Advanced Machine Learning

Bogdan Alexe,

bogdan.alexe@fmi.unibuc.ro

University of Bucharest, 2nd semester, 2021-2022

Recap - Shattering

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 $\mathcal{X}$, $C = \{c_1, c_2, \ldots, c_m\}$. 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:

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

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 $|\mathcal{H}_C| = 2^{|C|}$.

Recap - The VC-dimension

Definition (VC-dimension)

The VC - dimension of a hypothesis class $\mathcal{H}$, denoted 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.

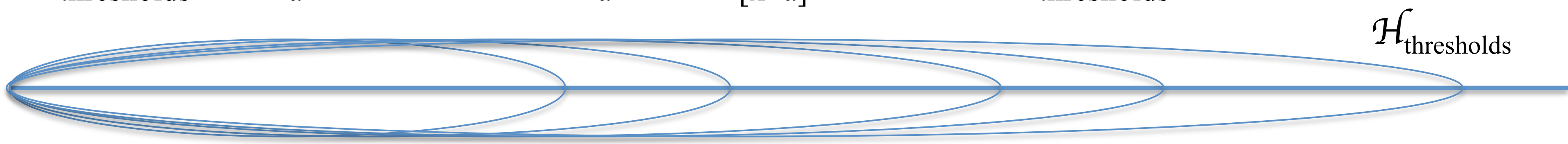
In order to show that the VC-dimension of a hypothesis class $\mathcal{H}$ is *d*, we need to show that:

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

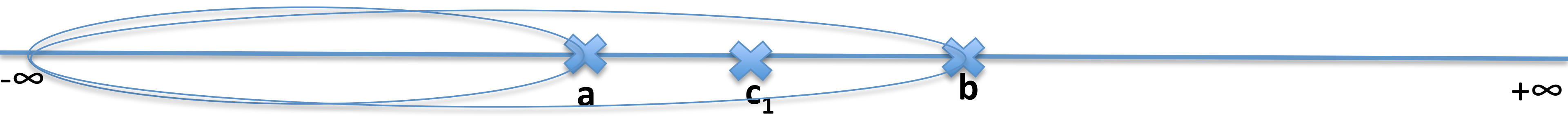
We will see in the next lecture that the converse is also true: ***a finite VC-dimension guarantees learnability***. ***Hence, the VC-dimension characterizes PAC learnability. VC-dimension is a combinatorial measure, does not imply computing probabilities.***

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

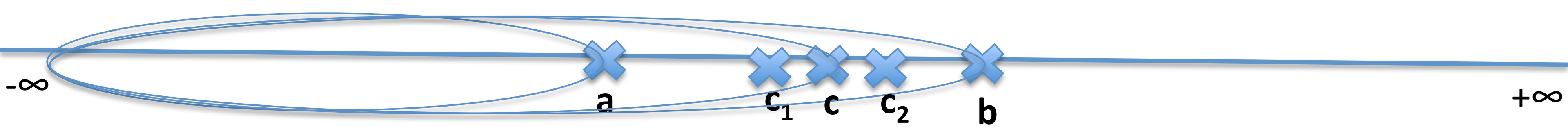
Consider $\mathcal{H} = \mathcal{H}_{\text{thersholds}}$ be the set of threshold functions over the real line.
$\mathcal{H}_{\text{thresholds}} = \{h_a\colon \mathrm{R} \to \{0, 1\}, h_a(\mathrm{x}) = \mathbf{1}_{[\mathrm{x}<\mathrm{a}]}, \mathrm{a} \in \mathbf{R}\}, |\mathcal{H}_{\text{thresholds}}|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



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

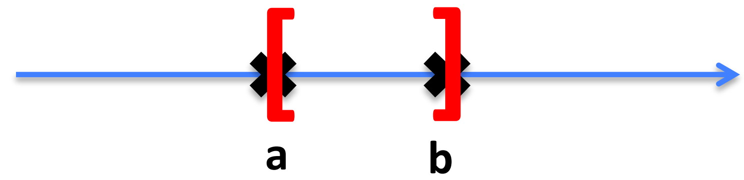


So, **VCdim($\mathcal{H}_{\text{thresholds}}$) = 1**

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

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

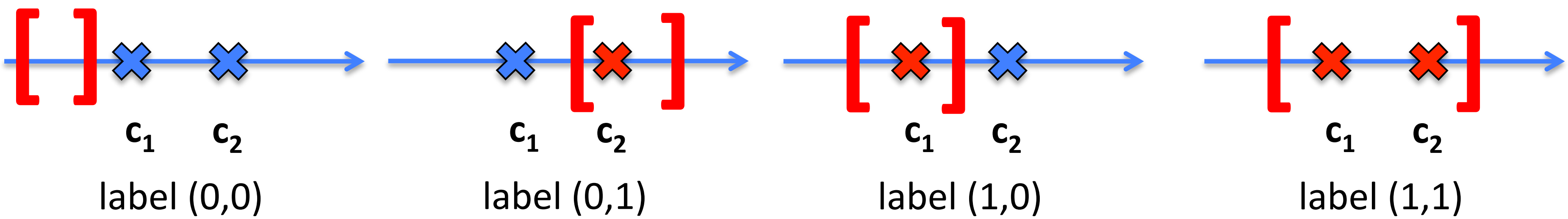
$\mathcal{H}_{\text{intervals}} = \{[a,b] \mid a \leq b,\ a, b \in \mathbf{R}\}$



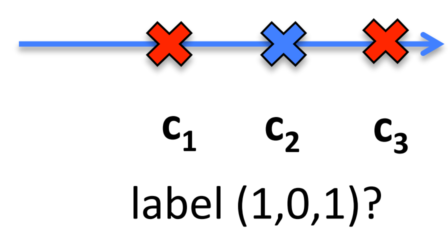
Can also view $\mathcal{H}_{\text{intervals}}$ as:

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

$\mathcal{H}_{\text{intervals}}$ 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($\mathbf{H}_{\text{intervals}}$) = 2**

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

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

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

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

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

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

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

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

$\mathcal{H}_{\text{rec}}{}^2$ shatters the set A of 4 points arranged as a diamond. So VCdim($\mathcal{H}_{\text{rec}}{}^2$) ≥ 4

$\mathcal{H}_{\text{rec}}{}^2$ doesn't shatter any set of A of five pints in $\mathbf{R}^2$ (geometric argument).

