# Gems of TCS

## VC Dimension

Sasha Golovnev

May 4, 2021

# Definitions

- $X$—set of all possible instances/examples

$X$ = set of all possible emails

# Definitions

- $X$—set of all possible instances/examples
- $\mathcal{D}$—target distribution over $X$

# Definitions

- $X$—set of all possible instances/examples
- $\mathcal{D}$—target distribution over $X$
- $c$—target concept

# Definitions

- $X$—set of all possible instances/examples
- $\mathcal{D}$—target distribution over $X$
- $c$—target concept
- $H$—set of concept hypotheses

emails $\xrightarrow{\text{extract features}}$ $\begin{array}{c|c|} x_1 & 1.1 \\ x_2 & 2 \\ x_3 & 3 \\ x_4 & 7 \end{array} \in X$

Example: H = the class of linear funs

$x_1 \cdot 5 + x_2 \cdot 1 + x_3 \cdot 0.1 + x_4 \cdot 7 \geq 10$

H = the class of quadratic funs

$x_1 \cdot x_3 + 5x^2 - 0.1 x_3 \cdot x_1 \geq x_4$

$\begin{array}{c|c|} x_1^2 & \\ x_1x_2 & \\ x_1x_3 & \\ x_1x_4 & \\ x_2^2 & \\ x_2x_3 & \\ \vdots & \end{array}$

H = all linear funs

# Definitions

- $X$—set of all possible instances/examples
- $\mathcal{D}$—target distribution over $X$
- $c$—target concept
- $H$—set of concept hypotheses
- Goal: given training set, select $h \in H$ that approximates $c$ well

# Generalization Error

## Generalization Error

For hypothesis $h$, target concept $c$, and target distribution $D$:

$$R(h) = \Pr_{\boxed{x\sim D}}\left[\overset{=1}{h(x)} \neq \overset{=0}{c(x)}\right].$$

Goal: pick $h$ minimizes Generalization Error

## PAC (Probably Approximately Correct)

Concept class $C$ is PAC-learnable if there exists learning algorithm s.t.

- for all $c \in C$, $\boxed{\varepsilon > 0}$, $\boxed{\delta > 0}$ all distributions $D$:

apx. confidence

Alg
- receives $m$ examples $S$ from $X$ dist $D$
- outputs $\boxed{h_S} \in H$

$$\Pr_{S \sim D^m}[R(h_S) \leq \varepsilon] \geq \boxed{1 - \delta},$$

$\varepsilon = 0.01$ $\delta = 0.01$

# PAC Learning

## PAC (Probably Approximately Correct)

Concept class $C$ is PAC-learnable if there exists learning algorithm s.t.

- for all $c \in C, \varepsilon > 0, \delta > 0$, all distributions $D$:

$$\Pr_{\underline{S\sim D^m}}[R(h_S) \le \underline{\varepsilon}] \ge 1 - \underline{\delta},$$

Alg. doesn't know $D$

- for random samples of size $\varepsilon = \frac{1}{100}$ $\delta = \frac{1}{100}$

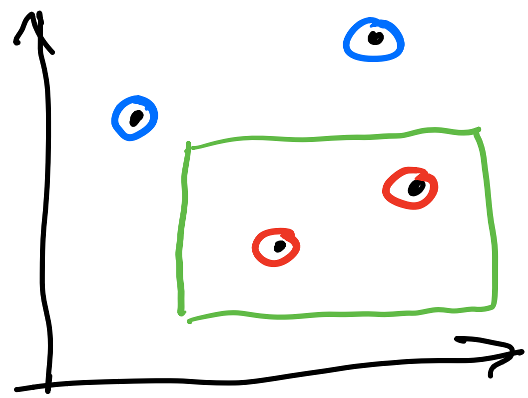
$$\boxed{m} \le \boxed{\text{poly}(1/\varepsilon, 1/\delta)}.$$

