

Advanced Machine Learning



Bogdan Alexe,

bogdan.alexe@fmi.unibuc.ro

University of Bucharest, 2nd semester, 2019-2020

Today's lecture: Overview

- Recap: Learnability - what do we know so far
- Shattering
- VC-dimension

Recap: Learnability - what do we know so far?

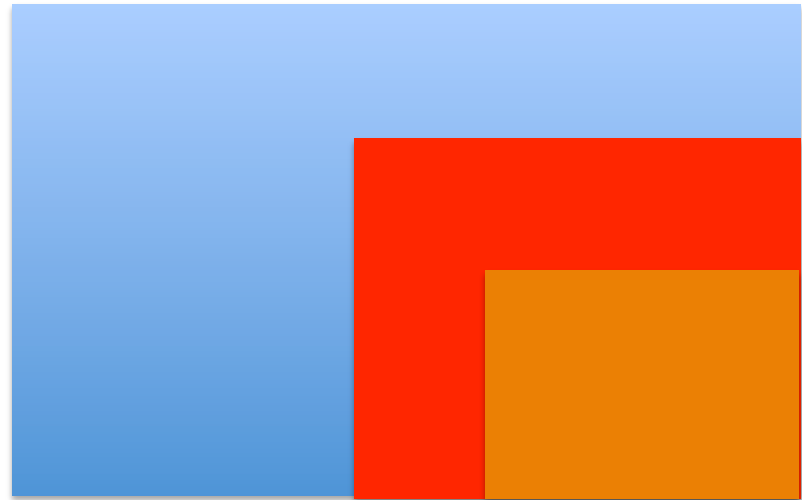
Let \mathcal{H} a hypothesis class.

- is it PAC learnable?
- is it agnostic PAC learnable?
 - we do know that agnostic PAC learnability \rightarrow PAC learnability

all hypothesis classes

PAC learnable

agnostic PAC learnable



Definition 3.1 (PAC Learnability). A hypothesis class \mathcal{H} is PAC learnable if there exist a function $m_{\mathcal{H}} : (0, 1)^2 \rightarrow \mathbb{N}$ and a learning algorithm with the following property: For every $\epsilon, \delta \in (0, 1)$, for every distribution \mathcal{D} over \mathcal{X} , and for every labeling function $f : \mathcal{X} \rightarrow \{0, 1\}$, if the realizable assumption holds with respect to $\mathcal{H}, \mathcal{D}, f$, then when running the learning algorithm on $m \geq m_{\mathcal{H}}(\epsilon, \delta)$ i.i.d. examples generated by \mathcal{D} and labeled by f , the algorithm returns a hypothesis h such that, with probability of at least $1 - \delta$ (over the choice of the examples), $L_{(\mathcal{D}, f)}(h) \leq \epsilon$.

Definition 3.4 (Agnostic PAC Learnability for General Loss Functions). A hypothesis class \mathcal{H} is agnostic PAC learnable with respect to a set Z and a loss function $\ell : \mathcal{H} \times Z \rightarrow \mathbb{R}_+$, if there exist a function $m_{\mathcal{H}} : (0, 1)^2 \rightarrow \mathbb{N}$ and a learning algorithm with the following property: For every $\epsilon, \delta \in (0, 1)$ and for every distribution \mathcal{D} over Z , when running the learning algorithm on $m \geq m_{\mathcal{H}}(\epsilon, \delta)$ i.i.d. examples generated by \mathcal{D} , the algorithm returns $h \in \mathcal{H}$ such that, with probability of at least $1 - \delta$ (over the choice of the m training examples),

$$L_{\mathcal{D}}(h) \leq \min_{h' \in \mathcal{H}} L_{\mathcal{D}}(h') + \epsilon,$$

where $L_{\mathcal{D}}(h) = \mathbb{E}_{z \sim \mathcal{D}} [\ell(h, z)]$.

Recap: Learnability - what do we know so far?

Let \mathcal{H} a hypothesis class.

- is it PAC learnable?
- is it agnostic PAC learnable?
 - we do know that agnostic PAC learnability \rightarrow PAC learnability
 - we don't know that PAC learnability \rightarrow agnostic PAC learnability

Size of class \mathcal{H} :

- finite
- infinite

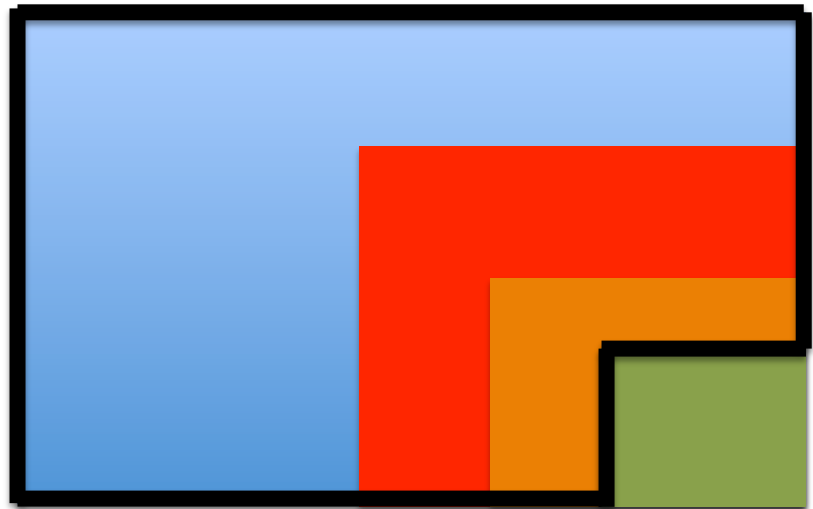
all hypothesis classes

PAC learnable

agnostic PAC learnable

finite size classes

infinite size classes



Recap: Learnability - what do we know so far?

Let \mathcal{H} a hypothesis class.

- is it PAC learnable?
- is it agnostic PAC learnable?
 - we do know that agnostic PAC learnability \rightarrow PAC learnability
 - we don't know that PAC learnability \rightarrow agnostic PAC learnability

Size of class \mathcal{H} :

