Computational Statistics and Data Analysis 85 (2015) 1–22

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

# Robust nonnegative garrote variable selection in linear regression

I. Gijbels *, I. Vrinssen

*KU Leuven, Department of Mathematics and Leuven Statistics Research Center (LStat), Celestijnenlaan 200B, Box 2400, B-3001 Leuven, Belgium*

ARTICLE INFO

*Article history:*
Received 6 November 2013
Received in revised form 17 November 2014
Accepted 18 November 2014
Available online 2 December 2014

*Keywords:*
Multiple linear regression
MM-estimation
Nonnegative garrote
S-estimation
Variable selection

ABSTRACT

Robust selection of variables in a linear regression model is investigated. Many variable selection methods are available, but very few methods are designed to avoid sensitivity to vertical outliers as well as to leverage points. The nonnegative garrote method is a powerful variable selection method, developed originally for linear regression but recently successfully extended to more complex regression models. The method has good performances and its theoretical properties have been established. The aim is to robustify the nonnegative garrote method for linear regression as to make it robust to vertical outliers and leverage points. Several approaches are discussed, and recommendations towards a final good performing robust nonnegative garrote method are given. The proposed method is evaluated via a simulation study that also includes a comparison with existing methods. The method performs very well, and often outperforms existing methods. A real data application illustrates the use of the method in practice.



## 1. Introduction

Variable selection has become a key issue in applied data analysis, since often many variables are measured. However, models including all the covariates are difficult to interpret and irrelevant variables increase the variance. Consider, for example, data of 60 U.S. Standard Metropolitan Statistical Areas, collected by researchers at General Motors to study whether air pollution contributes to mortality. The response is age adjusted mortality and the 14 covariates measure demographic characteristics of the cities, climate characteristics and the pollution potential of three air pollutants, namely hydrocarbon, nitrous oxide and sulfur dioxide. Our interest is to find out which of the many characteristics influence mortality and in particular, whether air pollution is significantly related to mortality. See further Section 6.

In the literature, different variable selection methods are proposed for multiple linear regression models

$$Y_i = \sum_{j=1}^{p} X_{ij}\beta_j + \epsilon_i,$$

with $\left(Y_i, X_{i1}, \ldots, X_{ip}\right), i = 1, \ldots, n$, i.i.d. observations from $\left(Y, X_1, \ldots, X_p\right)$, satisfying the model $Y = \sum_{j=1}^{p} X_j\beta_j + \epsilon$, where $Y$ is the response, $X_1, \ldots, X_p$ are the $p$ covariates, and $\epsilon$ is the error term with mean 0 and variance $\sigma^2$. We denote $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})'$, for $i = 1, \ldots, n$, with $\mathbf{A}'$ denoting the transpose of a matrix or vector $\mathbf{A}$.

* Corresponding author. Tel.: +32 (0)16 322018.
*E-mail addresses:* irene.gijbels@wis.kuleuven.be (I. Gijbels), inge.vrinssen@wis.kuleuven.be (I. Vrinssen).

http://dx.doi.org/10.1016/j.csda.2014.11.009
0167-9473/

One approach for variable selection is least angle regression (LARS, Efron et al., 2004): this method sequences the candidate predictors in order of importance. Another approach is to add a penalty term $n\lambda \sum_{j=1}^{p} g(\beta_j)$ to the objective function of least squares regression to enforce sparsity of the model. For example, the Least Absolute Shrinkage and Selection Operator (LASSO, Tibshirani, 1996) and Bridge (Frank and Friedman, 1993; Fu, 1998) have an $L_q$-type of penalty on the regression coefficients, i.e. $g(\theta) = |\theta|^q$, with $q = 1$ and $q < 1$ respectively. Fan and Li (2001) use another penalty function, namely the Smoothly Clipped Absolute Deviation (SCAD) penalty. This penalty function $g_\lambda(|\cdot|) = \lambda g(|\cdot|)$ satisfies $g_\lambda(0) = 0$ and has the first-order derivative

$$g'_\lambda(\theta) = \lambda \left\{ I(\theta \le \lambda) + \frac{(a\lambda - \theta)_+}{(a-1)\lambda} I(\theta > \lambda) \right\},$$

for some $a > 2$ and $\theta > 0$. The nonnegative garrote (Breiman, 1995) uses a penalty on shrinkage factors of the regression coefficients. This method starts from an initial estimator, the ordinary least squares estimator (OLS), and then it shrinks or puts some coefficients of the OLS equal to zero using the nonnegative garrote shrinkage factors. Let $\widehat{\beta}_j^{\mathrm{OLS}}$ denote the OLS estimator of the coefficient $\beta_j$, then the nonnegative garrote shrinkage factors $\widehat{\mathbf{c}} = \left(\widehat{c}_1, \ldots, \widehat{c}_p\right)'$ are found by solving