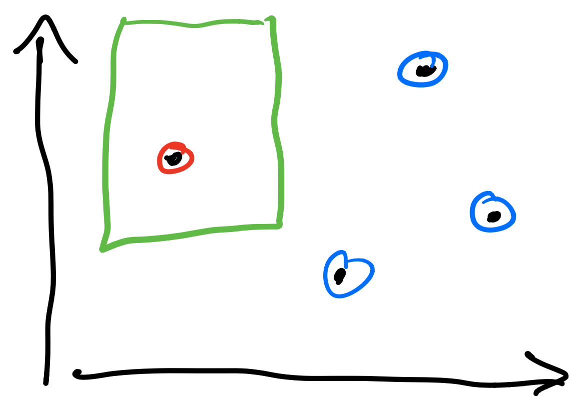
# Last Time

$X \in \mathbb{R}^2$

$\boxed{H}$ = set of axis-aligned rectangles

H is PAC-learnable

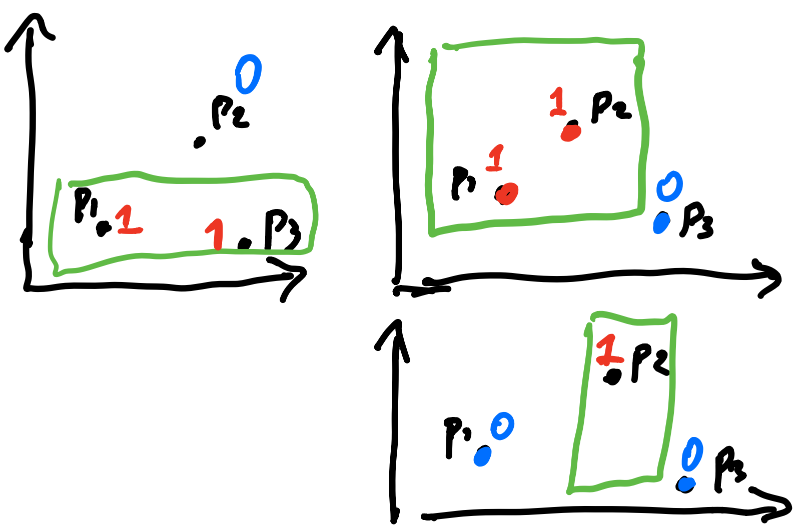




# Shattering

## Dichotomy

Given a set of $m$ instances/examples $S \in X^m$, a dichotomy $\{0,1\}^m$ is one of the possible ways of labeling examples $S$ using a hypothesis from $H$




$m=3$

$S = \{p_1, p_2, p_3\}$

$H =$ axis aligned rectangles

110 – dichotomy of S

010 – dichotomy of S

101 – dichotomy of S

# SHATTERING

## Dichotomy

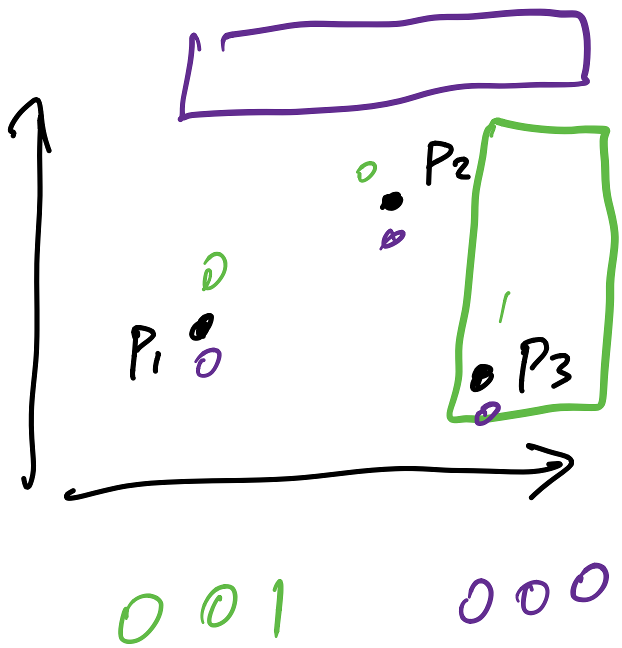
Given a set of $m$ instances/examples $S \in X^m$, a dichotomy $\{0, 1\}^m$ is one of the possible ways of labeling examples $S$ using a hypothesis from $H$

## Shattering

A set of $m$ instances/examples $S \in X^m$ is shattered if all $2^m$ are realizable by hypotheses from $H$.

