Advanced Machine Learning

Bogdan Alexe,

bogdan.alexe@fmi.unibuc.ro

University of Bucharest, 2nd semester, 2019-2020

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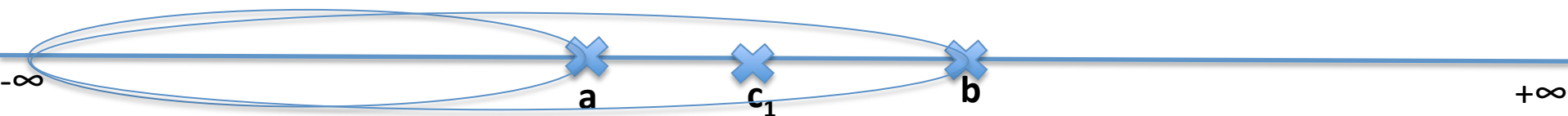
Recap - VCdim($\mathcal{H}_{\text{thresholds}}$)

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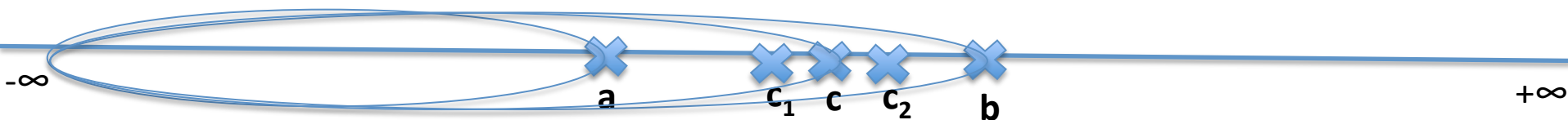$\mathcal{H}_{\text{thresholds}} = \{h_a : \mathbb{R} \to \{0, 1\}, h_a(x) = \mathbf{1}_{[x<a]}, a \in \mathbf{R}\}, |\mathcal{H}_{\text{thresholds}}| = \infty.$

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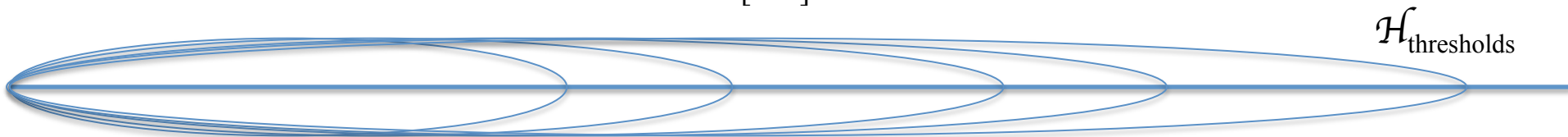
Consider $C = \{c_1\}$. Then $\mathcal{H}_C = \{h : C \to \{0, 1\} | h \in \mathcal{H}\}$ has two elements $\{h_a, h_b\}$ with $a \leq c_1$ and $b > c_1$ so $\mathcal{H}$ shatters C. $\mathcal{H}_C = \{(0),(1)\}, |\mathcal{H}_C| = 2^{|C|} = 2^1$

