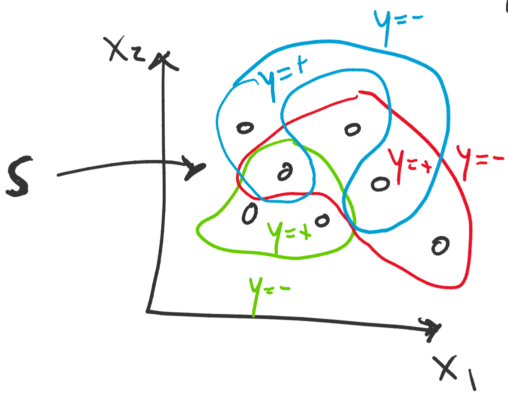
Section B

Wednesday, March 15, 2023 12:29 PM

Ex: Shattering for Binary Classification



$\#$ points $= |S| = 7$

$\#$ labelings of $S = 2^7 = 2^{|S|}$

$\mathcal{H}$ = all circular decision boundaries

$\mathcal{H}[S] = \#$ labelings of S by $\mathcal{H} < 2^7$ ← strictly less than

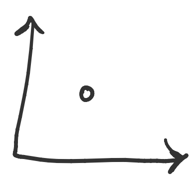
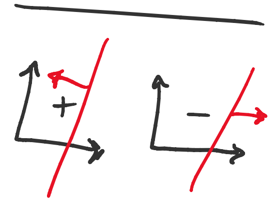
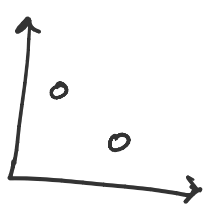
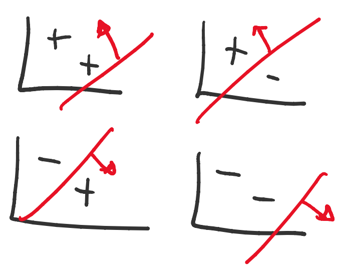
Ex: VC Dimension of Linear Separators

$\mathcal{H}$ = linear separators in 2D

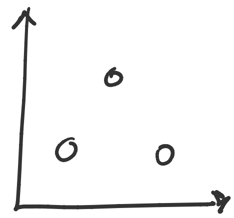
To prove $VC(\mathcal{H}) = d$:

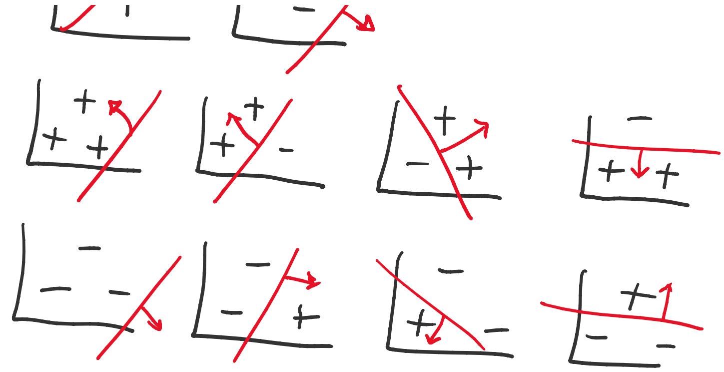
1. $\exists S$ s.t. $|S| = d$ and $\mathcal{H}$ shatters S
2. $\nexists S$ s.t. $|S| = d+1$ and $\mathcal{H}$ shatters S

To show 1:

	Pick a dataset S (unlabeled) for d	List all labelings of S	Show that $\mathcal{H}$ shatters S
$d = 1$			
$d = 2$			

$d=3$



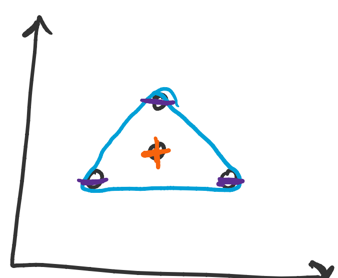


$\Rightarrow VC(\mathcal{H}) \geq 3$

To show (2):

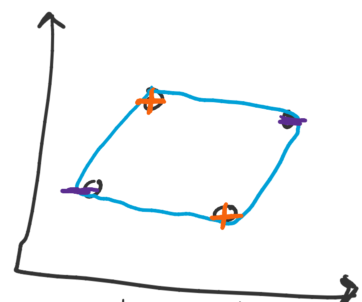
Divide all datasets S of size 4 into two categories:

Case 1



some point inside the convex hull

Case 2



no point inside convex hull

$\Rightarrow VC(\mathcal{H}) < 4$

Conclude: $VC(\mathcal{H}) = 3$

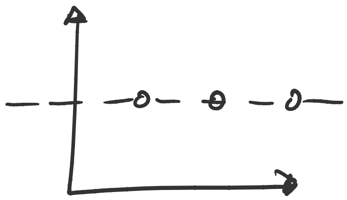
General Case:

If $\mathcal{H}$ = lin sep in M dimensions

then $VC(\mathcal{H}) = M+1$

Q: If $M=2$, is there a dataset of size 3 we cannot shatter?

A:



MLE Bernoulli

① Model: $x^{(i)} \sim \text{Bernoulli}(\phi) \qquad \phi \in [0,1]$

$$p(x^{(i)}|\phi) = \begin{cases} \phi & \text{if } x^{(i)} = 1 \\ (1-\phi) & \text{if } x^{(i)} = 0 \end{cases} = \phi^{(x^{(i)})}(1-\phi)^{(1-x^{(i)})}$$

② $\xrightarrow{\text{Log}}$ Likelihood $\quad D = \{x^{(1)}, \dots, x^{(N)}\}$

$$\begin{aligned} \ell(\phi) &= \log p(D|\phi) \\ &= \log \prod_{i=1}^{N} p(x^{(i)}|\phi) \\ &= \log \prod_{i=1}^{N} \phi^{(x^{(i)})}(1-\phi)^{(1-x^{(i)})} \\ &= \log\left[\phi^{N_1}(1-\phi)^{N_0}\right] \\ &= N_1 \log(\phi) + N_0 \log(1-\phi) \end{aligned}$$

$$N_1 = \#(x^{(i)} = 1)$$
$$N_0 = \#(x^{(i)} = 0)$$

③ Derivative

$$\begin{aligned} \frac{\partial \ell(\phi)}{\partial \phi} &= \frac{\partial}{\partial \phi}\left[N_1 \log(\phi) + N_0 \log(1-\phi)\right] \\ &= \frac{N_1}{\phi} + \frac{N_0}{1-\phi} \end{aligned}$$

④ Set to zero and solve

(4) Set to zero and solve

$$\frac{N_1}{\phi} - \frac{N_0}{1-\phi} = 0$$

oops! actually minus

$$\Rightarrow \phi_{MLE} = \frac{N_1}{N_0 + N_1} = \frac{N_1}{N}$$

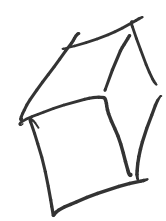
$\boxed{\text{OH}} \in \mathbb{R}^K$

$$\frac{\partial l}{\partial \mathbf{b}} \in \mathbb{R}^K$$

$$\frac{\partial l}{\partial \beta} = \frac{\partial \mathbf{b}}{\partial \beta} \frac{\partial l}{\partial \mathbf{b}}$$

$\mathbb{R}^{K \times M \times K}$

$\mathbb{R}^{K \times M}$



$$\frac{\partial b_k}{\partial \beta} \in \mathbb{R}^{K \times M}$$

$$\frac{\partial l}{\partial \beta_{km}} = \sum_{j=1}^{K} \partial$$