Computational Statistics and Data Analysis 62 (2013) 136–148

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Bandwidth selection for backfitting estimation of semiparametric additive models: A simulation study

Jenny Häggström *

Department of Statistics, USBE, UmeåUniversity, SE-90187 Umeå, Sweden

ARTICLE INFO

Article history:
Received 20 December 2011
Received in revised form 18 December 2012
Accepted 8 January 2013
Available online 23 January 2013

Keywords:
Bandwidth selection
Semiparametric additive model
Backfitting estimation
Mean squared error
Nonparametric regression

ABSTRACT

A data-driven bandwidth selection method for backfitting estimation of semiparametric additive models, when the parametric part is of main interest, is proposed. The proposed method is a double smoothing estimator of the mean-squared error of the backfitting estimator of the parametric terms. The performance of the proposed method is evaluated and compared with existing bandwidth selectors by means of a simulation study.

1. Introduction

The semiparametric additive model (SAM) is a regression model which consists of a sum of unknown smooth functions and parametric terms; see e.g. Hastie and Tibshirani (1990), Fan and Jiang (2005) and references therein. Formally, the model is written as

$$y_i = \sum_{d=1}^{D} \beta_d(x_{di}) + \sum_{p=1}^{P} \tau_p z_{pi} + \epsilon_i = \sum_{d=1}^{D} \mu_d(x_{di}) + \mathbf{z}_i^T \boldsymbol{\gamma} + \epsilon_i, \tag{1}$$

with y_i the response, $i = 1, \ldots, n$, $\beta_d(x_{di})$, $d = 1, \ldots, D$, and τ_p, $p = 1, \ldots, P$, unknown smooth functions and parameters, respectively, where x_d:s and z_p:s are covariates. We assume $\mathrm{E}(\epsilon_i) = 0$, $\mathrm{Var}(\epsilon_i) = \sigma^2$ and have $\mu_d(x_{di}) = \beta_d(x_{di}) - \mathrm{E}\big(\beta(x_{di})\big)$, $\mathbf{z}_i = (1, z_{1i}, \ldots, z_{Pi})^T$ and $\boldsymbol{\gamma} = (\alpha, \tau_1, \ldots, \tau_P)^T$, with $\alpha = \mathrm{E}\big(\sum_{d=1}^{D} \beta_d(x_{di})\big)$. Let $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^T$ and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ with $\mathbf{x}_i = (x_{1i}, \ldots, x_{Di})^T$ be the vector and matrices containing the data.

This model is often preferable, as opposed to the fully nonparametric additive model, if the relationship between $\mathbf{z}_i$ and y_i is assumed to be linear while the shape of the relationships between the x_{di}:s and y_i is more uncertain. Another scenario is if $\mathbf{z}_i$ contains categorical variables in which case smooth functions are not an option.

An often used method for estimating additive models is the classical backfitting, readily available, e.g., in the R-package `gam`, (Buja et al., 1989; Hastie and Tibshirani, 1990). For other estimation methods than classical backfitting, see, e.g., Linton and Nielsen (1995), Eilers and Marx (1996), Mammen et al. (1999), Fahrmeir et al. (2004), Lin and Zhang (1999) and Wood (2000).

In order to estimate the smooth functions in (1), irrespective of the choice of the estimator, some kind of smoothing parameters (or bandwidths) must be selected. If the classical backfitting estimator of (1) is employed together with the

* Tel.: +46 90 7869318; fax: +46 90 7866114.
E-mail address: jenny.haggstrom@stat.umu.se.

0167-9473/$ – see front matter
doi:10.1016/j.csda.2013.01.010

"usual" optimal bandwidth, which most bandwidth procedures, e.g., cross-validation attempts to select, $\hat{\boldsymbol{\gamma}}$ does not achieve the parametric rate of convergence, it is not $\sqrt{n}$-consistent (Rice, 1986; Speckman, 1988; Robinson, 1988; Linton, 1995; Opsomer and Ruppert, 1999). The purpose of this paper is to investigate bandwidth selection when (1) is fitted by backfitting and the parametric part is of main interest. This was shown by Speckman (1988) for SAM with $D = 1$, often called the partially linear model, and by Opsomer and Ruppert (1999) for $D > 1$.

For a SAM with $D = 1$ Speckman (1988) proposed an alternative estimator which is $\sqrt{n}$-consistent. That estimator is equivalent to the profile likelihood estimator, treated in e.g. Maity et al. (2007), if $\epsilon_i \sim$ Normal. He et al. (2007) extend this estimator to $D > 1$ but no distributional results are derived for this case.

In this paper, we propose a, for this setting, new, intuitive and easily implemented bandwidth selection procedure for backfitting estimation of SAMs. As Opsomer and Ruppert (1999), we propose to select the smoothing parameters by using an estimate of $\text{MSE}(\hat{\boldsymbol{\gamma}}|\mathbf{X}, \mathbf{Z})$. In contrast with Opsomer and Ruppert (1999), our estimator does not use an asymptotic approximation of the bias of $\hat{\boldsymbol{\gamma}}$. Properties of the new procedure are examined and compared with competing selection methods in a simulation study. Up to our knowledge this is the first attempt to study by simulation the properties of estimators of $\boldsymbol{\gamma}$ in (1) obtained with different bandwidth selectors. We show that both the new procedure and the procedure proposed by Opsomer and Ruppert (1999) improve on cross-validation in the sense that estimates of $\boldsymbol{\gamma}$ obtained with these procedures have lower mean squared errors than estimates obtained with cross-validation. Additionally, we show that the new procedure is, in most cases, preferable to the one in Opsomer and Ruppert (1999).

