VC Dimension: Examples

STATS 303 Statistical Machine Learning

Spring 2022

Lecture 21

recall: VC dimension

Definition (Restriction of $\mathcal{H}$ to C) can be replaced by other $\mathcal{Y}$

Let $\mathcal{H}$ be a class of functions from $\mathcal{X}$ to $\{0,1\}$ and let $C = \{c_1, \cdots, c_m\} \subset \mathcal{X}$ .The restriction of $\mathcal{H}$ to C is the set of functions from C to $\{0,1\}$ that can be derived from $\mathcal{H}$. That is,

$$\mathcal{H}_C = \{(h(c_1), \cdots, h(c_m)) : h \in \mathcal{H}\}$$

Definition (Shattering)

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

recall: VC dimension

- The **VC-dimension** of a hypothesis class $\mathcal{H}$, denoted by $\mathrm{VCdim}(\mathcal{H})$, is the maximal size of a set $C \subset \mathcal{X}$ that can be shattered by $\mathcal{H}$.

- To show $\mathrm{VCdim}(\mathcal{H}) = d$ we need to show that
  1. There exists a set C of size d that is shattered by $\mathcal{H}$
  2. Every set C of size $d+1$ cannot be shattered by $\mathcal{H}$

VC dim: example 1

- Let $\mathcal{H}$ be the set of threshold functions
  - $\mathcal{H} = \{h_a : a \in \mathbb{R}\}$ where $h_a(x) = \mathbf{1}_{\{x<a\}}(x)$
- Take $C = \{c_1\}$ for some c_1
  - Take $a = c_1 + 1$,then $h_a(c_1) = 1$
  - Take $a = c_1 - 1$,then

$\mathcal{H}$ shatters C

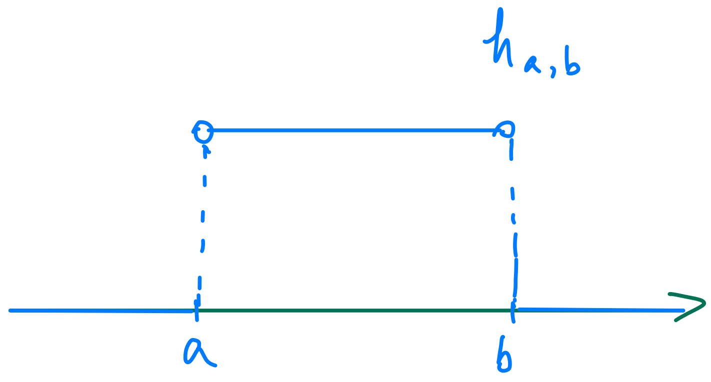
- Consider $C' = \{c_1, c_2\}$ for any $c_1 \leq c_2$
  - No h_a can maps c_1 to 0 and c_1 to 1

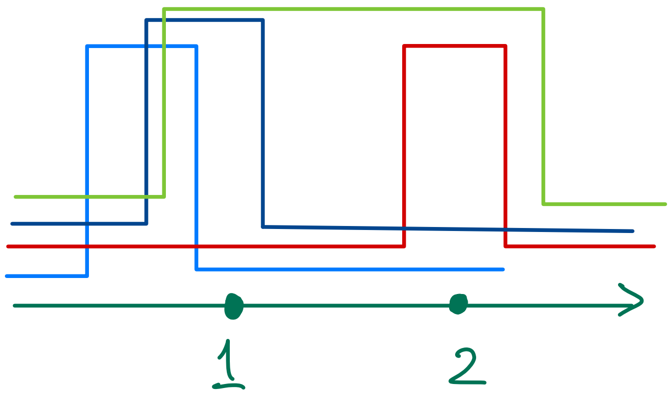
$\mathcal{H}$ does not shatter C'