000 100
001 101
010 110
011 111

are dichotomies of $S$ $\Rightarrow$ $S$ is shattered by $H$




$\Rightarrow$ this set of 3 points is shattered by axis-aligned rectangles

## VC Dimension

Vapnik - Chervonenkis

The VC dimension of $H$ is the size of the largest set that can be shattered by $H$.

Ex. $H$ = axis-aligned rectangles

$\Rightarrow$ VC-dim$(H) \geq 3$

Even if there are some sets of size 3 that cannot be shattered, VC-dim can still be $\geq 3$

Ex. 1 0 1

this set cannot be shattered because this dichotomy 101 cannot be realized by a rectangle

**VC Dimension**

The VC dimension of $H$ is the size of the largest set that can be shattered by $H$.

If we want to prove VC-dim(H) = d

- To prove that VC dimension of $H$ is $d$ we need to

# VC DIMENSION

## VC Dimension

The VC dimension of $H$ is the size of the largest set that can be shattered by $H$.

- To prove that VC dimension of $H$ is $d$ we need to
  - Show a set of $d$ examples that can be shattered by $H$ $\quad$ $\text{VC-dim}(H) \geq d$

# VC DIMENSION

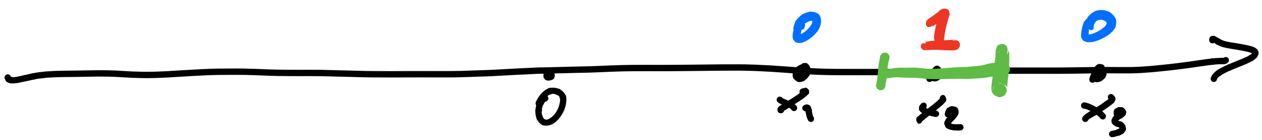
**VC Dimension**

The VC dimension of $H$ is the size of the largest set that can be shattered by $H$.

- To prove that VC dimension of $H$ is $d$ we need to
  - Step 1 — Show a set of $d$ examples that can be shattered by $H$ $\quad$ VC-dim$(H) \geq d$
  - Step 2 — Porve that every set of $d+1$ examples cannot be shattered by $H$ $\quad$ VC-dim$(H) \leq d$

# Intervals on the Line

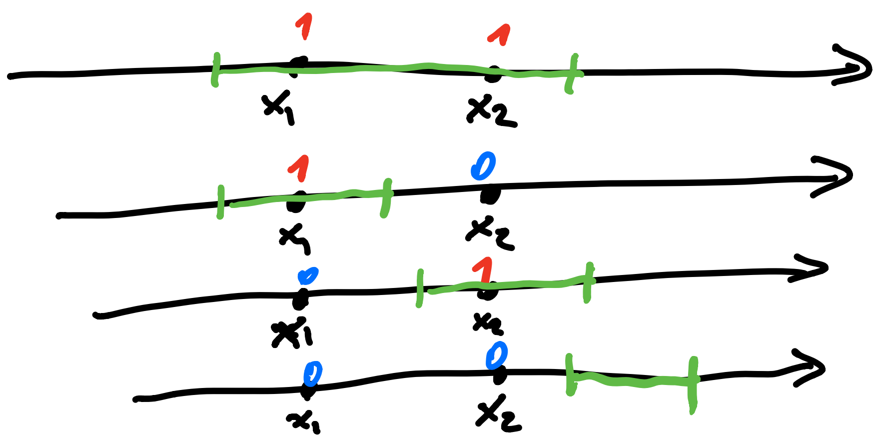
$x \in \mathbb{R}$ H = set of all intervals