The paper is organized as follows. The next section presents backfitting estimation of SAMs and a closed-form expression of $\text{MSE}(\hat{\boldsymbol{\gamma}}|\mathbf{X}, \mathbf{Z})$. Section 3 focuses on bandwidth selection in SAMs. We review some asymptotics regarding $\sqrt{n}$-consistency of $\hat{\boldsymbol{\gamma}}$, describe existing bandwidth selectors and introduce a new bandwidth selection method. These selectors are compared by simulation in Section 4. The paper is concluded in Section 5.

2. Estimating SAM

2.1. Backfitting

Backfitting is an iterative estimation method. For estimation of a SAM, starting with some initial estimates $\hat{\boldsymbol{\mu}}_d^0$, of $\boldsymbol{\mu}_d = (\mu(x_{d1}), \dots, \mu(x_{dn}))^T$, for $d = 1, \dots, D$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\mu}}_d$ are repeatedly computed, until the estimates converge, according to

$$\begin{cases} \hat{\boldsymbol{\gamma}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\left(\mathbf{y} - \sum_{d=1}^{D}\hat{\boldsymbol{\mu}}_d\right), \\ \hat{\boldsymbol{\mu}}_d = \mathbf{M}_d\left(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}} - \sum_{k\neq d}\hat{\boldsymbol{\mu}}_k\right), \quad d = 1, \dots, D, \end{cases} \tag{2}$$

where $\mathbf{M}_d = (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\mathbf{S}_d$. $\mathbf{I}$ denotes the $n \times n$ identity matrix, $\mathbf{1}$ a vector of 1:s with length n and $\mathbf{S}_d$ is a smoothing matrix which is only a function of the design points x_{di}, $i = 1, \dots, n$, and a smoothing parameter h_d. For instance, $\hat{\boldsymbol{\beta}}_d = \mathbf{S}_d\mathbf{y}$ is a linear smoother of $\tilde{\beta}_d(x_{di}) = \text{E}(y_i|x_{di})$. Typical examples of linear smoothers are kernel regression, splines and local polynomial regression; see e.g. Fan and Gijbels (1996, pp. 14–45).

Provided that the backfitting procedure converges to a unique solution (see Opsomer, 2000, for sufficient conditions) the backfitting estimator is equivalent to the following non-iterative solution

$$\hat{\boldsymbol{\gamma}} = \left(\mathbf{Z}^T(\mathbf{I} - \mathbf{Q}_\mu)\mathbf{Z}\right)^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{Q}_\mu)\mathbf{y}, \tag{3}$$
$$\hat{\boldsymbol{\mu}} = \sum_{d=1}^{D}\hat{\boldsymbol{\mu}}_d = \mathbf{Q}_\mu\left(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\right),$$

with $\mathbf{Q}_\mu = \sum_{d=1}^{D}\mathbf{Q}_d$. For $D = 1$, $\mathbf{Q}_\mu = \mathbf{Q}_1 = \mathbf{M_1}$ and, for $D = 2$, $\mathbf{Q}_d = \left(\mathbf{I} - (\mathbf{I} - \mathbf{M}_d\mathbf{M}_{k\neq d})^{-1}(\mathbf{I} - \mathbf{M}_d)\right)$, $k = 1, 2$, Hastie and Tibshirani (1990, p. 119). For $D > 2$, Opsomer (2000) provides the following recursive definition: $\mathbf{Q}_d = \left(\mathbf{I} - (\mathbf{I} - \mathbf{M}_d\mathbf{Q}_\mu^{[-d]})^{-1}(\mathbf{I} - \mathbf{M}_d)\right)$, where $\mathbf{Q}_\mu^{[-d]}$ is the additive smoother matrix for the $(D-1)$-variate function $\mu = \mu_1 + \cdots + \mu_{d-1} + \mu_{d+1} + \cdots + \mu_D$. Now, letting $\boldsymbol{\mu} = \sum_{d=1}^{D}\boldsymbol{\mu}_d$, we can derive:

$$\text{Bias}(\hat{\boldsymbol{\gamma}}|\mathbf{X}, \mathbf{Z}) = \left(\mathbf{Z}^T(\mathbf{I} - \mathbf{Q}_\mu)\mathbf{Z}\right)^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{Q}_\mu)\boldsymbol{\mu}, \tag{4}$$

and

$$\text{Var}(\hat{\boldsymbol{\gamma}}|\mathbf{X}, \mathbf{Z}) = \sigma^2\text{diag}\left(\left[\left(\mathbf{Z}^T(\mathbf{I} - \mathbf{Q}_\mu)\mathbf{Z}\right)^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{Q}_\mu)\right]\left[\left(\mathbf{Z}^T(\mathbf{I} - \mathbf{Q}_\mu)\mathbf{Z}\right)^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{Q}_\mu)\right]^T\right), \tag{5}$$