$$\begin{cases} \widehat{\mathbf{c}} = \underset{\mathbf{c}}{\operatorname{argmin}} \left\{ \dfrac{1}{2n} \displaystyle\sum_{i=1}^{n} \left( Y_i - \sum_{j=1}^{p} c_j \widehat{\beta}_j^{\mathrm{OLS}} X_{ij} \right)^2 + \lambda \sum_{j=1}^{p} c_j \right\} \\ \text{s.t. } 0 \le c_j \quad (j = 1, \ldots, p), \end{cases} \tag{1}$$

for given $\lambda > 0$. Breiman (1995) recommends to choose the regularization parameter $\lambda$ with five fold cross-validation. The nonnegative garrote estimator of the coefficient $\beta_j$, $j = 1, \ldots, p$, is then given by

$$\widehat{\beta}_j^{\mathrm{NNG}} = \widehat{c}_j \widehat{\beta}_j^{\mathrm{OLS}}.$$

However, none of these variable selection methods are robust to outliers. Robust versions of the LARS, LASSO and SCAD methods have been considered in the literature. Khan et al. (2007) proposed a robust version of LARS (called RLARS) by replacing the mean, variance and correlation by their robust counterparts to sequence and select the important covariates. A robust regression estimator is then applied to the selected covariates. See also Agostinelli and Salibian-Barrera (2010). Different robust versions of the LASSO have been developed in the literature. See for example Owen (2006). The LAD-LASSO of Wang et al. (2007) is a penalized least absolute deviation estimator, but this method is not robust to leverage points. To overcome this drawback, Arslan (2012) proposed the WLAD-LASSO. In this method the LAD-LASSO is applied to the transformed data set $\left(w_i Y_i, w_i X_{i1}, \ldots, w_i X_{ip}\right)$, $i = 1, \ldots, n$, where the weights $w_i$ are computed using robust distances. The Sparse LTS of Alfons et al. (2013) is a trimmed version of the LASSO and is also robust with respect to vertical outliers and leverage points. A robust version of the SCAD is obtained by Wang and Li (2009) and Wang et al. (2013) proposed penalized robust regression estimators based on the exponential squared loss function, where the penalty function can be of any type.

During the review process of this paper, our attention was drawn to a technical report of Medina and Ronchetti (2014). This paper deals with robust and consistent variable selection for generalized linear and additive models, using as a basis the nonnegative garrote method. In the framework of these models one assumes that the error distribution belongs to the exponential family. Medina and Ronchetti (2014) rely on a quasi-likelihood quantity that is robustified by using a Huber function, and by introducing a weight function (depending on the covariates) that could possibly deal with bad leverage points. The choice of the weight function needs to be made in some way; in their simulation study the authors take this function to be the identity function.

In this paper we consider a linear regression model with unspecified error distribution. Unlike the LARS, the LASSO and the Bridge, the standard nonnegative garrote method (i.e. in case a least squares estimator is used as an initial estimator) is not directly applicable to high-dimensional data (i.e. the case that $p$ is much larger than $n$). When an initial estimator is used that can deal with the high-dimensional case, this disadvantage is also resolved. See Medina and Ronchetti (2014). The theoretical properties of the nonnegative garrotte method are well studied in the literature (Yuan and Lin, 2007) and are extended for variable selection in additive regression models and varying coefficient models by Antoniadis et al. (2012a,b). Extensive simulation studies in these papers, including comparisons with among others the LASSO and SCAD methods, revealed that the nonnegative garrote method overall performs quite well. We therefore, in this paper, robustify the nonnegative garrote for outliers in the response and in the covariates by using robust alternatives to the least squares regression estimator, such as the S-estimator (Rousseeuw and Yohai, 1984) and the least trimmed squares (LTS) estimator (Rousseeuw, 1984). We also develop a reweighting step that is related with the MM-estimator of Yohai (1987), to increase the efficiency of the proposed robust nonnegative garrote method under the normal error model. Our study indicates that a carefully designed robust nonnegative garrote method performs quite well and often outperforms other available methods. This paper thus contributes in a thorough study of the different approaches to robustify the nonnegative garrote method (on all levels of the estimation procedure) for linear regression models.

The rest of the paper is organized as follows. In Section 2, we present three robust versions of the nonnegative garrote, namely the M-, S- and LTS-nonnegative garrote. In Section 3, we give a first simulation study to compare these proposed methods. In Section 4 an extra reweighting step is proposed to improve the results of the S-nonnegative garrote. Section 5