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

Some basic properties of the 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

Today’s lecture: Overview

- Computing the VC-dimension for some particular $\mathcal{H}$
- Assignment 1

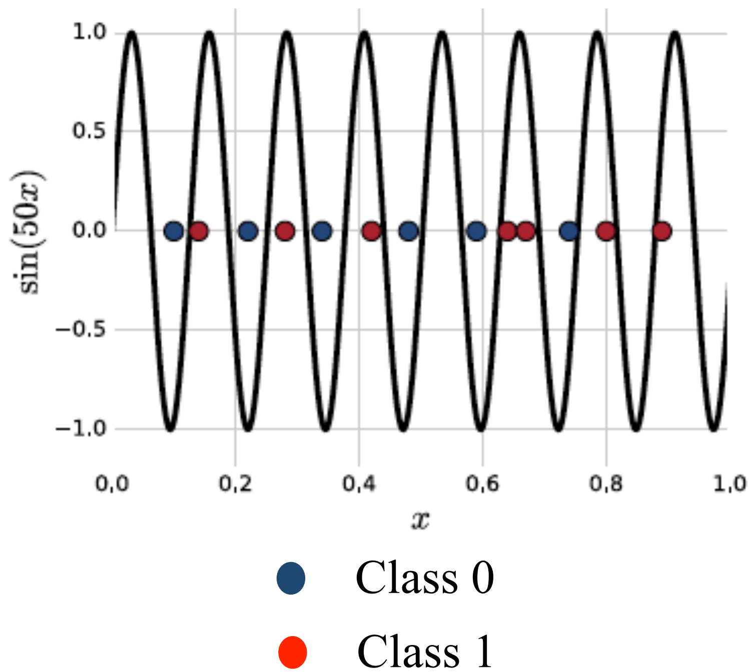
VCdim($\mathcal{H}_{\text{sin}}$)

VCdim($\mathcal{H}_{\text{thresholds}}$) = 1, VCdim($\mathcal{H}_{\text{intervals}}$) = 2, VCdim($\mathcal{H}_{\text{lines}}$) = 3
VCdim($\mathcal{H}_{\text{rec}^2}$) = 4

Consider $\mathcal{H} = \mathcal{H}_{\text{sin}}$ be the set of sin functions:
$\mathcal{H}_{\text{sin}} = \{h_\theta: \mathbf{R} \rightarrow \{0,1\} \mid h_\theta(x) = \lceil \sin(\theta x) \rceil, \theta \in \mathrm{R}\}$, $\lceil -1 \rceil = 0$

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

We will show that VCdim($\mathcal{H}_{\text{sin}}$) = ∞



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

Lemma

Let $x \in (0,1)$ and let $0.x_1x_2x_3\ldots$ 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:

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

Example of binary representation:

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

$x = 0.75 = ½ + ¼ = (0.110000\ldots.)_2$

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

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

Lemma

Let x $\in$ (0, 1) and let $0.x_1x_2x_3\ldots$ 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:

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

Proof

$\sin(2^m\pi x) = \sin(2^m\pi(0.x_1x_2x_3\ldots)) = \sin(2\pi * 2^{m-1}(0.x_1x_2x_3\ldots)) =$

$= \sin(2\pi*(x_1x_2x_3\ldots x_{m-1}.x_m\, x_{m+1}\ldots))$ (left shift with m-1 position)

$= \sin(2\pi*(x_1x_2x_3\ldots x_{m-1}.x_m\, x_{m+1}\ldots) - 2\pi*(x_1x_2x_3\ldots x_{m-1}.0))$ (sin has period 2π)

$= \sin(2\pi*(0.x_mx_{m+1}\ldots))$

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

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

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

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

To prove VCdim($\mathcal{H}_{\sin}$) = ∞, we need to pick n points which are shattered by $\mathcal{H}_{\sin}$, for any n. To do so, we construct n points $x_1, x_2, \ldots, x_n \in [0, 1]$, 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 $x_1, x_2, \ldots, x_n$.

x_1 = 0. 0 0 0 0 ... 1 1

x_2 = 0. 0 0 0 0 ... 1 1

...

x_{n-1} = 0. 0 0 1 1 ... 1 1

x_n = 0. 0 1 0 1 ... 0 1

m=1

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

$$h(x) = \lceil \sin(2^1 \pi x) \rceil = 1 - first\_bit\_of\_binary\_repres\_of\_x$$

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

which returns 1 - the first bit (column) in the binary expansion.

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

To prove VCdim($\mathcal{H}_{\sin}$) = ∞, we need to pick n points which are shattered by $\mathcal{H}_{\sin}$, for any n. To do so, we construct n points $x_1, x_2, \ldots, x_n \in [0, 1]$, 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 $x_1, x_2, \ldots, x_n$.

$$
\begin{array}{lcl}
x_1 & = & 0.\ 0\ 0\ 0\ 0 \ldots\ \ 1\ 1 \\
x_2 & = & 0.\ 0\ 0\ 0\ 0 \ldots\ \ 1\ 1 \\
 & \ldots & \\
x_{n-1} & = & 0.\ 0\ 0\ 1\ 1 \ldots\ \ 1\ 1 \\
x_n & = & 0.\ 0\ 1\ 0\ 1 \ldots\ \ 0\ 1
\end{array}
$$

m=2

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

$$h(x) = \lceil \sin(2^2 \pi x) \rceil = 1 - second\_bit\_of\_binary\_repres\_of\_x$$

which returns 1 - the second bit (column) in the binary expansion.

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

To prove VCdim($\mathcal{H}_{\sin}$) = ∞, we need to pick *n* points which are shattered by $\mathcal{H}_{\sin}$, for any *n*. To do so, we construct *n* points $x_1, x_2, \ldots, x_n \in [0, 1]$, 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 $x_1, x_2, \ldots, x_n$.

x_1 = 0. 0 0 0 0 ... 1 1

x_2 = 0. 0 0 0 0 ... 1 1

...

x_{n-1} = 0. 0 0 1 1 ... 1 1

x_n = 0. 0 1 0 1 ... 0 1

$m=2^n$

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

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

which returns 1 - the bit on position 2^n (column) in the binary expansion.

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

To prove VCdim($\mathcal{H}_{\sin}$) = ∞, we need to pick *n* points which are shattered by $\mathcal{H}_{\sin}$, for any *n*. To do so, we construct *n* points $x_1, x_2, \ldots, x_n \in [0, 1]$, 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 $x_1, x_2, \ldots, x_n$.

$x_1 \;=\; 0.\,0\,0\,0\,0 \ldots \;\; 1\,1$

$x_2 \;=\; 0.\,0\,0\,0\,0 \ldots \;\; 1\,1$

...

