

plsRglm, PLS generalized linear models for

Frédéric Bertrand^{1,2,*}, Jérémy Magnanensi^{1,2,3}, Nicolas Meyer^{1,3}, Myriam Bertrand^{1,2}

1. Université of Strasbourg

2. Centre national de la recherche scientifique

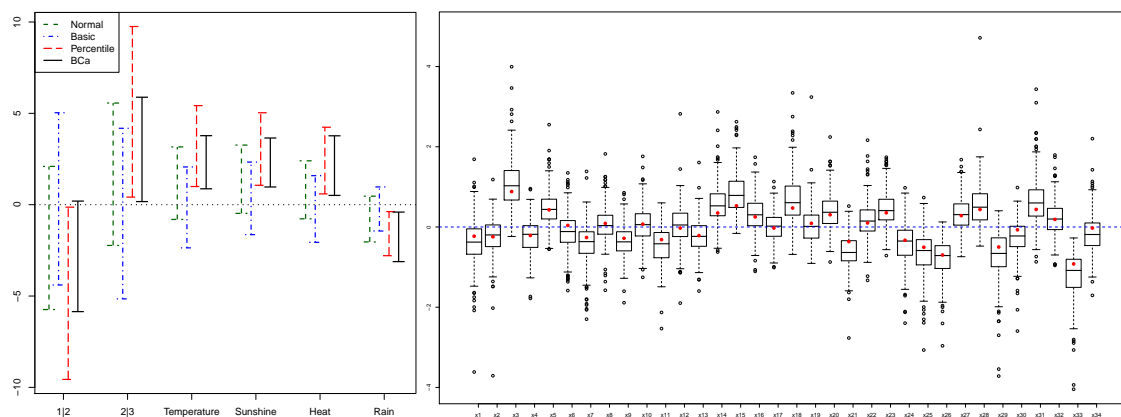
3. Institut national de la santé et de la recherche médicale

*Contact author: bertran@math.unistra.fr

Keywords: Partial least squares regression, generalized linear models, bootstrap, high dimensional data, R software package

There are mainly two aims for the **plsRglm** library written by the authors for the R language. The extension of PLS regression to generalized linear models, and for instance to logistic regression models [2], and the need to provide tools to PLS users to deal with incomplete datasets using cross-validation.

These models were successfully applied to datasets of various kind We will provide six examples of use: PLSR model fitted to a Mixture design dataset (Cornell,[3]), PLSR-multinomial logistic model fitted to a Bordeaux wine quality dataset (bordeaux, [2]) and to a Chemotaxonomy dataset (hyptis from library **chemometrics**, [5]), PLSR-binary logistic model fitted to a Microarray Colon Cancer dataset (ColonCA, [1]), to a Chemometrics dataset (phenyl dataset from library **chemometrics**,[5]) and to allelotyping data (aze_comp1, [4] ; Figure 1 features bootstrap distribution of the coefficients of the predictors and Figure 2 balanced bootstrap confidence intervals).



References

- [1] Alon, U., N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and L. A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissue probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* 96, 6745–6750.
- [2] Bastien, P., V. Esposito Vinzi, and M. Tenenhaus (2005). Pls generalised linear regression. *Computational Statistics & Data Analysis* 48, 17–46.
- [3] Kettaneh-Wold, N. (1992). Analysis of mixture data with partial least squares. *Chemometrics & Intelligent Laboratory Systems* 14, 57–69.
- [4] Meyer, N., M. Maumy-Bertrand, and F. Bertrand (2010). Comparaison de variantes de régressions logistiques pls et de régression pls sur variables qualitatives : application aux données d’allélotypage. *Chemometrics & Intelligent Laboratory Systems* 151, 1–18.
- [5] Varmuza, K. and P. Filzmoser (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics*. Boca Raton: CRC Press.