Lecture 24:

$\mathcal{H} = \{ h_1, \dots, h_m \}$

PAC learning:

$$n \geq n_0(\epsilon, \delta) = \frac{2}{\epsilon^2} \log \frac{2m}{\delta}$$

Then: For any data distribution, with probability $1-\delta$, the outcome of $\text{ERM}_{\mathcal{H}}$, which we denote by h_S^*, satisfies:

$$\min_{h \in \mathcal{H}} L_D(h) \leq L_D(\underline{h_S^*}) \leq \min_{h \in \mathcal{H}} L_D(h) + \epsilon$$

training data $S = \{(x_i, y_i)\}_{i=1}^{n}$ $\longrightarrow$ learning Alg. $\text{ERM}_{\mathcal{H}}$ $\longrightarrow$ $\underline{h_S^*}$

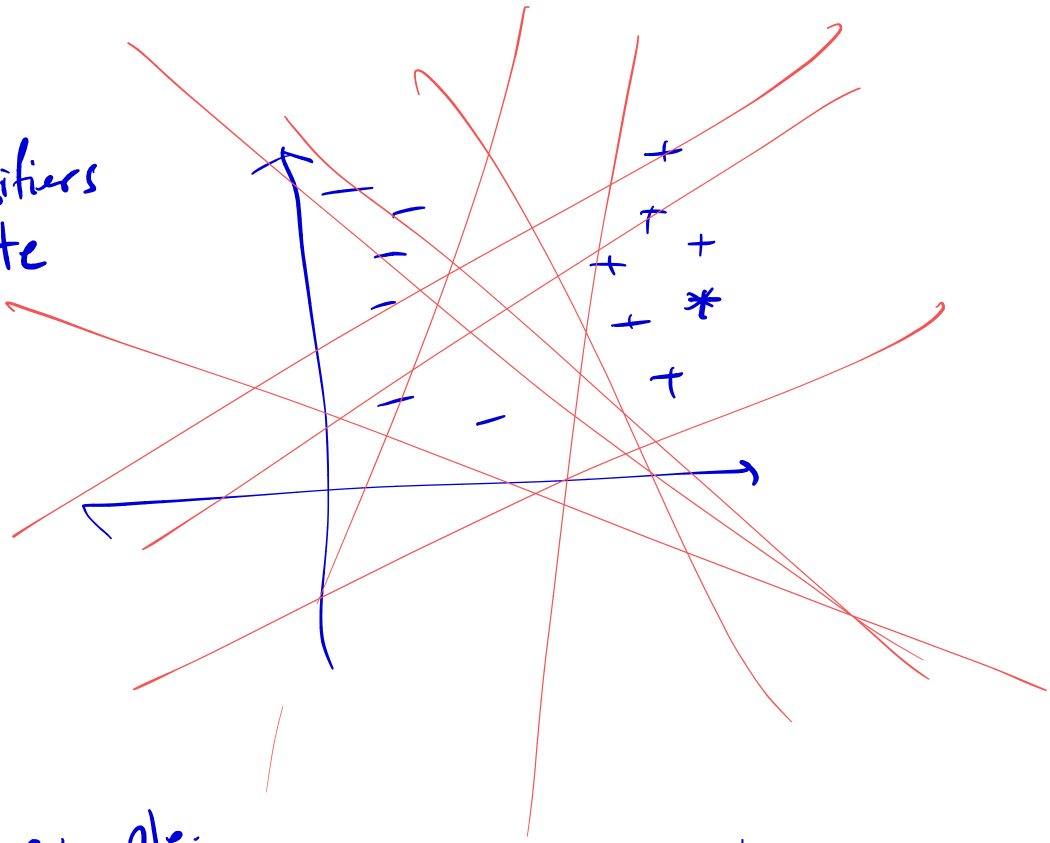
$$\min_{h \in \mathcal{H}} \underbrace{L_S(h)}_{\text{training error}}$$

tractable $\longleftarrow$

So far, we've show that classes $\mathcal{H}$ with finitely many functions are PAC learnable.

But ~~imp~~ in practice, all the function classes that we use are not finite.