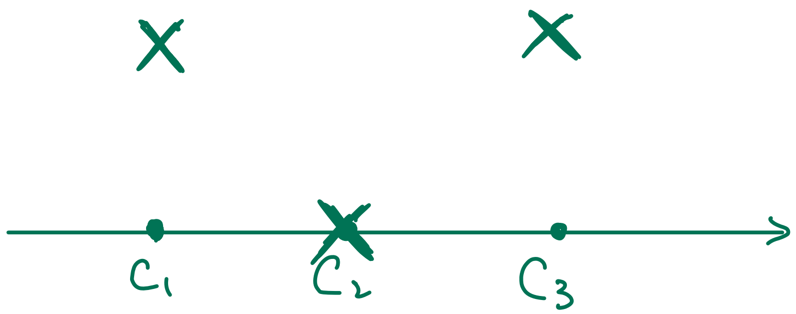
$\mathrm{VCdim}(\mathcal{H}) = 1$

VC dim: example 2

- Let $\mathcal{H}$ be the set of intervals
  - $\mathcal{H} = \{h_{a,b} : a < b \in \mathbb{R}\}$ where $h_{a,b}(x) = \mathbf{1}_{\{a<x<b\}}(x)$
- Take $C = \{1, 2\}$
  - $\mathcal{H}$ shatters C
- Take $C = \{c_1, c_2, c_3\}$ for any $c_1 < c_2 < c_3$
  - $\mathcal{H}$ does not shatter C



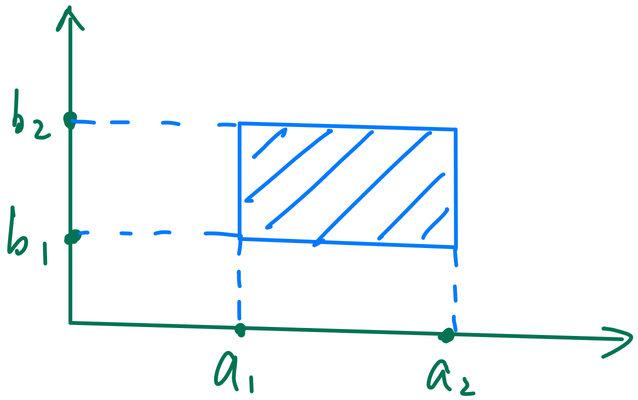




$\mathrm{VCdim}(\mathcal{H}) = 2$

VC dim: example 3

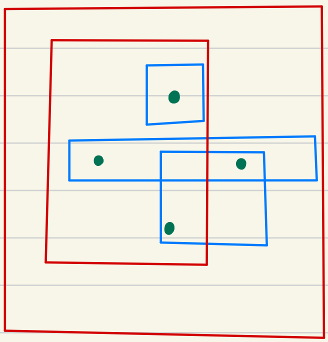
"Class of rectangles"



- Let $\mathcal{H} = \{h_{(a_1,a_2,b_1,b_2)}: a_1 \leq a_2, b_1 \leq b_2\}$ where

$$h_{(a_1,a_2,b_1,b_2)}(x_1, x_2) = \mathbf{1}_{\{a_1 \leq x_1 \leq a_2, b_1 \leq x_2 \leq b_1\}}(x_1, x_2)$$

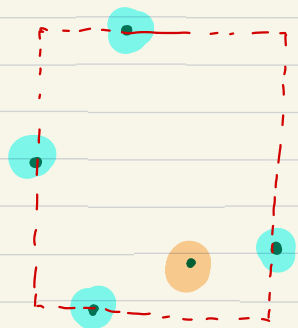
$\text{VC dim}(\mathcal{H}) = 4.$



$\rightarrow \quad VC\dim(\mathcal{H}) \geq 4$

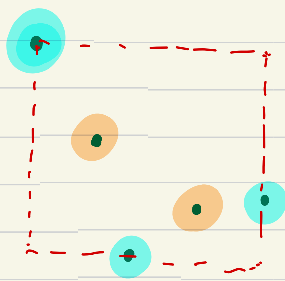
How do we show $VC\dim(\mathcal{H}) < 5$?

Consider any set C that contains 5 points, say



Locate the uppermost point,
the lowermost point,
the leftmost point,
the rightmost point.

The remaining point(s) must lie in the rectangle determined by the above points.



It is impossible to assign value 1 to the points in , while assigning value 0 to the remaining points in .

VC dim: example 4

- Suppose $|\mathcal{H}|$ is finite.
- Any set C with $2^{|C|} > |\mathcal{H}|$ cannot be shattered.

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

VC dim: example 5 - halfspaces

- Let $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Y} = \{-1, +1\}$. The class of halfspaces is defined by

$$\mathcal{H}_d = \{\mathbf{x} \mapsto \mathbf{w}^\mathrm{T}\mathbf{x} + b : \mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}\}$$

- This is the hypothesis class used in, for instance, the support vector machine (SVM)
- The class of homogeneous halfspaces is defined by

$$\mathcal{H}_d' = \{\mathbf{x} \mapsto \mathbf{w}^\mathrm{T}\mathbf{x} : \mathbf{w} \in \mathbb{R}^d \cancel{, b \in \mathbb{R}}\}$$

- What are their VC dimensions?

