

klaR: A Package Including Various Classification Tools

Christian Röver, Nils Raabe, Karsten Luebke, and Uwe Ligges

Fachbereich Statistik,
Universität Dortmund, 44221 Dortmund, Germany

Abstract. Classification is a ubiquitous challenge for statisticians. The R-package **klaR** includes functions to build, check and visualize classification rules. Since the package is being developed at the Dept. of Statistics at the University of Dortmund, the name of the package is based on the German word “Klassifikation” and R. In this paper, some of the functions included in the package are described and presented in the context of classification of the German economy’s business cycle phases (Heilemann and Münch, 1996).

Among these functions, there is a wrapper to call **SVMlight** (by Thorsten Joachims, <http://svmlight.joachims.org/>) as well as an implementation of an Regularized Discriminant Analysis (*RDA*; Friedmann, 1989) which may also optimize the parameters needed for an RDA.

Another function, **stepclass()**, performs “stepwise classification”, i.e. it selects variables of the data that are used for the classification method by minimizing the cross-validated error. The classification result for any two variables can be visualized by drawing the partitions of the data.

In order to check the results various performance measures like the one described by Garczarek (2002) can be calculated. If the data consists of 3 or 4 classes, the membership values of different classifiers can be compared by visualization in a barycentric coordinate system.

Moreover, part of the package are functions for *EDAM*. In this approach a set of vectors is visualized in a fixed solution space by an “Eight-Directions-Arranged-Map” (*EDAM*; Raabe, 2003). If the visualized vectors are equal to the centroids of a set of clusters, the result can directly be compared to that of a Self-Organizing-Map.

References

- Friedman, J.H. (1989): Regularized Discriminant Analysis. *Journal of the American Statistical Association*, 84, 165–175.
- Garczarek, U.M. (2002): *Classification rules in standardized partition spaces*. Phd Thesis, University of Dortmund.
- Heilemann, U. and Münch, J.M. (1996): West german business cycles 1963-1994: A multivariate discriminant analysis. *CIRET-Conference in Singapore, CIRET-Studien 50*.
- Raabe, N. (2003): *Vergleich von Kohonen Self-Organizing-Maps mit einem nicht-simultanen Klassifikations- und Visualisierungsverfahren*. Diploma Thesis, Department of Statistics, University of Dortmund.
<http://www.statistik.uni-dortmund.de/de/content/einrichtungen/lehrstuehle/personen/raabe/Diplomarbeit.pdf>.

Keywords

CLASSIFICATION, VISUALIZATION