- finite
- infinite

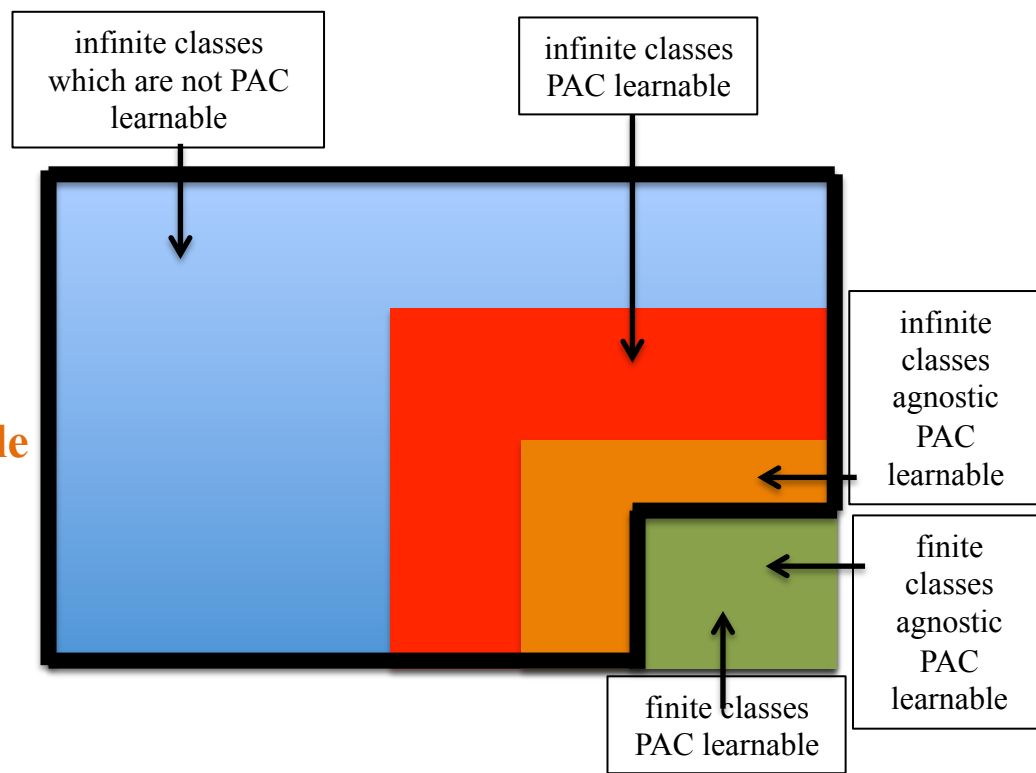
all hypothesis classes

PAC learnable

agnostic PAC learnable

finite size classes

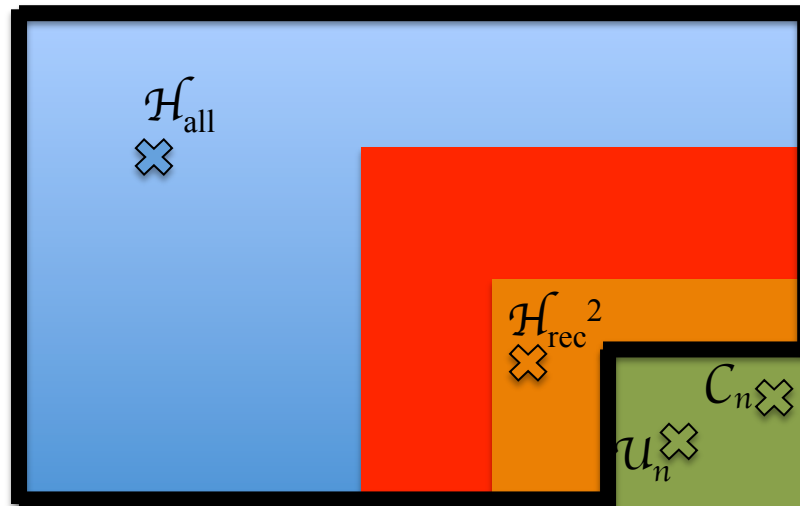
infinite size classes



Recap: Learnability - what do we know so far?

Hypothesis classes \mathcal{H} encountered until now:

- finite \mathcal{H}
 - C_n concept class of conjunctions of at most n Boolean literals x_1, \dots, x_n
 - \mathcal{U}_n universal concept class
- infinite \mathcal{H}
 - $\mathcal{H}_{\text{rec}}^2$ set of all axis-aligned rectangle lying in \mathbb{R}^2 (with positive labels unaffected – PAC or affected by noise – agnostic PAC)
 - \mathcal{H}_{all} all functions from X to $\{0,1\}$ (No Free-Lunch theorem)



all hypothesis classes

PAC learnable

agnostic PAC learnable

finite size classes

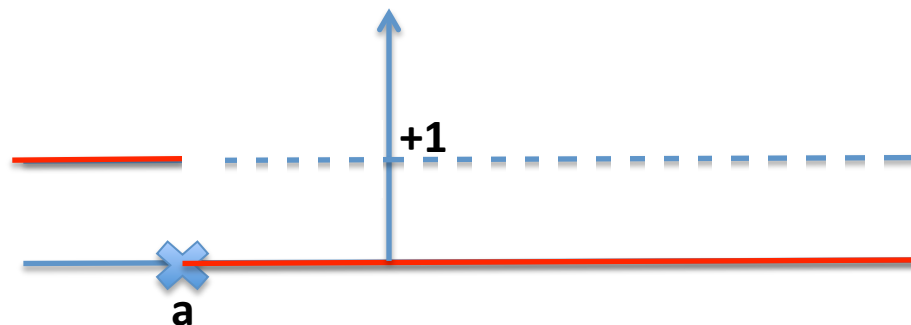
infinite size classes

Recap: Another class example - $\mathcal{H}_{\text{thresholds}}$

Consider $\mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line

$$\mathcal{H}_{\text{thresholds}} = \{h_a: \mathbb{R} \rightarrow \{0, 1\}, h_a(x) = \mathbf{1}_{[x < a]}, a \in \mathbb{R}\}, |\mathcal{H}_{\text{thresholds}}| = \infty$$

$$\mathbf{1}_{[x < a]} = \begin{cases} 1, x < a \\ 0, x \geq a \end{cases},$$



is the indicator function of the set $\{x \in \mathbb{R} \mid x < a\}$

Lemma

$\mathcal{H}_{\text{thresholds}}$ is PAC learnable, using the ERM learning rule, with sample complexity of

$$m_{\mathcal{H}_{\text{thresholds}}} \leq \left\lceil \frac{1}{\varepsilon} \log \frac{1}{\delta} \right\rceil$$

Recap: No Free Lunches vs. $\mathcal{H}_{\text{thresholds}}$

Why is $\mathcal{H}_{\text{thresholds}}$ = set of threshold classifiers not a victim of the No Free Lunch theorem? (we can PAC learn them)

The reason is simple:

- the class of threshold classifiers is so simple that an adversary has no room to create an adversarial distribution

In fact, as our discussion above shows:

- if two threshold classifiers agree on a large enough sample
- their respective thresholds will be close to each other
- there is no way you can force them to behave completely differently on unseen examples.

If that would have been possible then:

- we would have been able to create an adversarial distribution.

So, it seems necessary for PAC learnability that the general class \mathcal{H} considered isn't too expressive

Shattering

How expressive is \mathcal{H} ?

In our binary classification context, a hypothesis is a function $h: \mathcal{X} \rightarrow \{0, 1\}$

Hence the expressiveness of \mathcal{H} :

- is necessary a measure of how many functions \mathcal{H} can express
- in the light of the No Free Lunch theorem, not only functions on \mathcal{X} , but also functions on (finite) subsets C of \mathcal{X}

Definition (restriction of \mathcal{H} to C)

Let \mathcal{H} be a set hypothesis, i.e., set of functions from \mathcal{X} to $\{0, 1\}$, and let C be a (finite) subset of \mathcal{X} , $C = \{c_1, c_2, \dots, c_m\}$. The restriction of \mathcal{H} to C , denoted by \mathcal{H}_C , is the set of functions from C to $\{0, 1\}$ that can be derived from \mathcal{H} .

That is:

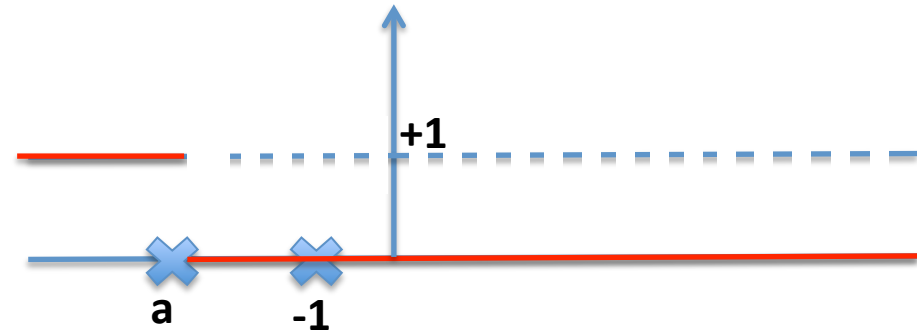
$$\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$$

Example: restriction of \mathcal{H} to C

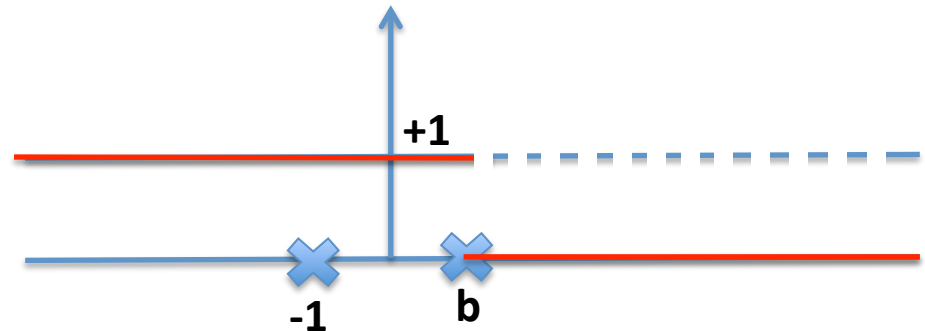
Consider $\mathcal{H} = \mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line
 $\mathcal{H}_{\text{thresholds}} = \{h_a: \mathbb{R} \rightarrow \{0, 1\}, h_a(x) = \mathbf{1}_{[x < a]}, a \in \mathbb{R}\}, |\mathcal{H}_{\text{thresholds}}| = \infty.$