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

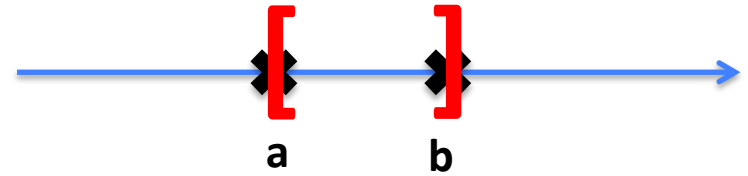

So, **VCdim($\mathcal{H}_{\text{thresholds}}$) = 1**

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

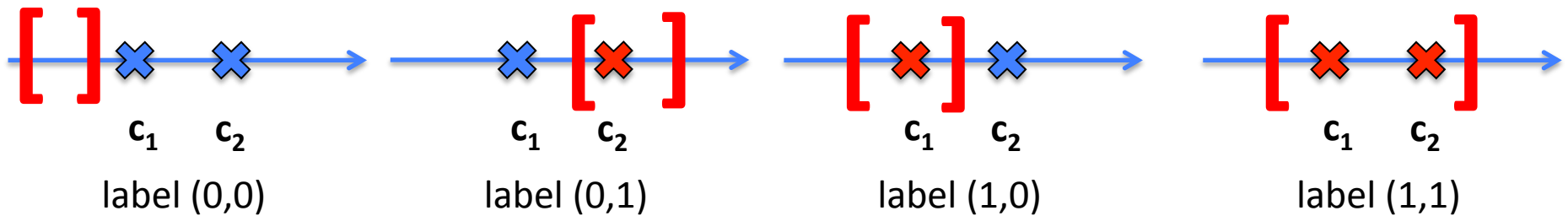Consider $\mathcal{H} = \mathcal{H}_{\text{intervals}}$ be the set of intervals over the real line.

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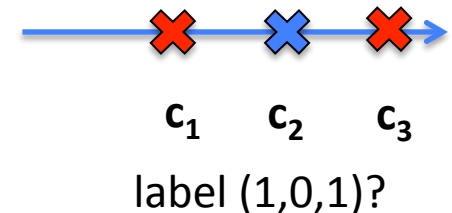
Can also view $\mathcal{H}_{\text{intervals}}$ as:

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

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

label (0,0)   label (0,1)   label (1,0)   label (1,1)

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

label (1,0,1)?

So, **VCdim(H_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(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**

Today's lecture: Overview

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

- Assignment 1

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

1. VCdim($\mathcal{H}$) $\leq \log_2 |\mathcal{H}|$

2. If $\mathcal{H}_1 \subseteq \mathcal{H}_2$ then VCdim($\mathcal{H}_1$) $\leq$ VCdim($\mathcal{H}_2$)

3. If VCdim($\mathcal{H}$) $= \infty$ then $\mathcal{H}$ is not PAC learnable
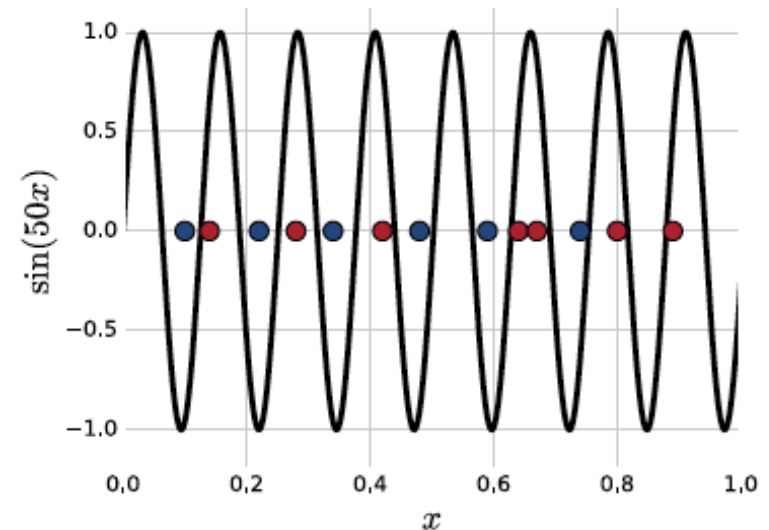
VCdim($\mathcal{H}_{sin}$)

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

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

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

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

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

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

Class 0
Class 1

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:

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

Example of binary representation:

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

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

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

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

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

To prove VCdim($\mathcal{H}_{\text{sin}}$) = ∞, we need to pick n points which are shattered by $\mathcal{H}_{\text{sin}}$, 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:

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

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

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 $x_1, x_2, \ldots, x_n$.

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

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:

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

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

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

To prove VCdim($\mathcal{H}_{\text{sin}}$) $= \infty$, we need to pick n points which are shattered by $\mathcal{H}_{\text{sin}}$, 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 $x_1, x_2, \ldots, x_n$.

$x_1 \quad = \quad 0.\ 0\ 0\ 0\ 0\ \ldots\ 1\ 1$

$x_2 \quad = \quad 0.\ 0\ 0\ 0\ 0\ \ldots\ 1\ 1$

$\quad\quad \ldots$

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

$x_n \quad = \quad 0.\ 0\ 1\ 0\ 1\ \ldots\ 0\ 1$

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, …, $x_n \in [0, 1]$, 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, …, x_n.

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

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

$\quad$ ...

