Journal of Statistical Planning and Inference 174 (2016) 52–67

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Nonnegative adaptive lasso for ultra-high dimensional regression models and a two-stage method applied in financial modeling

Yuehan Yang [a,*], Lan Wu [b]

[a] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, PR China

[b] School of Mathematical Sciences, Peking University, Beijing 100871, PR China

ARTICLE INFO

Article history:
Received 8 January 2016
Accepted 30 January 2016
Available online 17 February 2016

Keywords:
Nonnegative adaptive lasso
Multiplicative updates
Index tracking
Variable selection consistency
Asymptotic unbiasedness
Asymptotic normality

ABSTRACT

This paper proposes the nonnegative adaptive lasso method for variable selection both in the classical fixed p setting (OLS initial estimator) and the ultra-high dimensional setting (root-n-consistent initial estimator). This method is an extension of the adaptive lasso with nonnegative constraint on the coefficients. It is shown to have asymptotic unbiasedness, asymptotic normality and variable selection consistency and its mean squared error decays fast too. Comparing with other procedures, nonnegative adaptive lasso satisfies oracle properties and can select the true variables under fewer assumptions. To get the solution of the nonnegative adaptive lasso, we extend the multiplicative approach for computing. This algorithm is valid for the general framework where the number of regression parameters p is allowed to very large. Simulations are performed to illustrate above results.

The constrained index tracking problem in the stock market without short sales is studied in the empirical part. A two-stage method, nonnegative adaptive lasso+nonnegative LS, is applied in the financial modeling. The tracking results indicate that nonnegative adaptive lasso and the two-stage method can both get small tracking error and is successful in assets selection.

1. Introduction

The advent of computer technology makes it easier to collect massive data-sets, especially in the financial field. An enduring interest in the financial field is doing portfolio management by using high dimensional data or even ultra-high dimensional data (where the dimensionality can grow exponentially with the sample size), that is, the dimensionality (hundreds even thousands of stocks) always exceeds the number of observations. Furthermore, nonnegative constraint appears frequently, e.g. time measurements, histograms data, intensity values of an image and economic quantities such as prices, incomes and growth rates. Yet, there are still gaps in our understanding that famous regularizations cannot handle this type of data very well. Similarly, related algorithms (e.g. Lars, Glmnet, etc.) cannot be used on this situation directly.

We consider a linear regression model.

$$\boldsymbol{y}_n = \boldsymbol{X}_n \boldsymbol{\beta}_n^* + \boldsymbol{\epsilon}_n, \tag{1}$$

* Corresponding author.
E-mail address: yyh@cufe.edu.cn (Y. Yang).

http://dx.doi.org/10.1016/j.jspi.2016.01.011
0378-3758/

where $(\boldsymbol{x}_{i,n}, y_{i,n}) \in \mathbb{R}^p \times \mathbb{R},\ i = 1, 2, \ldots, n$ are the predictors and response. $\boldsymbol{X}_n$ is a $n \times p$ design matrix which is assumed to be a deterministic one. $\boldsymbol{y}_n = (y_{1,n}, y_{2,n}, \ldots, y_{n,n})'$ is a n response vector and $\boldsymbol{\epsilon}_n = (\epsilon_{1,n}, \epsilon_{2,n}, \ldots, \epsilon_{n,n})'$ is a vector of independent identically distributed random variables with mean 0 and variance σ^2. $\boldsymbol{\beta}^*$ is the true regression. The data and model parameters are indexed by n to allow them to change as n grows. For simplicity of presentation. We replace $\boldsymbol{y}_n$ $\boldsymbol{\beta}_n^*$, $\boldsymbol{X}_n$ by $\boldsymbol{y}$, $\boldsymbol{\beta}^*$, $\boldsymbol{X}$ in the following.

Regularizations are vastly popular and successful in statistical modeling, especially in high-dimensional data. Lasso (Tibshirani, 1996), regularized with an l_1 penalty, is largely popularize since it generates sparse solutions. Elastic net (Zou and Hastie, 2005), regularized with a combination of the l_1 and l_2, is also a successful idea in this category. However, two methods both have drawbacks. One is the bias problem. To fix this problem, Zou (2006) proposed the adaptive lasso, where adaptive weights are used for penalizing different coefficients in the l_1 penalty. It has been showed that the adaptive lasso enjoys the oracle properties[1] (Fan and Li, 2001).