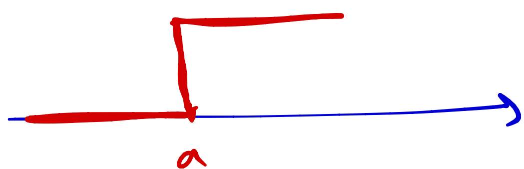
examples:
linear classifiers
~~inf~~ infinite



another example:
Consider the following simple class

$$\mathcal{H} = \{ h_a(x), a \in \mathbb{R} \}$$

$$h_a(x) = \begin{cases} 1 & \text{if } x > a \\ 0 & \text{if } x \leq a \end{cases}$$



The fundamental Theorem of Learning Theory:

For any function class $\mathcal{H}$, if the number of training data points is larger than $n_0(\epsilon,\delta) = 4 \cdot \frac{\log 1/\delta + \overbrace{\text{VC-dim}(\mathcal{H})}^{\text{quantifies the "complexity" of } \mathcal{H}}}{\epsilon^2}$, then

$\mathcal{H}$ is PAC-learnable: For any distribution D on data, with $1-\delta$:

$$L_D(h_S^*) \leq \overbrace{\min_{h \in \mathcal{H}} L_D(h)}^{\text{intractable}} + \epsilon$$

where $h_S^* = \underbrace{\operatorname{argmin}_{h \in \mathcal{H}} L_S(h)}_{\text{tractable}}$.

As long as we have sufficiently many data points, we'll be fine with solving the "tractable" $\text{ERM}_{\mathcal{H}}$ problem and we lose at most a small value ϵ with respect to the best attainable accuracy.

$\Longrightarrow$ Machine learning is possible as long as we have sufficient data!

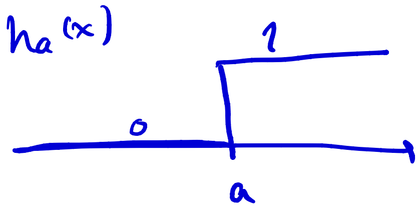
VC - Dimension :

V → Vapnik

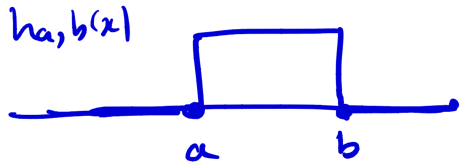
C → chernovskii

Example: Consider the following function classes:

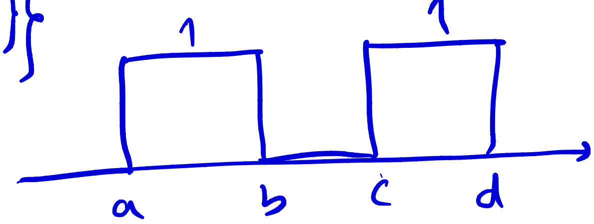
$H_1 = \{ h_a(x), a \in \mathbb{R} \}$ →



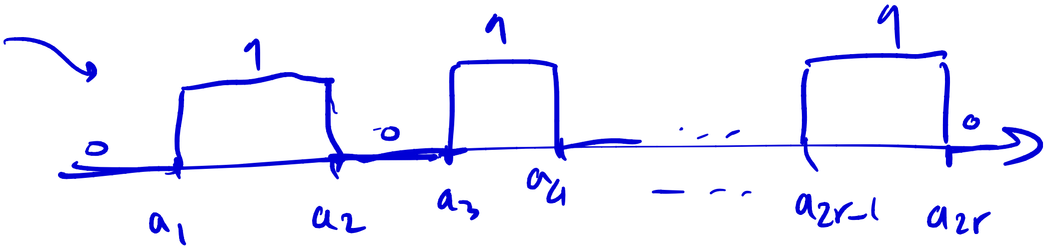
$H_2 = \{ h_{a,b}(x), a,b \in \mathbb{R} \}$ →



$H_3 = \{ h_{abcd}(x), a,b,c,d \in \mathbb{R} \}$



$H_4 = \{ h_{a_1,\ldots,a_{2r}}(x), a_1,\ldots,a_{2r} \in \mathbb{R} \}$



Intuitively, in terms of complexity we should have

$$H_1 < H_2 < H_3 < H_4$$

VC-dim is a precise mathematical notion to quantify "complexity".