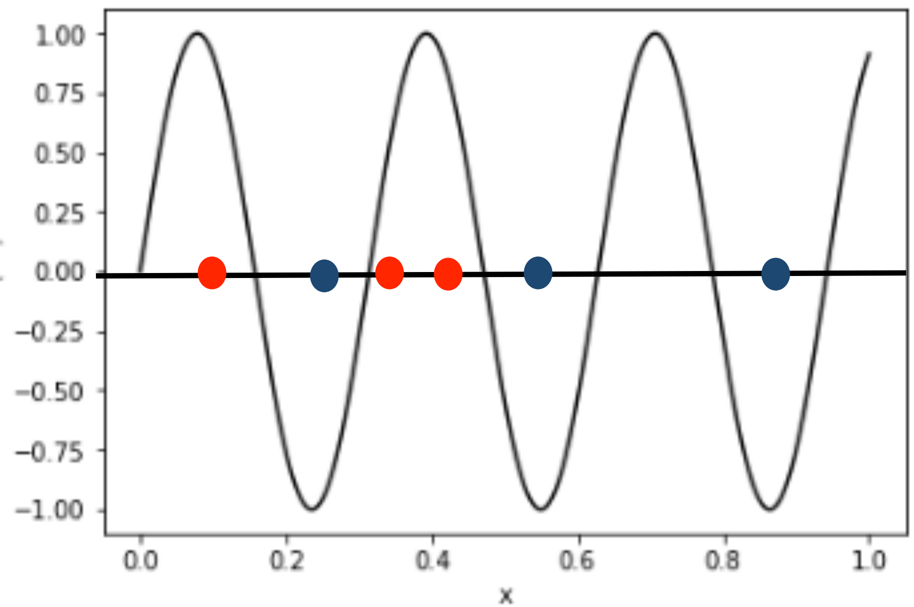
$x_{n-1} = \; 0.\,0\,0\,1\,1 \ldots \;\; 1\,1$

$x_n \;=\; 0.\,0\,1\,0\,1 \ldots \;\; 0\,1$

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

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

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



Training points

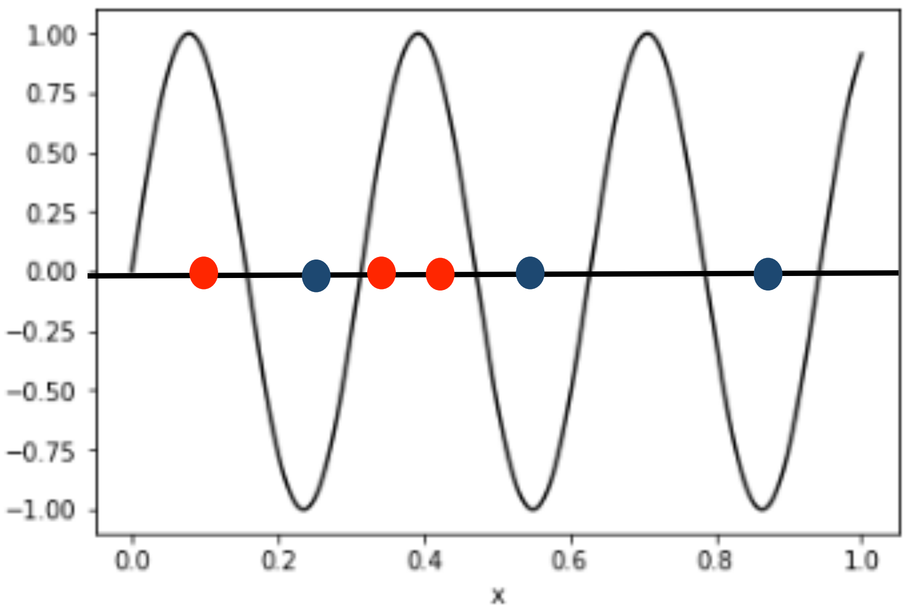
Class 0

Class 1

$$h(x) = \left[\sin(20x)\right]$$

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

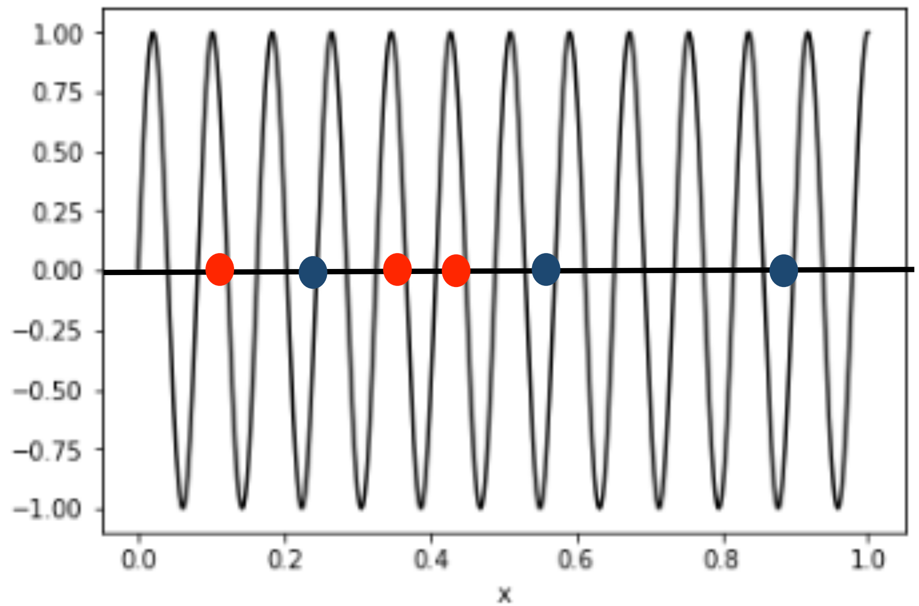
We conclude that $x_1, x_2, \ldots, x_{n-1}, x_n$ can be given any labeling by some $h \in \mathcal{H}_{\sin}$, so it is shattered. This can be done for any n, so VCdim($\mathcal{H}_{\sin}$) = ∞.