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

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

We conclude that x_1, x_2, …, $x_{n-1,}$ 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($\mathcal{H}_{\text{sin}}$)

We conclude that x_1, x_2, …, 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 VCdim($\mathcal{H}_{\text{sin}}$) = ∞.
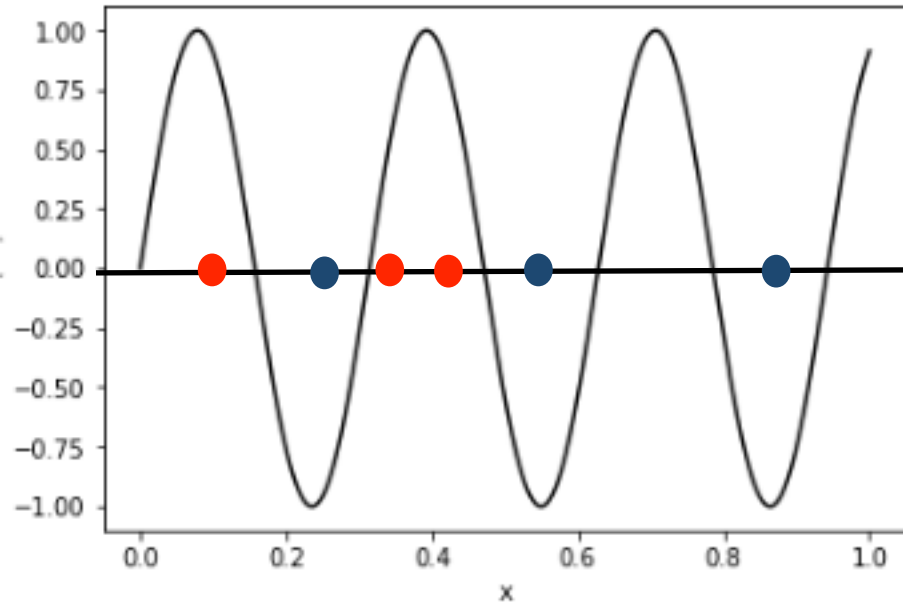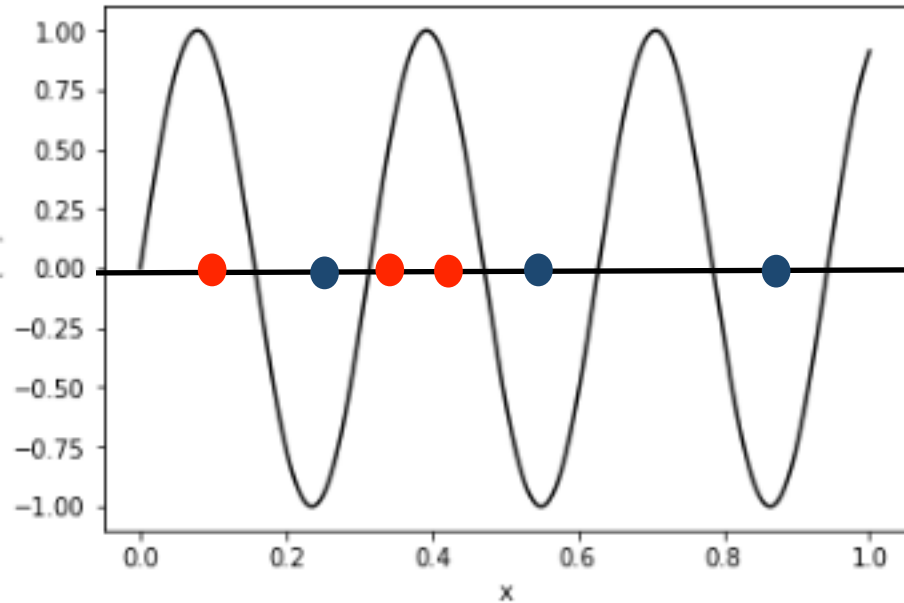

