Comparison of Multivariate Estimation Methods

Raju Rimal[a,*], Trygve Almøy[a], Solve Sæbø[b]

[a]*Faculty of Chemistry and Bioinformatics, Norwegian University of Life Sciences, Ås, Norway*
[b]*Prorector, Norwegian University of Life Sciences, Ås, Norway*

Abstract

Prediction performance often does not reflect the estimation behaviour of a method. High error in estimation not necessarily results in high prediction error but can leads to an unreliable prediction when test data are in a different direction than the training data. In addition, the effect of a variable becomes unstable and can not be interpreted in such situations. Many research fields are more interested in these estimates than performing prediction. This study compares some newly-developed (envelope) and well-established (PLS, PCR) prediction methods using simulated data with specifically designed properties such as multicollinearity, correlation between multiple responses and position of principal components of predictor that are relevant for the response. This study aims to give some insight on these methods and help researcher to understand and use them for further study. *Write some specifics from the results to show what we have found.*

Keywords: model-comparison,multi-response,simrel,estimation,estimation error,meta modeling

1. Introduction

Estimation of parameters in a regression model is an integral part in many research study. Research fields such as social science, econometric, psychology and medical study are

*Corresponding Author
Email addresses: raju.rimal@nmbu.no (Raju Rimal), trygve.almoy@nmbu.no (Trygve Almøy), solve.sabo@nmbu.no (Solve Sæbø)

more interested in measuring the impact of certain indicator or variable rather than performing prediction. Such studies has large influence in people's perception and also help in policy making and decisions.

Technology has facilitated researcher to collected large amount of data however often times, such data either contains irrelevant information or are highly collinear. Researchers are devising new estimators to extract information and identify their interrelationship. Some estimators are robust towards fixing multicollinearity problem while some are targeted to model only the relevant information content in response variable.

This study extends the (Rimal et al.) and compares some well established estimators such as Principal Components Analysis (PCA), Partial Least Squares (PLS) together with two new methods based on envelope estimation: Envelope estimation in predictor space (Xenv) (Cook et al., 2010) and simultaneous estimation of envelope (Senv) (Cook and Zhang, 2015). The estimation process of these methods are discussed in [Methods] section. The comparison tests the estimation performance of these methods using multi-response simulated data from linear model with controlled properties. The properties includes the number of predictors, level of multicollinearity, correlation between different response variables and the position of relevant predictor components. These properties are explained in [Experimental Design] section together with the strategy behind the simulation and data model.

References

Cook, R.D., Li, B., Chiaromonte, F., 2010. Envelope Models for Parsimonious and Efficient Multivariate Linear Regression. Statistica Sinica 20, 927–1010.

Cook, R.D., Zhang, X., 2015. Simultaneous envelopes for multivariate linear regression. Technometrics 57, 11–25. doi:doi:10.1080/00401706.2013.872700.

Rimal, R., Almøy, T., Sæbø, . Comparison of multivariate prediction methods.