

Untersucht wurden Projektionen in einen linearen Unterraum, der die Daten gut separiert. Als Methode zur Dimensionsreduktion wurde Heteroskedastische Diskriminanzanalyse angewendet (Szepanek et al., 2009), wie sie im R Paket `hda` (Szepanek, 2010) implementiert ist. Diese identifiziert im Vergleich zur klassischen linearen Diskriminanzanalyse zusätzlich ungleiche Kovarianzstrukturen der Klassen, wie das folgende Beispiel zeigt (Mitte: LDA, rechts: HDA)

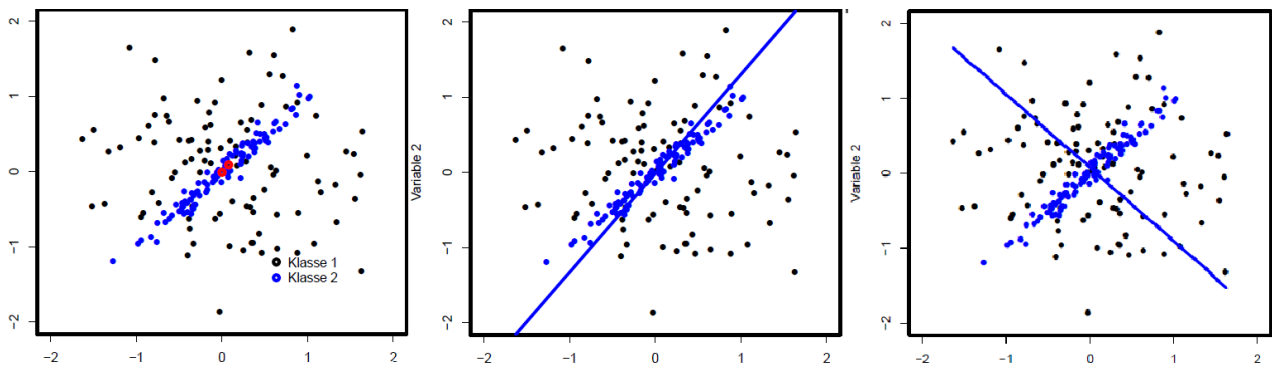


Abb.: Unterschied von HDA und LDA.

Preprocessing:

- Elimination nicht-metrisch skalierten Variablen
- Imputation der fehlenden Werte mit Regressionsbäumen
- Standardisierung
- Elimination von Klasse 4 (wenig Beobachtungen).

Anwendung von HDA auf die archäologischen Daten liefert folgende interessanten Komponenten (schwarz: group1, rot: group2, grün: group3).

