Assignment 4

Due: 27.06.2024, 23:59

Points: 14

The solutions have to be handed in via Moodle. We do not accept late submissions.

We would recommend using LaTeX for writing your submission but also accept handwritten solutions, but please note that if we can not read or understand it, we cannot grade it.

To get full points, always provide the steps in your derivation/proofs and make clear when/how you use known results, for example, from the lecture (e.g. already proven concentration inequalities).

Exercise 4.1: Validation

This exercise provides an example where leave-one-out error is a poor estimate of the generalisation error. Consider the 0-1 loss and assume that the distribution $\mathcal{D}$ is such that

$$\mathbb{P}(y = 1|x) = \mathbb{P}(y = 0|x) = \frac{1}{2} \qquad \text{for every } x \in \mathcal{X}$$

Given a training sample $S = \{(x_1, y_1), \ldots, (x_m, y_m)\} \sim \mathcal{D}^m$, we consider the following classification rule

$$h_S(x) := \begin{cases} 0 & \text{if } \sum_{i=1}^{m} y_i \text{ is odd,} \\ 1 & \text{if } \sum_{i=1}^{m} y_i \text{ is even} \end{cases} \qquad \text{for every } x$$

1. Compare the expected generalisation error for h_S, $\mathbb{E}_S[L_{\mathcal{D}}(h_S)]$, and the expected leave-one-out error, $\mathbb{E}_S[L_{loo}(h_S)]$.
2. Now compute $\mathbb{E}_S\big[|L_{\mathcal{D}}(h_S) - L_{loo}(h_S)|\big]$, and comment on your result.
3. Give reasons behind the similarities / dissimilarities in these findings.

(2+3+1=6 points)

Exercise 4.2: Rademacher Complexity

For any $p \geq 1$ let B_p denote the set

$$B_p = \Big\{x \in \mathbb{R}^d : \|x\|_p \leq 1\Big\}$$

Prove that the Rademacher complexity of B_p satisfies

$$\mathcal{R}(B_p) = d^{-1/p}$$

(3 points)

Exercise 4.3: Universality of the Gaussian kernel

Let $\mathcal{X} = \{x \in \mathbb{R}^p : \|x\|_2 \leq 1\}$. In the lecture, we saw that the exponential kernel $k(x, y) = \exp(\langle x, y\rangle)$ is universal on $\mathcal{X}$. Conclude that the Gaussian kernel $k(x, y) = \exp(-\frac{1}{2}\|x - y\|^2)$ is also universal on $\mathcal{X}$.

Hint: Start with approximating a function f using the exponential kernel. From there, work your way to the Gaussian kernel.

(5 points)