Training points

● Class 0

● Class 1
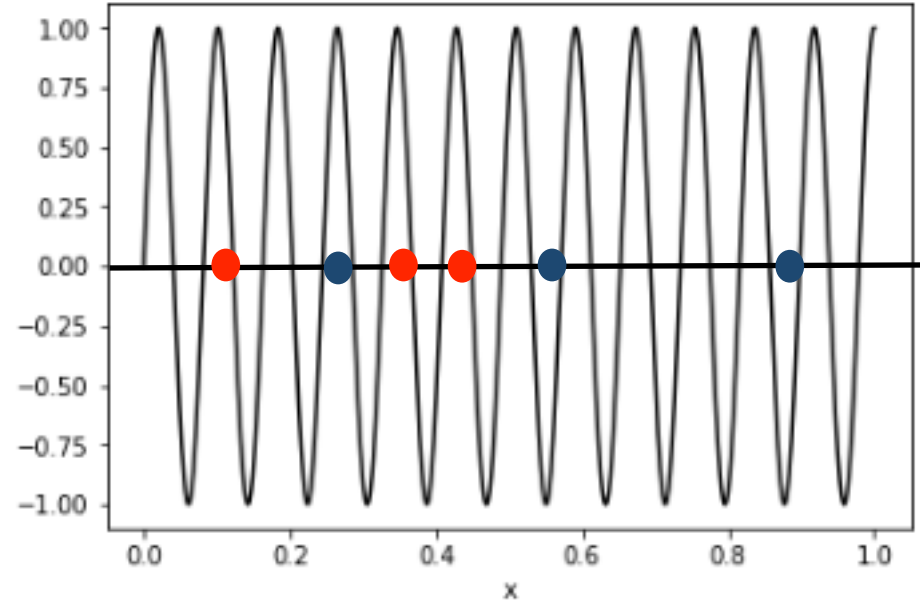
$$h(x) = \lceil \sin(20x) \rceil$$

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

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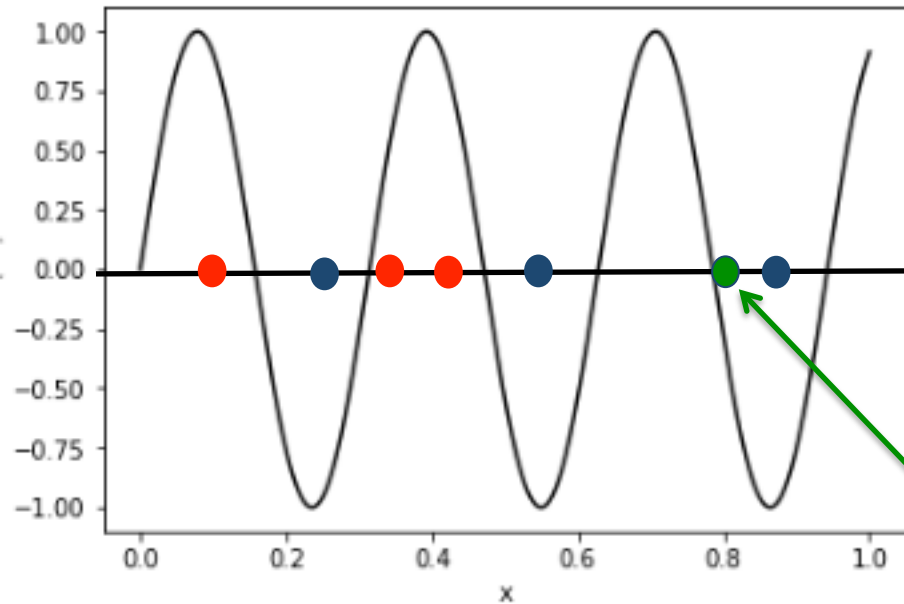

