

sda: an R package for shrinkage discriminant analysis

Miika Ahdesmäki^{1,2,*}, Korbinian Strimmer¹

1. Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Härtelstr. 16-18, D-04107 Leipzig, Germany

2. Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland

* Contact author: miika.ahdesmaki@imise.uni-leipzig.de

Keywords: James-Stein shrinkage, linear classification, feature selection, prediction

We have developed a package (“sda”) for R that implements multi-class linear discriminant analysis especially for high dimensional problems in omics-data. The package implements feature selection in a natural way without resorting to Monte Carlo or cross validation methods, thus achieving computational efficiency and also comparable or better performance than competing state-of-the-art high dimensional classifiers. The implementation relies on James-Stein-type shrinkage and uses local false nondiscovery rate (fndr) methodology to control the number of important features. The performance of our approach has been shown favourable in several applications including gene expression microarrays, metabolomic data and mass spec data - all of which are noisy and high dimensional in nature.

The package is easy to use and readily available in CRAN <http://cran.r-project.org/>. The implementation performs the data analyses and visualises the results automatically in an intuitive way, so that it is approachable also for the nonstatisticians.

References

Strimmer, K. (2009). R-package “sda”,

<http://cran.r-project.org/web/packages/sda/>.

Ahdesmäki, M. and Strimmer, K. (2009). A natural feature selection criterion for high-dimensional multi-class linear discriminant analysis. *Submitted*.