University of Salzburg | Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **C**

VC-Dimension

Exercise 1. 3 P.

Show the following monotonicity property of the VC dimension: For every two hypothesis classes $\mathcal{H}', \mathcal{H}$, the following holds:

$$\mathcal{H}' \subseteq \mathcal{H} \Rightarrow \text{VCdim}(\mathcal{H}') \leq \text{VCdim}(\mathcal{H})$$

Exercise 2. 3 P.

Let $\mathcal{H}_{\text{si}}$ be the class of **signed intervals**:

$$\mathcal{H}_{\text{si}} = \{h_{a,b,s} : a \leq b, s \in \{+1, -1\}\}$$

where

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

What is the VC-dimension $\text{VCdim}(\mathcal{H}_{\text{si}})$? Do not just list the VC dimension, but provide a rigorous argumentation.

Exercise 3. 4 P.

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

1. Find an example of a class $\mathcal{H}$ of functions over the real interval $\mathcal{X} = [0,1]$ such that $\mathcal{H}$ is <u>infinite</u> while $\text{VCdim}(\mathcal{H}) = 1$. In other words, you have an infinite number of hypothesis, but you can only shatter sets of size ≤ 1 (Note: the empty set is always shattered).

2. Give an example of a finite hypothesis class $\mathcal{H}$ over the domain $\mathcal{X} = [0,1]$, where

$$\text{VCdim}(\mathcal{H}) = \lfloor \log_2(|\mathcal{H}|) \rfloor \enspace .$$

Both items give 2 points each.

Exercise 4. 6 P.

Let $\mathcal{H}_1, \ldots, \mathcal{H}_r$ be hypothesis classes over some fixed domain $\mathcal{X}$. Let

$$d = \max_i \text{VCdim}(\mathcal{H}_i)$$

and assume, for simplicity, that $d \geq 3$. Prove (using Lemma 1) that

$$\text{VCdim}(\cup_i \mathcal{H}_i) \leq 4d \log_2(2d) + 2\log_2(r) \enspace .$$

Lemma 1. *Let $a \geq 1$ and $b > 0$. Then*

$$x \geq 4a \log(2a) + 2b \Rightarrow x \geq a \log(x) + b$$

It follows that a necessary condition for $x < a \log(x) + b$ to hold is $x < 4a \log(2a) + 2b$.

Proof. *(Note: This is not fully rigorous, but suffices for our purposes).* If we can show that (1) $x \geq 2a\log(x)$ and (2) $x \geq 2b$, then

$$2x \geq 2a\log(x) + 2b$$

hence $x \geq a\log(x) + b$ and we are done. First, as $a \geq 1, b > 0$, we have

$$x \geq \underbrace{4a\log(2a)}_{>0} + \underbrace{2b}_{>0}$$

and hence $x \geq 2b$. Second, we also get that $x \geq 4a\log(2a)$. The question then is if $x \geq 4a\log(2a)$ implies $x \geq 2a\log(x)$. Equivalently, we can ask if $x \geq 2a\log(a)$ implies $x \geq a\log(x)$. To answer this, we look at solutions of $f(x) = x - a\log(x) > 0$ in terms of x. In particular, for $a \in (0, \sqrt{e}]$, the implication holds trivially. A simple check with Mathematica shows that, in fact, $x \geq 2a\log(a)$ implies $x \geq a\log(x)$ and our claim follows. $\square$

Total #points: 16 P.