Class 0

Class 1

$$h(x) = \lceil \sin(20x) \rceil$$



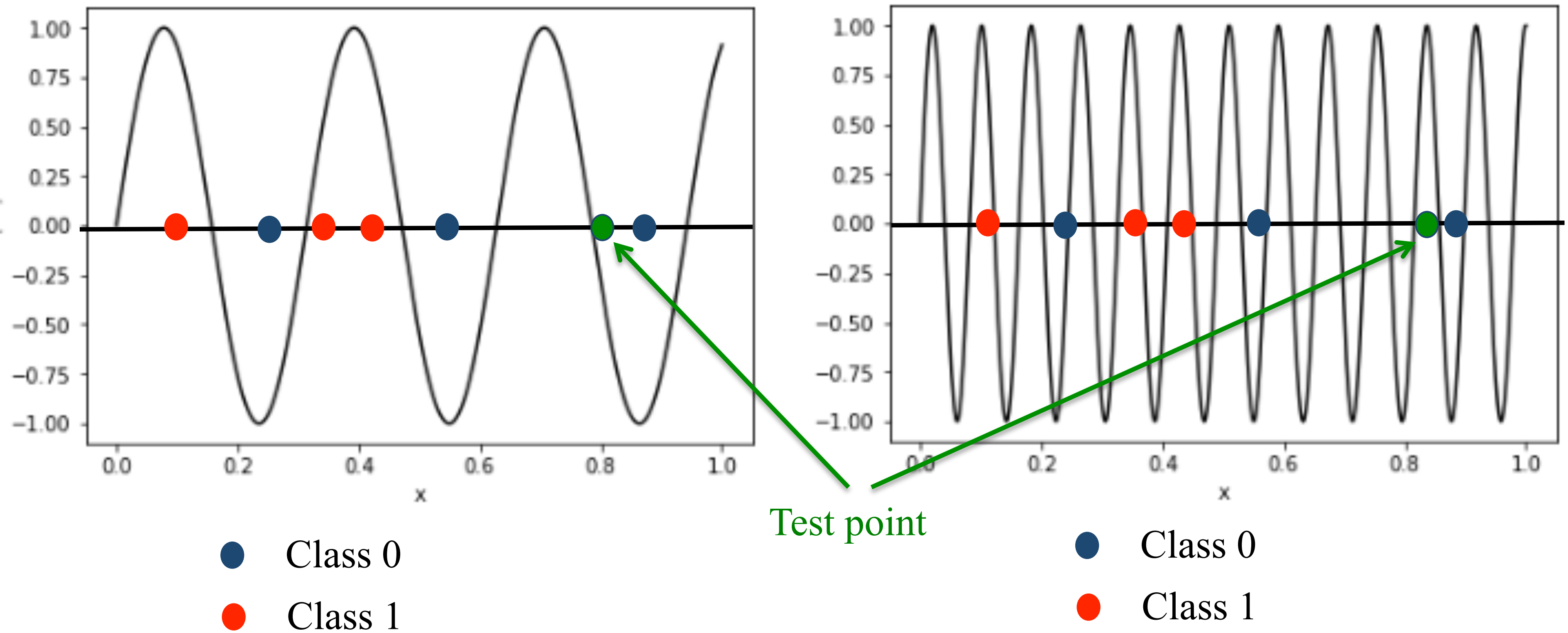
Class 0

Class 1

$$h(x) = \lceil \sin(77x) \rceil$$

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

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

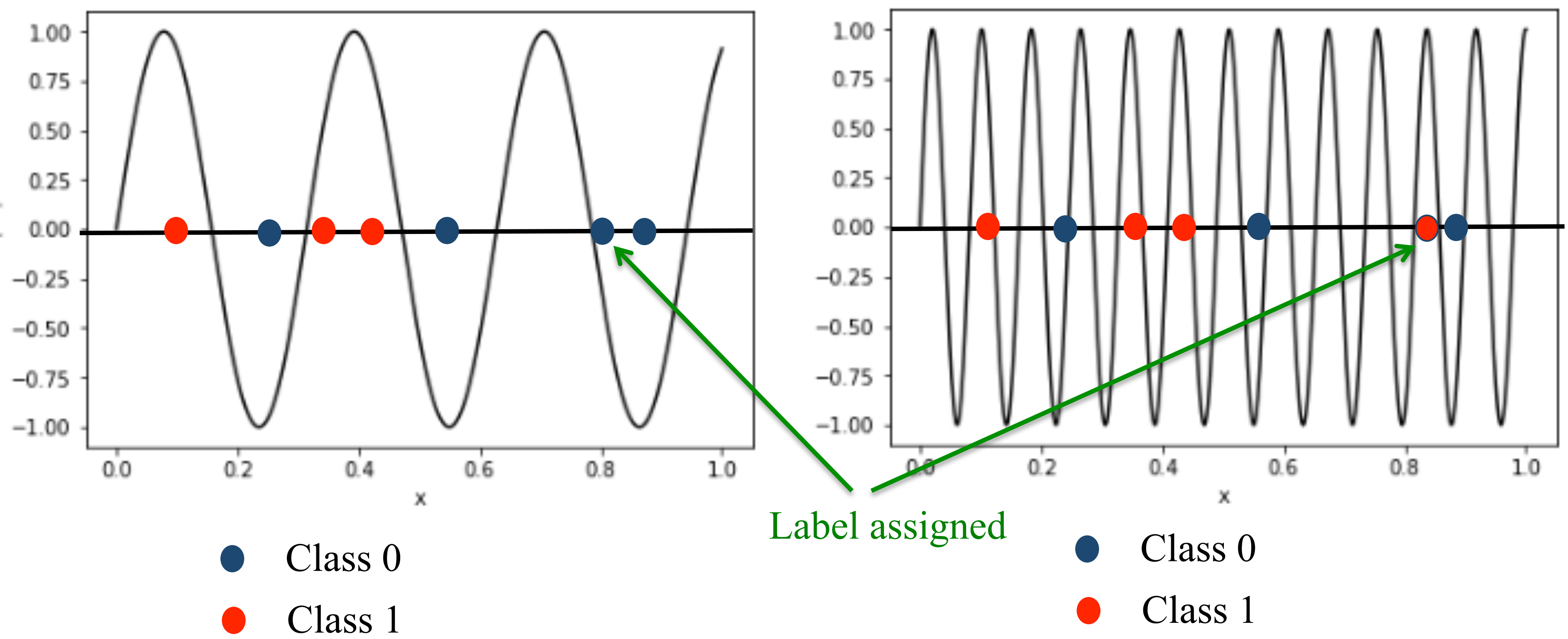


$h(x) = \left[\sin(20x)\right]$

$h(x) = \left[\sin(77x)\right]$

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

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



$$h(x) = \lceil \sin(20x) \rceil \qquad h(x) = \lceil \sin(77x) \rceil$$

VCdim($\mathcal{HS}_0^n$)

Consider $\mathcal{H} = \mathcal{HS}^n$ be the set of halfspaces (linear classifiers) in R^n

$$\mathcal{H} = \mathcal{HS}^n = \{h_{w,b}: \mathbf{R}^n \rightarrow \{-1, 1\},\ h_{w,b}(x) = sign\left(\sum_{i=1}^{n} w_i x_i + b\right) \mid w \in \mathbf{R}^n,\ b \in \mathbf{R}\}$$

Consider label -1 to correspond to label 0, they are basically the same.

