University of Salzburg Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **C**

Exercise 1. 3 P.

In the NFL theorem from the lecture, we assumed $m < |\mathcal{X}|/2$. If one would now assume $|X| \geq km$ with integer $k \geq 2$ and let $m < |\mathcal{X}|/k$, we would get

$$\mathbb{E}_{S \sim \mathcal{D}^m}[L_{\mathcal{D}}(A(S))] \geq \frac{1}{2} - \frac{1}{2k} \ ,$$

i.e., the lower bound of $1/4$ is replaced by $1/2 - 1/2k$.

In the NFL proof, which part has to be modified (and how) to show this? What happens as k gets large?

Exercise 2. 3 P.

Consider the hypothesis class of intervals $[a, b]$ on the real line ($\mathbb{R}$), i.e., $\mathcal{H} = \{h_{a,b} : a < b, a \in \mathbb{R}, b \in \mathbb{R}\}$ with $h_{a,b}(x) = \mathbf{1}_{x \in [a,b]}$. That is, for a given point x, a hypothesis $h_{a,b}$ returns 1 if the point is inside $[a, b]$ and 0 otherwise. What is the size of the largest set that is *shattered* and what is $|\mathcal{H}|$? Provide an argument for your answer.

Exercise 3. 5 P.

Claim: Let $A \subseteq \mathcal{X}$ and let L be an arbitrary set of labelings (from $\{-1, +1\}$) of A. Then, L shatters *at least* $|L|$ subsets of A. Proof this claim.

Hint: The statement means that there are at least $|L|$ distinct sets $A' \subseteq A$ such that we can label A' in all $2^{|A'|}$ distinct ways using our labelings from L. Establish the statement by induction on $|L|$, i.e., first consider the base case $|L| = 1$ (remembering that the empty set is always shattered), and then look at the case $|L| > 1$. For the latter step, the idea is to partition the labelings into to sets such that they differ on a point $x \in A$. Hence, that point can not be in the shattered sets that correspond to the two partitions. Then apply the induction hypothesis and go on from there ...

Total #points: 11 P.