University of Salzburg | Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **C**

Exercise 1. 5 P.

In the NFL lecture, we enumerated all $k = (2m)^m$ possible training sets of size m (out of the set C of size $2m$), i.e., $S_1, \ldots, S_k$. If labeled by f_i, we added the superscript i, e.g., $S_1^i, \ldots, S_k^i$. Later in the proof of the NFL theorem, we wrote

$$\mathbb{E}_{S\sim\mathcal{D}_i^m}[L_{\mathcal{D}_i}(A(S))] = \frac{1}{k}\sum_{j=1}^{k} L_{\mathcal{D}_i}(A(S_j^i))$$

for the expected generalization error wrt. $\mathcal{D}_i$. As an argument for why we can actually do this, we remarked that "…all sets are equally likely" (meaning the sets $S_1^i, \ldots, S_k^i$ are equally likely to be drawn). *Where do we actually see this? Provide a (formal) argument, remembering how we constructed the $\mathcal{D}_i$'s.*

Exercise 2. 5 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?

Total #points: 10 P.