For n = 2 we have:

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

Let us restrict our attention to “homogenous” linear classifiers, the ones that go through origin, b = 0.

$$\mathcal{HS}_0^n = \{h_{w,0}: \mathbf{R}^n \rightarrow \{-1, 1\},\ h_{w,0}(x) = sign\left(\sum_{i=1}^{n} w_i x_i\right) \mid w \in \mathbf{R}^n\}$$

What is the VCdim($\mathcal{HS}_0^n$)?

VCdim($\mathcal{H}S_0^{\,n}$)

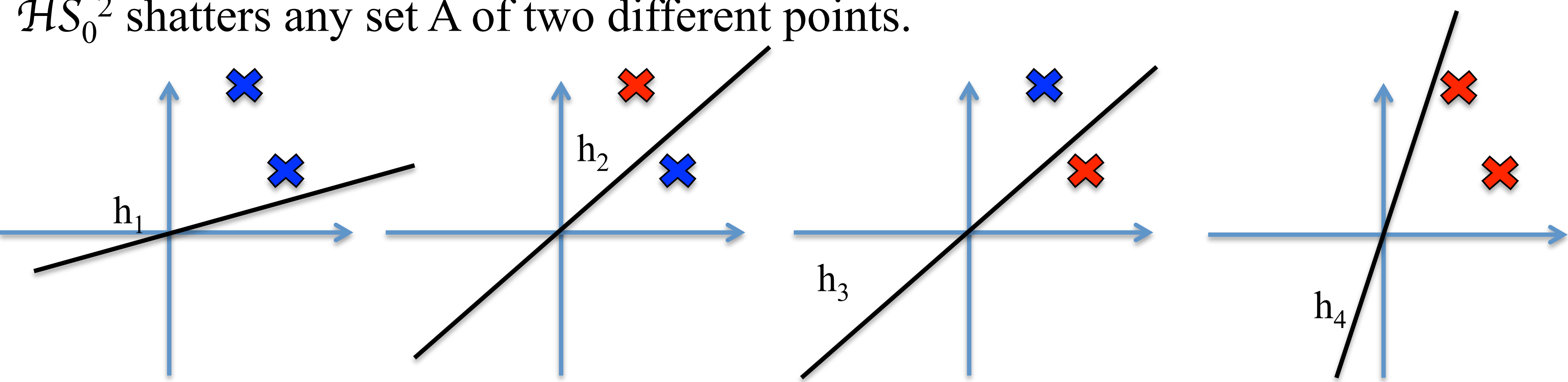
$\mathcal{H}S_0^{\,n} = \{h_{w,0}: \mathbf{R}^n \rightarrow \{-1, 1\},\ h_{w,0}(x) = sign\left(\sum_{i=1}^{n} w_i x_i\right) \mid w \in \mathbf{R}^n\}$

For n = 2 we have:

$\mathcal{H}S_0^{\,2} = \{h_{w1,w2}: \mathbf{R}^2 \rightarrow \{-1, 1\},\ h_{w1,w2}(x) = sign(w_1x_1 + w_2x_2) \mid (w_1,w_2) \in \mathbf{R}^2\}$

What is the VCdim($HS_0^{\,2}$) ?

$\mathcal{H}S_0^{\,2}$ shatters any set A of two different points.



label (-1,-1)

label (1,-1)

label (-1,1)
Change the signs for the coefficient in h_2

label (1,1)

Does $HS_0^{\,2}$ shatter a set A of three points?

Difficult to reason geometrically… choose the algebraic proof.

VCdim($\mathcal{H}S_0^n$)

We will show that VCdim($\mathcal{H}S_0^n$) = n.

Proof: 1st part

We first show that VCdim($\mathcal{H}S_0^n$) $\geq n$.

We find a set A consisting of n points in $\mathbf{R}^n$ that is shattered by $\mathcal{H}S_0^n$.

Take A = $\{e_1, e_2, \ldots, e_n\}$ to be the orthonormal basis of $\mathbf{R}^n$.

$e_1 = (1, 0, 0, \ldots, 0)$; $e_2 = (0, 1, 0, \ldots, 0)$; …. ; $e_n = (0, 0, 0, \ldots, 1)$

We want to proof that $\mathcal{H}S_0^n$ shatters A, so that VCdim($\mathcal{H}S_0^n$) $\geq n$. This is equivalent to proof that for every B $\subseteq$ A, there is a function $h_B \in \mathcal{H}S_0^n$ such that h_B gives label +1 to all elements in B and label -1 to all elements of A \ B.

Pick B subset of A, B $\subseteq$ $\{e_1, e_2, \ldots, e_n\}$. Choose $w = (w_1, w_2, \ldots, w_n)$ such that:

$$w_i = \begin{cases} 1, \text{ if } e_i \in B \\ -1, \text{ if } e_i \notin B \end{cases}$$

Then, $h_B(e_i) = \text{sign}(\langle w, e_i \rangle) = w_i$ will generate the labels +1 for elements in B, -1 for elements not in B

VCdim($\mathcal{H}S_0^{\,n}$)

Proof: 2nd part

We now show that VCdim($\mathcal{H}S_0^{\,n}$) < n + 1.

We will prove that given any set A = {x_1, x_2, …, x_{n+1}} of n + 1 points in $\mathbf{R}^n$, A cannot be shattered by $\mathcal{H}S_0^{\,n}$.

The points {x_1, x_2, …, x_{n+1}} "live" in $\mathbf{R}^n$, a vector space with dimension n. So, {x_1, x_2, …, x_{n+1}} are linearly dependent and there exist coefficients a_1, a_2, … a_{n+1} not all of them 0 such that:

$$\sum_{i=1}^{n+1} a_i x_i = 0$$

Take P ⊆ { 1, 2, …, n +1} the set of strictly positive coefficients a_i and N ⊆ { 1, 2, …, n +1} the set of negative coefficients of a_i. So we have:

$$\sum_{i\in P} a_i x_i = \sum_{j\in N} |a_j| x_j$$

VCdim($\mathcal{H}S_0^n$)

Take P $\subseteq$ { 1, 2, …, n +1} the set of strictly positive coefficients a_i and N $\subseteq$ { 1, 2, …, n +1}the set of negative coefficients of a_i. Both P and N cannot be at the same time empty. So we have:

$$\sum_{i \in P} a_i x_i = \sum_{j \in N} |a_j| x_j$$

Assume that A is shattered by $\mathcal{H}S_0^n$ and take B = {x_i| i $\in$ P}. In particular, there exist h_B such that it realizes the label consisting of +1 for all $x_i \in$ B and -1 for all $x_i \notin$ B.

