Approximation regions, regularization and non-parametric regression

Laurie Davies
Universität Essen, Germany

Abstract:

We propose a unified approach to the problem of non-parametric regression on the unit interval. It is based on an honest non-asymptotic approximation (confidence) region $\mathcal{A}_n$ which is defined by a set of linear inequalities involving the values of the functions at the design points. Interest typically centers on certain simplest functions in $\mathcal{A}_n$ where simplicity can be defined in terms of shape (number of intervals of monotonicity or convexity/concavity) or smoothness (bounds on derivatives) or a combination of both (the smoothest function with four intervals of convexity/concavity). Once some form of regularization has been decided upon the approximation region can be used to provide honest non-asymptotic approximation bounds which are conceptually much simpler but also less informative. Although the procedure makes no attempt to minimize any loss function such as MISE the resulting estimates have optimal rates of convergence in the supremum norm both for shape and smoothness regularization. We show that rates of convergence can be misleading even for samples of size $n = 10^6$ and propose a different form of asymptotics which allows model complexity to increase with sample size.