Consider $C = \{-1\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\} = \{h: \{-1\} \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has 2 elements h_a, h_b where:

$$h_a(-1) = 0, a < -1$$



$$h_b(-1) = 1, b \geq -1$$

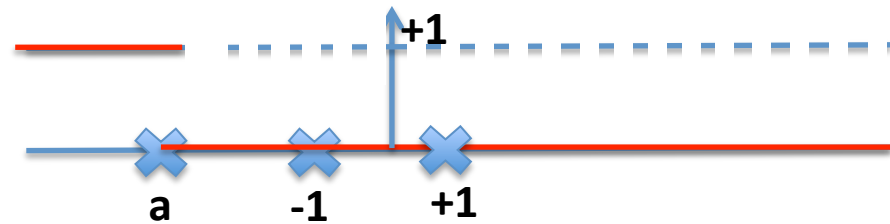


Example: restriction of \mathcal{H} to C

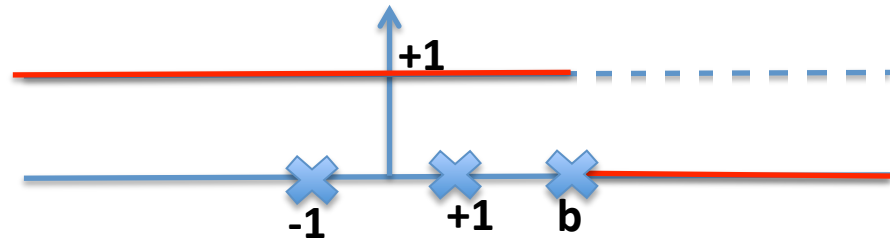
Consider $\mathcal{H} = \mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line
 $\mathcal{H}_{\text{thresholds}} = \{h_a: \mathbb{R} \rightarrow \{0, 1\}, h_a(x) = \mathbf{1}_{[x < a]}, a \in \mathbb{R}\}, |\mathcal{H}_{\text{thresholds}}| = \infty.$

Consider $C = \{-1, 1\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\} = \{h: \{-1, 1\} \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has 3 elements h_a, h_b, h_c where:

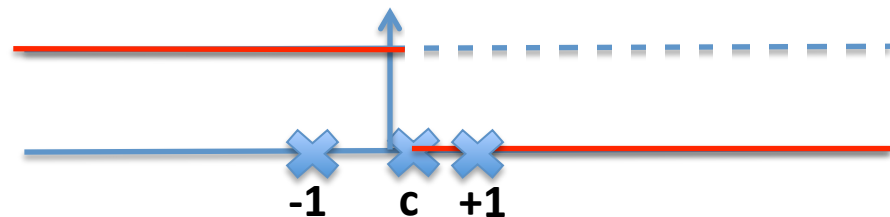
$$h_a(-1) = 0, h_a(1) = 0, a < -1$$



$$h_b(-1) = 1, h_b(1) = 1, b \geq 1$$



$$h_c(-1) = 1, h_c(1) = 0, -1 \leq c < 1$$



Example: restriction of \mathcal{H} to C

Consider $\mathcal{H} = \mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line

$$\mathcal{H}_{\text{thresholds}} = \{h_a: \mathbb{R} \rightarrow \{0, 1\}, h_a(x) = \mathbf{1}_{[x < a]}, a \in \mathbb{R}\}, |\mathcal{H}_{\text{thresholds}}| = \infty.$$

Consider $C = \{-1, 1\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\} = \{h: \{-1, 1\} \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has 3 elements h_a, h_b, h_c where:

$$h_a(-1) = 0, h_a(1) = 0, a \leq -1$$

$$h_b(-1) = 1, h_b(1) = 1, b > 1$$

$$h_c(-1) = 1, h_c(1) = 0, -1 < c \leq 1$$

There is no function h_d in \mathcal{H}_C such that $h_d(-1) = 0$ and $h_d(1) = 1$ (as we work with the threshold functions)

Alternative view of functions

There is equivalence between functions from \mathcal{X} to $\{0,1\}$ and subsets C of \mathcal{X}

- given $h: \mathcal{X} \rightarrow \{0, 1\}$, we can define $C = \{x \in \mathcal{X} \mid \text{such that } h(x) = 1\}$
- given C subset of \mathcal{X} , we can define $h: \mathcal{X} \rightarrow \{0,1\}$, $h(x) = \mathbf{1}_C(x)$ indicator function of C

Can represent a subset C of \mathcal{X} as a vector, put 1 for element in C , 0 otherwise.

Can represent each function from C to $\{0,1\}$ as a vector in $\{0,1\}^{|C|}$.

The restriction of \mathcal{H} to C , denoted by \mathcal{H}_C , is the set of functions from C to $\{0, 1\}$ that can be derived from \mathcal{H} . It can be also seen as the set of all possible vectors that can be generated by $h \in \mathcal{H}$ with elements from C . That is:

$$\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\} = \{(h(c_1), h(c_2), \dots, h(c_m)) \mid h \in \mathcal{H}\}$$

Consider $C = \{-1, 1\}$. Then $\mathcal{H}_{thresholds} = \mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\} = \{h: \{-1, 1\} \rightarrow \{0, 1\} \mid h \in \mathcal{H}\} = \{(0,0), (1,0), (1,1)\}$. **The vector (0,1) is not realizable by \mathcal{H}_C .**

Shattering

Definition (Shattering)

A hypothesis class \mathcal{H} *shatters* a finite set C of \mathcal{X} , if the restriction of \mathcal{H} to C is the set of all functions from C to $\{0, 1\}$. That is $|\mathcal{H}_C| = 2^{|C|}$.

Examples:

Consider $\mathcal{H} = \mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line.

Consider $C = \{c_1\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has two elements $\{h_a, h_b\}$ with $a < c_1$ and $b \geq c_1$ so \mathcal{H} shatters C . $\mathcal{H}_C = \{(0), (1)\}$, $|\mathcal{H}_C| = 2^{|C|} = 2^1$

Consider $C = \{c_1, c_2 \mid c_1 \leq c_2\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has at most three elements, there is no function that realizes the labeling (0,1) and so \mathcal{H} does not shatter C .

Alternative view of Shattering

Definition (Shattering)

A hypothesis class \mathcal{H} *shatters* a finite set C of \mathcal{X} , if the restriction of \mathcal{H} to C is the set of all functions from C to $\{0, 1\}$. That is $|\mathcal{H}_C| = 2^{|C|}$.

Similar Definition (Shattering)

Let \mathcal{H} be a collection of subsets of \mathcal{X} and C a finite subset of \mathcal{X} .