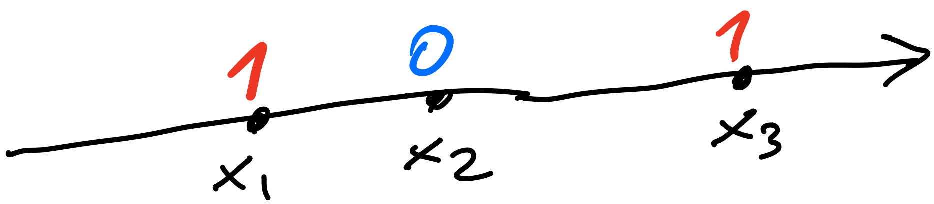

What is VC-dim of H?

$\text{VC-dim}(H) = 2$

- $\text{VC-dim}(H) \geq 2 \iff \exists$ set of 2 points that can be shattered



- $\text{VC-dim}(H) < 3 \iff$ For <u>every</u> set of 3 distinct points, $\exists$ $\{0,1\}^3$ labeling that cannot be realized




No interval contains $x_1$ & $x_3$ but not $x_2$ $\Rightarrow$ no interval realizes this labeling

$\Rightarrow$ any set of 3 points cannot be shattered

$\Downarrow$

$\text{VC-dim}(H) = 2$

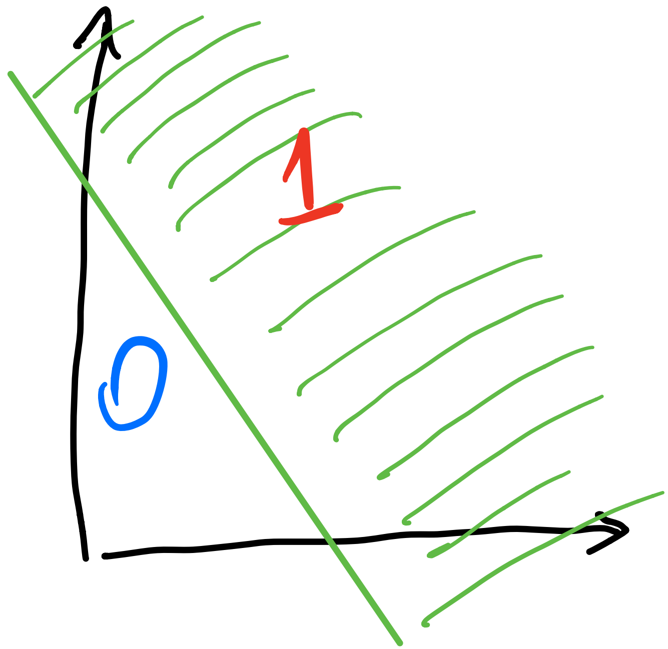
# Half-planes

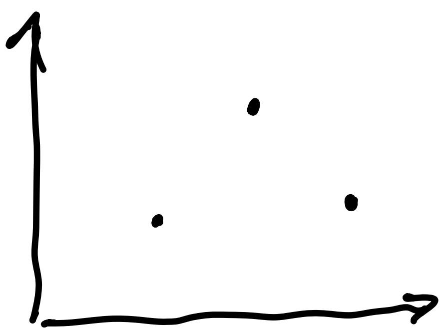
$x \in \mathbb{R}^2$

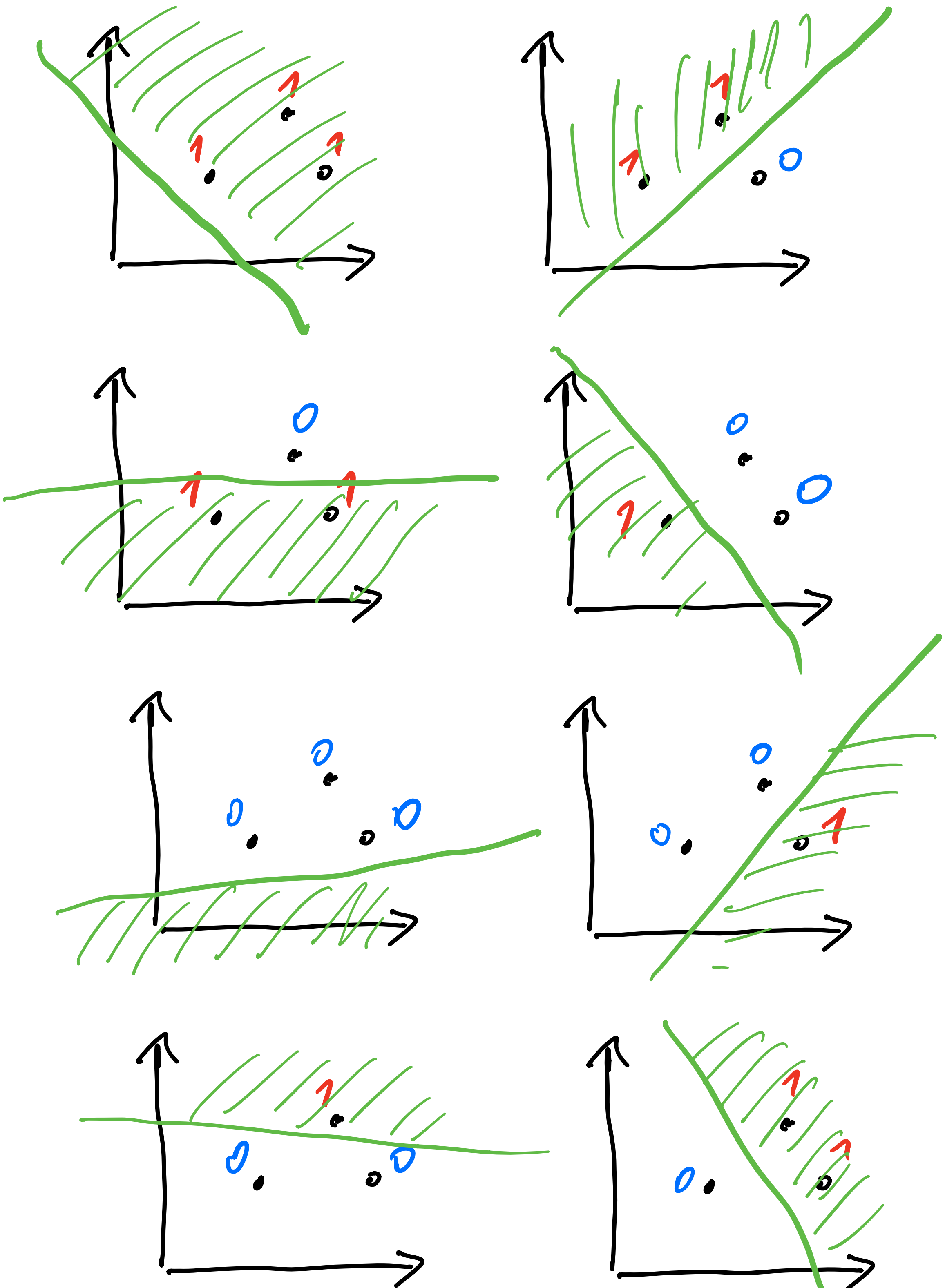
$H$ = all half-planes

$\text{VC-dim}(H) = 3$

- $\text{VC-dim}(H) \geq 3 \iff$ we want to find a set of 3 points that can be shattered

