Variable selection with nonnegative garrote in additive models

Anneleen Verhasselt

Katholieke Universiteit Leuven, Department of Mathematics and Leuven Statistics Research Center (LStat)
Celestijnenlaan 200B box 2400, 3001 Heverlee, Belgium

Abstract

The nonnegative garrote was originally proposed by Breiman (1995) for variable selection in a multiple linear regression context. The procedure starts from the ordinary least squares estimator (OLS) and shrinks or puts some coefficients of the OLS equal to zero.

In this work we consider a functional additive model and use P-splines as a basic estimation method. P-splines were introduced by Eilers and Marx (1996) as a univariate flexible smoothing technique. We therefore combine this technique with a backfitting algorithm to deal with the additive modelling. A nonnegative garrote step then takes care of the variable selection issue.

Keywords: additive models; nonnegative garrote; P-splines; variable selection.

AMS: 62G08, 62H99

1. Introduction

We consider an additive model

$$Y_i = f_0 + \sum_{j=1}^{d} f_j(X_{ij}) + \varepsilon_i \quad \text{for } i = 1, \ldots, n, \tag{1}$$

with observations $(Y_i, X_{i1}, \ldots, X_{id})$ from $(Y, X_1, \ldots, X_d)$ where Y is the response, $X_1, \ldots, X_d$ the d explanatory variables and ε the noise term with mean 0 and variance σ^2. Often only a few components f_j are different from 0 and thus important in explaining the response. Therefore we want to select and estimate the non-zero components.

The original nonnegative garrote was proposed by Breiman (1995) for variable selection in multiple linear regression models. Other methods for variable selection include Least Absolute Shrinkage and Selection Operator (LASSO, Tibshirani, 1996) and the Component Selection and Smoothing Operator (COSSO, Lin and Zhang, 2006). The nonnegative garrote has been extended to nonparametric additive models by Cantoni *et al.* (2006) and Yuan (2007). Cantoni *et al.* (2006) formulate the nonnegative garrote for additive models with smoothing splines. The nonnegative garrote uses an initial estimator and searches for shrinkage factors which will lead to a sparser representation. In this work we consider additive models and use P-splines combined with backfitting, to initially estimate the components f_j.

2. Nonnegative garrote

2.1. Original nonnegative garrote

Breiman (1995) proposed the nonnegative garrote for subset regression. It starts from an initial estimator, namely the ordinary least squares estimator (OLS) and it shrinks or puts some coefficients of the OLS equal to zero. The multiple linear regression model is

$$Y_i = \beta_0 + \sum_{j=1}^{d} \beta_j X_{ij} + \varepsilon_i, \quad \text{for } i = 1, \dots, n.$$

The nonnegative garrote shrinkage factors $\hat{c}_j$ are found by solving the following optimization problem

$$\begin{cases} (\hat{c}_1, \dots, \hat{c}_d) = \operatorname{argmin}_{c_1,\dots,c_d} \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0^{\text{OLS}} - \sum_{j=1}^{d} c_j \hat{\beta}_j^{\text{OLS}} X_{ij} \right)^2 \\ \text{s.t. } 0 \le c_j \ (j = 1, \dots, d), \quad \sum_{j=1}^{d} c_j \le s \end{cases}$$

or equivalently

$$\begin{cases} (\hat{c}_1, \dots, \hat{c}_d) = \operatorname{argmin}_{c_1,\dots,c_d} \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0^{\text{OLS}} - \sum_{j=1}^{d} c_j \hat{\beta}_j^{\text{OLS}} X_{ij} \right)^2 + \theta \sum_{j=1}^{d} c_j \\ \text{s.t. } 0 \le c_j \ (j = 1, \dots, d), \end{cases}$$

where $\hat{\beta}_j^{\text{OLS}}$ is the ordinary least squares estimator of the regression coefficient of the j-th component and θ and s are regularization parameters. The nonnegative garrote estimator of the regression coefficient is then

$$\hat{\beta}_j^{\text{NNG}} = \hat{c}_j \hat{\beta}_j^{\text{OLS}}.$$

In the special case that the design is orthogonal, i.e. $X'X = I_{n \times n}$ (where X is the design matrix), the nonnegative garrote estimates are the following

$$\hat{c}_j = \left(1 - \frac{\theta}{2(\hat{\beta}_j^{\text{OLS}})^2} \right)_+ .$$

This nonnegative garrote function $c(\theta) = \left(1 - \frac{\theta}{2(\hat{\beta}^{\text{OLS}})^2} \right)_+$ is presented in Figure 1 for different values of θ.

2.2. Functional nonnegative garrote

The nonnegative garrote of Breiman was extended to additive models of the form (1) by Cantoni *et al.* (2006) and Yuan (2007).

In the functional nonnegative garrote procedure we start with an initial estimator $\hat{f}_j^{\text{init}}(X_j)$ for the function describing the j-th component. This replaces $\hat{\beta}_j^{\text{OLS}} X_j$ in the original nonnegative garrote. The nonnegative garrote shrinkage factors are then found by solving

$$\begin{cases} \min_{c_1,\dots,c_d} \sum_{i=1}^{n} \left(Y_i - \hat{f}_0^{\text{init}} - \sum_{j=1}^{d} c_j \hat{f}_j^{\text{init}}(X_{ij}) \right)^2 \\ \text{s.t. } 0 \le c_j \ (j = 1, \dots, d), \ \sum_{j=1}^{d} c_j \le s. \end{cases}$$

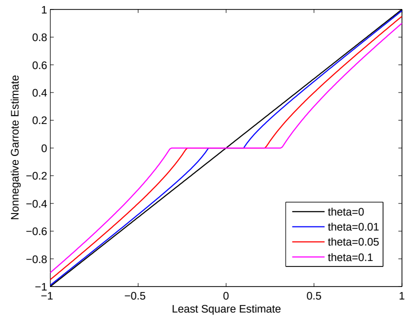

Figure 1: Shrinkage effect of the nonnegative garrote

The resulting nonnegative garrote estimate of the j-th component is then

$$\hat{f}_j^{\mathrm{NNG}} = \hat{c}_j \hat{f}_j^{\mathrm{init}}.$$

The consistency of the nonnegative garrote with P-splines is established in Antoniadis *et al.* (2009). The proof relies on a consistency result for (univariate) P-splines, on an extension of a univariate smoothing estimator to additive models via backfitting and on a consistency result for the functional nonnegative garrote. Simulation studies and real data applications are given in Antoniadis *et al.* (2009).

Acknowledgements: This research is joint work with Anestis Antoniadis (Université Joseph Fourier) and Irène Gijbels (Katholieke Universiteit Leuven) and supported by the IAP Research Network P6/03 of the Belgian State (Belgian Science Policy) and the Research Fund of the KULeuven (GOA/07/04-project) .

REFERENCES

[1] Antoniadis, A., Gijbels, I. & Verhasselt, A. (2009), Nonnegative garrote in additive models using P-splines, *Manuscript*.

[2] Breiman, L. (1995), Better subset regression using the nonnegative garrote, *Technometrics*, 37, 373-384.

[3] Cantoni, E., Flemming, J. & Ronchetti E. (2006), Variable Selection in Additive Models by Nonnegative Garrote, *Cahier du Département d'Econométrie 2006.02*, University of Geneva (Switzerland).

[4] Eilers, P. & Marx, B. (1996), Flexible smoothing with B-splines and penalties, *Statistical Science*, 11, 89-102.

[5] Yuan, M. (2007), Nonnegative Garrote Component Selection in Functional ANOVA models, *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics March 21-24, 2007, San Juan, Puerto Rico*, 2, 660-666.