MLT Homework 4

Ana Borovac
Jonas Haslbeck
Bas Haver

October 7, 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($\mathcal{H}$) $= 1$.

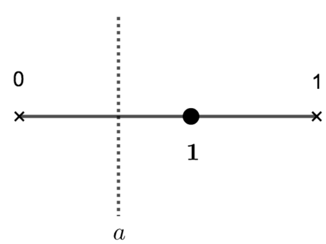
Solution

Let's define hypothesis class as:

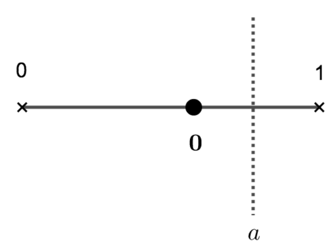
$$\mathcal{H} = \{h_a : \ a \in [0,1]\}$$

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

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($\mathcal{H}$) ≥ 1.

- VCdim($\mathcal{H}$) ≥ 1: The proof we can see from the figure 1.
- VCdim($\mathcal{H}$) ≤ 1: From the figure 2 it is seen that hypothesis class $\mathcal{H}$ does not shatter a set of two points (no matter how we position them).

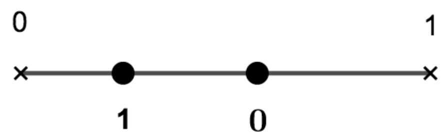


Figure 2: The problem we have when trying to shatter a set of two points.

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

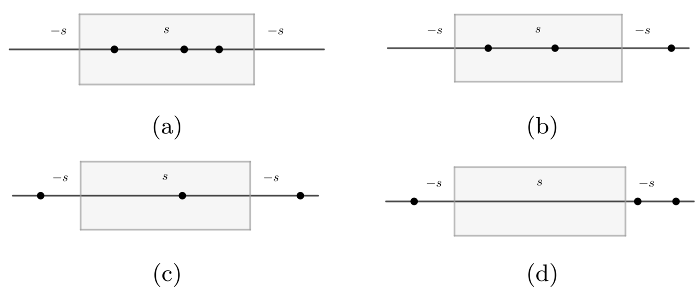


Figure 3: Proof that VCdim($\mathcal{H}$) $\geq$ 3.

- VCdim($\mathcal{H}$) $\leq$ 3: There does not exist a hypothesis $h \in \mathcal{H}$ that lables the situation on figure 4. From here we can conclude VCdim($\mathcal{H}$) $\leq$ 3.

Figure 4: Proof that VCdim($\mathcal{H}$) $\leq$ 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

First, denote a union class as $\mathcal{H}_\cup = \cup_{i=1}^r \mathcal{H}_i$. Second, assume that $\text{VCdim}(\mathcal{H}_\cup) = k$ and therefore $\mathcal{H}_\cup$ shatters a set of k elements . Furthermore, the union class can produce all 2^k possible labelings on these elements.

Let's recall Sauer's lemma: Let $\mathcal{H}$ be a hypothesis class with $\text{VCdim}(\mathcal{H}) \leq d < \infty$. Then for all m,

$$\Pi_{\mathcal{H}}(m) \leq m^d$$

From our assumption it follows:

$$\Pi_{\mathcal{H}_\cup}(k) = 2^k$$

The definition of shatter function gives as the following inequality:

$$\Pi_{\mathcal{H}_\cup}(k) \leq \Pi_{\mathcal{H}_1}(k) + \cdots + \Pi_{\mathcal{H}_r}(k)$$

Now, we can use Sauer's lemma on each summand:

$$2^k = \Pi_{\mathcal{H}_\cup}(k) \leq \Pi_{\mathcal{H}_1}(k) + \cdots + \Pi_{\mathcal{H}_r}(k) \leq \underbrace{k^d + \cdots k^d}_{r} = rk^d$$

If we use logarithm on the inequality, we get:

$$k \leq d \log k + \log r \tag{1}$$

In the next step we are going to use Lemma A.2 from the book, which says: Let $a \geq 1$ and $b > 0$. Then $x \geq 4a \log(2a) + 2b \implies x \geq a \log(x) + b$.

Let's assume that $\text{VCdim}(\mathcal{H}_\cup) > 4d \log(2d) + 2\log(r)$. From our first assumption we get:

$$k > 4d \log(2d) + 2\log(r)$$

Now, we can use Lemma A.2 (where $a = d \geq 3$, $b = \log r > 0$):

$$k > d \log k + \log r$$

We got into a contradiction with (1), this means that our assumption was not correct and it holds:

$$\text{VCdim}(\mathcal{H}_\cup) \leq 4d \log(2d) + 2\log(r)$$

Subquestion 4.2

Prove that for $r = 2$ it holds that

$$VCdim(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 2d + 1$$

Solution

This question was solved with the help of [1].

As same as before:

$$\Pi_{\mathcal{H}_1 \cup \mathcal{H}_2}(m) \leq \Pi_{\mathcal{H}_1}(m) + \Pi_{\mathcal{H}_2}(m)$$

Now we use Sauer's lemma:

$$\Pi_{\mathcal{H}_1 \cup \mathcal{H}_2}(m) \leq \Pi_{\mathcal{H}_1}(m) + \Pi_{\mathcal{H}_2}(m) \leq \sum_{i=0}^{d} \binom{m}{i} + \sum_{i=0}^{d} \binom{m}{i}$$

If we use the fect $\binom{m}{i} = \binom{m}{m-i}$, we get:

$$\begin{aligned}
\sum_{i=0}^{d} \binom{m}{i} + \sum_{i=0}^{d} \binom{m}{i} &= \sum_{i=0}^{d} \binom{m}{i} + \sum_{i=0}^{d} \binom{m}{m-i} \\
&= \sum_{i=0}^{d} \binom{m}{i} + \sum_{i=m-d}^{d} \binom{m}{i} \\
&= \underbrace{\binom{m}{0} + \cdots + \binom{m}{d}}_{d+1} + \underbrace{\binom{m}{m-d} + \cdots + \binom{m}{m}}_{d+1}
\end{aligned}$$

If $m > 2d + 1$:

$$\binom{m}{0} + \cdots + \binom{m}{d} + \binom{m}{m-d} + \cdots + \binom{m}{m} \leq \sum_{i=0}^{m} \binom{m}{i} - \binom{m}{d+1} < 2^m$$

Let's sum up what we just calculated:

$$\Pi_{\mathcal{H}_1 \cup \mathcal{H}_2}(m) < 2^m$$

So, if $m > 2d + 1$ the set with m elements can not be shattered, therefore:

$$\text{VCdim}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 2d + 1$$

References

[1] Mehryar Mohri. Solution assignment 2. `cs.nyu.edu/~mohri/ml/ml10/sol2.pdf`, 2010. Online; accessed 7 October 2018.