MLT Homework 3

Ana Borovac
Jonas Haslbeck
Bas Haver

September 27, 2018

Question 1

Show the following monotonicity property of VC-dimension: For every two hypothesis classes if $\mathcal{H}' \subset \mathcal{H}$ then VCdim$(\mathcal{H}') \leq$ VCdim$(\mathcal{H})$.

Solution

Let assume:

$$C' \subseteq \mathcal{H}' : \text{ VCdim}(\mathcal{H}') = |C'|$$

We also know:

$$C' \subseteq \mathcal{H}' \subset \mathcal{H} \quad \implies \quad C' \subset \mathcal{H}$$

From that we can conclude:

$$\text{VCdim}(\mathcal{H}) \geq |C'| = \text{VCdim}(\mathcal{H}') \quad \implies \quad \text{VCdim}(\mathcal{H}') \leq \text{VCdim}(\mathcal{H})$$

Question 2

Given some finite domain set, $\mathcal{X}$, and a number $k \leq |\mathcal{X}|$, figure out the VC-dimension of each of the following classes (and prove your claims).

Subquestion 2.1

$\mathcal{H}^{\mathcal{X}}_{=k} = \{h \in \{0,1\}^{\mathcal{X}} : |\{x : h(x) = 1\}| = k\}$. *That is, the set of all functions that assign the value 1 to exactly k elements of $\mathcal{X}$.*

Solution

Claim: $\text{VCdim}(\mathcal{H}_{=k}^{\mathcal{X}}) = \min\{k, |\mathcal{X}| - k\}$.
Proof: Let C denote subset of $\mathcal{X}$ such that $|C| = \min\{k, |\mathcal{X}| - k\}$.

In the "worst" case we want to label all elements of C with 0. In order to do that we need at least k elements in a set $\mathcal{X} \setminus C$. From there we can conclude that the biggest subset of $\mathcal{X}$ which can be shattered by $\mathcal{H}_{=k}^{\mathcal{X}}$ is of size $|\mathcal{X}| - k$.

On the other hand if we want to label all elements of C with 1, the size of C can not be bigger than k.

If we combine both explanations above, we get:

$$\text{VCdim}(\mathcal{H}_{=k}^{\mathcal{X}}) = \min\{k, |\mathcal{X}| - k\}$$

Subquestion 2.2

$\mathcal{H}_{at-most-k} = \{h \in \{0,1\}^{\mathcal{X}} : |\{x : h(x) = 1\}| \leq k \;\; or \;\; |\{x : h(x) = 0\}| \leq k\}$.

Solution

Claim: $\text{VCdim}(\mathcal{H}_{at-most-k}) = \min\{2k, |\mathcal{X}|\}$
Proof: Let's analyse the case that we do not want to happen. We do not want to have a subset C where on of labelings contains $> k$ elements that are labeled 0 and $> k$ elements labeled 1. So, subsets bigger that $2k$ can not be shattered by $\mathcal{H}_{at-most-k}$. From this and the fact that VC-dimension can not bi bigger than $|\mathcal{X}|$ we conclude:

$$\text{VCdim}(\mathcal{H}_{at-most-k}) = \min\{2k, |\mathcal{X}|\}$$

Question 3

Solution

Question 4

Solution