Class 0
Class 1

$$h(x) = \lceil \sin(20x) \rceil$$
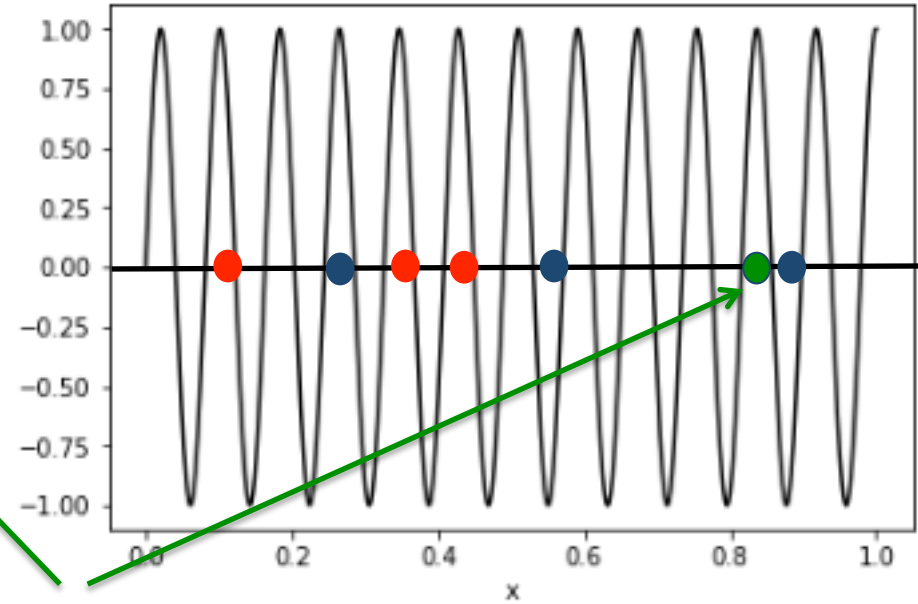
Class 0
Class 1

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

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

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

Test point

● Class 0
● Class 1

● Class 0
● Class 1

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

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

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

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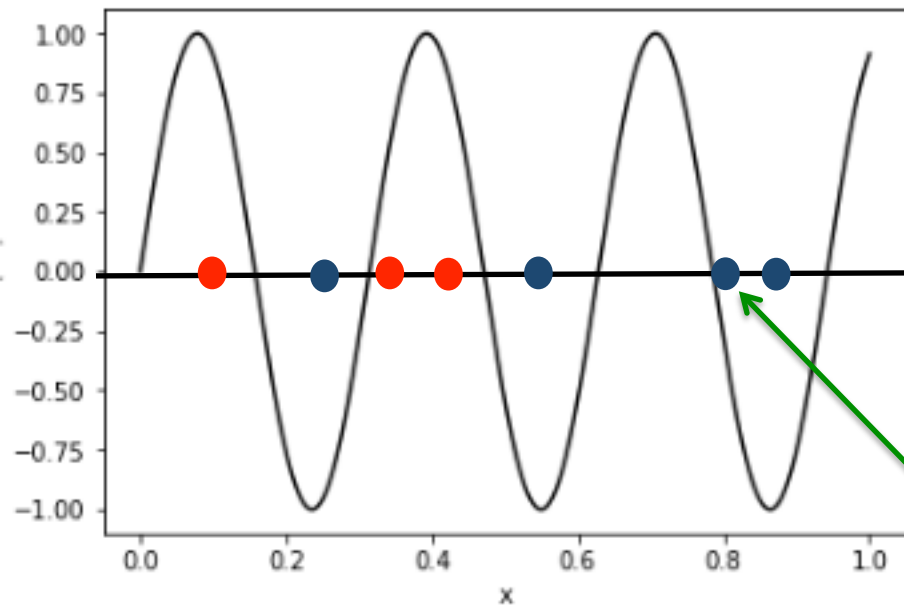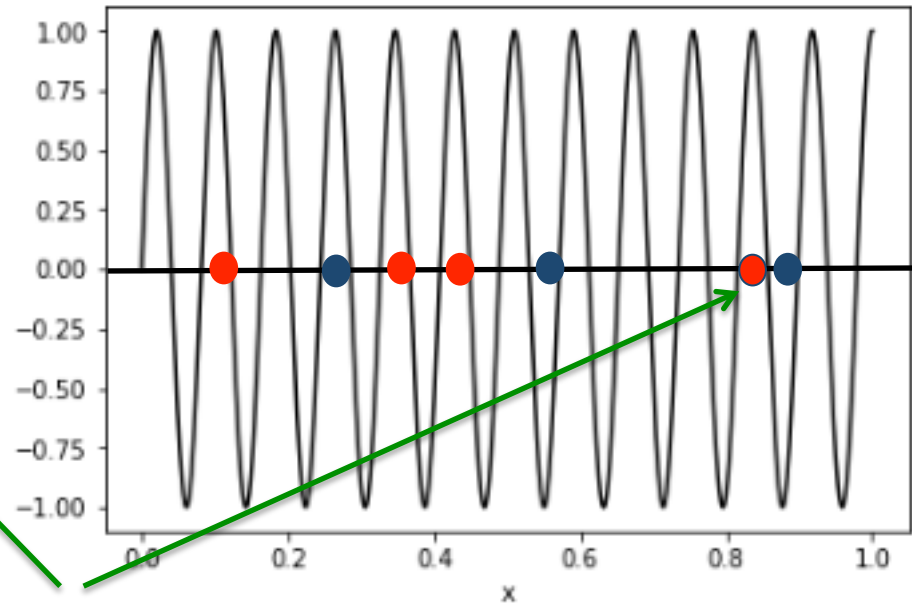

