Package ‘hda’

March 4, 2016

Version 0.2-14
Date 2016-03-04
Title Heteroscedastic Discriminant Analysis
Author Gero Szepannek
Maintainer Gero Szepannek <gero.szepannek@web.de>
Suggests mvtnorm, klaR, MASS
Imports e1071
Description Functions to perform dimensionality reduction for classification if the covariance matrices of the classes are unequal.
License GPL (>= 2)
NeedsCompilation no
Repository CRAN
Date/Publication 2016-03-04 22:44:09

R topics documented:

   hda ................................................................. 1
   plot.hda .......................................................... 5
   predict.hda ....................................................... 6
   showloadings ...................................................... 8

Index

hda Heteroscedastic discriminant analysis

Description

Computes a linear transformation loadings matrix for discrimination of classes with unequal covariance matrices.
Usage

\texttt{hda(x, \ldots)}

## Default S3 method:
\texttt{hda(x, grouping, newdim = 1:(ncol(x)-1), crule = FALSE,}
\hspace{1cm} \texttt{reg.lamb = NULL, reg.gamm = NULL, initial.loadings = NULL,}
\hspace{1cm} \texttt{sig.levs = c(0.05,0.05), noutit = 7, ninit = 10, verbose = TRUE, \ldots)}

## S3 method for class 'formula'
\texttt{hda(formula, data = NULL, \ldots)}

Arguments

\texttt{x} \hspace{1cm} A matrix or data frame containing the explanatory variables. The method is restricted to numerical data.

\texttt{grouping} \hspace{1cm} A factor specifying the class for each observation.

\texttt{formula} \hspace{1cm} A formula of the form \texttt{grouping ~ x1 + x2 + \ldots}. That is, the response is the grouping factor and the right hand side specifies the (non-factor) discriminators.

\texttt{data} \hspace{1cm} Data frame from which variables specified in formula are to be taken.

\texttt{newdim} \hspace{1cm} Dimension of the discriminative subspace. The class distributions are assumed to be equal in the remaining dimensions. Alternatively, a vector of integers can be specified which is then computed until for the first time both tests on equal means as well as homoscedasticity do not reject. This option is to be be applied with care and the resulting dimension should be checked manually.

\texttt{crule} \hspace{1cm} Logical specifying whether a \texttt{naiveBayes} classification rule should be computed. Requires package \texttt{e1071}.

\texttt{reg.lamb} \hspace{1cm} Parameter in [0,1] for regularization towards equal covariance matrix estimations of the classes (in the original space): 0 means equal covariances, 1 (default) means complete heteroscedasticity.

\texttt{reg.gamm} \hspace{1cm} Similar to \texttt{reg.lamb}: parameter for shrinkage towards diagonal covariance matrices of equal variance in all variables where 0 means diagonality. Default is no shrinkage.

\texttt{initial.loadings} \hspace{1cm} Initial guess of the matrix of loadings. Must be quadratic of size \texttt{ncol(x)} Default is the identity matrix. By specification of \texttt{initial.loadings = "random"} a random orthonormal matrix will be generated using \texttt{qr.Q(qr())} of a random matrix with uniformly distributed elements.

\texttt{sig.levs} \hspace{1cm} Vector of significance levels for \texttt{eqmean.test} (position 1) and \texttt{homog.test} (pos. 2) to stop search for an appropriate dimension of the reduced space.

\texttt{noutit} \hspace{1cm} Number iterations of the outer loop, i.e. iterations of the likelihood. Default is 7.

\texttt{ninit} \hspace{1cm} Number of iterations of the inner loop, i.e. reiterations of the loadings matrix within one iteration step of the likelihood.

\texttt{verbose} \hspace{1cm} Logical indicating whether iteration process should be displayed.

\texttt{\ldots} \hspace{1cm} For \texttt{hda.formula}: Further arguments passed to function \texttt{hda.default} such as \texttt{newdim}. For \texttt{hda.default}: currently not used.
Details

The function returns the transformation that maximizes the likelihood if the classes are normally
distributed but differ only in a newdim dimensional subspace and have equal distributions in the
remaining dimensions (see Kumar and Androu, 1998). The scores are uncorrelated for all classes.
The algorithm is implemented as it is proposed by Burget (2006). Regularization is computed as
proposed by Friedman et al. (1989) and Szepannek et al. (2009).

