Partial Least Squares for binary data and its associated biplot applied to the classification of Colletotrichum Graminicola strains

Laura Vicente-Gonzalez[1], *Jose Luis Vicente-Villardon*[2]

[1]laura20vg@usal.es, Departmento de Estadística, Universidad de Salamanca
[2]villardon@usal.es, Departmento de Estadística, Universidad de Salamanca

In this work we propose a generalization of Partial Least Squares Regression where all the variables, responses and predictors, are binary. The method is named Binary Partial Least Squares (BPLS). A representation for BPLS, that combines two logistic biplots for responses and predictors, is also described.

The final algorithm is based on a generalization of NIPALS to handle binary variables, extending also a procedure recently proposed by the authors.

The method is applied to the classification of several strains of Colletotrichum Graminicola using RNA data. The differences among nine strains, corresponding to their countries of origin, and the genes that characterize them are studied.

For the calculations we have used the `R` software. New functions have been included in the package `MultBiplotR`.

Keywords: Binary Data, PLS, Biplot