ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Monotone splines lasso

Linn Cecilie Bergersen, Kukatharmini Tharmaratnam, Ingrid K. Glad [*]

Department of Mathematics, University of Oslo, Norway

ARTICLE INFO

Article history:
Received 27 March 2013
Received in revised form 22 March 2014
Accepted 23 March 2014
Available online xxxx

Keywords:
Cooperative lasso
High-dimensional data
I-splines
Lasso
Monotone regression
Nonparametric additive models

ABSTRACT

The important problems of variable selection and estimation in nonparametric additive regression models for high-dimensional data are addressed. Several methods have been proposed to model nonlinear relationships when the number of covariates exceeds the number of observations by using spline basis functions and group penalties. Nonlinear *monotone* effects on the response play a central role in many situations, in particular in medicine and biology. The monotone splines lasso (MS-lasso) is constructed to select variables and estimate effects using monotone splines (I-splines). The additive components in the model are represented by their I-spline basis function expansion and the component selection becomes that of selecting the groups of coefficients in the I-spline basis function expansion. A recent procedure, called cooperative lasso, is used to select sign-coherent groups, i.e. selecting the groups with either exclusively non-negative or non-positive coefficients. This leads to the selection of important covariates that have nonlinear monotone increasing or decreasing effect on the response. An adaptive version of the MS-lasso reduces both the bias and the number of false positive selections considerably. The MS-lasso and the adaptive MS-lasso are compared with other existing methods for variable selection in high dimensions by simulation and the methods are applied to two relevant genomic data sets. Results indicate that the (adaptive) MS-lasso has excellent properties compared to the other methods both in terms of estimation and selection, and can be recommended for high-dimensional monotone regression.

1. Introduction

Along with the massive production of large data sets within most areas of science and technology, methods for high dimensional regression have become increasingly important. When the number of predictors P is large compared to the sample size n, penalized regression methods handle the dimensionality problem by adding a penalty to the negative log-likelihood to be minimized. The lasso (Tibshirani, 1996) and its many variants (Zou, 2006; van de Geer et al., 2011; Yuan and Lin, 2006; Zou and Hastie, 2005; Meinshausen, 2007) have the advantage of setting some of the regression coefficients to zero, thus producing a sparse solution. Recently, nonparametric methods for high-dimensional regression have started to emerge. Recent papers (Avalos et al., 2007; Meier et al., 2009; Huang et al., 2010; Ravikumar et al., 2009) consider a generalized additive model (GAM) (Hastie and Tibshirani, 1990) in combination with spline approximations. Given the observations $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, where y_i is the response and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iP})^t$ is the vector of covariates for observation

[*] Correspondence to: Department of Mathematics, University of Oslo, PB 1053 Blindern, NO-0316 Oslo, Norway. Tel.: +47 22855879; fax: +47 22854349.
E-mail address: glad@math.uio.no (I.K. Glad).

http://dx.doi.org/10.1016/j.csda.2014.03.013
0167-9473/© 2014 Elsevier B.V. All rights reserved.

i, the additive model is given as

$$y_i = \beta_0 + \sum_{j=1}^{P} g_j^0(x_{ij}) + \epsilon_i. \tag{1}$$

Here β_0 is the intercept, the g_j^0s are unknown functions to be estimated and ϵ_i is the independent random error with mean zero and variance σ^2. We assume $Eg_j^0(\boldsymbol{x}_j) = \mathbf{0}$, for $1 \le j \le P$, where now $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})^t$, to ensure unique identification of the g_j^0s. In Avalos et al. (2007), Meier et al. (2009), Bühlmann and van de Geer (2011), Huang et al. (2010) and Ravikumar et al. (2009), each nonparametric component g_j^0 is represented by a linear combination of spline basis functions and the problem can be viewed as a group lasso problem (Yuan and Lin, 2006) by selecting groups of spline basis functions representing relevant covariates. Covariates are often represented by B-splines due to their flexibility and minimal assumptions with respect to the form of function to be estimated. Combined with the group lasso, the framework becomes a highly flexible alternative to (standard) linear lasso-type methods.