Value

Returns an object of class hda.

hda.loadings Transformation matrix to be post-multiplied to new data.
hda.scores Input data after hda transformation. Reduced discriminative space are the first
newdim dimensions.
grouping Corresponding class labels for hda.scores data. Identical to input grouping.
class.dist Estimated class means and covariance matrices in the transformed space.
reduced.dimension Input parameter: dimension of the reduced space.
naivebayes Object of class naiveBayes trained on input data in the reduced space for clas-
sification of new (transformed) data. Its computation must be specified by input 
the parameter crule.
comp.acc Matrix of accuracies per component and class: reports up to which degree each
class k can be classified ($P(f_k > f_{\neq k})$) correctly according to the estimated 
(normal) distribution in any single component in the identified subspace. Meaningful for reasons of interpretability as HDA is invariant to reordering of the 
components.
vlift Returns the variable importance in terms of ratio between the accuracy comp.acc and 
and the resulting accuracy that results if single variable loadings are set to 0. The first element describes overall accuracy lift where the second element is an array of dimension (number of classes, number of components in reduced space, 
number of variables) specifying the lifts for recognition each class separately.
reg.lambda Input regularization parameter.
reg.gamm Input regularization parameter.
eqmean.test Test on equal means of the classes in the remaining dimensions like in manova 
based on Wilk’s lambda.
homog.test Test on homoscedasticity of the classes in the remaining dimensions (see e.g. 
Fahrmeir et al., 1984, p.75.)
hda.call (Matched) function call.
initial.loadings Initialization of the loadings matrix.
trace.dimensions Matrix of p values for different subspace dimensions (as specified in newdim).

Author(s)

Gero Szepannek
References


See Also

predict.hda, showloadings, plot.hda

Examples

library(mvtnorm)
library(MASS)

# simulate data for two classes
n <- 50
meana <- meanb <- c(0,0,0,0,0)
cova <- diag(5)
cova[1,1] <- 0.2
for(i in 3:4){
  for(j in (i+1):5){
    cova[i,j] <- cova[j,i] <- 0.75*(j-i)
  }
}
covb <- cova
diag(covb)[1:2] <- c(1,0.2)

xa <- rmvnorm(n, meana, cova)
xb <- rmvnorm(n, meanb, covb)
x <- rbind(xa, xb)
classes <- as.factor(c(rep(1,n), rep(2,n)))

# rotate simulated data
symmat <- matrix(runif(5^2),5)
symmat <- symmat + t(symmat)
even <- eigen(symmat)$vectors
rotatedspace <- x %*% even
plot(as.data.frame(rotatedspace), col = classes)

# apply linear discriminant analysis and plot data on (single) discriminant axis
lda.res <- lda(rotatedspace, classes)
plot(rotatedspace %*% lda.res$scaling, col = classes,
     ylab = "discriminant axis", xlab = "Observation index")

# apply heteroscedastic discriminant analysis and plot data in discriminant space
hda.res <- hda(rotatedspace, classes)
```
plot(hda.res$hda.scores, col = classes)

# compare with principal component analysis
pca.res <- prcomp(as.data.frame(rotatedspace), retx = TRUE)
plot(as.data.frame(pca.res$x), col=classes)

# Automatically build classification rule
# this requires package e1071
hda.res2 <- hda(rotatedspace, classes, crule = TRUE)
```

---

**plot.hda**

*Plot transformed data*

**Description**

Visualizes the scores on selected components of the discriminant space of reduced dimension.

**Usage**

```r
## S3 method for class 'hda'
plot(x, comps = 1:x$reduced.dimension, scores = TRUE, col = x$grouping, ...)
```

**Arguments**

- **x**: An object of class hda.
- **comps**: A vector of component ids for which the data should be displayed.
- **scores**: Logical indicating whether the scores in the projected space should be plotted. If FALSE estimated densities are plotted.
- **col**: Color vector for the data to be displayed. Per default, different colors represent the classes.
- **...**: Further arguments to be passed to the plot function.

**Details**

Scatterplots of the scores or estimated densities.

**Value**

No value is returned.

**Author(s)**

Gero Szepannek
References


See Also

hda, predict.hda, showloadings

Examples

library("mvtnorm")
library("MASS")

# simulate data for two classes
n <- 50
meana <- meanb <- c(0,0,0,0)
cova <- diag(5)
cova[1,1] <- 0.2
for(i in 3:4){
  for(j in (i+1):5){
    cova[i,j] <- cova[j,i] <- 0.75*(j-i)
  }
}
covb <- cova
diag(covb)[1:2] <- c(1,0.2)

xa <- rmvnorm(n, meana, cova)
xb <- rmvnorm(n, meanb, covb)
x <- rbind(xa,xb)
classes <- as.factor(c(rep(1,n), rep(2,n)))

## rotate simulated data
symmat <- matrix(runif(5^2),5)
symmat <- symmat + t(symmat)
even <- eigen(symmat)$vectors
rotatedspace <- x %*% even
plot(as.data.frame(rotatedspace), col = classes)

# apply heteroscedastic discriminant analysis and plot data in discriminant space
hda.res <- hda(rotatedspace, classes)

# plot scores
plot(hda.res)
Description

Computes linear transformation of new data into lower dimensional discriminative space using some model produced by `hda`.