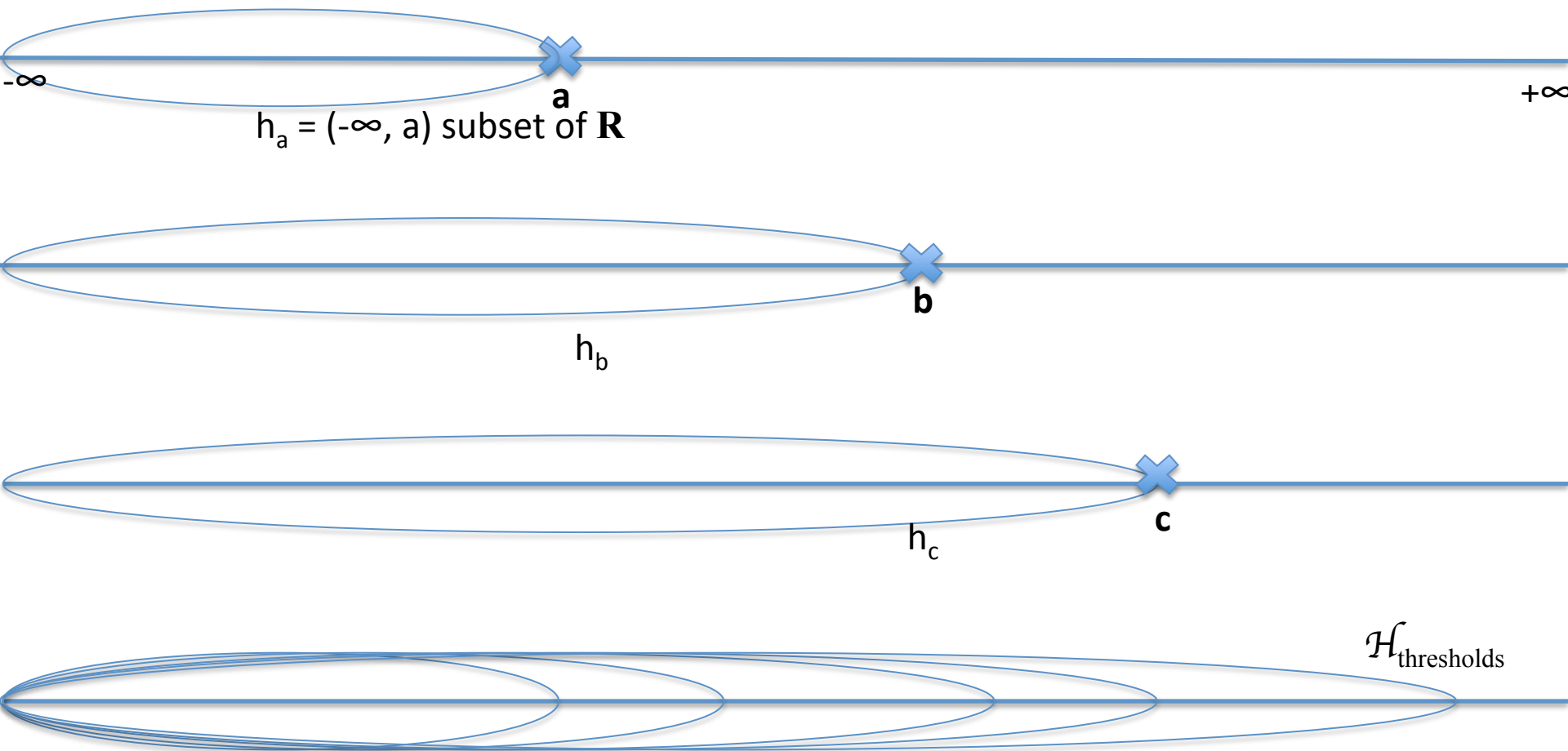
\mathcal{H} shatters C if for every subset B of C there exist some subset $h_B \in \mathcal{H}$ such that $B = h_B \cap C$

In other words, using the elements of \mathcal{H} , we can cut C in every possible way.

Shattering – graphical representation

Examples:

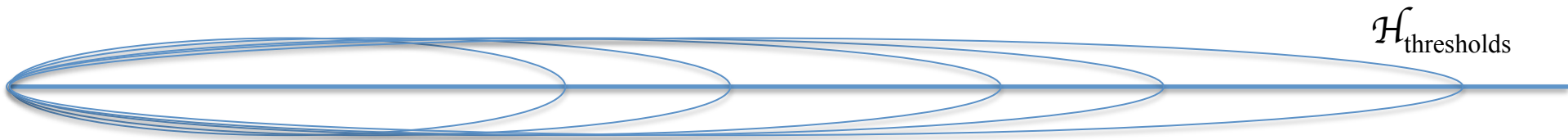
Consider $\mathcal{H} = \mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line.



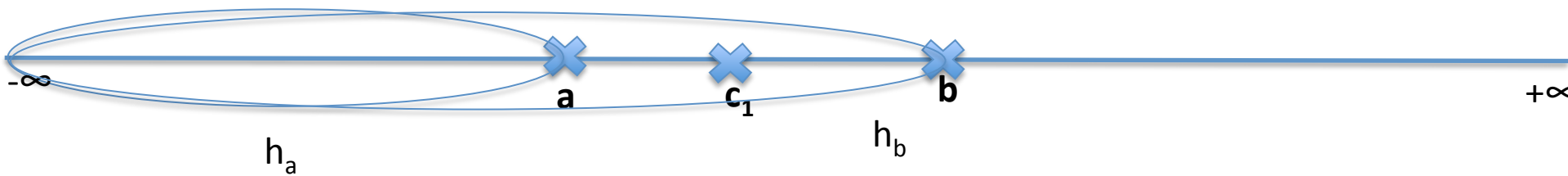
Shattering – graphical representation

Examples:

Consider $\mathcal{H} = \mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line.



Consider $C = \{c_1\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has two elements $\{h_a, h_b\}$ with $a < c_1$ and $b \geq c_1$ so \mathcal{H} shatters C . $\mathcal{H}_C = \{(0), (1)\}$, $|\mathcal{H}_C| = 2^{|C|} = 2^1$



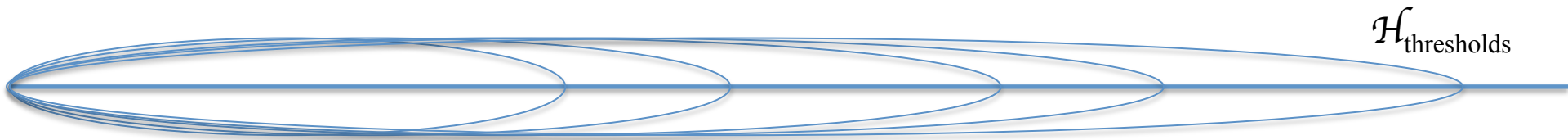
h_a generates label 0

h_b generates label 1

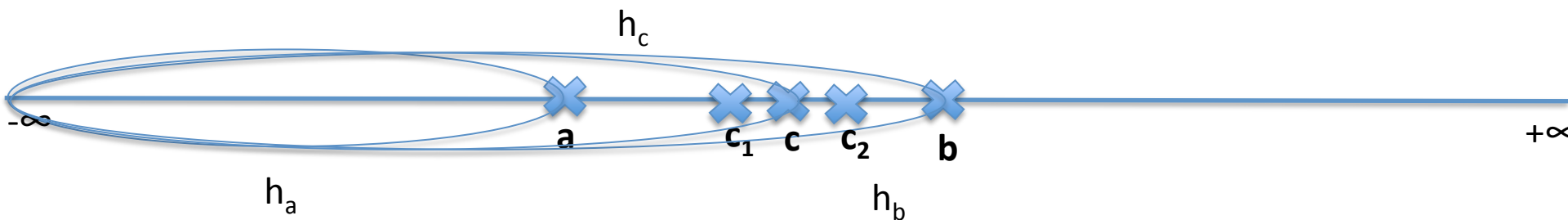
Shattering – graphical representation

Examples:

Consider $\mathcal{H} = \mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line.



Consider $C = \{c_1, c_2 \mid c_1 \leq c_2\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has at most three elements, there is no function that realizes the labeling (0,1) and so \mathcal{H} does not shatter C .



h_a generates label (0,0)

h_b generates label (1,1)

h_c generates label (1,0)

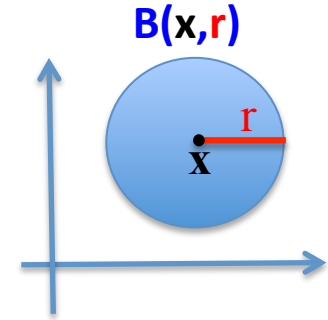
cannot generate the label (0,1)

Shattering – example $\mathcal{H}_{\text{balls}}$

Consider $\mathcal{H} = \mathcal{H}_{\text{balls}}$ be the set of all balls in \mathbf{R}^2 :

$$\mathcal{H}_{\text{balls}} = \{B(\mathbf{x}, r), \mathbf{x} \in \mathbf{R}^2, r \geq 0\},$$

$$B(\mathbf{x}, r) = \{y \in \mathbf{R}^2 \mid \|y - \mathbf{x}\|_2 \leq r\}$$

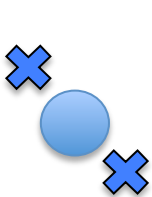


Can also view $\mathcal{H}_{\text{balls}}$ as:

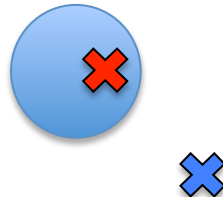
$$\mathcal{H}_{\text{balls}} = \{h_{\mathbf{x}, r}: \mathbf{R}^2 \rightarrow \{0, 1\}, h_{\mathbf{x}, r} = \mathbf{1}_{B(\mathbf{x}, r)}, \mathbf{x} \in \mathbf{R}^2, r \geq 0\}$$

Is there a set A in \mathbf{R}^2 of size 2 shattered by $\mathcal{H}_{\text{balls}}$?

