R algorithms for the calculation of markers to be used in the construction of predictive and interpolative biplot axes in routine multivariate analyses.

M. Rui Alves[1,2] and M. Beatriz Oliveira[1]

[1]REQUIMTE, Serviço de Bromatologia, Faculdade de Farmácia, Universidade do Porto
R. Aníbal Cunha, 164, 4099-030 PORTO, Portugal
[2]ESTG / Instituto Politécnico de Viana do Castelo, Av. Atlântico, s/n, 4901-908 Viana do Castelo, Portugal. mruialves@clix.pt

Multivariate analyses are used to search for the main data structures of data sets, and whenever possible provide the means to display those structures as graphs. However, for inexperienced users of statistics, such outputs are sometimes difficult to interpret.

Predictive biplots[1] can be used to carry out interpretations in relation to initial values and variables instead of latent variables, without loosing the benefits of the multivariate modulation. Once multivariate graphs are produced and printed, the interpolative biplots[1] can be used in routine laboratory practice to position new samples in the graph.

To achieve the objectives of biplots, one biplot axis for each initial variable must be drawn in the graphs. First one decides on convenient scale values and calculate the coordinates of each scale value in the multidimensional subspaces; then, one marker is drawn in the graphs to indicate the position of the scale value; afterwards, all scale values are linked together by a straight line to represent the variable's axis, all markers are labelled with the original scale values, and the axis is also labelled to identify the variable in question.

Biplots can be used with advantage in the solution of many multivariate problems[2] and also in complex situations where different multivariate techniques are coupled together[3].
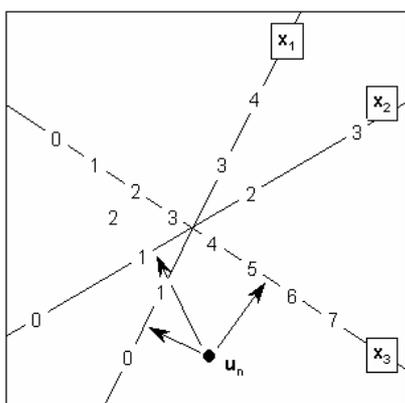
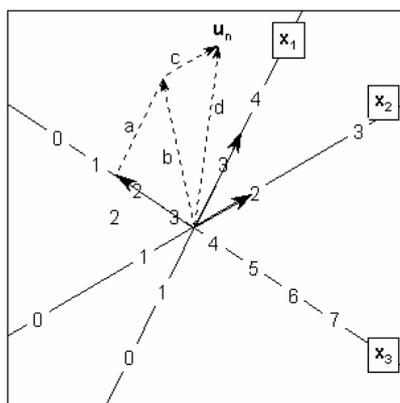

Fig. 1: Predictive biplot

Fig. 2: Interpolative biplot

Fig. 1 shows how a predictive biplot is used to read off from the graph the sample $\mathbf{u}_n$ initial values: $\mathbf{x}_1 \approx 0.5$, $\mathbf{x}_2 \approx 1.1$ and $\mathbf{x}_3 \approx 5.4$.

Fig. 2 shows how a given sample $\mathbf{u}_n$ with coordinates $\mathbf{x}_1 = 3.5$, $\mathbf{x}_2 = 2.0$ and $\mathbf{x}_3 = 1.3$ is interpolated in the two-dimensional space by the complete parallelogram's method.

This paper presents complete algorithms in R, adapted from our algorithms initially written in Genstat 5.3.1, to carry out principal components and canonical variate analyses, and produce the markers for predictive and interpolative biplots, exemplified by applications to real laboratory data. The graphical outputs are still being studied in order to obtain interactive ways of adding/deleting biplot axis to achieve the final biplot representations.

[1] Gower, J.C.; Hand, D.J. *Biplots*. Chapman and Hall: London, 1996.
[2] Alves, M.R.; Oliveira, M.B. (2004). Predictive and interpolative biplots applied to canonical variate analysis in the discrimination of vegetable oils by their fatty acid composition. *Journal of Chemometrics*, **18**: 393-401.
[3] Alves, M.R.; Cunha, S.C.; Amaral, J.S.; J.A. Pereira; Oliveira, M.B. (2005). Classification of PDO Olive Oils on the Basis of Their Sterol Composition by Multivariate Analysis. *Analytica Chimica Acta*, **549**: 166–178.