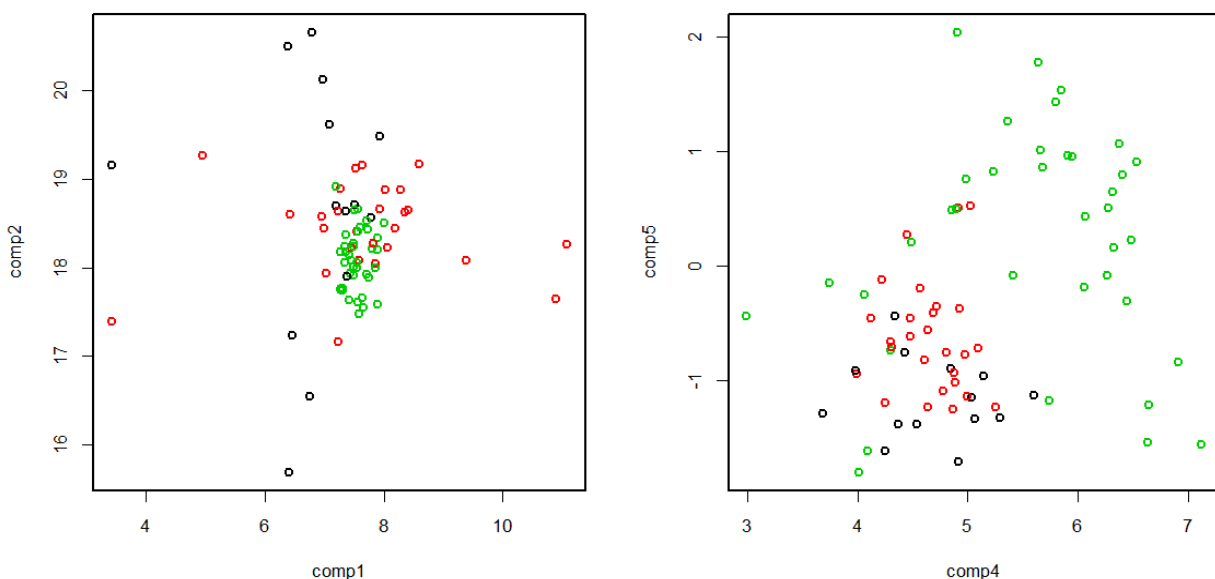


Abb.:HDA Diskriminanzkomponenten (Auszug).

Gruppe 3 ist in den ersten beiden Komponenten stark konzentriert, und hat Lageunterschiede bei größerer Streuung in Komponenten 4.

Die Gruppe 2 streut stark in Komponente 1, Gruppe 1 streut stark in Komponente 2.

Betrachtung der Ladungen veranschaulicht, welche der Variablen gemessen an der durchschnittlichen Streuung der Klassen stark auf die entsprechenden Komponenten laden.

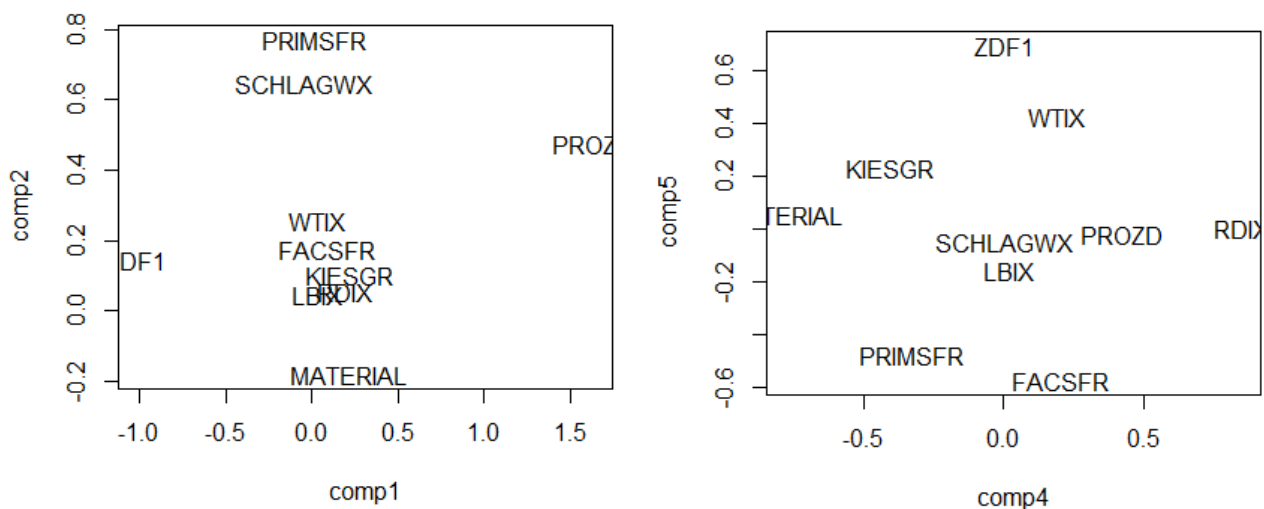


Abb.: Ladungen.

Interpretation der Komponenten anhand der stärksten Ladungen:

Komponente 1: PROZD, ZDF1

Komponente 2: PRIMSFR, SCHLAGWX, MATERIAL

Komponente 4: MATERIAL, RDIX

Komponente 5: ZDF1, WTIX, PRIMSFR, FACSFR

Referenzen:

Szepannek G., Harczos, T., Klefenz, F. and Weihs, C. (2009): Extending features for automatic speech recognition by means of auditory modelling. In: *Proceedings of European Signal Processing Conference (EUSIPCO) 2009, Glasgow*, pp.1235-1239.

Szepannek, G. (2012): hda: R package Version 0.1-18, <http://CRAN.R-project.org/package=hda>.