Simultaneous Response and Predictor Selection Model and Estimation to Multivariate Linear Regression

Jianhua Hu[1], **Jian Huang**[2] **and Feng Qiu**[3]

[1]*Shanghai University of Finance and Economics, China*
[2] *University of Iowa, Iowa City, Iowa USA*
[3] *Zhejiang Agriculture and Forestry University, China*

Abstract

The response best subset selection model [4] addresses the problems in which predictors are known and only responses need to be selected in multivariate linear models. In practice, we couldn't make sure not only responses but also predictors and need variable selection simultaneously for responses and predictors, for which no research has been found. In this paper, we propose a novelty simultaneous response and predictor selection (SRPS) model, which is motivated by applications where some responses or predictors are unimportant in multivariate linear regression analysis. We simultaneously investigate variable selection both for responses and predictors and estimation to regression coefficients in the standard multivariate linear regression, group adaptive lasso and the response best subset selection contexts. We also establish model consistency, consisting of response selection, predictor selection and coefficient estimation, and the oracle property of coefficient estimators. Our simulation studies suggest that the proposed method is pronouncedly efficient. We also apply our methodology to study a real data set.

Keywords: Group adaptive lasso, model consistency, multivariate linear regression, response best subset selection model, response selection, simultaneous response and predictor selection model

References

[1] Anderson, T.W. (2003). An Introduction to Multivariate Statistical Analysis (3rd Edition). John Wiley & Sons, Inc., Publication.

[2] Breheny, P. and Huang, J. (2009). Penalized methods for bi-level variable selection. *Statistics and Its Interface*, 2(3):369-380.

[3] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348-1360.

[4] Hu, J., Huang, J. and Qiu, F. (2018). Response best subset selection model and efficient estimation in multivariate linear regression. *Technical Report*, Shanghai University of Finance and Economics.

[5] Huang, J., Ma, S., Xie, H., and Zhang, C.-H. (2009). A group bridge approach for variable selection. *Biometrika*, 96(2):339-355.

[6] Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics*, 28(5):1356-1378.

[7] Li, R., Zhong, W., and Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107(499):1129-1139.

[8] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. Hohn Wiley & Sons, Inc.

[9] Su, Z., Zhu, G., Chen, X. and Yang, Y. (2016). Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. *Biometrika*, 103(3), 579-593.

[10] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):267-288.

[11] Wang, H. and Leng, C. (2008). A note on adaptive group lasso. *Computational Statistics and Data Analysis*, 52: 5277-5286.

[12] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49-67.

[13] Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894-942.

[14] Zhang, H.H. and Lu, W. (2007). Adaptive lasso for Coxs proportional hazard model. *Biometrika*, 94:691-703.

[15] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418-1429.

[16] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301- 320.