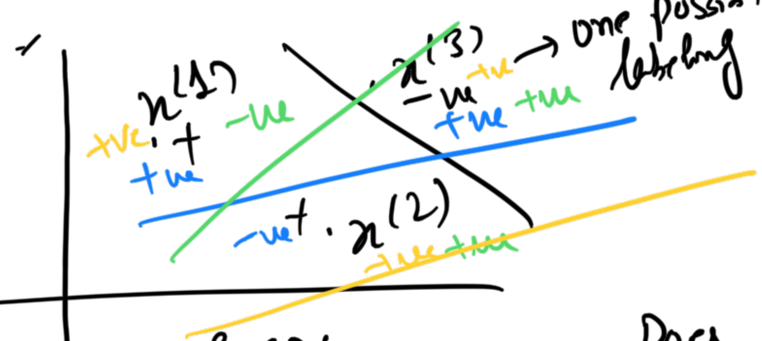
↳ Shattering :- Given a set of points

$S = \{x^{(1)} \ldots x^{(d)}\}$ $x^{(i)} \in \mathcal{X}$, $\mathcal{H}$,

we say that S is shattered by $\mathcal{H}$ if for every possible l (labeling) to the points in S, $\exists h \in \mathcal{H}$, such h realizes l

i.e $h(x^{(i)}) = l^{(i)}$

$l = (l^{(1)} \ldots l^{(d)})$ $l^{(i)} \in \{0, 1\}$



R^2 — (1)

$\mathcal{H}$:- set of all linear classifiers in R^2.

Does $\mathcal{H}$ shatter $S = \{x^{(1)}, x^{(2)}, x^{(3)}\}$

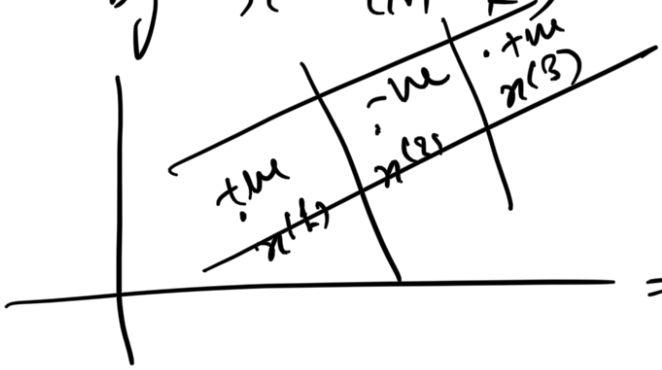
$\forall$ labeling l, $\exists h \in \mathcal{H}$

$h(x^{(i)}) = l^{(i)}$

Yes!

Yes!

Question:- Can any set of 3 points be shattered by $\mathcal{H}$ (in $\mathbb{R}^2$)



$\Rightarrow$ $S = \{x^{(1)}, x^{(2)}, x^{(3)}\}$ → collinear

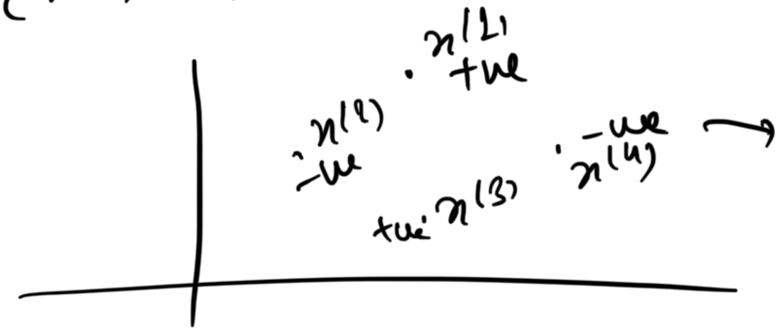
can not be shattered by $\mathcal{H}$.

