Ann Inst Stat Math (2010) 62:249–275
DOI 10.1007/s10463-008-0185-1

Oracle inequality for conditional density estimation and an actuarial example

Sam Efromovich

Received: 22 January 2007 / Revised: 29 February 2008 / Published online: 16 July 2008

Abstract Conditional density estimation in a parametric regression setting, where the problem is to estimate a parametric density of the response given the predictor, is a classical and prominent topic in regression analysis. This article explores this problem in a nonparametric setting where no assumption about shape of an underlying conditional density is made. For the first time in the literature, it is proved that there exists a nonparametric data-driven estimator that matches performance of an oracle which: (i) knows the underlying conditional density, (ii) adapts to an unknown design of predictors, (iii) performs a dimension reduction if the response does not depend on the predictor, (iv) is minimax over a vast set of anisotropic bivariate function classes. All these results are established via an oracle inequality which is on par with ones known in the univariate density estimation literature. Further, the asymptotically optimal estimator is tested on an interesting actuarial example which explores a relationship between credit scoring and premium for basic auto-insurance for 54 undergraduate college students.

Keywords Dimension reduction · Fixed and random design · MISE · Nonparametric regression