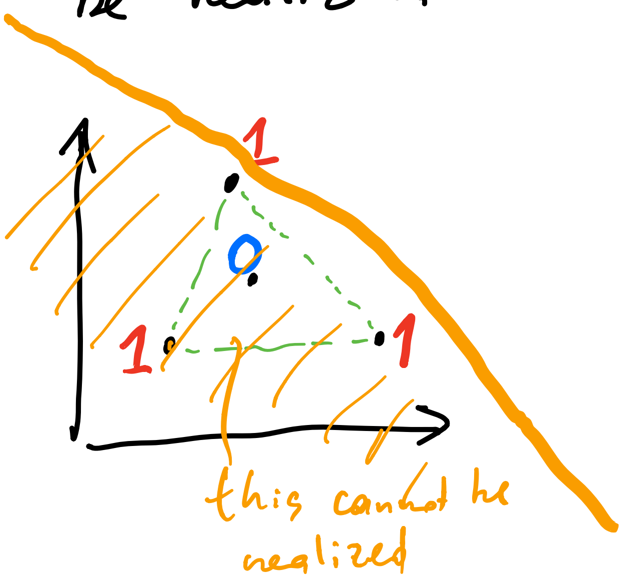


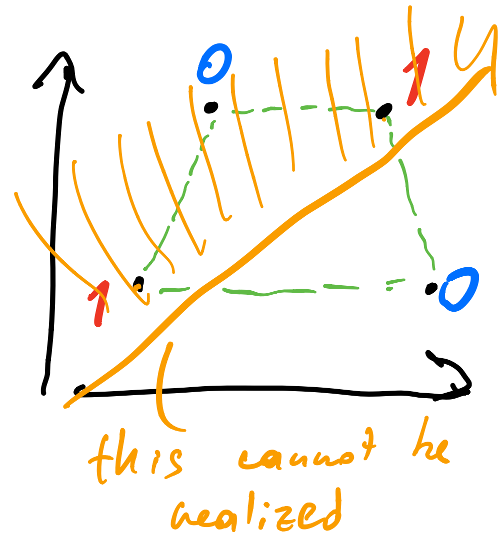





This set of 3 points is shattered $\Rightarrow VC \geq 3$

– VC-dim(H) < 4 $\iff$ $\forall$ set of 4 points $\exists$ labeling that cannot be realized.




No set of 4 points can be shattered

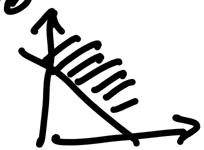
$\Downarrow$

VC-dim(H) = 3

# Half-spaces

$x \in \mathbb{R}^d$ $\quad H$ = linear classifiers = Half-spaces

$x_1 \cdot 0.1 + x_2 \cdot 5 - 7x_3 \geq 10$



**Thm** $\quad \text{VC-dim}(H) = \boxed{d+1}$

**Ex.** $\quad x \in \mathbb{R}^2 \quad H$ = Half-planes $\quad \text{VC-dim}(H) = 3$

**Proof:** Radon's theorem:

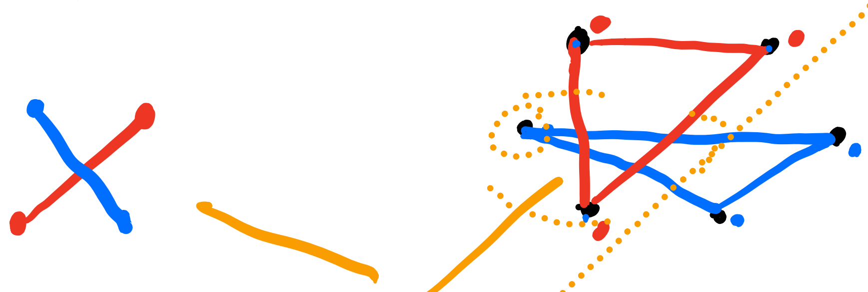
$x_1, \dots, x_{d+2} \in \mathbb{R}^d \Rightarrow$ this set can be partitioned into two sets of points s.t. their convex hulls intersect.