$\Rightarrow$ Give a labeling ℓ, that can not be realized by any $h \in \mathcal{H}$

$$h_\theta(x) = 1\{\theta^T x \geq 0\}$$

Prove:- $\Rightarrow$ Every set $S = \{x^{(1)}, x^{(2)}, x^{(3)}\}$ in $\mathbb{R}^2$ s.t. $x^{(1)}, x^{(2)}, x^{(3)}$ are not collinear can be shattered by $\mathcal{H}$

$S = \{x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}\}$

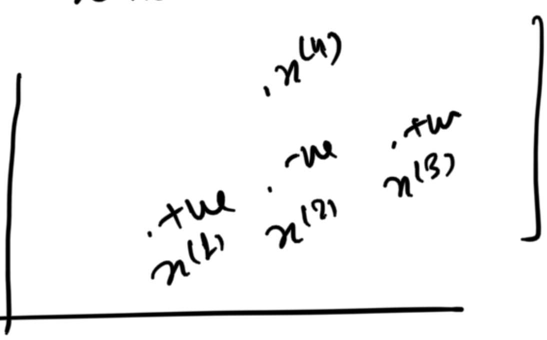


→ $\mathcal{H}$:- set of all linear classifiers

labeling ℓ can not be realized by any $h \in \mathcal{H}$

Question: Is there any set $S \subset \mathbb{R}^2$, $|S| = 4$ which can be shattered by $\mathcal{H}$

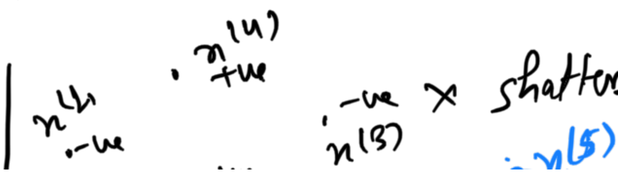
(a)



$S = \{x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}\}$

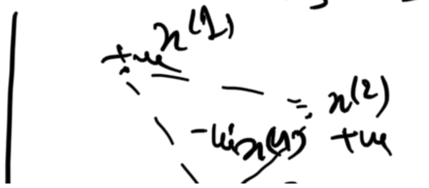
× shattered

(b)

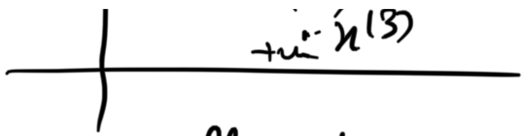


× shattered

(c)



× shattered



⇒ Think about whether above cases suffice to cover "all possible" setting of 4 points in R^2?.

⇒ ∄ $S \subset R^2$, $|S| = 4$ & S can be shattered by $\mathcal{H}$ ($\mathcal{H}$:- set of all linear classifiers in R^2). — (2)

⇒ VC-Dim($\mathcal{H}$) is <u>3.</u>

⇓ set of all linear classifiers in R^2

<u>Definition:-</u> Given an instance space $\mathcal{X}$, $\mathcal{H}$, VC-D($\mathcal{H}$) is defined to be the size of the largest set $S \subseteq \mathcal{X}$ which can be shattered by $\mathcal{H}$. (Captures the "representational" capability of $\mathcal{H}$).

$\mathcal{H}$:- set of all quadratic classifiers in R^2

⇒ 16 labeling
↳ 6 labeling (2+ve, 2-ve) ✓
↳ 4 labeling (3+ve, 1-ve)
↳ 4 labeling (3-ve, 1+ve)
↳ 2 labeling → All the +ve
↳ All -ve

set of all quadratic classifiers in

VC-Dim($\mathcal{H}$) ≥ 4

⇒ VC-Dim($\mathcal{H}$) = <u>d</u> ✓

↳ (a) $\text{VC-Dim}(\mathcal{H}) \geq d$

$\Rightarrow \exists S,\ S \subseteq \mathcal{X}$ s.t. $|S| = d$ & S can be shattered by $\mathcal{H}$.

(b) $\text{VC-Dim}(\mathcal{H}) < d+1$

$\Rightarrow \nexists S \subseteq \mathcal{X}$, s.t. $|S| = d+1$ & S can be shattered by $\mathcal{H}$.