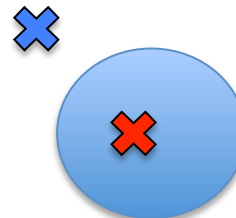
Any set A of two distinct points in \mathbf{R}^2 is shattered by $\mathcal{H}_{\text{balls}}$



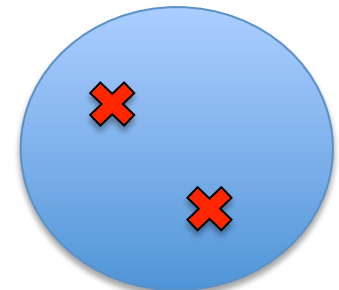
label (0,0)



label (1,0)



label (0,1)

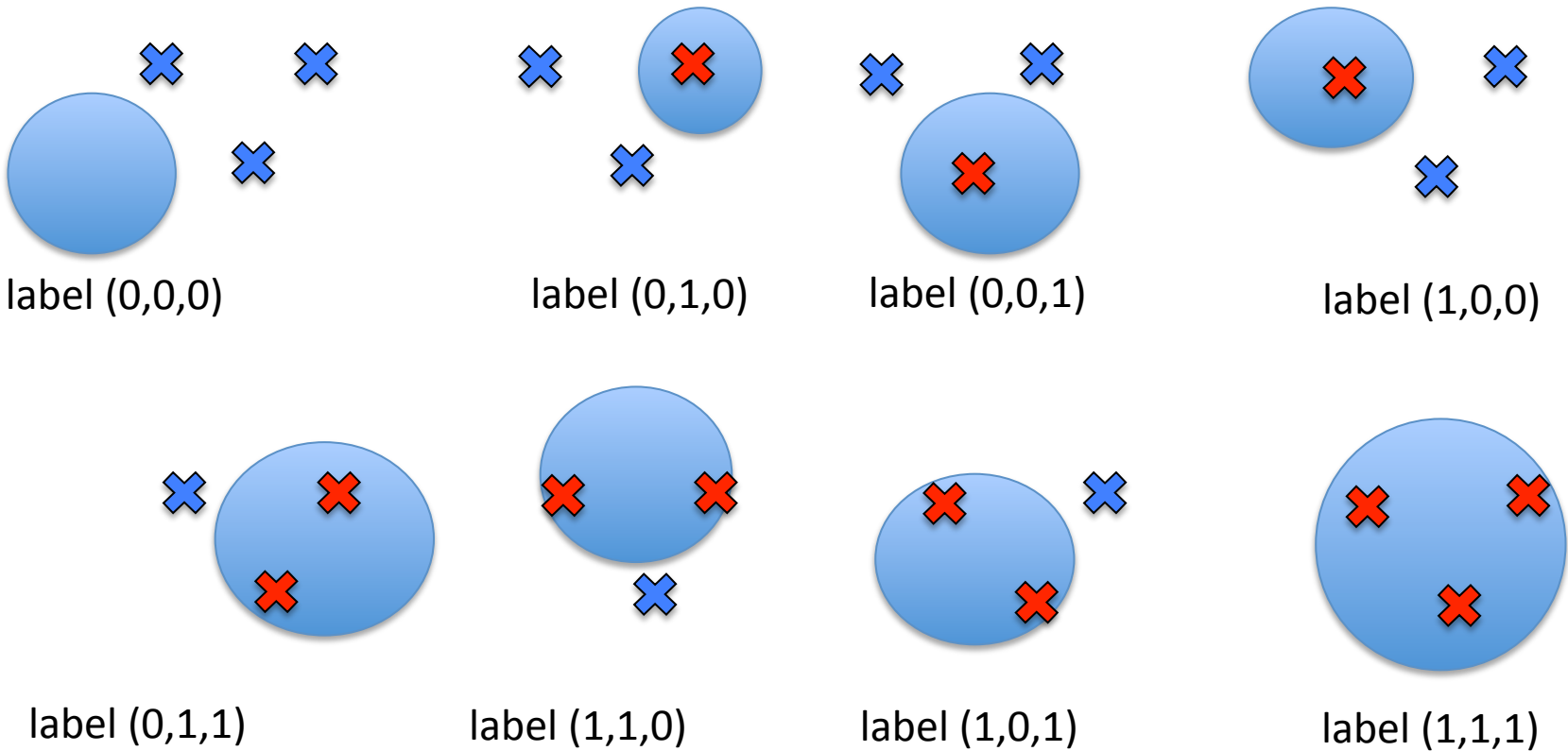


label (1,1)

Shattering – example $\mathcal{H}_{\text{balls}}$

Is there a set A in \mathbf{R}^2 of size 3 shattered by $\mathcal{H}_{\text{balls}}$?

Any set A of three distinct points in \mathbf{R}^2 that are not collinear is shattered by $\mathcal{H}_{\text{balls}}$



Shattering – example $\mathcal{H}_{\text{balls}}$

Is there a set A in \mathbf{R}^2 of size 3 shattered by $\mathcal{H}_{\text{balls}}$?

Any set A of three distinct points in \mathbf{R}^2 that are collinear is not shattered by $\mathcal{H}_{\text{balls}}$



cannot realize the label $(1, 0, 1)$

What are the conditions for which a set A in \mathbf{R}^2 of size 4 is shattered by $\mathcal{H}_{\text{balls}}$?

No Free lunches - revisited

- from the proof of the No Free Lunch theorem, we saw that we can create an adversarial distribution if \mathcal{H} shatters a too large class.
- let \mathcal{H} be a hypothesis class of functions $h: X \rightarrow \{0, 1\}$ and S a training set of size m . If there exists a set $C \subseteq X$ of size $2m$ that is shattered by \mathcal{H} , then for any learning algorithms A there exists a distribution \mathcal{D} over $X \times \{0, 1\}$ such that:
 - there exists a function $f: X \rightarrow \{0, 1\}$ with $L_{\mathcal{D}}(f) = 0$
 - with probability of at least $1/7$ over the choice of $S \sim \mathcal{D}^m$ we have that $L_{\mathcal{D}}(A(S)) \geq 1/8$
- the labels of the m instances give us no information about the labels of the rest of the instances in C – every possible labeling of the rest of the instances can be explained by some hypothesis in \mathcal{H} .
- *“If someone can explain every phenomenon, his explanations are worthless”*
- shattering is good, but don’t shatter too much.

VC-dimension

The VC-dimension

Definition (VC-dimension)

The VC - dimension of a hypothesis class \mathcal{H} , denoted $\text{VCdim}(\mathcal{H})$, is the maximal size of a set $C \subseteq X$ that can be shattered by \mathcal{H} . If \mathcal{H} can shatter sets of arbitrarily large size we say that \mathcal{H} has infinite VC-dimension.

Theorem

Let \mathcal{H} be a class of infinite VC-dimension. Then, \mathcal{H} is not PAC learnable.

Proof: Since \mathcal{H} has an infinite VC-dimension, for any training set S of size m , there exists a shattered set of size $2m$, and the claim follows for the No Free Lunch theorem.

We will see in the next lecture that the converse is also true: *a finite VC-dimension guarantees learnability. Hence, the VC-dimension characterizes PAC learnability. VC-dimension is a combinatorial measure, does not imply computing probabilities.*

Determining the VC-dimension of \mathcal{H}

Definition (VC-dimension)

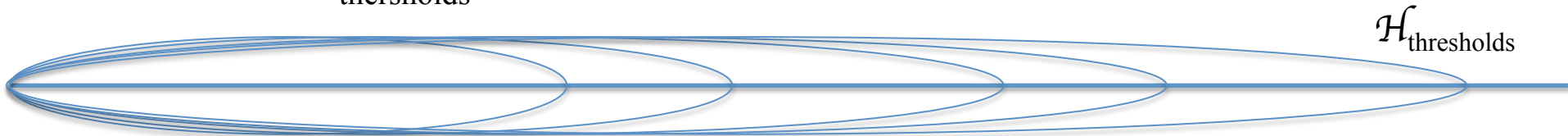
The VC - dimension of a hypothesis class \mathcal{H} , denoted $\text{VCdim}(\mathcal{H})$, is the maximal size of a set $C \subset X$ that can be shattered by \mathcal{H} . If \mathcal{H} can shatter sets of arbitrarily large size we say that \mathcal{H} has infinite VC-dimension.

