Computational Statistics and Data Analysis 108 (2017) 70–83

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Adaptive penalized splines for data smoothing☆

Lianqiang Yang [a,b,*], Yongmiao Hong [b]

[a] School of Mathematical Sciences, Anhui University, Hefei 230601, China
[b] Department of Economics, Cornell University, Ithaca, NY, 14853, USA

HIGHLIGHTS

- An adaptive penalized splines model for data smoothing is proposed.
- The model is locally adaptive to the inhomogeneity of observations.
- Local penalization is constructed without complicated pre-estimation.

ARTICLE INFO

Article history:
Received 6 February 2016
Received in revised form 18 October 2016
Accepted 24 October 2016
Available online 4 November 2016

Keywords:
Nonparametric regression
Data smoothing
Penalized splines
Adaptivity
Local penalty

ABSTRACT

Data driven adaptive penalized splines are considered via the principle of constrained regression. A locally penalized vector based on the local ranges of the data is generated and added into the penalty matrix of the classical penalized splines, which remarkably improves the local adaptivity of the model for data heterogeneity. The algorithm complexity and simulations are studied. The results show that the adaptive penalized splines outperform the smoothing splines, l_1 trend filtering and classical penalized splines in estimating functions with inhomogeneous smoothness.

1. Introduction

Several non- or semi-parametric regression methods have been developed in recent years, including classical methods, such as kernels, local polynomials, smoothing splines, sieves and wavelets, and relative newcomers such as l_1 trend filtering, lasso, and generalized lasso (Wahba, 1990; Green and Silverman, 1994; Fan and Gijbels, 1996; Eubank, 1999; Ruppert et al., 2003; Hastie et al., 2001). The principle of maximizing likelihood functions under some constraints is obeyed in certain regression methods, e.g., smoothing splines, penalized splines, l_k trend filtering and generalized lasso, where the models are commonly designed as

$$y_i = f(x_i) + \epsilon_i.$$

With given observations (x_i, y_i), the design x is arranged such that $x_i \leqslant x_{i+1}, i = 1, \ldots, n$. The regression function to be estimated is $f(x)$, and the error ϵ_i is generally assumed to be independent and identically subject to a Gaussian distribution with mean 0 and variance σ^2. Once $f(x)$ is represented by a linear combination of basis functions $B_j(x)$ as $\sum_{j=1}^{p} B_j(x)\beta_j$, the

☆ The R codes are attached to the electronic version of this paper.
* Corresponding author at: School of Mathematical Sciences, Anhui University, Hefei 230601, China.
E-mail address: yanglq@ahu.edu.cn (L. Yang).

http://dx.doi.org/10.1016/j.csda.2016.10.022
0167-9473/

estimate is defined by a penalized least-square optimization problem:

$$\hat{\boldsymbol{\beta}} = arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \{\|Y - X\boldsymbol{\beta}\|_2^2 + \lambda \text{ penalty}(\boldsymbol{\beta})\}, \quad (1)$$

where $Y = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, the design matrix X is generated by $X(i,j) = B_j(x_i)$, $\|\cdot\|_2$ is the Euclid norm, λ is a tuning parameter, and penalty($\boldsymbol{\beta}$) is a positive convex function of $\boldsymbol{\beta}$.

Constrained regression based on splines or piecewise polynomials is applied in some popular methods. First, several types of spline bases are used for $B_j(x)$ such as the truncated power basis (Ruppert, 2002), the B-spline basis (De Boor, 1978; Eilers and Marx, 1996), the natural spline basis (Wahba, 1990) and other piecewise polynomial bases (Tibshirani et al., 2014). Second, several types of penalties are used, such as smoothing splines with $\int f''(x)^2 dx$ (Wahba, 1990; Gu, 2002), penalized splines with $\|D^{(k)}\boldsymbol{\beta}\|_2^2$ (O'Sullivan, 1988; Eilers and Marx, 1996), l_1 trend filtering with $\|D^{(1)}\boldsymbol{\beta}\|_1$, where $D^{(k)}$ is the difference matrix of order k (Kim et al., 2009), and generalized lasso based on splines (Bazerque et al., 2011; Lin and Zhang, 2006) with $\|D\boldsymbol{\beta}\|_1$, where D is a particular matrix (Tibshirani and Taylor, 2011). Intuitively, the first term in target function (1) measures the estimator's loss, and the second term denotes the estimator's undesirable characters; the weight λ controls the trade-off between the first and second terms. It can also be said that the target function maintains the balance between the goodness of fit and the degrees of freedom of the model.

