SEEDCCA: An Integrated R-Package for Canonical Correlation Analysis and Partial Least Squares

by Bo-Young Kim, Yunju Im, and Jae Keun Yoo

Abstract Canonical correlation analysis (CCA) has a long history as an explanatory statistical method in high-dimensional data analysis and has been successfully applied in many scientific fields such as chemometrics, pattern recognition, genomic sequence analysis, and so on. The so-called seedCCA is a newly developed R package that implements not only the standard and seeded CCA but also partial least squares. The package enables us to fit CCA to large-p and small-n data. The paper provides a complete guide. Also, the seeded CCA application results are compared with the regularized CCA in the existing R package. It is believed that the package, along with the paper, will contribute to high dimensional data analysis in various science field practitioners and that the statistical methodologies in multivariate analysis become more fruitful.

Bo-Young Kim

Yunju Im

Jae Keun Yoo