Usage

```r
## S3 method for class 'hda'
predict(object, newdata, alldims = FALSE, task = c("dr", "c"), ...)
```

Arguments

- **object**: Model resulting from a call of `hda`.
- **newdata**: A matrix or data frame to be transformed into lower dimensional space of the same dimension as the data used for building the model.
- **alldims**: Logical flag specifying whether the result should contain only the reduced space (default) or should also include the redundant dimensions and thus be of the same dimension as the input data. In this case the reduced space is given by the first `newdim` columns.
- **task**: "dr" for standard application of the `hda` model to `newdata`. Choose "c" for classification of new data. This is an interface to predict function of `naiveBayes`. The option can be chosen if `crule = TRUE` has been specified in the `hda()` call.
- ... Further arguments to be passed to the `naiveBayes` predict function.

Value

If option `type = "dr"` the transformed data are returned. For `type = "c"` both the transformed data as well as the resulting object of the naive Bayes prediction are returned.

Author(s)

Gero Szepannek

References


See Also

`hda, showloadings, plot.hda`
Examples

library(mvtnorm)
library(MASS)

# simulate data for two classes
n <- 50
meana <- meanb <- c(0,0,0,0)
cova <- diag(5)
cova[1,1] <- 0.2
for(i in 3:4)
  for(j in (i+1):5){cova[i,j] <- cova[j,i] <- 0.75*(j-i)}
covb <- cova
diag(covb)[1:2] <- c(1,0.2)

xa <- rmvnorm(n,meana,cova)
xb <- rmvnorm(n,meanb,covb)
x <- rbind(xa,xb)
classes <- as.factor(c(rep(1,n),rep(2,n)))

# rotate simulated data
symmat <- matrix(runif(5^2),5)
symmat <- symmat + t(symmat)
even <- eigen(symmat)$vectors
rotatedspace <- x %*% even

# apply heteroscedastic discriminant analysis and plot data in discriminant space
hda.res <- hda(rotatedspace, classes)

# simulate new data
xanew <- rmvnorm(n,meana,cova)
xbnew <- rmvnorm(n,meanb,covb)
xnew <- rbind(xanew,xbnew)
classes <- as.factor(c(rep(1,n),rep(2,n)))
newrotateddata <- x %*% even
plot(as.data.frame(newrotateddata), col = classes)

# transform new data
prediction <- predict(hda.res, newrotateddata)
plot(as.data.frame(prediction), col = classes)

# predict classes for new data on automatically computed naive Bayes classification rule
# this requires package eQ07Q
hda.res2 <- hda(rotatedspace, classes, crule = TRUE)
prediction2 <- predict(hda.res2, newrotateddata, task = "c")
prediction2

showloadings
Loadings plot for heteroscedastic discriminant analysis
**Description**

Visualizes the loadings of the original variables on the components of the transformed discriminant space of reduced dimension.

**Usage**

```r
showloadings(object, comps = 1:object$reduced.dimension, loadings = TRUE, ...)
```

**Arguments**

- `object` An object of class `hda`.
- `comps` A vector of component ids for which the loadings should be displayed.
- `loadings` Logical indicating whether loadings or variable importance lifts should be plotted.
- `...` Further arguments to be passed to the plot functions.

**Details**

Scatterplots of loadings (or lifts) of any variable on any hda component to give an idea of what variables do mainly contribute to the different discriminant components (see corresponding values of `object`). Note that as opposed to linear discriminant analysis not only location but also scale differences contribute to class discrimination of the hda components.

**Value**

No value is returned.

**Author(s)**

Gero Szepannek

**References**


**See Also**

`hda`, `predict.hda`, `plot.hda`
Examples

library(mvtnorm)
library(MASS)

# simulate data for two classes
n  <- 50
meana <- meanb <- c(0,0,0,0,0)
cova <- diag(5)
cova[1,1] <- 0.2
for(i in 3:4){
  for(j in (i+1):5){
    cova[i,j] <- cova[j,i] <- 0.75*(j-i)
  }
}
covb <- cova
diag(covb)[1:2] <- c(1,0.2)

xa <- rmvnorm(n, meana, cova)
xb <- rmvnorm(n, meanb, covb)
x <- rbind(xa, xb)
classes <- as.factor(c(rep(1,n), rep(2,n)))
# rotate simulated data
symmat <- matrix(runif(5^2),5)
symmat <- symmat + t(symmat)
even <- eigen(symmat)$vectors
rotatedspace <- x %*% even
plot(as.data.frame(rotatedspace), col = classes)

# apply heteroscedastic discriminant analysis and plot data in discriminant space
hda.res <- hda(rotatedspace, classes)

# visualize loadings
showloadings(hda.res)
Index

*Topic **classif**
  hda, 1
  plot.hda, 5
  predict.hda, 6
  showloadings, 8

*Topic **multivariate**
  hda, 1
  plot.hda, 5
  predict.hda, 6
  showloadings, 8

hda, 1, 6, 7, 9

naiveBayes, 2, 7

plot.hda, 4, 5, 7, 9
predict.hda, 4, 6, 6, 9
print.hda (hda), 1

showloadings, 4, 6, 7, 8