Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 4605–4621
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Knot selection by boosting techniques

Florian Leitenstorfer*, Gerhard Tutz

Ludwig-Maximilians-Universität München, Akademiestraße 1, 80799 München, Germany

Received 23 August 2005; received in revised form 1 August 2006; accepted 1 August 2006
Available online 24 August 2006

Abstract

A novel concept for estimating smooth functions by selection techniques based on boosting is developed. It is suggested to put radial basis functions with different spreads at each knot and to perform selection and estimation simultaneously by a componentwise boosting algorithm. The methodology of various other smoothing and knot selection procedures (e.g. stepwise selection) is summarized. They are compared to the proposed approach by extensive simulations for various unidimensional settings, including varying spatial variation and heteroskedasticity, as well as on a real world data example. Finally, an extension of the proposed method to surface fitting is evaluated numerically on both, simulation and real data. The proposed knot selection technique is shown to be a strong competitor to existing methods for knot selection.

Keywords: Nonparametric regression; Knot selection; Radial basis functions; Boosting; Surface fitting

1. Introduction

In the last decades, a tremendous amount of methods has been developed for the estimation of smooth functions f in an unidimensional regression setting $y = f(x) + \varepsilon$. Besides localized approaches (see e.g. Fan and Gijbels, 1996), one distinguishes between methods based on smoothing splines or regression splines. The former (see Eubank, 1988 or Wahba, 1990) uses many knots (up to the sample size n) which are placed in the design space, and the roughness of the estimate is controlled by a specific penalty term. The latter, on which we will focus in the present paper, is based on an expansion of f into basis functions, $f = \sum \alpha_j B_j$. In this setting, the number of actually chosen knots is much less than n. To avoid overfitting, one uses penalization strategies (P-splines, see Eilers and Marx, 1996) as well as knot selection strategies which are based on well-known variable selection techniques. Stone et al. (1997) and He and Ng (1999) use stepwise selection, whereas Osborne et al. (1998) propose knot selection by Lasso (see Tibshirani, 1996). Knot selection from a Bayesian perspective has been treated by Smith and Kohn (1996), Denison et al. (1998) and Lang and Brezger (2004).

In the present paper, we aim at knot selection by employing recent developments in variable selection based on boosting techniques. Bühlmann and Yu (2003) propose a boosting algorithm constructed from the L_2-loss, which is suitable for high dimensional predictors in an additive model context. Bühlmann (2006) extends L_2 Boost to the special issue of fitting high-dimensional linear models, where the number of covariates may exceed the sample size. This approach can straightforwardly be adapted to a regression spline context. It is possible to work with a very high

* Corresponding author. Tel.: +49 89 2180 2925; fax: +49 89 2180 5308.
E-mail addresses: leiten@stat.uni-muenchen.de (F. Leitenstorfer), tutz@stat.uni-muenchen.de (G. Tutz).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.08.008

number of basis functions, which are selected componentwise in a stepwise fashion. In order to obtain high flexibility in the resulting fits, we recommend to use radial basis functions (e.g. Ripley, 1996) with spreads chosen data-adaptively by componentwise boosting. As simulation will show, this leads to superior performance for the estimation of functions with high spatial variation as well as to robustness against violations of model assumptions.

The outline of the paper is as follows: in Section 2, we give an outline of the boosting algorithm. Section 3 contains a brief review over some alternative smoothing methods. The procedures are compared by a simulation study in the style of Wand (2000). In Section 4, the approach is extended to surface fitting.

2. A smoothing procedure based on componentwise L_2-boosting

We consider the problem of conventional unidimensional nonparametric regression. For a dependent variable y_i and a covariate x_i, $i = 1, \ldots, n$, the model

$$y_i = f(x_i) + \sigma(x_i)\,\varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1) \tag{1}$$

is assumed, where $f(.)$ is a smooth function and $\sigma(.)$ is a positive function. A very popular approach to this problem is the expansion of f into basis functions, i.e.

$$f(x_i) = \alpha_0 + \sum_{j=1}^{m} \alpha_j B_j(x_i), \tag{2}$$

where the α_j are unknown coefficients, m is the number of knots and B_j denote the basis functions. Basis functions that have often been used in the literature are e.g. the truncated power series basis (see Ruppert and Carroll, 2000 or Wand, 2000), the B-spline basis (Eilers and Marx, 1996) and the natural spline basis (Green and Silverman, 1994). Alternative basis functions that are suggested in the neural network community are the so-called radial basis functions (e.g. Ripley, 1996). An example for the latter are localized Gaussian densities, given by

$$B_j(x) = \exp\left(-\frac{|x - \tau_j|^2}{2h^2}\right), \tag{3}$$

where τ_j is the center of the basis function, and h determines the spread.

In the following, we will focus on radial basis functions, since they have some properties which are useful for the proposed procedure based on componentwise boosting. First, if we assume a sequence of knots $\{\tau_j\}_{j=1}^{m}$, the $B_j(x)$ as given in (3) are only linked to one knot τ_j. Furthermore, radial basis functions provide support on the entire real line. In contrast, B-splines—which are widely-used due to their numerical stability—are determined by $q + 2$ knots if they are of degree q. They have local support, i.e. they take values greater than 0 only on $q + 2$ consecutive knots. This implies for knot selection that the whole B-spline basis has to be recomputed, if a certain knot is added or deleted. It entails further a re-estimation of all coefficients.

The use of basis functions as given in (3) raises the question how to choose the spread h appropriately. A simple concept would be to take the same h at each knot, and to determine it e.g. by a data driven choice. However, this strategy is doubtful if one aims at constructing a flexible smoothing procedure, which should also be able to handle the estimation of functions with high spatial variation (see for example Ruppert and Carroll, 2000). We suggest to put at each knot τ_j, $j = 1, \ldots, m$, several radial basis functions with different spread h. Suppose we have a sequence of r distinct spread variables $h_1 < \cdots < h_r$, then the expansion into basis functions from (2) is given as

$$f(x_i) = \alpha_0 + \sum_{j=1}^{m} \sum_{k=1}^{r} \alpha_{jk} B_{jk}(x_i), \tag{4}$$

where $B_{jk}(x_i) = \exp\left(-|x_i - \tau_j|^2 \Big/ 2h_k^2\right)$. From (4), it is seen that the number of parameters is large, if the number of knots m is high enough and the grid of h_k is subtle enough to get satisfying flexibility. Thus, a procedure is needed which is able to estimate high-dimensional problems and avoids overfitting. In our simulation studies we found that the use of more than one spread improves the performance distinctively. We found that it is useful to center the basis functions by their means, i.e. we suggest to use $\tilde{B}_{jk}(x_i) = B_{jk}(x_i) - 1/n \sum_{i=1}^{n} B_{jk}(x_i)$.