Our aim is to construct a new method for high dimensional regression which is nonparametric and flexible as above, but which can be restricted to select and estimate monotone functions g_j^0 only. In certain bio-medical applications it is important to assume that the relationship between an explanatory variable and the outcome is monotonically increasing or decreasing. Actually, every time linear regression is applied, an implicit assumption of monotonicity is made. For example, monotone, but not necessarily linear relations typically appear for dose–response data. It is also reasonable to assume that the relationship between a disease and a risk factor is monotone, but not necessarily linear (Raftery and Richardson, 1996).

There has been a major effort in developing methods for monotone regression beyond the strictly linear regression models. In simple regression problems, monotone increasing relationships are often modeled through isotonic regression (Barlow et al., 1972; Robertson et al., 1988). Additive isotonic models, assuming that each component effect in the additive model is isotonic, were presented in Bacchetti (1989). However, most literature on monotone and isotonic regressions is limited to low dimensions. Very recently, one important contribution has appeared for monotone regression in high dimensions. Fang and Meinshausen (2012) propose Lasso Isotone (liso), combining estimation of nonparametric isotonic functions with ideas from sparse high-dimensional regression in an additive isotonic regression model. This is, to our knowledge, the only method feasible for monotone high-dimensional problems. Using an adaptive liso approach, Fang and Meinshausen (2012) also present a way of fitting the model without assuming that all effects are either increasing or decreasing, thus allowing for component effects of different signs. In this paper we develop another, substantially different, tool for the same purpose.

Isotonic regression is probably the best known method for preserving monotonicity, but has the disadvantage of producing step functions, which often have little biological plausibility, instead of smooth functions. For simple regression, it is possible to use an additional smoothing procedure in a second step to obtain a smooth function (He and Shi, 1998). Tibshirani et al. (2011) proposed nearly-isotonic regression which involves a penalty term controlling the level of monotonicity compared to the goodness of fit.

Another way of preserving monotonicity is to fit a smooth monotone function via monotone regression splines (Ramsay, 1988; He and Shi, 1998). While He and Shi (1998) proposed monotone B-spline smoothing based on a constrained least absolute deviation principle, Ramsay (1988) introduced integrated splines (I-splines), which essentially are integrated versions of M-splines that in combination with strictly positive coefficients will produce monotone increasing smooth functions. I-splines have previously been used in connection with a boosting technique to do monotonic regression in a multivariate model in Tutz and Leitenstorfer (2007). Meyer (2008) also considers shape-restricted regression splines by means of I-splines, but only in the one-dimensional case.

In this paper a new approach to fit nonparametric additive models under the assumption that each component effect $g_j^0(x)$ is monotone is proposed. The *monotone splines lasso* (MS-lasso) combines the idea of I-splines with the cooperative lasso (Chiquet et al., 2012), and is feasible in high-dimensional settings where the number of covariates P can exceed the number of observations n. The cooperative lasso is a lasso method where known groups of covariates are treated together, but differs from the standard group lasso (Yuan and Lin, 2006) in that it assumes that the groups are sign-coherent. That is, the covariates inside a group are cooperating, so either the linear coefficients are all nonpositive, all nonnegative or all null inside a group. This can be combined with monotone I-splines by letting each covariate, represented via an I-spline basis, constitute a group in the cooperative lasso. Thus the MS-lasso fits the additive nonparametric regression model with components that can be either nondecreasing, nonincreasing or of no effect. The important advantages of the MS-lasso are that the monotone functions g_j^0 can be either monotone *increasing* or *decreasing* in the same model, and that it is fitting *smooth* monotone functions to each g_j^0. In this way it is more flexible than the linear model, but more restrictive than purely nonparametric methods without any shape constraints. The method is also biologically more relevant than the adaptive liso, in that smooth representations of the functions are immediately obtainable. A two-step estimator is also proposed, the adaptive MS-lasso, which leads to less bias and fewer false positives in the final model.

This paper is organized as follows. In Section 2 we present the MS-lasso and discuss some of its properties. The adaptive MS-lasso is also presented, and connections to related methods are discussed. Section 3 is dedicated to simulation studies. In Section 4 the use of the method is illustrated in genomic data, before a final discussion is presented in Section 5.

Please cite this article in press as: Bergersen, L.C., et al., Monotone splines lasso. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.03.013