In order to define VC-dim we need two other definitions (restriction, shattering).

Definition (restriction):

Let $\mathcal{H}$ be a class of functions from a domain X to $\{0,1\}$ ($\forall h \in \mathcal{H} : h: X \longrightarrow \{0,1\}$).

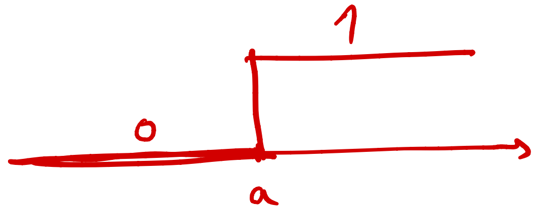
Let $C = \{x_1, \dots, x_k\} \subseteq X$. Then, the restriction of $\mathcal{H}$ to C is the set of all the k-tuples of the following form:

$$\mathcal{H}_C = \{ \underbrace{(h(x_1), h(x_2), \dots, h(x_k))}_{k\text{-tuple}} ; h \in \mathcal{H} \}$$

Example:

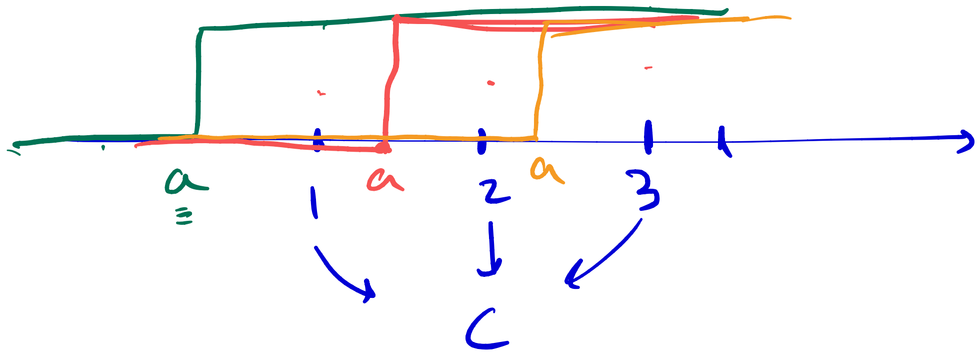
$$\mathcal{H} = \{ h_a(x) ; a \in \mathbb{R} \}$$

$h_a(x)$



Let $C = \{1, 2, 3\}$. What is $\mathcal{H}_C$?

To find $\mathcal{H}_C$ we need to go over all the functions $h \in \mathcal{H}$.



if $a \leq 1 \longrightarrow (h_a(1), h_a(2), h_a(3))$

$= (1, 1, 1)$

if $1 < a \leq 2 \longrightarrow (h_a(1), h_a(2), h_a(3))$

$= (0, 1, 1)$

if $2 < a \leq 3 \quad = (h_a(1), h_a(2), h_a(3))$

$= (0, 0, 1)$

if $a > 3 \longrightarrow (0, 0, 0)$

$$\mathcal{H}_C = \{(1,1,1), (0,1,1), (0,0,1), (0,0,0)\}.$$

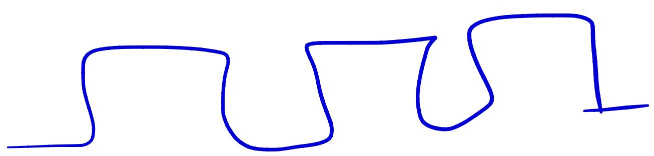
We note that if $C = \{x_1, \dots, x_k\}$

then $|\mathcal{H}_C| \leq 2^k$

↳ recall that there are 2^k binary k-tuples.

$$1 \leq |\mathcal{H}_C| \leq 2^k$$

complexity ↑



Definition (shattering): We say that a function class $\mathcal{H}$ shatters a set C of size k if $|\mathcal{H}_C| = 2^k$.