So, we have that $h_B(x_i) = 1$, if $x_i \in$ B, meaning that $\langle w_B, x_i \rangle \geq 0$ if $x_i \in$ B and $h_B(x_i) = -1$, if $x_i \notin$ B, meaning that $\langle w_B, x_i \rangle < 0$ if $x_i \notin$ B

So, we have that $h_B\left(\sum_{i \in P} a_i x_i\right) = sign(\left\langle w_B, \sum_{i \in P} a_i x_i \right\rangle) = sign(\sum_{i \in P} a_i \langle w_B, x_i \rangle)$

But $a_i > 0$ (because i $\in$ P) and also $\langle w_B, x_i \rangle \geq 0$ as $x_i \in$ B, so we obtain that:

$$h_B\left(\sum_{i \in P} a_i x_i\right) = sign(\left\langle w_B, \sum_{i \in P} a_i x_i \right\rangle) = sign(\sum_{i \in P} a_i \langle w_B, x_i \rangle) = 1$$

VCdim($\mathcal{H}S_0^{\,n}$)

Take P ⊆ { 1, 2, …, n +1} the set of strictly positive coefficients a_i and N ⊆ { 1, 2, …, n +1}the set of negative coefficients of a_i. Both P and N cannot be at the same time empty. So we have:

$$\sum_{i \in P} a_i x_i = \sum_{j \in N} |a_j| x_j$$

Assume that A is shattered by $\mathcal{H}S_0^{\,n}$ and take B = $\{x_i | \ i \in P\}$. In particular, there exist h_B such that it realizes the label consisting of +1 for all $x_i \in$ B and -1 for all $x_i \notin$ B.

So, we have that $h_B(x_i) = 1$, if $x_i \in$ B, meaning that $\langle w_B, x_i\rangle \geq 0$ if $x_i \in$ B and $h_B(x_i) = -1$, if $x_i \notin$ B, meaning that $\langle w_B, x_i\rangle < 0$ if $x_i \notin$ B

On the other hand, we have that $h_B\left(\sum_{j \in N} |a_j| x_j\right) = sign\left(\left\langle w_B, \sum_{j \in N} |a_j| x_j\right\rangle\right) = sign\left(\sum_{j \in N} |a_j| \langle w_B, x_j\rangle\right)$

But $|a_j| > 0$ and also $\langle w_B, x_j\rangle < 0$ as $x_j \notin$ B, so we obtain that:

$$h_B\left(\sum_{j \in N} |a_j| x_j\right) = sign\left(\left\langle w_B, \sum_{j \in N} |a_j| x_j\right\rangle\right) = sign\left(\sum_{j \in N} |a_j| \langle w_B, x_j\rangle\right) = -1$$

VCdim($\mathcal{H}S_0^{\mathrm{n}}$)

Take P $\subseteq$ { 1, 2, …, n +1} the set of strictly positive coefficients a_i and N $\subseteq$ { 1, 2, …, n +1}the set of negative coefficients of a_i. Both P and N cannot be at the same time empty. So we have:

$$\sum_{i\in P} a_i x_i = \sum_{j\in N} |a_j| x_j$$

Assume that A is shattered by $\mathcal{H}S_0^{\mathrm{n}}$ and take B = $\{x_i| \ i \in P\}$. In particular, there exist h_B such that it realizes the label consisting of +1 for all $x_i \in$ B and -1 for all $x_i \notin$B. So $h_B(x_i)$ = 1, if $x_i \in$ B and $h_B(x_i)$ = 1, if $x_i \notin$ B

$$\sum_{i\in P} a_i x_i = \sum_{j\in N} |a_j| x_j$$

$$h_B\left(\sum_{i\in P} a_i x_i\right) = sign(\left\langle w_B, \sum_{i\in P} a_i x_i \right\rangle) = sign(\sum_{i\in P} a_i \langle w_B, x_i \rangle) = 1$$

$$h_B\left(\sum_{j\in N} |a_j| x_j\right) = sign\left(\left\langle w_B, \sum_{j\in N} |a_j| x_j \right\rangle\right) = sign\left(\sum_{j\in N} |a_j| \langle w_B, x_j \rangle\right) = -1$$

So, this is a contradiction.

VCdim($\mathcal{HS}_0^n$)

Proof:

1st part – show that VCdim($\mathcal{HS}_0^n$) $\geq$ n

A = {e_1, e_2, …, e_n}, the orthonormal basis of $\mathbf{R}^n$ is shattered by $\mathcal{HS}_0^n$.

2nd part – show that VCdim($\mathcal{HS}_0^n$) < n + 1

Any set A = {x_1, x_2, …, x_{n+1}} of *n* + 1 points in $\mathbf{R}^n$ cannot be shattered by $\mathcal{HS}_0^n$. Provide an algebraic proof, based on the fact that {x_1, x_2, …, x_{n+1}} are linearly dependent in $\mathbf{R}^n$.

So, VCdim($\mathcal{HS}_0^n$) = n

Similarly, it can be shown that VCdim($\mathcal{HS}^n$) = n + 1

Assignment 1

Assignment 1 – good to know

- 6 problems = 5 points + 0.5 points (bonus) = 5.5 points
- deadline: in ~ 3 weeks time, Friday, 22nd of April 2022, 23:59
  - late submission policy: maximum 3 days allowed, -10% (= 0.5 points) for each day
  - upload a pdf written in a scientific editor (Word, Latex, LyX) containing your solution here: ***https://tinyurl.com/AML-2022-ASSIGNMENT1***
  - *is mandatory that you write your solution with a scientific editor, otherwise your solution would not be taken into account*
  - you can insert drawings for your proofs
  - after uploading your solution you should get a confirmation email; if not send me an email.