Nonnegative constraint, $\boldsymbol{\beta}^* \in \mathbb{R}_+^p$ is assumed to be nonnegative, is common in various fields especially the financial problem. (Histograms data, time measurements, for instance, also need sparse solution). Since nonnegative constraint on the coefficients is commonly used due to its relative simplicity and often performs well in practice. Breiman (1995) proposed the nonnegative garrotte estimator, which showed to be a stable variable selection method that outperforms its competitors including subset regression and ridge regression. Yuan and Lin (2007) studied the consistency, computation and flexibility of this procedure. Slawski and Hein (2013) introduced the performance of the non-negative least squares (NNLS) and showed that NNLS has better l_∞-rate in estimation and hence advantages with respect to support recovery when combined with thresholding. Similarly, Meinshausen (2013) showed that sign-constrained least squares estimation is an effective regularization technique for a certain class of high-dimensional regression problems. Wu et al. (2014) proposed the nonnegative lasso and proved that it is variable selection consistency and estimation consistency in high dimensional sparse linear regression models with weaker assumptions than original lasso. Wu and Yang (2014) proposed the nonnegative elastic net compared with both original elastic net and nonnegative lasso.

These outstanding statistical techniques also can be used to solve financial problems, e.g. investment portfolio, bankruptcy prediction, etc. Investment portfolio, for instance, we can regard the time dimension as sample size which is lower than the spatial dimension. It plays the same rules as the parameters in high-dimensional data analysis.

Furthermore, several literatures have studied the asymptotic properties of these regularizations. For the lasso part, Knight and Fu (2000) showed that the lasso is consistent for estimating the regression parameter when p is fixed as $n \to \infty$. Zhao and Yu (2006) proved a single condition called the Irrepresentable Condition (Meinshausen and Bühlmann, 2006; Tropp, 2004; Wainwright, 2009), which is almost necessary and sufficient for the lasso to select the true model. Consequently, there exist certain scenarios where the lasso is inconsistent for variable selection. For the adaptive lasso, it was shown to have the oracle properties even in generalized linear models (Zou, 2006). Jian et al. (2008) studied the asymptotic properties of the adaptive lasso estimators in sparse, high-dimensional, linear regression models when the number of covariates increase with the sample size.

A large number of researchers work on the improvement of above methods and its applications in other fields. Zou and Zhang (2009) combined the ideas of the adaptive weighted l_1 penalty and the Elastic Net regularization to obtain a method which can improve the lasso in both direction, called Adaptive elastic net. Candes and Tao (2007) established the Dantzig selector and Bickel et al. (2009) established the asymptotic equivalence of the lasso and the Dantzig selector. Zhao et al. (2009) and Yuan and Lin (2006) studied in group and hierarchical lasso estimators. In the financial field, Wang et al. (2007) extended lasso to the REGAR (regression model with autoregressive errors) time series model and got better performance in terms of accuracy of prediction.

Original adaptive lasso have been shown very successful of regularizations and lots of researchers work on the improvement of the adaptive lasso (Lee et al., 2010; Wang and Leng, 2008; Zou and Zhang, 2009). However, there is no literature discussed about the adaptive lasso with nonnegative constraint before. Thus, we propose nonnegative adaptive lasso and study its properties.

In this paper, **our contributions** are summarized as follows

1. We propose a valid procedure, called nonnegative adaptive lasso, where adaptive weights are used for penalizing different nonnegative coefficients in l_1 penalty. More specifically, we consider two types of initial estimator for the weight vectors: $\hat{\boldsymbol{\beta}}(ols)$ and $\widetilde{\boldsymbol{\beta}}$ ($\widetilde{\boldsymbol{\beta}}$ is r_n-consistent).
2. We review results from literatures on crucial condition for each method, e.g. Irrepresentable Condition for the lasso, Elastic Irrepresentable Condition for the elastic net. And establish a new one, the Nonnegative Adaptive Irrepresentable Condition (NAIC) to guarantee nonnegative adaptive lasso variable selection consistency under ultra-high dimensional setting. Compared with other conditions, NAIC is much weaker since it can be satisfied by a proper choice of $\hat{\boldsymbol{w}}$.

[1] This procedure is proposed by Fan and Li (2001), which considered that a good penalty function should result in an estimator with three properties: unbiasedness, sparsity, continuity. They proposed oracle properties as follows (in language similar to Fan and Li, 2001 and Zou, 2006).

Oracle Property: $\hat{\boldsymbol{\beta}}$ is an oracle estimator if it is satisfy

1. Identifies the right subset model.
2. Asymptotic normality.