In order to show that the VC-dimension of a hypothesis class \mathcal{H} is d , we need to show that:

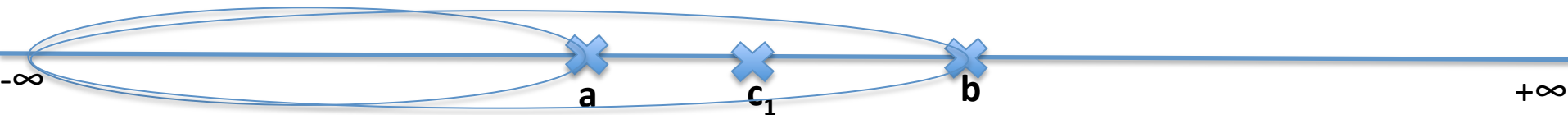
1. **There exists a set C of size d that is shattered by \mathcal{H} . ($\text{VCdim}(\mathcal{H}) \geq d$)**
2. **Every set C of size $d + 1$ is not shattered by \mathcal{H} . ($\text{VCdim}(\mathcal{H}) < d+1$)**

VCdim($\mathcal{H}_{\text{thresholds}}$)

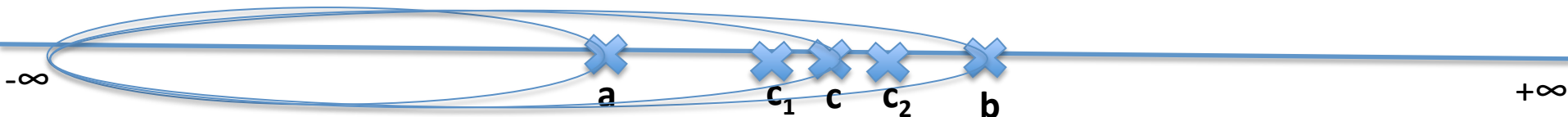
Consider $\mathcal{H} = \mathcal{H}_{\text{thresholds}}$ be the set of threshold functions over the real line.



Consider $C = \{c_1\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has two elements $\{h_a, h_b\}$ with $a \leq c_1$ and $b > c_1$ so \mathcal{H} shatters C . $\mathcal{H}_C = \{(0), (1)\}$, $|\mathcal{H}_C| = 2^{|C|} = 2^1$



Consider $C = \{c_1, c_2 \mid c_1 \leq c_2\}$. Then $\mathcal{H}_C = \{h: C \rightarrow \{0, 1\} \mid h \in \mathcal{H}\}$ has at most three elements, there is no function that realizes the labeling (0,1) and so \mathcal{H} does not shatter C .



So, **VCdim($\mathcal{H}_{\text{thresholds}}$) = 1**

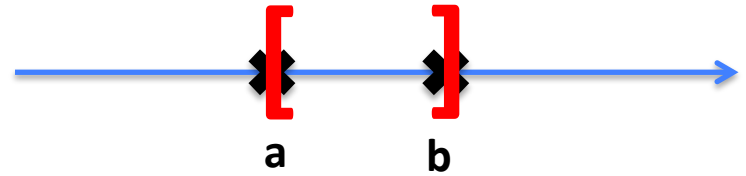
VCdim($\mathcal{H}_{\text{intervals}}$)

Consider $\mathcal{H} = \mathcal{H}_{\text{intervals}}$ be the set of intervals over the real line.

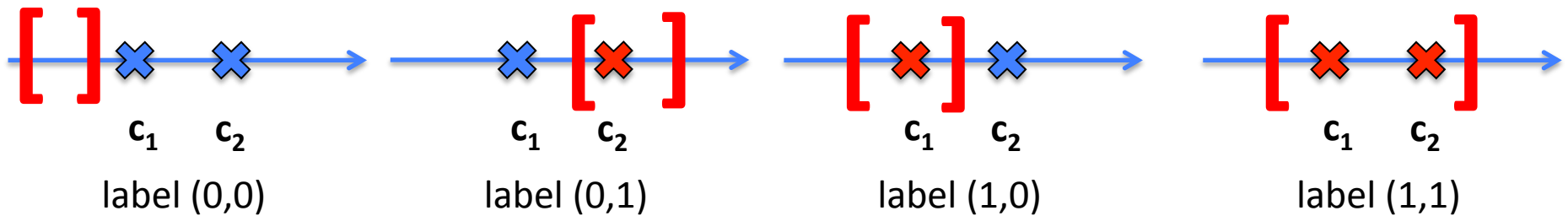
$$\mathcal{H}_{\text{intervals}} = \{[a,b] \mid a \leq b, a, b \in \mathbf{R}\}$$

Can also view $\mathcal{H}_{\text{intervals}}$ as:

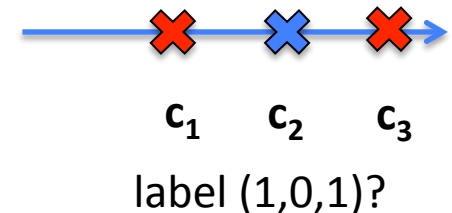
$$\mathcal{H}_{\text{intervals}} = \{h_{a,b}: \mathbf{R} \rightarrow \{0, 1\}, h_{a,b} = \mathbf{1}_{[a,b]}, a \leq b, a, b \in \mathbf{R}\}$$



$\mathcal{H}_{\text{intervals}}$ shatters any set A of two different points in \mathbf{R} .



$\mathcal{H}_{\text{intervals}}$ cannot shatter any set A of three points in \mathbf{R} .



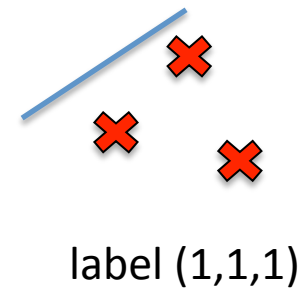
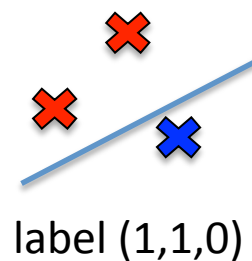
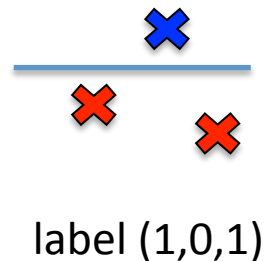
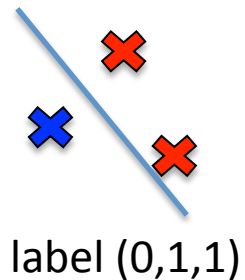
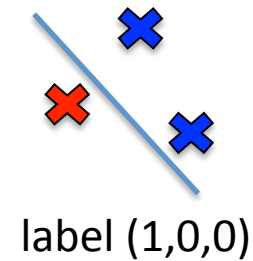
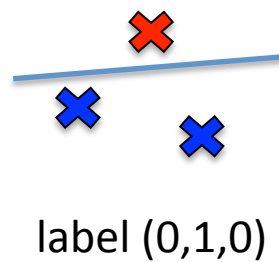
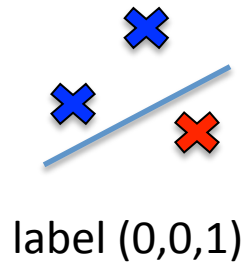
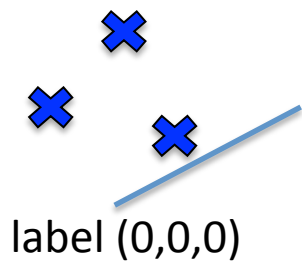
So, **VCdim($\mathcal{H}_{\text{intervals}}$) = 2**

VCdim($\mathcal{H}_{\text{lines}}$)

Consider $\mathcal{H} = \mathcal{H}_{\text{lines}}$ be the set of lines in \mathbf{R}^2 .

$\mathcal{H}_{\text{lines}} = \{h_{a,b,c}: \mathbf{R}^2 \rightarrow \{0, 1\}, h_{a,b,c}((x,y)) = \mathbf{1}_{[ax+by+c>0]}((x,y)), a, b, c \in \mathbf{R}\}$

$\mathcal{H}_{\text{lines}}$ shatters any set A of three non-collinear points in \mathbf{R}^2 .



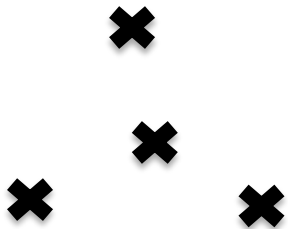
$\text{VCdim}(\mathcal{H}_{\text{lines}})$

Consider $\mathcal{H} = \mathcal{H}_{\text{lines}}$ be the set of lines in \mathbf{R}^2 .

