Machine Learning

VC-Dimension

Fabio Vandin

November 11^{th}, 2022

PAC Learning

Question: which hypothesis classes $\mathcal{H}$ are PAC learnable?

Up to now: if $|\mathcal{H}| < +\infty \Rightarrow \mathcal{H}$ is PAC learnable.

What about $\mathcal{H}$: $|\mathcal{H}| = +\infty$? Not PAC learnable?

We focus on:

- *binary classification*:$\mathcal{Y} = \{0, 1\}$
- 0-1 loss

but similar results apply to other learning tasks and losses.

Restrictions

Definition (Restriction of $\mathcal{H}$ to C)

Let $\mathcal{H}$ be a class of functions from $\mathcal{X}$ to $\{0,1\}$ and let $C = \{c_1, \ldots, c_m\} \subset \mathcal{X}$. The restriction $\mathcal{H}_C$ of $\mathcal{H}$ to C is:

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

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

if $|\mathcal{H}| \geq 1$

$\downarrow$

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

Restrictions

Definition (Restriction of $\mathcal{H}$ to $\mathcal{C}$)

Let $\mathcal{H}$ be a class of functions from $\mathcal{X}$ to $\{0,1\}$ and let $C = \{c_1, \ldots, c_m\} \subset \mathcal{X}$. The restriction $\mathcal{H}_C$ of $\mathcal{H}$ to C is:

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

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

Note: $\mathcal{H}_C$ is the set of functions from C to $\{0,1\}$ that can be derived from $\mathcal{H}$.

VC-dimension and Shattering

Definition (Shattering)

Given $C \subset \mathcal{X}$, $\mathcal{H}$ *shatters* C if $\mathcal{H}_C$ contains all $2^{|C|}$ functions from C to $\{0,1\}$.

if $|C| = m$: $2^{|C|} = 2^{|m|}$

$\qquad\qquad\qquad\qquad\qquad\quad \| $

$\qquad\qquad\qquad\qquad\qquad |\mathcal{H}_C|$

VC-dimension and Shattering

Definition (Shattering)

Given $C \subset \mathcal{X}$, $\mathcal{H}$ *shatters* C if $\mathcal{H}_C$ contains all $2^{|C|}$ functions from C to $\{0, 1\}$.

Definition (VC-dimension)

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

Notes:

- VC = Vapnik-Chervonenkis, that introduced it in 1971
- if $\mathcal{H}$ can shatter sets of arbitrarily large size then we say that $VCdim(\mathcal{H}) = +\infty$;
- if $|\mathcal{H}| < +\infty \Rightarrow VCdim(\mathcal{H}) \leq \log_2 |\mathcal{H}|$ $(\star)$

(HW: prove $(\star)$)

VC-dimension and Shattering

Definition (Shattering)

Given $C \subset \mathcal{X}$, $\mathcal{H}$ *shatters* C if $\mathcal{H}_C$ contains all $2^{|C|}$ functions from C to $\{0, 1\}$.

Definition (VC-dimension)

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

Notes:

- VC $=$ Vapnik-Chervonenkis, that introduced it in 1971
- if $\mathcal{H}$ can shatter sets of arbitrarily large size then we say that $VCdim(\mathcal{H}) = +\infty$;
- if $|\mathcal{H}| < +\infty \Rightarrow VCdim(\mathcal{H}) \leq \log_2 |\mathcal{H}|$

Intuition: the VC-dimension measures the *complexity* of $\mathcal{H}$ ($\approx$ how large a dataset that is perfectly classified using the functions in $\mathcal{H}$ can be)

Example

$h_1(\vec{x}_7) = 1$

$\mathcal{X} = \{\vec{x}_1, \dots, \vec{x}_9\}$

$\mathcal{H} = \{h_1, \dots, h_8\}$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
h_1	0	0	**1**	0	0	0	**1**	0	0
h_2	0	**1**	0	0	0	**1**	0	0	0
h_3	**1**	0	0	0	**1**	**1**	0	0	0
h_4	0	0	0	**1**	**1**	0	0	0	**1**
h_5	0	0	**1**	0	0	0	0	**1**	0
h_6	0	**1**	0	0	0	0	**1**	0	0
h_7	**1**	0	0	0	0	**1**	0	0	0
h_8	0	0	0	0	0	0	0	0	0

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

We need to find the "largest" set $C \subseteq \mathcal{X}$ s.t. C is shattered by $\mathcal{H}$.

