MLT Homework 4

Ana Borovac
Jonas Haslbeck
Bas Haver

October 4, 2018

Question 1

We have shown that for a finite hypothesis class $\mathcal{H}$, VCdim$(\mathcal{H}) \leq \lfloor \log(|\mathcal{H}|) \rfloor$. However, this is just an upper bound. The VC-dimension of a class can be much lower than that.

Subquestion 1.1

Find an example of a class $\mathcal{H}$ of functions over the real interval $\mathcal{X} = [0, 1]$ such that $\mathcal{H}$ is infinite while VCdim$(\mathcal{H}) = 1$.

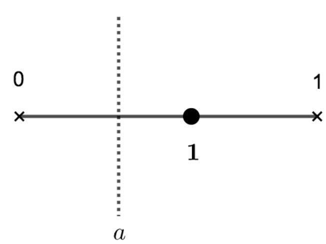
Solution

Let's define hypothesis class as:

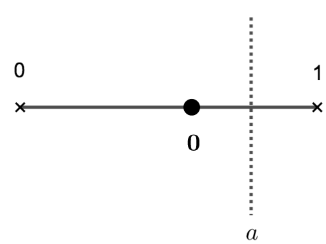
$$\mathcal{H} = \{h_a : \ a \in [0, 1]\}$$

$$h_a(x) = \begin{cases} 1; & x \geq a \\ 0; & x < a \end{cases}$$

From definition we know $|\mathcal{H}| = \infty$. Now we need to prove that VCdim$(\mathcal{H}) = 1$.



(a) If point is labeled "1".



(b) If point is labeled "0".

Figure 1: Proof that VCdim$(\mathcal{H}) \geq 1$.

- VCdim($\mathcal{H}$) ≥ 1: The proof we can see from the figure **??**.
- VCdim($\mathcal{H}$) ≤ 1: From the figure **??** it is seen that hypothesis class $\mathcal{H}$ does not shatter a set of two points (no matter how we position them).

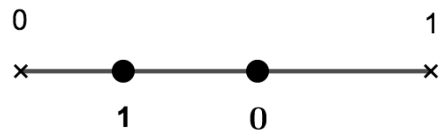


Figure 2: The problem we have when trying to shatter a set of two points.

Subquestion 1.2

Give an example of a finite hypothesis class $\mathcal{H}$ over domain $\mathcal{X} = [0, 1]$, where VCdim($\mathcal{H}$) $= \lfloor \log_2(|\mathcal{H}|) \rfloor$

Solution

Let's define hypothesis class:

$$\mathcal{H} = \{h_0, h_1\}$$

where $h_0(x) = 0$ $(\forall x)$ and $h_1(x) = 1$ $(\forall x)$. We would like to prove that VCdim($\mathcal{H}$) $= \lfloor \log_2(|\mathcal{H}|) \rfloor = \lfloor \log_2(2) \rfloor = 1$.

- VCdim($\mathcal{H}$) ≥ 1: If we want to label a $x \in [0, 1]$ as 1, we pick h_1 as hypothesis, otherwise we pick h_0. So, VCdim($\mathcal{H}$) ≥ 1.
- VCdim($\mathcal{H}$) ≤ 1: Let say that we have a set of two points. If we want to label one of the point with 1 and the other with 0, there does not exist a hypothesis in hypothesis class which can label two points differently.

We can conclude that VCdim($\mathcal{H}$) $= 1$.

Question 2

6.8

Solution

Question 3

Let $\mathcal{H}$ be the class of signed intervals, that is, $\mathcal{H} = \{h_{a,b,s} : a \leq b, s \in \{-1, 1\}\}$ where

$$h_{a,b,s}(x) = \begin{cases} s & \text{if } x \in [a, b] \\ -s & \text{if } x \notin [a, b] \end{cases}$$

Calculate VCdim($\mathcal{H}$).

Solution

Claim: VCdim($\mathcal{H}$) = 3.

- VCdim($\mathcal{H}$) ≥ 3: On figure **??** it is seen that set of three points can be shattered. Furthermore VCdim($\mathcal{H}$) ≥ 3.

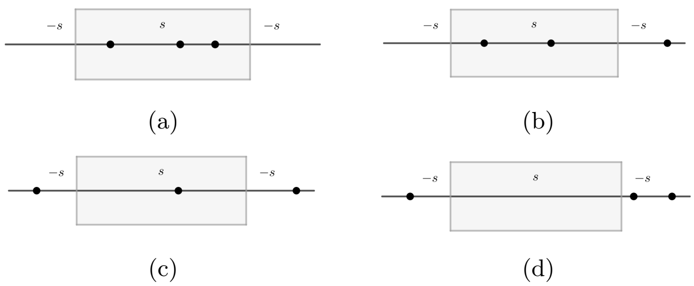


Figure 3: Proof that VCdim($\mathcal{H}$) ≥ 3.

- VCdim($\mathcal{H}$) ≤ 3: There does not exist a hypothesis $h \in \mathcal{H}$ that lables the situation on figure **??**. From here we can conclude VCdim($\mathcal{H}$) ≤ 3.

Figure 4: Proof that VCdim($\mathcal{H}$) ≤ 3.

Question 4

VC of union: *Let $\mathcal{H}_1, \ldots, \mathcal{H}_r$ be hypothesis classes over some fixed domain set $\mathcal{X}$. Let $d = \max_i$ VCdim($\mathcal{H}_i$) and assume for simplicity that $d \geq 3$.*

Subquestion 4.1

Prove that

$$VCdim(\cup_{i=1}^{r} \mathcal{H}_i) \leq 4d \log(2d) + 2\log(r).$$

Solution