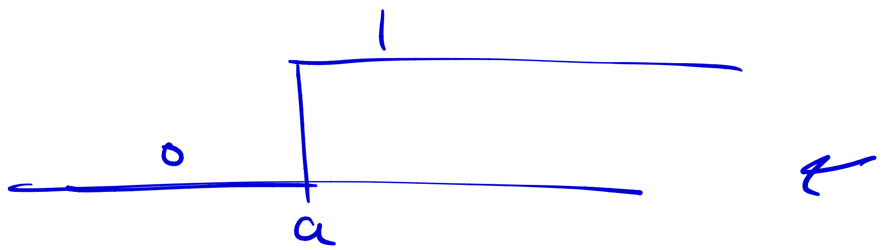
Definition (VC-dimension):

The VC-dimension of a function class $\mathcal{H}$ is the largest k such that there exists a set C of size k which is shattered by $\mathcal{H}$.

Example:

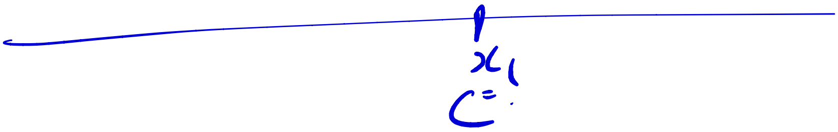
$$\mathcal{H} = \{ h_a(x), \quad a \in \mathbb{R} \}$$

$h_a =$



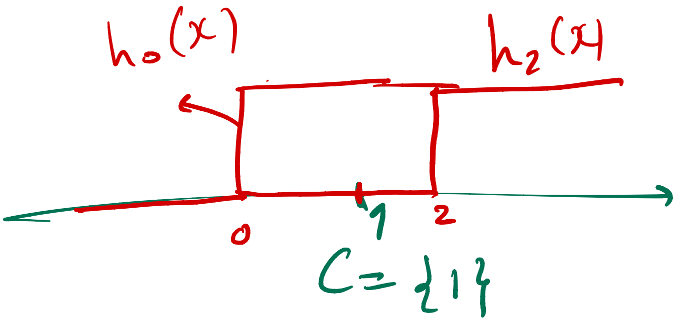
In order to find the VC-dim$(\mathcal{H})$ we need to find the largest k for which $\mathcal{H}$ shatters a set C of size k:

$k=1 \longrightarrow$ Is there a set $C \subseteq X$ of size k that is shattered by $\mathcal{H}$?



$|C|=1$, $\quad H_C = \{ (h(x_1)) \; ; \; h \in \mathcal{H} \}$

$C=\{x_1\}$ $\qquad = \{ 0, 1 \}$

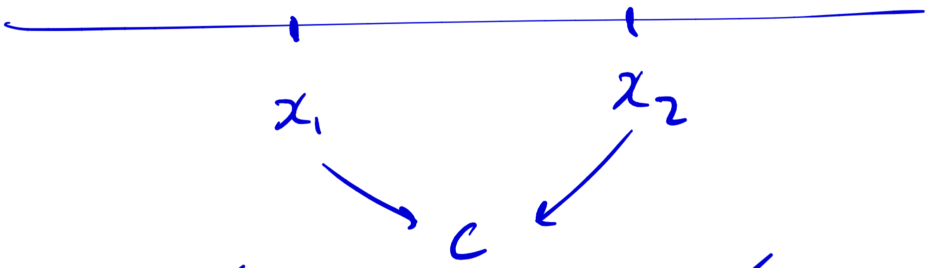


$$\begin{cases} h_0(1) = 1 \\ h_2(1) = 0 \end{cases} \longrightarrow \mathcal{H}_C = \{0, 1\}$$

$\Rightarrow C=\{1\}$ is shattered by $\mathcal{H}$.