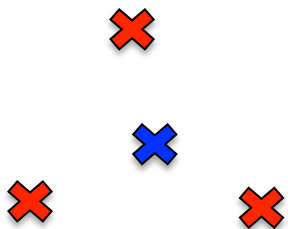
$\mathcal{H}_{\text{lines}} = \{h_{a,b,c}: \mathbf{R}^2 \rightarrow \{0, 1\}, h_{a,b,c}((x,y)) = \mathbf{1}_{[ax+by+c>0]}((x,y)) , a, b, c \in \mathbf{R}\}$

$\mathcal{H}_{\text{lines}}$ doesn't shatter any set A of four points in \mathbf{R}^2 .

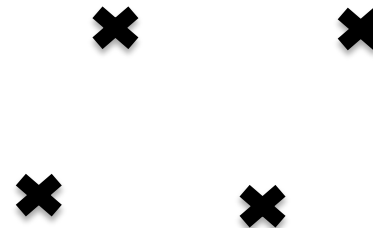
Case a: one point is interior to the convex hull of the other 3



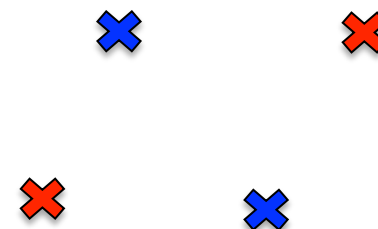
label $(1,1,1,0)$ is un-realizable



Case b: no point is interior to the convex hull of the other 3



label $(1,0,1,0)$ is un-realizable

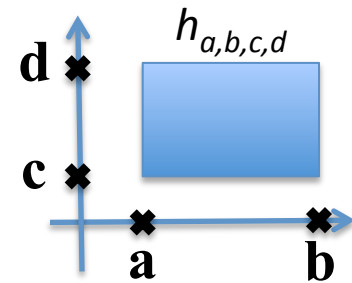


So, $\text{VCdim}(\mathcal{H}_{\text{lines}}) = 3$

VCdim($\mathcal{H}_{\text{rec}}^2$)

Consider $\mathcal{H} = \mathcal{H}_{\text{rec}}^2$ be the set of axis aligned rectangles in the \mathbf{R}^2 .

$$\mathcal{H}_{\text{rec}}^2 = \{[a,b] \times [c,d] \mid a \leq b, c \leq d, a, b, c, d \in \mathbf{R}\}$$



Can also view $\mathcal{H}_{\text{rec}}^2$ as:

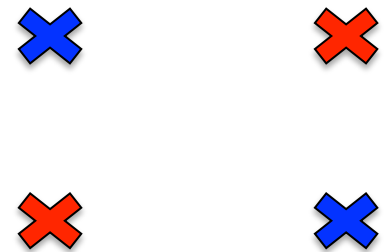
$$\mathcal{H}_{\text{rec}}^2 = \{h_{a,b,c,d}: \mathbf{R}^2 \rightarrow \{0, 1\}, h_{a,b,c,d} = \mathbf{1}_{[a,b] \times [c,d]}, a \leq b, c \leq d, a, b, c, d \in \mathbf{R}\}$$

Does $\mathcal{H}_{\text{rec}}^2$ shatters any set A of four different points in \mathbf{R}^2 ?

Take A the set of 4 vertices of a rectangle with axis

aligned in \mathbf{R}^2 . Then $\mathcal{H}_{\text{rec}}^2$ doesn't shatter A (the (0,1,0,1)

labeling is not realizable).



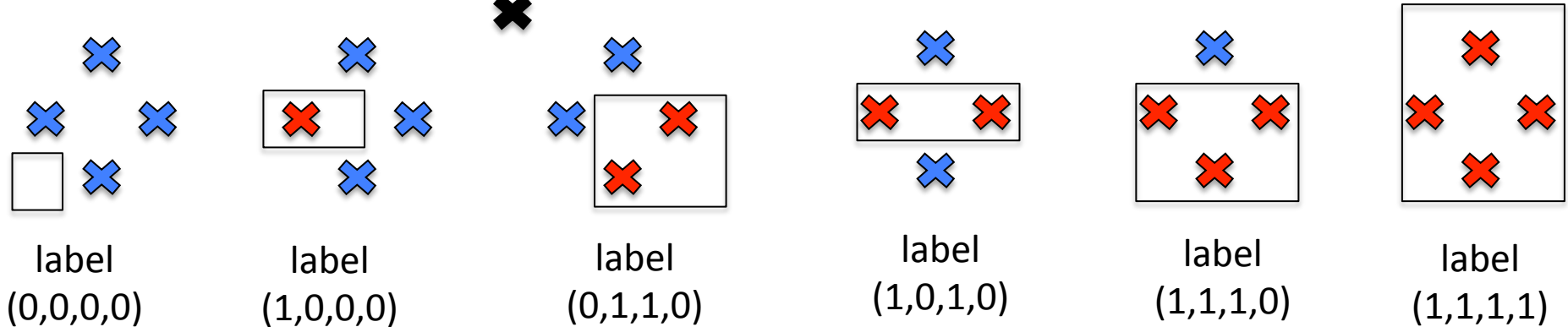
Does this mean that $\text{VCdim}(\mathcal{H}_{\text{rec}}^2) < 4$?

No! Either we show that all sets A of size 4 are not shattered by $\mathcal{H}_{\text{rec}}^2$

($\text{VCdim}(\mathcal{H}_{\text{rec}}^2) < 4$) or find a set A of size 4 that is shattered ($\text{VCdim}(\mathcal{H}_{\text{rec}}^2) \geq 4$).

VCdim($\mathcal{H}_{\text{rec}}^2$)

Consider the set A:  Then A is shattered by $\mathcal{H}_{\text{rec}}^2$.

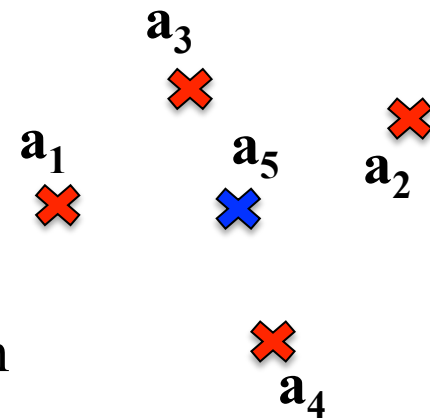


Can realize all 16 possible labels. So $\text{VCdim}(\mathcal{H}_{\text{rec}}^2) \geq 4$

Show now that all sets A of 5 points are not shattered by $\mathcal{H}_{\text{rec}}^2$.

A = $\{a_1, a_2, a_3, a_4, a_5\}$. Consider a_1 – the leftmost point (smaller x), a_2 – the rightmost point (larger x), a_3 – the lowest point (smaller y), a_4 – the highest point (larger y).

Then every rectangle containing a_1, a_2, a_3, a_4 will also contain a_5 , so (1,1,1,1,0) is not realizable. So, **$\text{VCdim}(\mathcal{H}_{\text{rec}}^2) = 4$**



$\text{VCdim}(\mathcal{H}_{\sin})$

$$\text{VCdim}(\mathcal{H}_{\text{thresholds}}) = 1, \text{VCdim}(\mathcal{H}_{\text{intervals}}) = 2, \text{VCdim}(\mathcal{H}_{\text{lines}}) = 3$$

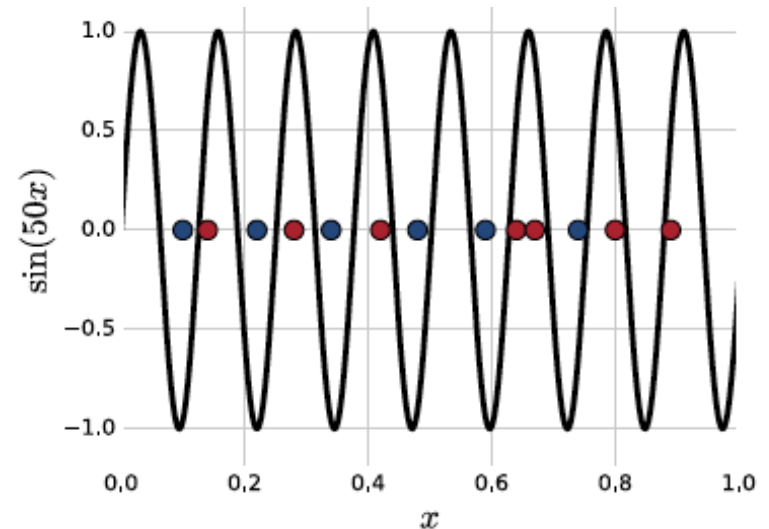
$$\text{VCdim}(\mathcal{H}_{\text{rec}}^2) = 4$$

Consider $\mathcal{H} = \mathcal{H}_{\sin}$ be the set of sin functions:

$$\mathcal{H}_{\sin} = \{h_{\theta}: \mathbf{R} \rightarrow \{0,1\} \mid h_{\theta}(x) = \lceil \sin(\theta x) \rceil, \theta \in \mathbf{R}\}, \lceil -1 \rceil = 0$$

$$\text{VCdim}(\mathcal{H}_{\sin}) = ?$$

It is possible to prove that $\text{VCdim}(\mathcal{H}) = \infty$,
namely, for every d , one can find d points
that are shattered by \mathcal{H} .