Thm. The VC dimension of $\mathcal{H}_d'$ is

$$\text{VCdim}(\mathcal{H}_d') = d.$$

Pf. • First, consider $\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_d$

where each $\vec{e}_i$ is the i-th standard basis

given by the one-hot vector $\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i\text{-th}$

• Claim: $C = \{\vec{e}_i\}_{i=1}^d$ is shattered by $\mathcal{H}_d'$.

Pf of Claim:

For any label $y_1, y_2, \cdots, y_d$ (each $y_i \in \{\pm 1\}$),

set $\vec{w} = (y_1, y_2, \cdots, y_d)^T$. Then

$$\vec{w}^T \vec{e}_i = y_i$$

• Next, let $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_{d+1}$ be any collection of $d+1$

vectors in $\mathbb{R}^d$. They have to be linearly dependent:

there exist $a_1, a_2, \cdots, a_{d+1}$, not all being zero.

s.t. $\sum_{i=1}^{d+1} a_i \vec{x}_i = \vec{0}$.

Let $I = \{ i \in [d+1] : a_i > 0 \}$,
$J = \{ j \in [d+1] : a_j < 0 \}$.

Let's first consider the case where both I and J are non-empty. Then $\sum_{i=1}^{d+1} a_i \vec{x}_i = \vec{0}$ implies

$$\sum_{i \in I} a_i \vec{x}_i = \sum_{j \in J} |a_j| \vec{x}_j .$$

Now assume on the contrary that $C = \{ \vec{x}_i \}_{i=1}^{d+1}$ is shattered by H_d'. Then there must be a $\vec{w} \in \mathbb{R}^d$ s.t. $\vec{w}^T \vec{x}_i > 0$ for all $i \in I$ and $\vec{w}^T \vec{x}_j < 0$ for all $j \in J$.

Then

$$0 < \sum_{i \in I} a_i \vec{x}_i^T \vec{w} = \sum_{j \in J} |a_j| \vec{x}_j^T \vec{w} < 0$$

This is a contradiction!

If either I or J is empty, then one of the "$<$" above will be replaced by "$=$". But we still have a contradiction.

Therefore, we have proved by contradiction that $\{\vec{x}_i\}_{i=1}^{d+1}$ cannot be shattered by $\mathcal{H}_d'$.

Combining the two parts (first and next), we conclude that $VC\dim(\mathcal{H}_d') = d$.

<u>Thm</u> The VC dimension of $\mathcal{H}_d$ is

$$\mathrm{VCdim}(\mathcal{H}_d) = d+1.$$

<u>Pf</u>: • <u>First</u>, $C = \{\vec{e}_i\}_{i=1}^{d} \cup \{\vec{0}\} = \{\vec{0}, \vec{e}_1, \vec{e}_2, \cdots, \vec{e}_d\}$

can be shattered by $\mathcal{H}_d$ (exercise).

• <u>Next</u>, assume $\{\vec{x}_i\}_{i=1}^{d+2}$ is shattered by $\mathcal{H}_d$.

Then define $$\vec{\xi}_i = \begin{pmatrix} \vec{x}_i \\ \hline 1 \end{pmatrix} \begin{matrix} \}\mathbb{R}^d \\ \}\mathbb{R} \end{matrix} \in \mathbb{R}^{d+1}$$

Then $\{\vec{\xi}_i\}_{i=1}^{d+2}$ is shattered by $\mathcal{H}'_{d+1}$

since $$\begin{pmatrix} \vec{w} \\ \hline b \end{pmatrix}^T \vec{\xi}_i = \vec{w}^T \vec{x}_i + b$$

This contradicts the fact that $\mathrm{VCdim}(\mathcal{H}'_{d+1}) = d+1$.

overview: the Fundamental Theorem of Statistical Learning (FTSL)

Let $\mathcal{H}$ be a hypothesis class of functions from a domain $\mathcal{X}$ to $\{0,1\}$ and let the loss function be 0-1 loss. Then the following statements are equivalent:

1. $\mathcal{H}$ has UCP
2. Any ERM is a successful agnostic PAC learner for $\mathcal{H}$
3. $\mathcal{H}$ is agnostic PAC learnable
4. $\mathcal{H}$ is PAC learnable
5. Any ERM is a successful PAC learner for $\mathcal{H}$
6. $\mathcal{H}$ has a finite VC-dimension

Questions?

Reference

- *VC dimension*
  - *[S-S] Ch 6.1-6.3*