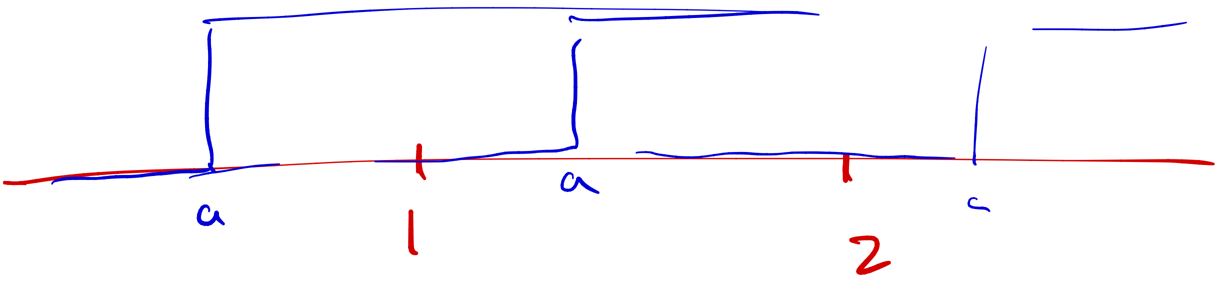
Now consider $k=2$.

Is there a set C of size 2 that is shattered by $\mathcal{H}$?



$\mathcal{H}_c \overset{?}{=} \underbrace{\{(0,0), (0,1), (1,0), (1,1)\}}$

let's take = $x_1 = 1$ and $x_2 = 2$



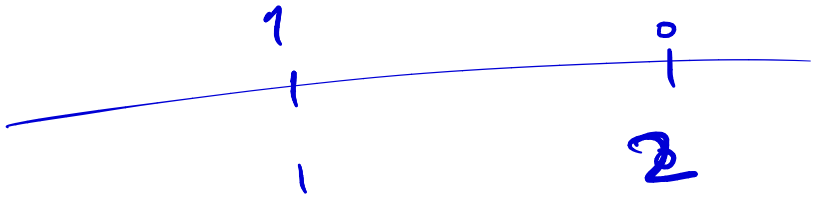
if $a > 2 \Rightarrow (h_a(1), h_a(2)) = (0,0)$

if $a < 1 \Rightarrow (h_a(1), h_a(2)) = (1,\underline{1})$

if $1 < a \leq 2 \Rightarrow (h_a(1), h_a(2)) = (0,\underline{1})$

is there any function that generates $\cancel{(1,0)}$?

→ No since all the functions in $\mathcal{H}$ are increasing.



We can show that for any set $C=\{x_1, x_2\}$, $x_1<x_2$, the 2-tuple $(1,0)$ can not be generated by any $h\in\mathcal{H}$.

$\Rightarrow$ $\nexists$ C of size 2 which is shattered by $\mathcal{H}$.

$k=1$ ✓

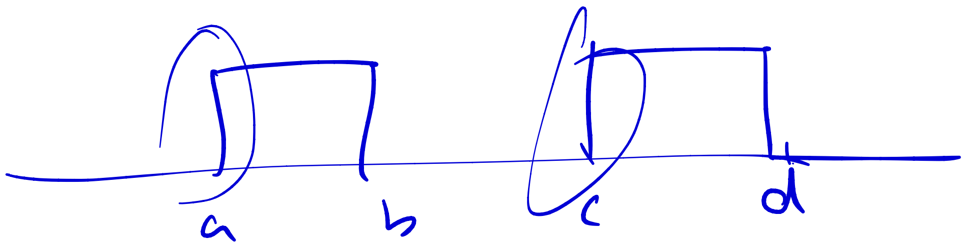
$k=\underline{2}$ ✗ $\longrightarrow$ $V_C\text{-dim}(\mathcal{H})=\underline{1}$.

$k=\underline{3}$ ✗

⋮

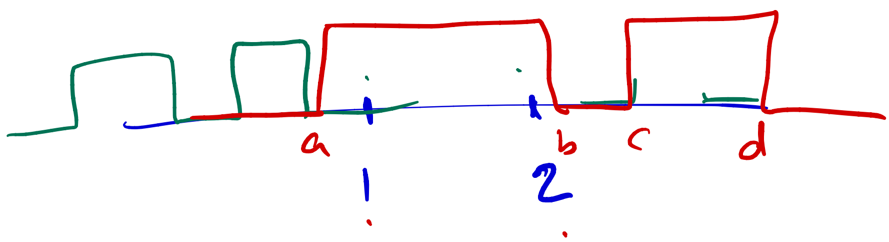
Example:

$\mathcal{H} = \{ h_{a,b,c,d}(x) \;:\; a,b,c,d \in \mathbb{R} \}$



$k=1 \rightarrow$ ✓

$k=2 \rightarrow$ is there a set C of size 2 that's shattered by $\mathcal{H}$?

