Statistical Foundations of Learning - Assignment 4

CIT4230004 (Summer Semester 2024)

Exercise 4.1: Validation

1. Compare the expected generalization error for h_S, $\mathbb{E}_S[L_D(h_S)]$, and the expected leave-one-out error, $\mathbb{E}_S[L_{\mathbf{loo}}(h_S)]$.

Given:

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

Classification rule:

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

For the generalization error $L_D(h_S)$:

$$L_D(h_S) = P(h_S(x) \neq y)$$

Since $y = 1$ or $y = 0$ with equal probability, the generalization error is:

$$L_D(h_S) = \frac{1}{2}$$

For the leave-one-out error $L_{\text{loo}}(h_S)$: - Each leave-one-out classifier is trained on $m - 1$ samples. - If $\sum_{i=1}^m y_i$ is odd, then $\sum_{i \neq j} y_i$ is even for $m - 1$ samples. - If $\sum_{i=1}^m y_i$ is even, then $\sum_{i \neq j} y_i$ is odd for $m - 1$ samples.

Hence, the leave-one-out error is:

$$L_{\text{loo}}(h_S) = \frac{m-1}{m} \cdot \frac{1}{2} + \frac{1}{m} \cdot \frac{1}{2} = \frac{1}{2}$$

Therefore:

$$\mathbb{E}_S[L_D(h_S)] = \frac{1}{2}$$

$$\mathbb{E}_S[L_{\text{loo}}(h_S)] = \frac{1}{2}$$

2. Compute $\mathbb{E}_S\left[|L_D(h_S) - L_{\mathbf{loo}}(h_S)|\right]$, and comment on your result.

Since both $L_D(h_S)$ and $L_{\text{loo}}(h_S)$ are equal to $\frac{1}{2}$, their absolute difference is zero:

$$|L_D(h_S) - L_{\text{loo}}(h_S)| = 0$$

Therefore:

$$\mathbb{E}_S\left[|L_D(h_S) - L_{\text{loo}}(h_S)|\right] = 0$$

3. Give reasons behind the similarities / dissimilarities in these findings.

The findings are similar because both the generalization error and the leave-one-out error are based on the same underlying probability distribution where $P(y = 1|x) = P(y = 0|x) = \frac{1}{2}$. This symmetry ensures that the expected errors are the same. The leave-one-out error does not provide additional information in this scenario due to the even split of class probabilities.

Exercise 4.2: Rademacher Complexity

Prove that the Rademacher complexity of B_p satisfies $R(B_p) = d^{-1/p}$.

Given:

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

The Rademacher complexity $R(B_p)$ is given by:

$$R(B_p) = \mathbb{E}\left[\sup_{x \in B_p} \frac{1}{n} \sum_{i=1}^{n} \sigma_i x_i\right]$$

where σ_i are Rademacher variables (i.e., $\mathbb{P}(\sigma_i = 1) = \mathbb{P}(\sigma_i = -1) = \frac{1}{2}$).

We need to bound the supremum:

$$\sup_{x \in B_p} \frac{1}{n} \sum_{i=1}^{n} \sigma_i x_i$$

By Hölder's inequality:

$$\left|\sum_{i=1}^{n} \sigma_i x_i\right| \leq \|\sigma\|_q \|x\|_p$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Since $\|x\|_p \leq 1$ for $x \in B_p$, we get:

$$\left|\sum_{i=1}^{n} \sigma_i x_i\right| \leq \|\sigma\|_q$$

For σ_i being Rademacher variables, $\|\sigma\|_q = \left(\sum_{i=1}^n |\sigma_i|^q\right)^{1/q}$.

By Jensen's inequality and properties of Rademacher variables:

$$\|\sigma\|_q = n^{1/q}$$

Thus:

$$R(B_p) = \mathbb{E}\left[\sup_{x\in B_p} \frac{1}{n}\sum_{i=1}^n \sigma_i x_i\right] \le \frac{n^{1/q}}{n} = n^{1/q-1}$$

Since $\frac{1}{q} = 1 - \frac{1}{p}$, we get:

$$R(B_p) \le n^{1-\frac{1}{p}-1} = n^{-1/p}$$

Therefore:

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

Exercise 4.3: Universality of the Gaussian Kernel

Prove that the Gaussian kernel $k(x, y) = \exp\left(-\frac{1}{2}\|x - y\|^2\right)$ is universal on X.

Given:

$$X = \{x \in \mathbb{R}^p : \|x\|_2 \le 1\}$$

We start by approximating a function f using the exponential kernel $k(x, y) = \exp(\langle x, y\rangle)$, which is known to be universal on X.

The Gaussian kernel can be expressed as:

$$k(x, y) = \exp\left(-\frac{1}{2}\|x - y\|^2\right)$$

Rewrite $\|x - y\|^2$:

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y\rangle$$

Since $\|x\|_2 \le 1$ and $\|y\|_2 \le 1$, we have:

$$k(x, y) = \exp\left(-\frac{1}{2}(\|x\|^2 + \|y\|^2 - 2\langle x, y\rangle)\right)$$

Simplifying:

$$k(x, y) = \exp\left(-\frac{1}{2}\|x\|^2\right)\exp\left(-\frac{1}{2}\|y\|^2\right)\exp(\langle x, y\rangle)$$

Since $\|x\| \le 1$ and $\|y\| \le 1$:

$$\exp\left(-\frac{1}{2}\|x\|^2\right)\exp\left(-\frac{1}{2}\|y\|^2\right)$$

is a positive scalar factor, and $\exp(\langle x, y\rangle)$ is universal.

Thus, the Gaussian kernel $k(x, y) = \exp\left(-\frac{1}{2}\|x - y\|^2\right)$ inherits the universality property from the exponential kernel, proving it is universal on X.