Mathematical methods for supervised learning

Temlyakov V. (Columbia, USA)

Let ρ be an unknown Borel measure defined on the space $Z := X \times Y$ with $X \subset \mathbb{R}^d$ and $Y = [-M, M]$. Given a set $\mathbf{z}$ of m samples $z_i = (x_i, y_i)$ drawn according to ρ, the problem of estimating a regression function f_ρ using these samples is considered. The main focus is to understand what is the rate of approximation, measured either in expectation or probability, that can be obtained under a given prior $f_\rho \in \Theta$, i.e. under the assumption that f_ρ is in the set Θ, and what are possible algorithms for obtaining optimal or semi-optimal (up to logarithms) results. The optimal rate of decay in terms of m is established for many priors given either in terms of smoothness of f_ρ or its rate of approximation measured in one of several ways. This optimal rate is determined by two types of results. Upper bounds are established using various tools in approximation such as entropy, widths, and linear and nonlinear approximation. Lower bounds are proved using Kullback-Leibler information and a certain type of entropy. A distinction is drawn between algorithms which employ knowledge of the prior in the construction of the estimator and those that do not. Algorithms of the second type which are universally optimal for a certain range of priors are given.