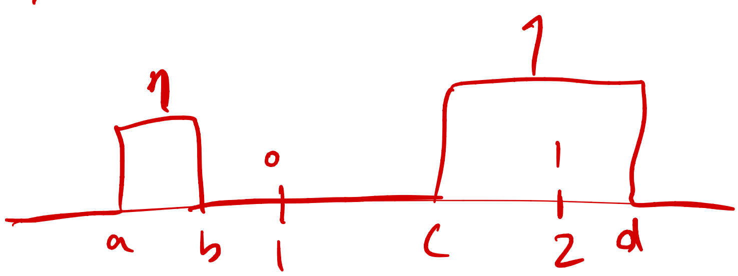


$H_C = \{(0,0),(0,1),(1,0),(1,1)\}$

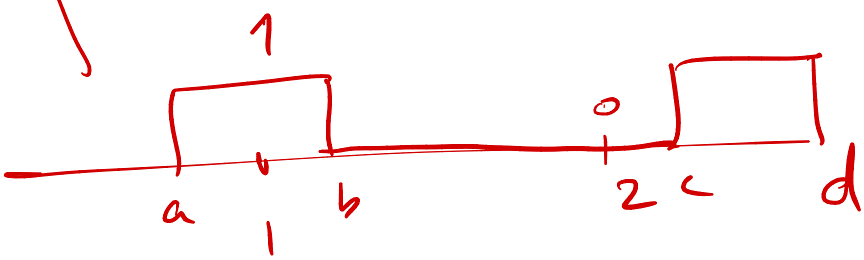
if $d<1$ then we get $(0,0)$

if $a<1<2<b$ then we get $(1,1)$

$a,b<1$ and $c<2<3 \rightarrow (0,1)$



this function gets $(1,0)$



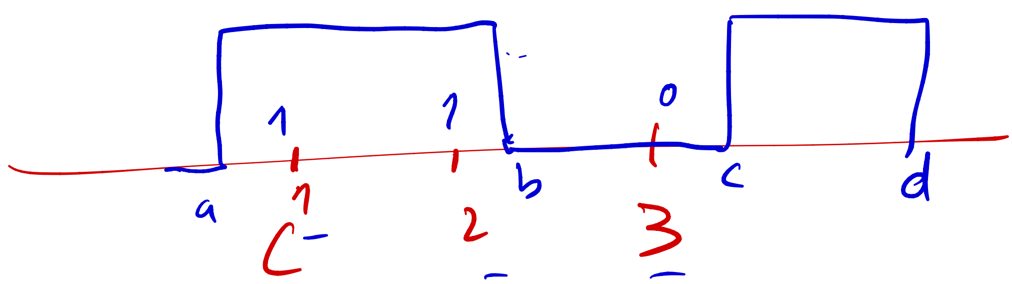
$\mathcal{H}$ shatters the set $C = \{1, 2\}$.

$K = 3$ ✓

$C = \{1, 2, 3\}$ will be shattered by $\mathcal{H}$.

