COMPLEXITY OF SUPERVISED LEARNING

PETER MATHÉ

ABSTRACT. In the simplest case of supervised learning we are given a sample $(x_i, y_i)_{i=1}^n \in (X \times Y)^n$, drawn from an unknown distribution $\rho_{X\times Y}$. The goal is to recover the dependency $y = f(x)$, which in the least-squares context means to detect the conditional expectation $\mathbb{E}(Y|x)$ of Y given x. Our approach shall be built upon the seminal studies [1, 2], and [3].

We shall indicate how this statistical problem may fit into the framework of Information-based Complexity. In particular we highlight the concept of *solution smoothness*, related to a suitably chosen reproducing kernel Hilbert space. An important ingredient will be the *covariance operator*, and we shall discuss its impact on the accuracy of reconstruction.

We shall provide lower bounds, and we indicate which numerical methods can be used to achieve the optimal order of reconstruction.

Part of this talk is based on joint work with S. Lu and S. V. Pereverzev [4].

REFERENCES

[1] Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)*, 39(1):1–49, 2002.

[2] Ronald DeVore, Gerard Kerkyacharian, Dominique Picard, and Vladimir Temlyakov. Approximation methods for supervised learning. *Found. Comput. Math.*, 6(1):3–58, 2006.

[3] A. Caponnetto and E. De Vito. Optimal rates for the regularized least-squares algorithm. *Found. Comput. Math.*, 7(3):331–368, 2007.

[4] Shuai Lu, Peter Mathé, and Sergei Pereverzev. Balancing principle in supervised learning for a general regularization scheme. RICAM-Report 2016-38, RICAM, Linz, Austria, 2016.