VC dimension?

Is the $\mathrm{VCdim}(\mathcal{H}) \geq 1$? $C = \{\vec{x}_1\} \Rightarrow \mathcal{H}_C = \{[0], [1]\} \rightarrow C$ is shattered by $\mathcal{H}$

" " " ≥ 2? $C = \{\vec{x}_1, \vec{x}_5\} \rightarrow \mathcal{H}_C = \{[0,0], [0,1], [1,0], [1,1]\} \Rightarrow$ "

Is the $\mathrm{VCdim}(\mathcal{H}) \geq 3$? No, because we need at least a column with ≥ 4 1's $\Rightarrow$ no set of 3 instances that can be shattered

Note

To show that $VCdim(\mathcal{H}) = d$ we need to show that:

1. $VCdim(\mathcal{H}) \geq d$
2. $VCdim(\mathcal{H}) \leq d$

that translates to

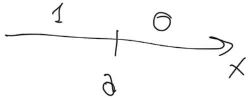
1. there exists a set C of size d which is shattered by $\mathcal{H}$
2. *every* set of size $d + 1$ is not shattered by $\mathcal{H}$

Question: why don't we need to consider sets of size $> d + 1$?

Example: Threshold Functions

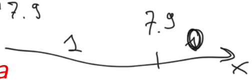
$|\mathcal{H}| = +\infty$

h_a :



$$\mathcal{H} = \{h_a : a \in \mathbb{R}\}$$

$h_{7.9}$



where $h_a : \mathbb{R} \to \{0, 1\}$ is

$$h_a(x) = \mathbb{1}[x < a] = \begin{cases} 1 & \text{if } x < a \\ 0 & \text{if } x \geq a \end{cases}$$

VC $\dim(\mathcal{H}) \geq 1$? YES

instance

h_{a_1}

$\Rightarrow h_{a_1}(c) = 0$

h_{a_2}

$\Rightarrow h_{a_2}(c) = 1$

Example: Threshold Functions

$$\mathcal{H} = \{h_a : a \in \mathbb{R}\}$$

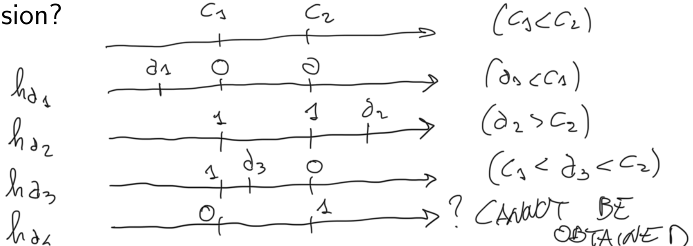
where $h_a : \mathbb{R} \to \{0, 1\}$ is

$$h_a(x) = \mathbb{1}[x < a] = \begin{cases} 1 & \text{if } x < a \\ 0 & \text{if } x \geq a \end{cases}$$

VC-dimension?

$\Rightarrow VCdim(\mathcal{H}) \leq 1$

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



Example: Intervals

$$\mathcal{H} = \{h_{a,b} : a, b \in \mathbb{R}, a < b\}$$

where $h_{a,b} : \mathbb{R} \to \{0, 1\}$ is

$$h_{a,b}(x) = \mathbb{1}[x \in (a, b)] = \begin{cases} 1 & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

VC-dimension?

Example: Axis Aligned Rectangles

$$\mathcal{H} = \{h_{(a_1,a_2,b_1,b_2)} : a_1, a_2, b_1, b_2 \in \mathbb{R}, a_1 \leq a_2, b_1 \leq b_2\}$$

$$h_{(a_1,a_2,b_1,b_2)}(x_1, x_2) = \begin{cases} 1 & \text{if } a_1 \leq x_1 < a_2, b_1 \leq x_2 \leq b_2 \\ 0 & \text{otherwise} \end{cases}$$

VC-dimension?

Example: Convex Sets

Model set $\mathcal{H}$ such that for $h \in \mathcal{H}$, $h : \mathbb{R}^2 \to \{0, 1\}$ with

$$h(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S \\ 0 & \text{otherwise} \end{cases}$$

where S is a convex subset of $\mathbb{R}^2$

VC-dimension?