Journal of Statistical Planning and Inference 142 (2012) 1733–1745

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

Profiled adaptive Elastic-Net procedure for partially linear models with high-dimensional covariates

Baicheng Chen [a], Yao Yu [b], Hui Zou [c], Hua Liang [b,*]

[a] Department of Statistics, Shanghai University of Finance and Economics, Shanghai, China
[b] Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA
[c] Department of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 8 July 2011
Accepted 15 February 2012
Available online 23 February 2012

Keywords:
Adaptive regularization
Elastic-Net
High dimensionality
Model selection
Oracle property
Presmoothing
Semiparametric model
Shrinkage methods

ABSTRACT

We study variable selection for partially linear models when the dimension of covariates diverges with the sample size. We combine the ideas of profiling and adaptive Elastic-Net. The resulting procedure has oracle properties and can handle collinearity well. A by-product is the uniform bound for the absolute difference between the profiled and original predictors. We further examine finite sample performance of the proposed procedure by simulation studies and analysis of a labor-market dataset for an illustration.

1. Introduction

Consider the partial linear models (PLM Härdle et al., 2000)

$$Y = X^T\beta + g(Z) + \varepsilon, \tag{1.1}$$

where $X = (x_1, \ldots, x_p)^T$ and Z are the linear and nonparametric components, $g(\cdot)$ is an unknown smooth function. We are interested in variable selection procedure for parametric components X when the dimension number p is large, which may depend upon the sample size. We propose to use the adaptive Elastic-Net (Zou and Zhang, 2009) for variable selection in the PLM using profile least squares approach to convert the partial linear models to the classical linear regression model.

In the past decade, we have witnessed great progress in variable selection for a variety of models, since two elegant penalized based methods, the least absolute shrinkage and selection operator (LASSO) penalty (Tibshirani, 1996) and the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li, 2001, 2002), had been proposed. A large body of penalized methods has been studied in the literature. See for example Zou and Hastie (2005), Meinshausen and Bühlmann (2006), Zou (2006, 2008), Zhao and Yu (2006), Huang et al. (2008, 2009), and van de Geer (2008). Recently, researchers have also considered applications of penalization methods in semiparametric and nonparametric models. For instance, Li and Liang (2008) for semiparametric models and Liang and Li (2009) for partially linear models with measurement errors. Huang et al. (2010) and Ravikumar et al. (2009) investigated high-dimensional nonparametric sparse additive models. Xie and

* Corresponding author.
E-mail address: hliang@bst.rochester.edu (H. Liang).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2012.02.035

Huang (2009) and Ni et al. (2009) studied variable selection for partially linear models with a divergent number of linear covariates. It is noteworthy that Xie and Huang (2009) required $p^2/n \to 0$ as $n \to \infty$.

The PLM, as a trade-off between linear models and additive models, that replace one linear component by a nonparametric function, have been studied well in literature and widely been used to explore the complicated relation between a response to treatment and predictors of interest (Härdle et al., 2000). See for example Opsomer and Ruppert (1999), Zeger and Diggle (1994), Severini and Staniswalis (1994), Robinson (1988), Speckman (1988), and Engle et al. (1986). An attractive feature of PLM in that a presmoothing procedure can transfer them to a standard linear model. Correspondingly, the linear parameters can be estimated at the root$-n$ rate under certain conditions. Note that redundant variables may enter the PLM when many covariates are collected, and should be excluded for a final model. The motivation for this paper is whether we can adopt variable selection procedure which was originally developed for linear models for the PLM. If so, variable selection for PLM becomes easier to handle. We confirm this conjecture for the adaptive Elastic-Net procedure. However, it is not trivial to give a theoretical justification since the "synthetic" data based on the presmoothed model are not independent.

The rest of the article is organized as follows. Section 2 introduces the presmoothing procedure, adaptive Elastic-Net procedure and its variant for the PLM. Section 3 presents the asymptotic properties for the proposed procedure. The resulting estimator is shown to have an oracle property. Monte Carlo simulations show the proposed procedure works well with moderate sample sizes. An empirical example is examined to illustrate the application of the method. All technical proofs are left to the Appendix.

2. Profiled adaptive Elastic-Net

2.1. Profiled responses and predictors

In partially linear models, the profile least squares approach has been used to convert the semiparametric model to the linear setting (Fan and Huang, 2005; Speckman, 1988). Note that $E(Y|Z) = \{E(X|Z)\}^T\beta + g(Z)$. It follows from this approach that

$$Y - E(Y|Z) = \{X - E(X|Z)\}^T\beta + \varepsilon, \tag{2.1}$$

which is a standard linear model if $E(Y|Z)$ and $E(X|Z)$ are known, and we may adopt the procedures developed in Zou and Zhang (2009) to study variable selection for the partially linear models. Our strategy is that we first nonparametrically estimate two conditional expectations $E(Y|Z)$ and $E(X|Z)$ and then substitute the two estimates in (2.1). Through this presmoothing technique, we can then develop a variable selection procedure for the partially linear models.

Let $(X_1,Z_1,Y_1),\ldots,(X_n,Z_n,Y_n)$ be an iid sample of size n from model (1.1). Let $\mathbf{X} = (X_1^T,\ldots,X_n^T)^T$, $\mathbf{Y} = (Y_1,\ldots,Y_n)^T$ and similarly for $\mathbf{Z}$. In matrix notation, (1.1) can be expressed as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{g} + \boldsymbol{\epsilon}. \tag{2.2}$$

In this paper, we use locally linear procedure to estimate $E(Y|Z)$ and $E(X|Z)$ (Fan and Gijbels, 1996), denote these estimates as $\widehat{E}(Y|Z)$ and $\widehat{E}(X|Z)$. In what follows, we define $m_x(z) = E(X|Z=z)$, $m_y(z) = E(Y|Z=z)$, $\tilde{X}_i = X_i - E(X_i|Z_i)$, $\tilde{Y}_i = Y_i - E(Y_i|Z_i)$, $\widehat{X}_i = X_i - \widehat{E}(X_i|Z_i)$, and $\widehat{Y}_i = Y_i - \widehat{E}(Y_i|Z_i)$.

Let $\mathcal{A} = \{j : \beta_j \neq 0, j = 1,2,\ldots,p\}$, called the size of $\mathcal{A}$ the intrinsic dimension of the underlying model. We wish to uncover the set $\mathcal{A}$ and estimate the corresponding coefficients.

2.2. The choice of penalty function

After profiling, we can apply the popular penalized least squares technique to the "synthetic" data. As discussed in the Introduction section, there are many nice penalty functions proposed in the literature. In this work we choose to use the adaptive Elastic-Net penalty. As nicely demonstrated in Zou and Zhang (2009), the adaptive Elastic-Net combines the strengths of adaptive ℓ_1 penalization (Zou, 2006) and the power of quadratic regularization to handle the collinearity problem which often appears in real data analysis.

The adaptive Elastic-Net procedure has two steps. First, we construct the Elastic-Net estimator defined as follows:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{EL}} = \left(1 + \frac{\lambda_2}{n}\right) \arg\min_{\boldsymbol{\beta}} \{\|\widehat{\mathbf{y}} - \widehat{\mathbf{X}}\boldsymbol{\beta}\|_2^2 + \lambda_2\|\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1\}. \tag{2.3}$$

The Elastic-Net does not possess the oracle properties of SCAD. In the second step, we use the Elastic-Net estimator to construct the adaptive weights by

$$\hat{w}_j = (|\hat{\beta}_{j,\mathrm{EL}}|)^{-\gamma}, \quad j = 1,2,\ldots,p, \tag{2.4}$$