Adaptability strongly influences model efficiency. Therefore, how to improve this aspect using a well-designed model is an interesting research topic in nonparametric regression analysis. Adaptive spline smoothing is developed in several methods. Generally, the first method focuses on the selection of knots (Friedman, 1991; Mammen et al., 1997; Zhou and Shen, 2001; Ruppert, 2002). These procedures always require complex computations to determine the number and positions of knots. The second method focuses on the construction of penalties (Ruppert and Carroll, 2000; Jullion and Lambert, 2007; Pintore et al., 2006; Liu and Guo, 2010; Storlie et al., 2010; Jang and Oh, 2011; Wang et al., 2013). In these procedures, the idea of generalizing the total penalty parameter λ from a constant to a function $\lambda(\cdot)$ plays the key role, and the function $\lambda(\cdot)$ is always pre-estimated by some elaborate techniques. Furthermore, Bayesian adaptive penalized splines have also been studied in some works. Baladandayuthapani et al. (2005) and Jullion and Lambert (2007) give heteroscedastic prior distributions for the penalty parameters, whereas Lang and Brezger (2004), Scheipl and Kneib (2009), Krivobokova et al. (2008) and Crainiceanu et al. (2007) give heteroscedastic prior distributions for the regression parameters. The estimates are normally obtained by Markov chain Monte Carlo (MCMC) or restricted maximum likelihood (REML). Krivobokova et al. (2008) further developed a fast and simple algorithm for the Bayesian P-splines based on Laplace approximation of the marginal likelihood.

In particular, Tibshirani et al. (2014) present a detailed study on l_1 trend filtering (Kim et al., 2009) and compare it with smoothing splines (Wahba, 1990) and locally adaptive regression splines (Mammen et al., 1997). The trend filtering has similar adaptability as the locally adaptive regression splines but outperforms the smoothing splines in estimating functions with inhomogeneous smoothness. However, greater local adaptivity appears necessary in certain cases, such as Fig. 1, where the Mexican hat function and its scatter points with random errors and l_1 trend filtering fitting with 15, 20, and 40 degrees of freedom are shown. Clearly, with more degrees of freedom, the cusp of the hat becomes well fitted; however, the flat part is over-fitted with more wiggles. With fewer degrees of freedom, the flat part is well fitted; however, the cusp is under-fitted.

Most of the aforementioned adaptive splines are based on smoothing splines and not penalized splines (Eilers and Marx, 1996; Ruppert et al., 2003), and complicated pre-computations are always required. In this paper, a new method called adaptive penalized splines is provided based on the classical penalized splines. The procedure only adds a weight vector to the differences in coefficients, the weight vector is fully data driven and does not need to be pre-estimated. Solving of the model is a simple quadratic convex optimization problem as generalized ridge regression. Empirical results and simulations show that this new method can remarkably improve the local adaptivity of the model compared to l_1 trend filtering, smoothing splines and classical penalized splines. The remainder of this paper is organized as follows: Section 2 constructs the adaptive penalized splines and the computation of the nonparametric estimator is considered in Section 3; a simulation study for comparing the adaptive penalized splines, l_1 trend filtering, smoothing splines and classical penalized splines is presented in Section 4; an application is shown in Section 5; and Section 6 contains some discussion.

2. Adaptive penalized splines

First, we provide a short review on classical penalized splines, i.e., P-splines, which was first named in Eilers and Marx (1996) but can be traced back to O'Sullivan (1988), where X is constructed using a l-degree B-spline basis and penalty($\boldsymbol{\beta}$) is equal to $\|D^{(k)}\boldsymbol{\beta}\|_2^2$. Then, (1) is equivalently written as

$$\hat{\boldsymbol{\beta}} = arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} B_j(x_i)\beta_j \right)^2 + \lambda \sum_{j=k+1}^{p} (\triangle^k \beta_j)^2 \right\}, \quad (2)$$

where $\triangle^k$ denotes the kth-order difference operator as $\triangle^1 \beta_j = \beta_j - \beta_{j-1}$ and $\triangle^k = \triangle(\triangle^{k-1})$. In the case of $k = 2$, the P-splines can be considered as the discrete version of smoothing splines. Additional details about P-splines can be found in Eubank (1999), Green and Silverman (1994) and Ramsay and Silverman (2005), etc.