- for every problem write clear explanations, proofs to justify your answer (if you write just some indications you will not get too many points)

- do not share/copy the solution with/from your colleagues: you + your colleague/s will get 0 points

Problems 1 and 2

Assignment 1

Deadline: Friday, 22nd of April 2022

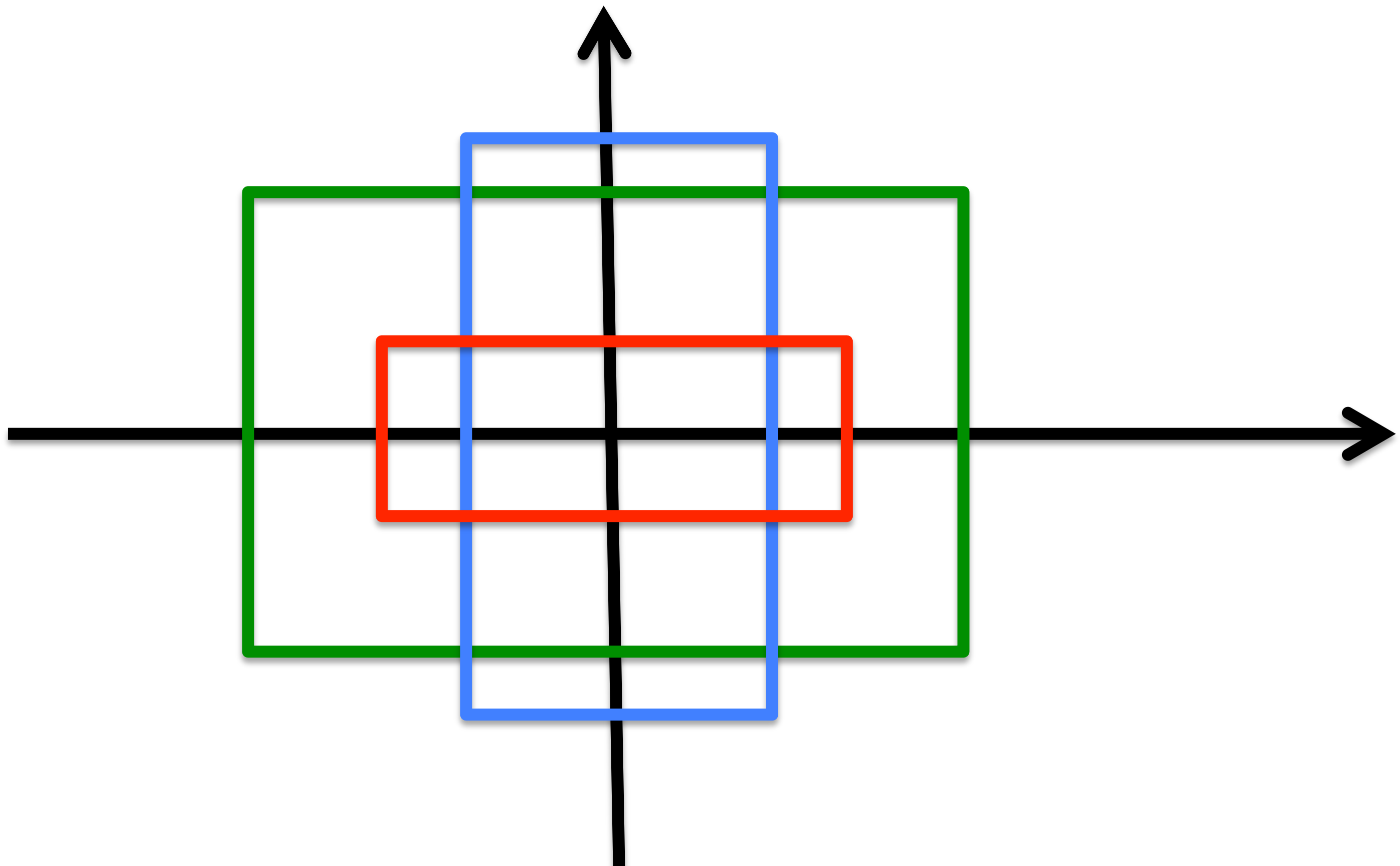
Upload your solutions at: https://tinyurl.com/AML-2022-ASSIGNMENT1

1. **(0.5 points)** Give an example of a finite hypothesis class $\mathcal{H}$ with $\text{VCdim}(\mathcal{H}) = 2022$. Justify your choice.

2. **(0.5 points)** What is the maximum value of the natural even number n, $n = 2m$, such that there exists a hypothesis class $\mathcal{H}$ with n elements that shatters a set C of $m = \frac{n}{2}$ points? Give an example of such an $\mathcal{H}$ and C. Justify your answer.