$\Downarrow$

$\text{VC-dim} < d+2$

$\Updownarrow$

$\text{VC-dim} \leq d+1$



this labeling cannot be implemented

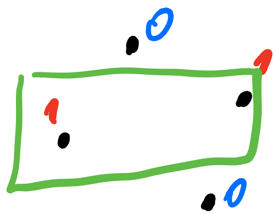
$\square$

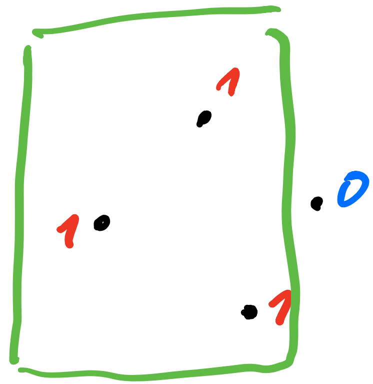
# Axis-aligned Rectangles

$x \in \mathbb{R}^2$

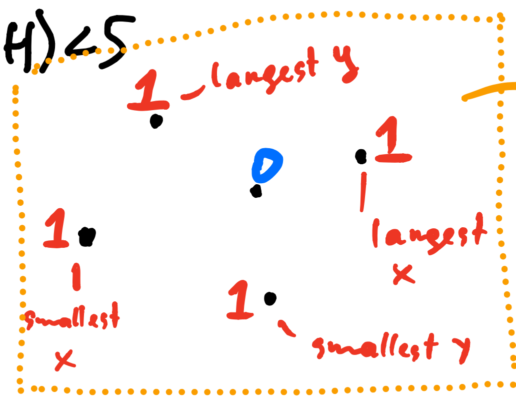
$H =$ Axis-Aligned Rectangles

$\text{VC-dim}(H) = 4$




- $\text{VC-dim}(H) \geq 4$
- $\text{VC-dim}(H) < 5$




impossible to implement

In general

$x \in \mathbb{R}^d$ H = generalized axis-aligned rectangles (boxes)

same argument $VC\text{-}dim(H) = 2 \cdot d$

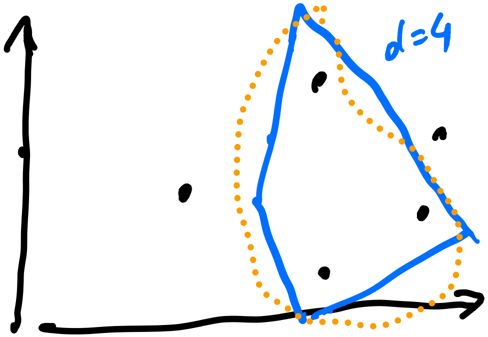
Ex. $d = 2$ (in the plane)

$VC\text{-}dim(H) = 4$

# Convex Polygons

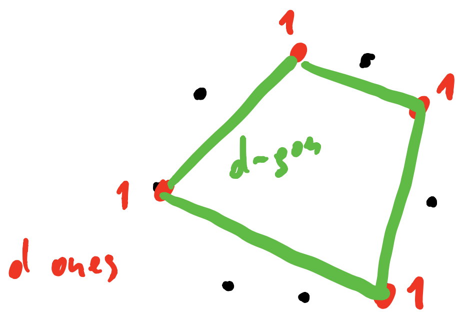
$x \in \mathbb{R}^2$

$H$ = convex $d$-gons $\equiv$ convex polygons with $d$ edges




VC-dim$(H)$ = $\boxed{2d+1}$.

Proof sketch




$2d+1$ points can be shattered

$\Downarrow$

$\le d$ labelled with 1s

OR

$\le d$ labelled with 0s

$\square$

In particular, convex polygons $\Rightarrow$ VC $= +\infty$

Steve Hanneke [2016]:

IF VC-dim$(H) = d \Rightarrow$

PAC-learn $H$ with

$m = O\left(\frac{d + \log(1/\delta)}{\varepsilon}\right)$ samples

Fix $\varepsilon, \delta$ = constants $\Rightarrow$

$m = O(d)$ labelled points to learn $H$.

---

- rectangles
- $d$-gons
- half-spaces
- intervals

can be learned

---

convex polygons cannot be learned

# Fundamental Theorem

**The Fundamental Theorem of Statistical Learning Theory**

- If $H$ has finite VC dimension, $d$ then $H$ is PAC-learnable. $O(d)$ samples
- If $H$ has infinite VC dimension, then $H$ is not PAC-learnable. you cannot learn