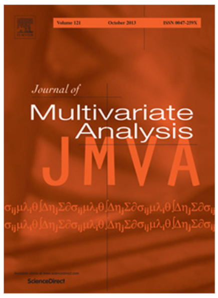
Accepted Manuscript



On the oracle property of a generalized adaptive elastic-net for multivariate linear regression with a diverging number of parameters

Xin Xin, Jianhua Hu, Liangyuan Liu

PII:	S0047-259X(17)30091-X
DOI:	http://dx.doi.org/10.1016/j.jmva.2017.08.005
Reference:	YJMVA 4282

To appear in: *Journal of Multivariate Analysis*

Received date : 14 February 2017

Please cite this article as: X. Xin, J. Hu, L. Liu, On the oracle property of a generalized adaptive elastic-net for multivariate linear regression with a diverging number of parameters, *Journal of Multivariate Analysis* (2017), http://dx.doi.org/10.1016/j.jmva.2017.08.005

On the oracle property of a generalized adaptive elastic-net for multivariate linear regression with a diverging number of parameters

Xin Xin[1], Jianhua Hu[2] and Liangyuan Liu[2]

[1]*School of Mathematics and Statistics, Henan University, Henan, China;*
[2]*School of Statistics and Management, and Key Laboratory of Mathematical Economics, Shanghai University of Finance and Economics, Shanghai, China*

Abstract

In this paper, we examine the problem of variable selection and coefficient estimation in multivariate linear regression with a diverging number of parameters. We propose a generalized adaptive elastic-net method that integrates elastic-net regularization, adaptively weighted penalty, and covariance of errors. Under some weak regularity conditions, we establish the oracle property of generalized adaptive elastic-net estimators. We also present an algorithm based on the subgradient method to solve the generalized adaptive elastic-net estimation. The numerical computation results show that the proposed method has variable selection accuracy that is comparable to those of the existing methods but that it also has more correct-fitting, less over-fitting, and is closest to the oracle estimator in terms of mean squares error. Moreover, we analyze a chemometrics data set to illustrate our methodology.

AMS 2000 subject classification: 62H12; secondary 62A01.

Keywords: Generalized adaptive elastic-net, group variable selection, multivariate linear regression, oracle property.

1. Introduction

Multivariate linear regression analysis is one of the most useful techniques in statistical science. It is widely used in chemometrics, econometrics, financial engineering, genetics, psychometrics, and many other areas of study to model the predictive relationships of multiple related responses on a set of common predictors. For example, in genetic, modeling gene expression levels at multiple time points using multiple transcription factors may yield interesting results. In stock markets, modeling the relationships and returns of multiple stocks using a set of common econometric variables is crucial. Because not all common predictor variables are essential for an analysis, variable selection is necessary in multivariate regression, especially as high-dimensional data become increasingly common.

Numerous researchers adopt sparsity-inducing regularization techniques under multivariate regression frameworks. In this technique, the rows of the regression coefficient matrix are regarded as special row groups of parameters. Turlach et al. [21] proposed a model for predictors as a natural extension of the Lasso technique developed by Tibshirani [20] who constrained the parameter estimates to within a suitable polyhedral region using the (joint) residual sum of squares. Simila and Tikka [18] addressed the penalized least squares problem by applying an ℓ_2-penalty to the row groups. Peng et al. [17] imposed both row- and element-wise sparsity of the coefficient matrix by using the penalized least squares problem and a penalty on the combination between distances of groups and the sum of the absolute values of components in the row groups. Chun and Keles [6] developed a sparse partial least squares by imposing constraints on the optimization criterion for partial least squares. Obozinski et al. [16] studied the behavior of the multivariate row group Lasso, in which block regularization based on the ℓ_1/ℓ_2-norm is used to support union recoveries. Hu et al. [11] proposed a model determination procedure with a SCAD (smoothly clipped absolute deviation) penalty to the row and column groups for the growth curve model, a generalized multivariate linear model.

Integrating the regularization technique and the reduced-rank method, Yuan et al. [25] proposed a regularized least squares approach for simultaneous conduct dimension reduction and coefficient estimation. The penalty they imposed encourages sparsity in the singular values of the coefficient matrix so that the rank can be automatically reduced and