Dimensional data fitting using sparse grid functions

Stephen G. Roberts

Australian National University, Canberra, Australia

In this paper we present a method for estimating probability density functions in high dimensions based on the sparse grid space, which scales linearly with data size and also has a weak dependence on the dimension of the space.

In particular we present an investigation of discrete histosplines using sparse grids spaces as the discretisation space. We provide a convergence analysis, which shows that for smooth enough underlying probability density functions, the complexity of the space is $O(|\log(h)|^{d-1}h^{-1})$ and the approximation error is $O(|\log(h)^{d-1}h^2)$. The use of classical sparse grids allows us to practically deal with data sets with up to 15 dimensions.

The method is demonstrated on a 10 dimensional data collection containing information on vegetation cover and various geographical features. We use the sparse grid histospline to generate a predictive model of vegetation type in terms of geographical data.