$$\mathcal{H}_C = \{(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1)\}$$

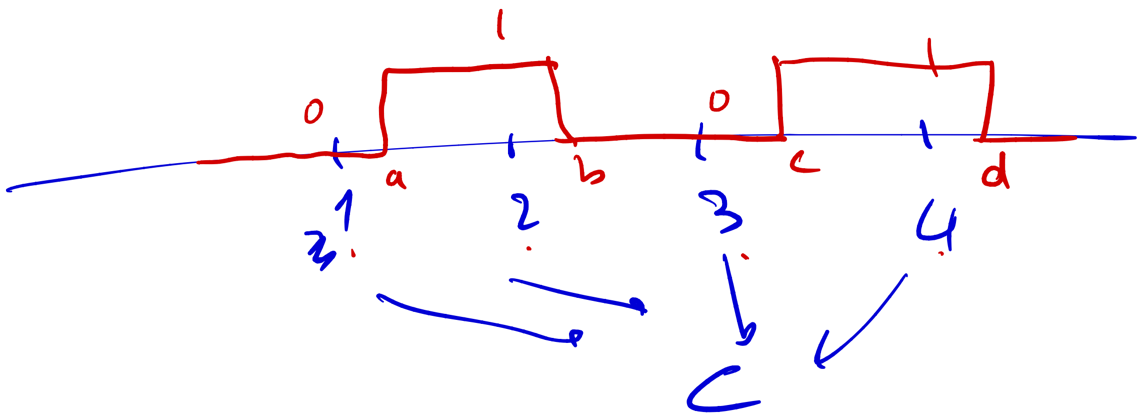
e.g. $(1,1,0)$



$k = 4$, the set $C = \{1, 2, 3, 4\}$ is shattered by $\mathcal{H}$.

$\mathcal{H}_C = \{$ all the 16 binary 4-tuples $\}$

e.g. $(0, 1, 0, 1)$



$\mathcal{H}$ shatters a set C of size 1 ✓

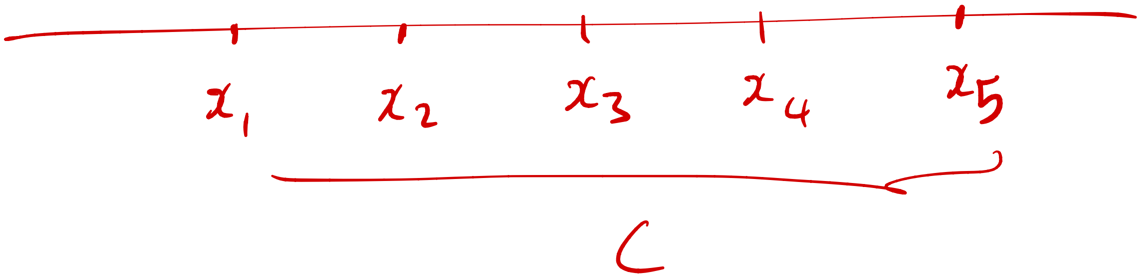
2 ✓

3 ✓

4 ✓

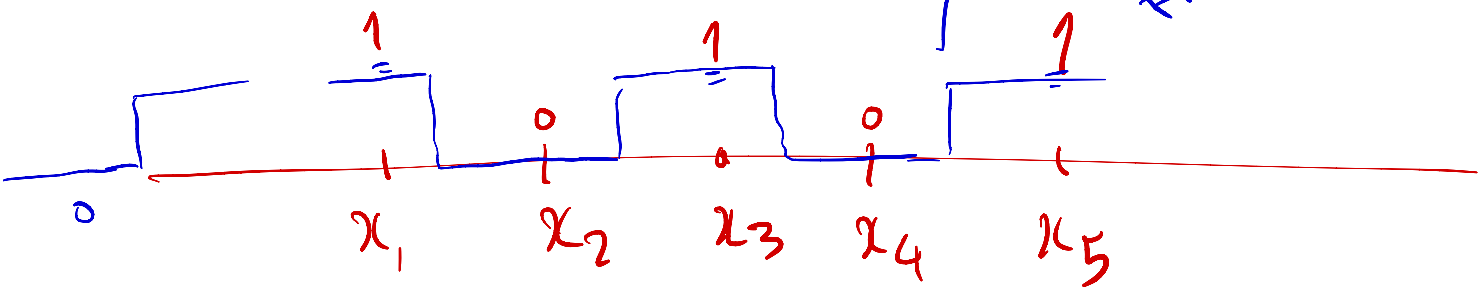
5 ? ✗

$K = 5$ does $\mathcal{H}$ shatter a set C of size 5?



We claim that the 5-tuple $(1,0,1,0,1)$ ~~is~~ can not be ~~to~~ generated by any function in $\mathcal{H}$.

the function has to have 3 jumps from 0 to 1.



$$\Longrightarrow \quad VC\text{-}dim(\mathcal{H}) = 4.$$