$\text{VCdim}(\mathcal{H}_{\sin})$

Lemma

Let $x \in (0, 1)$ and let $0.x_1x_2x_3\dots$ be the binary representation of x . Then, for any natural number m , provided that there exist $k \geq m$ such that $x_k = 1$, we have:

$$\left\lceil \sin(2^m \pi x) \right\rceil = 1 - x_m$$

Example of binary representation:

$$x = (0.x_1x_2x_3\dots)_2 = x_1 \times 2^{-1} + x_2 \times 2^{-2} + x_3 \times 2^{-3} + \dots$$

$$x = 0.75 = \frac{1}{2} + \frac{1}{4} = (0.110000\dots)_2$$

$$x = 0.3 = 0 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 0 \times 2^{-4} + 1 \times 2^{-5} + \dots = (0.01001\dots)_2$$

$VCdim(\mathcal{H}_{\sin})$

Lemma

Let $x \in (0, 1)$ and let $0.x_1x_2x_3\dots$ be the binary representation of x . Then, for any natural number m , provided that there exist $k \geq m$ such that $x_k = 1$, we have:

$$\left\lceil \sin(2^m \pi x) \right\rceil = 1 - x_m$$

Proof

$$\begin{aligned} \sin(2^m \pi x) &= \sin(2^m \pi (0.x_1x_2x_3\dots)) = \sin(2\pi * 2^{m-1} (0.x_1x_2x_3\dots)) = \\ &= \sin(2\pi * (x_1x_2x_3\dots x_{m-1}.x_m x_{m+1}\dots)) \text{ (left shift with } m-1 \text{ position)} \\ &= \sin(2\pi * (x_1x_2x_3\dots x_{m-1}.x_m x_{m+1}\dots) - 2\pi * (x_1x_2x_3\dots x_{m-1}.0)) \text{ (sin has period } 2\pi) \\ &= \sin(2\pi * (0.x_mx_{m+1}\dots)) \end{aligned}$$

Note that $0.x_mx_{m+1}\dots > 0$ as there exist $k \geq m$ such that $x_k = 1$

Case 1: $x_m = 0$, then $0 < 2\pi * (0.x_mx_{m+1}\dots) < 2\pi * 1/2 = \pi$. So $0 < \sin(2^m \pi x) < 1$, and from here it results that: $\left\lceil \sin(2^m \pi x) \right\rceil = 1 = 1 - 0 = 1 - x_m$

Case 2: $x_m = 1$, then $2\pi > 2\pi * (0.x_mx_{m+1}\dots) \geq 2\pi * 1/2 = \pi$. So $-1 \leq \sin(2^m \pi x) \leq 0$, and from here it results that: $\left\lceil \sin(2^m \pi x) \right\rceil = 0 = 1 - 1 = 1 - x_m$ (we consider $\lceil -1 \rceil = 0$)

VCdim(\mathcal{H}_{sin})

To prove $\text{VCdim}(\mathcal{H}_{\text{sin}}) = \infty$, we need to pick n points which are shattered by \mathcal{H}_{sin} , for any n . To do so, we construct n points $x_1, x_2, \dots, x_n \in [0, 1]$, such that the set of the m -th bits in the binary expansion, as m ranges from 1 to 2^n , ranges over all possible labelings of x_1, x_2, \dots, x_n .

$$\begin{array}{rcl} x_1 & = & 0.0000 \dots 11 \\ x_2 & = & 0.0000 \dots 11 \\ & \dots & \\ x_{n-1} & = & 00011 \dots 11 \\ x_n & = & 0.0101 \dots 01 \end{array}$$

m=1

For example, to give the labeling 1 for all instances, we just pick m=1:

$$h(x) = \left\lceil \sin(2^1 \pi x) \right\rceil$$

which returns the first bit (column) in the binary expansion.

VCdim(\mathcal{H}_{\sin})

To prove $\text{VCdim}(\mathcal{H}_{\sin}) = \infty$, we need to pick n points which are shattered by \mathcal{H}_{\sin} , for any n . To do so, we construct n points $x_1, x_2, \dots, x_n \in [0, 1]$, such that the set of the m -th bits in the binary expansion, as m ranges from 1 to 2^n , ranges over all possible labelings of x_1, x_2, \dots, x_n .

$$\begin{array}{rcl} x_1 & = & 0.00000\dots 11 \\ x_2 & = & 0.00000\dots 11 \\ & \dots & \\ x_{n-1} & = & 0.0011\dots 11 \\ x_n & = & 0.0101\dots 01 \end{array}$$

$m=2$

If we wish to give the labeling 1 for x_1, x_2, \dots, x_{n-1} , and the labeling 0 for x_n , we pick $m=2$:

$$h(x) = \left\lceil \sin(2^2 \pi x) \right\rceil$$

which returns the second bit (column) in the binary expansion.

VCdim(\mathcal{H}_{\sin})

To prove $\text{VCdim}(\mathcal{H}_{\sin}) = \infty$, we need to pick n points which are shattered by \mathcal{H}_{\sin} , for any n . To do so, we construct n points $x_1, x_2, \dots, x_n \in [0, 1]$, such that the set of the m -th bits in the binary expansion, as m ranges from 1 to 2^n , ranges over all possible labelings of x_1, x_2, \dots, x_n .

$$\begin{array}{rcl} x_1 & = & 0.0000\dots 11 \\ x_2 & = & 0.0000\dots 11 \\ & \dots & \\ x_{n-1} & = & 0.0011\dots 11 \\ x_n & = & 0.0101\dots 01 \end{array}$$

$m=2^n$

If we wish to give the labeling 0 for $x_1, x_2, \dots, x_{n-1}, x_n$ we pick $m=2^n$:

$$h(x) = \left\lfloor \sin(2^{2^n} \pi x) \right\rfloor$$

which returns the bit on position 2^n (column) in the binary expansion.

$\text{VCdim}(\mathcal{H}_{\text{sin}})$

To prove $\text{VCdim}(\mathcal{H}_{\text{sin}}) = \infty$, we need to pick n points which are shattered by \mathcal{H}_{sin} , for any n . To do so, we construct n points $x_1, x_2, \dots, x_n \in [0, 1]$, such that the set of the m -th bits in the binary expansion, as m ranges from 1 to 2^n , ranges over all possible labelings of x_1, x_2, \dots, x_n .

$$x_1 = 0.0000\dots 11$$

$$x_2 = 0.0000\dots 11$$

...

$$x_{n-1} = 0.0011\dots 11$$

$$x_n = 0.0101\dots 01$$

We conclude that $x_1, x_2, \dots, x_{n-1}, x_n$ can be given any labeling by some $h \in \mathcal{H}_{\text{sin}}$, so it is shattered. This can be done for any n , so $\text{VCdim}(\mathcal{H}_{\text{sin}}) = \infty$.

Some basic properties of the $\text{VCdim}(\mathcal{H})$

1. $\text{VCdim}(\mathcal{H}) \leq \log_2 |\mathcal{H}|$
2. If $\mathcal{H}_1 \subseteq \mathcal{H}_2$ then $\text{VCdim}(\mathcal{H}_1) \leq \text{VCdim}(\mathcal{H}_2)$
3. If $\text{VCdim}(\mathcal{H}) = \infty$ then \mathcal{H} is not PAC learnable