Label assigned

● Class 0

● Class 1

● Class 0

● Class 1

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

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

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

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

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

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

For n = 2 we have:

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

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

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

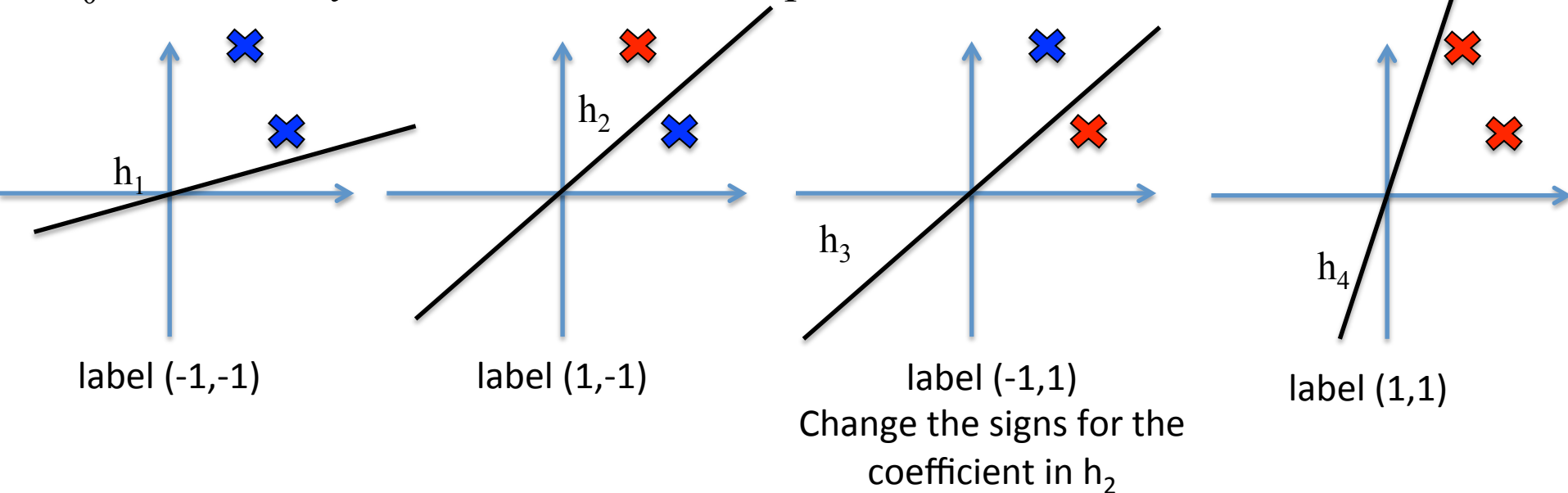What is the VCdim($\mathcal{H}S_0{}^n$)?

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

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

For n = 2 we have:

$\mathcal{HS}_0^2 = \{h_{w1,w2}: \mathbf{R}^2 \rightarrow \{-1, 1\}, h_{w1,w2}(x) = sign(w_1 x_1 + w_2 x_2) \mid (w_1, w_2) \in \mathbf{R}^2\}$

What is the VCdim(HS_0^2) ?

$\mathcal{HS}_0^2$ shatters any set A of two different points.

label (-1,-1) label (1,-1) label (-1,1) label (1,1)

Change the signs for the coefficient in h_2

Does HS_0^2 shatter a set A of three points?

Difficult to reason geometrically… choose the algebraic proof.

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

We will show that VCdim($\mathcal{HS}_0{}^n$) = n.

Proof: 1st part

We first show that VCdim($\mathcal{HS}_0{}^n$) $\geq n$.

We find a set A consisting of n points in $\mathbf{R}^n$ that is shattered by $\mathcal{HS}_0{}^n$.