A ↳ Propose a set S $\longrightarrow$ B Provide a labeling ℓ

$|S| = d$

Repeats

Gives $h \in \mathcal{H}$, $h(x^{(i)}) = \ell^{(i)}$ $\forall x^{(i)} \in S$

$\Rightarrow$ Result:- $\mathcal{H}$:- set of all linear classifiers in $\mathbb{R}^n$ ($\mathcal{X} = \mathbb{R}^n$)

$\therefore \text{VC-Dim}(\mathcal{H}) = n+1$]

$\Rightarrow$

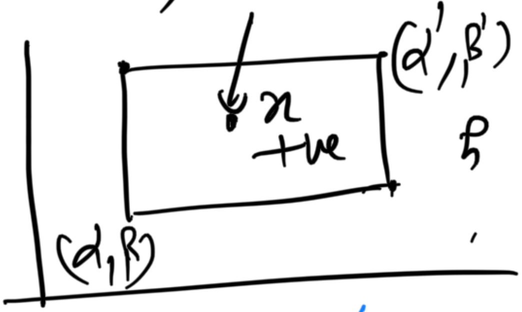
(a) $\mathcal{H}_{\parallel}$:- set of all possible axis $\parallel$ rectangles in $\mathbb{R}^2$

$h \in \mathcal{H}$:-

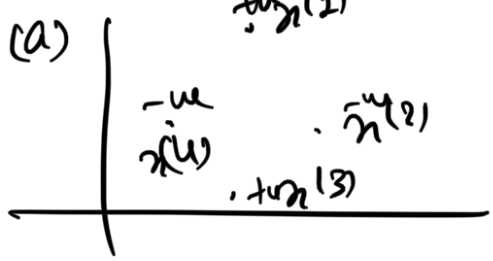
$$h_{(\alpha,\beta),(\alpha',\beta')}(x) = 1\{\alpha \leq x_1 \leq \alpha' \wedge \beta \leq x_2 \leq \beta'\}$$

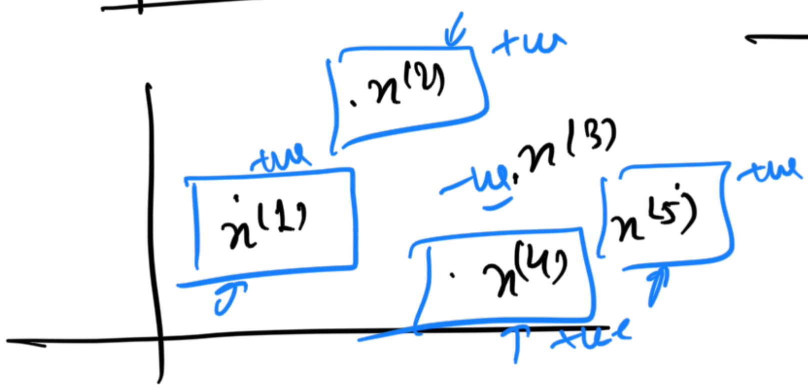
$\ell = ((\alpha,\beta)\ (\alpha',\beta'))$

$h = ((\alpha, \beta), (\alpha', \beta'))$



$VC\text{-}D(\mathcal{H}_{11})$
$= \underline{4}$ ✓

(a)




$\nexists S, |S| = 5$
s.t S can be shattered by $\underline{\mathcal{H}_{11}}$.

PAC: 1

$$\Pr[\forall h \in \mathcal{H}: \underbrace{|\varepsilon(h) - \hat{\varepsilon}(h)|}\leq \gamma] \geq (1 - \delta)$$

$\underbrace{|\mathcal{H}|}, \gamma, \delta, m$

$K \Rightarrow$ finite sized $\mathcal{H}$

VC-Dim($\mathcal{H}$):- infinite sized $\mathcal{H}$

$\underline{(d)}$

Finite Case	Infinite Case
$\lvert\mathcal{H}\rvert = K$	VC-Dim($\mathcal{H}$) $= d$
(a) $m \geq \frac{1}{2\gamma^2} \log \frac{2K}{\delta}$	(a) $m \geq \frac{1}{\gamma}\left(4 \log \frac{2}{\delta} + 8d \log \frac{13}{\gamma}\right)$
	$m = O_{\gamma,\delta}(d)$

Bias-Variance Tradeoff

(b) $r \geq \sqrt{\frac{1}{2m} \log \frac{2k}{\delta}}$

(b) $r \geq O\left(d/m \log m/d + 1/m \log 1/\delta \right)$