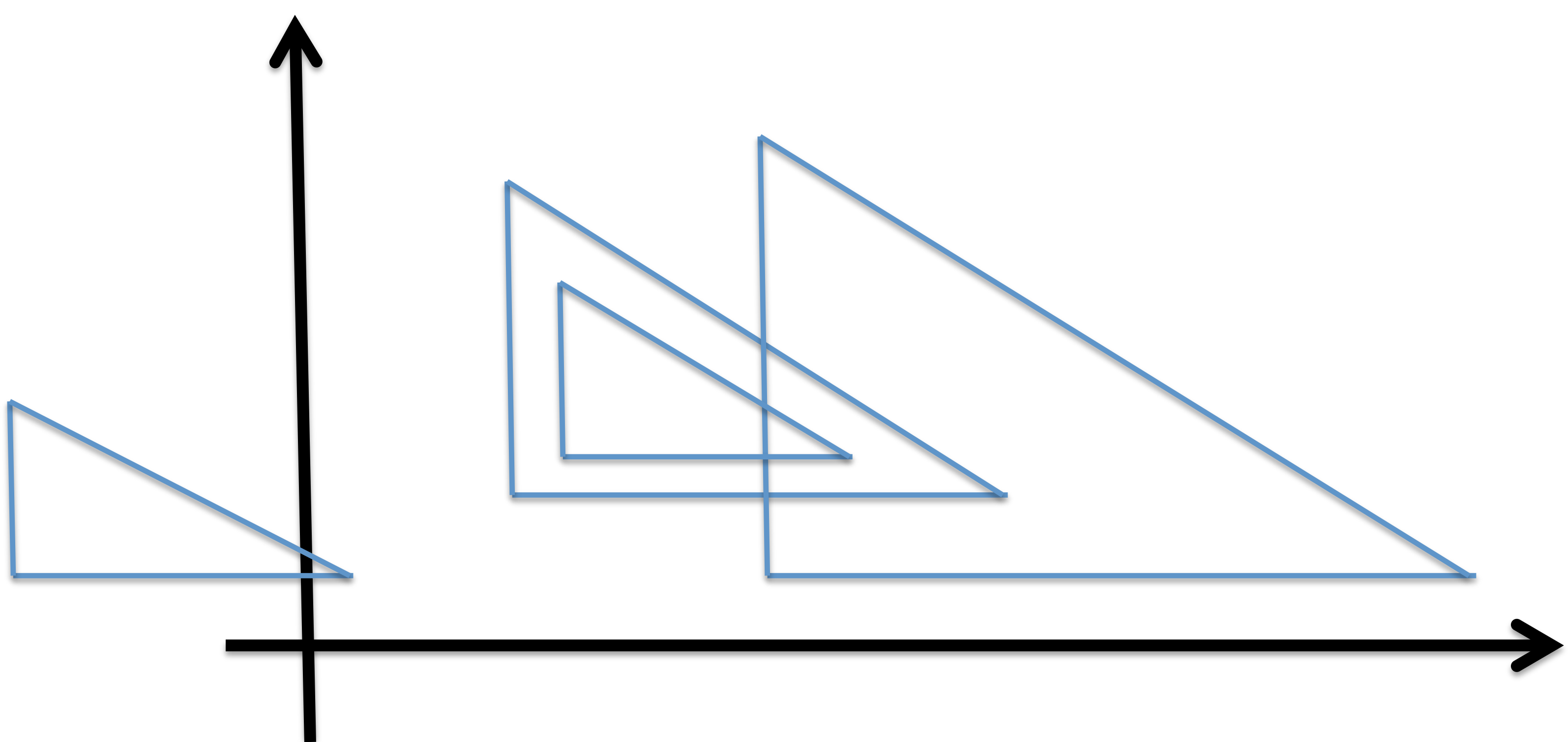
Problem 3

3. **(0.75 points)** Let $\mathcal{X} = \mathbb{R}^2$ and consider $\mathcal{H}$ the set of axis aligned rectangles with the center in origin O(0, 0). Compute the $VCdim(\mathcal{H})$.



Problem 4

4. **(1 point)** Let $\mathcal{X} = \mathbb{R}^2$ and consider $\mathcal{H}_\alpha$ the set of concepts defined by the area inside a right triangle ABC with two catheti AB and AC parallel to the axes (Ox and Oy), and with the ratio AB/AC $= \alpha$ (fixed constant > 0). Consider the realizability assumption. Show that the class $\mathcal{H}_\alpha$ is (ϵ, δ)-PAC learnable by giving an algorithm A and determining an upper bound on the sample complexity $m_H(\epsilon, \delta)$ such that the definition of PAC-learnability is satisfied.



Problem 5

5. **(1.25 points)** Consider $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3$, where:

$$\mathcal{H}_1 = \{h_{\theta_1} : \mathbb{R} \to \{0,1\} \mid h_{\theta_1}(x) = \mathbf{1}_{[x \geq \theta_1]}(x) = \mathbf{1}_{[\theta_1, +\infty)}(x), \theta_1 \in \mathbb{R}\},$$

$$\mathcal{H}_2 = \{h_{\theta_2} : \mathbb{R} \to \{0,1\} \mid h_{\theta_2}(x) = \mathbf{1}_{[x < \theta_2]}(x) = \mathbf{1}_{(-\infty, \theta_2)}(x), \theta_2 \in \mathbb{R}\},$$

$$\mathcal{H}_3 = \{h_{\theta_1,\theta_2} : \mathbb{R} \to \{0,1\} \mid h_{\theta_1,\theta_2}(x) = \mathbf{1}_{[\theta_1 \leq x \leq \theta_2]}(x) = \mathbf{1}_{[\theta_1, \theta_2]}(x), \theta_1, \theta_2 \in \mathbb{R}\}.$$

Consider the realizability assumption.

a) Compute VCdim($\mathcal{H}$).

b) Show that $\mathcal{H}$ is PAC-learnable.

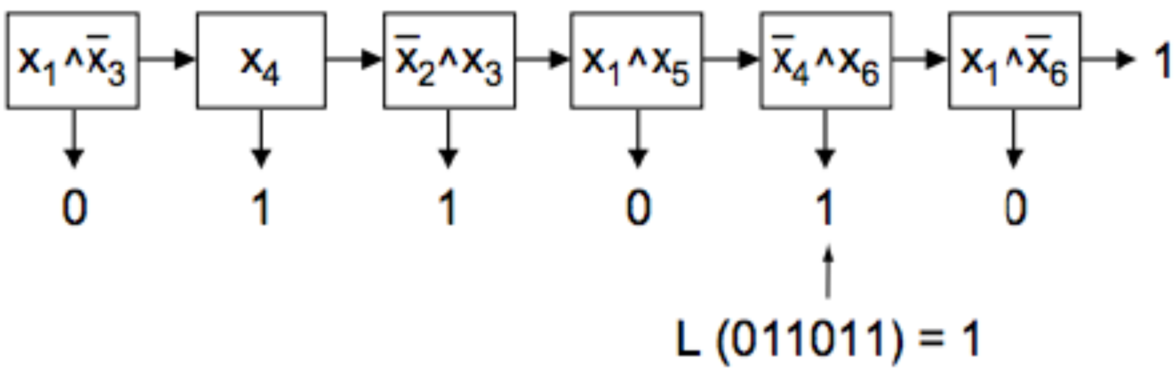
c) Give an algorithm A and determine an upper bound on the sample complexity $m_{\mathcal{H}}(\epsilon, \delta)$ such that the definition of PAC-learnability is satisfied.

Problem 6

6. **(1 point)** A decision list may be thought of as an ordered sequence of if-then-else statements. The sequence of conditions in the decision list is tested in order, and the answer associated with the first satisfied condition is output.

More formally, a *k-decision list* over the boolean variables $x_1, x_2, \ldots, x_n$ is an ordered sequence $L = \{(c_1, b_1), (c_2, b_2), \ldots, (c_l, b_l)\}$ and a bit b, in which each c_i is a conjunction of at most k literals over $x_1, x_2, \ldots, x_n$ and each $b_i \in \{0, 1\}$. For any input $a \in \{0,1\}^n$, the value $L(a)$ is defined to be b_j where j is the smallest index satisfying $c_j(a) = 1$; if no such index exists, then $L(a) = b$. Thus, b is the "default" value in case a falls off the end of the list. We call b_i the bit associated with the condition c_i.

The next figure shows an example of a *2-decision list* along with its evaluation on a particular input.



Show that the VC dimension of 1-decision lists over $\{0,1\}^n$ is lower and upper bounded by linear functions, by showing that there exists $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ such that:

$$\alpha \cdot n + \beta \leq VCdim(\mathcal{H}_{1-\text{decision list}}) \leq \gamma \cdot n + \delta$$

Hint: Show that 1-decision lists over $\{0,1\}^n$ compute linearly separable functions (halfspaces).