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

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

We want to proof that $\mathcal{HS}_0{}^n$ shatters A, so that VCdim($\mathcal{HS}_0{}^n$) $\geq n$. This is equivalent to proof that for every B $\subseteq$ A, there is a function $h_B \in \mathcal{HS}_0{}^n$ 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, $h_B(e_i) = \text{sign}(<w, e_i>) = w_i$ 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 $\subseteq$ { 1, 2, …, n +1} the set of positive coefficients a_i and N $\subseteq$ { 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 $P \subseteq \{1, 2, \ldots, n+1\}$ the set of positive coefficients a_i and $N \subseteq \{1, 2, \ldots, n+1\}$ the set of negative coefficients of a_i. So we have:

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

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 $x_i \notin B$.

$$h_B\left(\sum_{i \in P} a_i x_i\right) = \sum_{i \in P} a_i h_B(x_i) > 0$$

$$h_B\left(\sum_{i \in P} a_i x_i\right) = h_B\left(\sum_{j \in N} |a_j| x_j\right) = \sum_{j \in N} |a_j| h_B(x_j) < 0$$

So, this is a contradiction.

Assignment 1

Assignment 1 – good to know

- 4 problems = 3.5 points

- 1 bonus problem = 1 point

- deadline: in 3 weeks time, Friday, 17th of April 2020, 23:59
  - late submission policy: maximum 3 days allowed, -10% (= 0.35 points) for each day
  - upload a pdf written in a scientific editor (Word, Latex, LyX) containing your solution here: ***https://tinyurl.com/AML-2020-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

- 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

1. **(0.5 points)** Consider $\mathcal{H} = \{h_\theta: \mathbb{R} \to \{0,1\}, h_\theta(x) = \mathbf{1}_{[x \geq \theta]}(x), \theta \in \mathbb{R}\} \cup \{h_\theta: \mathbb{R} \to \{0,1\}, h_\theta(x) = \mathbf{1}_{[x < \theta]}(x), \theta \in \mathbb{R}\}$. Compute VCdim($\hat{\mathcal{H}}$).

2. **(0.75 points)** Consider $\mathcal{H}$ to be the class of all centered in origin sphere classifiers in the 3D space. A centered in origin sphere classifier in the 3D space is a classifier h_r that assigns the value 1 to a point if and only if it is inside the sphere with radius $r > 0$ and center given by the origin **O**(0,0,0).
   a. show that the class $\mathcal{H}$ can be $(\epsilon, \delta) - \text{PAC}$ learned by giving the algorithm A and determining the sample complexity $m_H(\epsilon, \delta)$ such that the definition of PAC-learnability is satisfied.
   b. compute VCdim($\hat{\mathcal{H}}$).

Problems 3 and 4

3. **(1 point)** What is the VC-dimension of the set of subsets I_θ of the real line parameterized by a single parameter θ where $I_\theta = [\,\theta, \theta + 1\,] \cup [\theta + 2, \infty)$.

4. **(1.25 points)** An axis aligned square classifier in the plane is a classifier that assigns the value 1 to a point if and only if it is inside a certain square. Formally, given the real numbers a_1, a_2, $r > 0 \in \mathbb{R}$ we define the classifier $h_{(a_1, a_2, r)}$ by

$$h_{(a1,a2,r)}(x_1, x_2) = \begin{cases} 1, \text{ if } a_1 \le x_1 \le a_1 + r,\, a_2 \le x_2 \le a_2 + r \\ \\ 0, \text{ otherwise} \end{cases}$$

The class of all axis aligned squares in the plane is defined as

$$\mathcal{H} = \{h_{(a_1, a_2, r)} : \mathbb{R}^2 \to \{0,1\} | a_1, a_2, r \in \mathbb{R}, r > 0\}$$

Consider the realizability assumption.

a. give a learning algorithm A that return a hypothesis $h_S = A(S)$ consistent with the training set S (h_S has empirical risk 0 on S);

b. find the sample complexity $m_H(\varepsilon, \delta)$ in order to show that $\mathcal{H}$ is PAC – learnable.

c. compute VCdim($\mathcal{H}$).

Bonus Problem

Bonus Problem (1 point)

Compute the VC-dimension of the class of convex d-gons (convex polygons with exactly d sides) in the plane. Provide a detailed proof of your result.

d = 4