Comparison of PLS algorithms when number of objects is much larger than number of explanatory variables

Aylin Alin

Dokuz Eylul University, Turkey

Abstract

Partial Least Squares Regression (PLSR) is a latent variable based multivariate technique which allows predicting multiple response variables, $\mathbf{Y}$, from collinear multiple explanatory variables, $\mathbf{X}$. Latent variables can be obtained using different algorithms among which NIPALS and SIMPLS are the mostly used. When the number of explanatory variables, K, is much more than the number of objects, N, NIPALS algorithm can be time consuming. Even though SIMPLS is not as time consuming as NIPALS and can be preferred over NIPALS, there are kernel algorithms developed especially for the cases N>>K. In this study, NIPALS algorithm, SIMPLS algorithm and kernel algorithms developed by Lindgren et al. (1993), De Jong and Ter Braak (1994) and Dayal and MacGregor (1997) have been compared in terms of total CPU time spent for the calculations of latent variables, one-fold cross validation and bootstrap test with 100 bootstrap samples. According to results kernel algorithms suggested by Dayal and MacGregor (1997) are faster than other algorithms under the considered designs.

Keywords

Multicollinearity, Multiple Linear Regression, Partial Least Squares (PLS).

References

B.S. Dayal, B.S. and J.F. MacGregor (1997). Improved PLS Algorithms. *J. Chemom. 11*, 73–85.

Lindgren, F., P. Geladi, and S. Wold (1993). The Kernel Algorithm for PLS. *J. Chemom. 7*, 45–59.

De Jong, S. and C.J.F. Ter Braak (1994). Comments on PLS Kernel Algorithm. *J. Chemom. 8*, 169–174.