UNIVERSITY OF SYDNEY

SCHOOL OF MATHEMATICS AND STATISTICS

Statistics Seminar

Friday, 30 May 2008, 2.00pm

Carslaw 375

Partially linear model selection by the bootstrap

Samuel Mueller
(University of Sydney)

Abstract

The purpose of model selection is to choose one or more models α from $\mathcal{A}$ with specified desirable properties, where $\mathcal{A}$ denotes a set of partially linear regression models for the relationship between a response vector $\mathbf{y}$ and a design matrix $\mathbf{X}$; that is $\mathcal{A} = \{\alpha \subseteq \{1, \ldots, p\}$, such that $\alpha = \{\alpha_1, \alpha_2\}, \alpha_1 \cap \alpha_2 = \emptyset\}$ where the functional relationship is

$$y_i = \theta_{\alpha_2}^T x_{\alpha_2,i} + g\left(\mathbf{x}_{\alpha_1,i}^T\right) + \epsilon_i, \quad i = 1, \ldots, n.$$

θ_{α_2} denotes an unknown p_{α_2}-vector of parameters, $g : \mathbb{R}^{p_{\alpha_1}} \mapsto \mathbb{R}$ is an unknown function, $\mathbf{X}_\alpha$ and $\epsilon_\alpha = (\epsilon_{1\alpha}, \ldots, \epsilon_{n\alpha})^T$ are independent, and the $\epsilon_{i\alpha}$ have location zero and variance σ^2. We use least squares (LS) estimation for estimating the two unknowns θ_{α_2} and g in the partially linear model and propose a new approach to the selection of partially linear models which in the spirit of Shao (1996; JASA) is based on the conditional expected prediction square loss function which is estimated using the bootstrap. Due to the different speeds of convergence of the linear and the nonlinear parts, a key idea is to select each part separately. In the first step we select the nonlinear components using a 'm-out-of-n' residual bootstrap which ensures good properties for the nonparametric bootstrap estimator. The second step selects the linear components among the remaining explanatory variables and the non-zero parameters are selected based on a two level residual bootstrap. We show that the model selection procedure is consistent under some conditions and simulations show that it performs well in comparison to other selection criteria.

Enquiries about the Statistics Seminar should be directed to
Jean Yang (`jeany@maths.usyd.edu.au`)