

Package ‘CompositionalHDDA’

July 11, 2025

Type Package

Title High Dimensional Discriminant Analysis with Compositional Data

Version 1.0

Date 2025-07-08

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports Compositional, HDclassif, Rfast, stats

Suggests Rfast2

Description High dimensional discriminant analysis with compositional data is performed. The compositional data are first transformed using the alpha-transformation of Tsagris M., Preston S. and Wood A.T.A. (2011) <doi:10.48550/arXiv.1106.1451>, and then the High Dimensional Discriminant Analysis (HDDA) algorithm of Bouveyron C. Girard S. and Schmid C. (2007) <doi:10.1080/03610920701271095> is applied.

License GPL (>= 2)

NeedsCompilation no

Author Michail Tsagris [aut, cre]

Repository CRAN

Date/Publication 2025-07-11 13:10:02 UTC

Contents

CompositionalHDDA-package	2
alfa.hdda	3
cv.alfahdda	4
Index	6

CompositionalHDDA-package

High Dimensional Discriminant Analysis with Compositional Data

Description

High dimensional discriminant analysis (HDDA) for compositional data using the alpha-transformation is performed.

Details

Package:	CompositionalHDDA
Type:	Package
Version:	1.0
Date:	2025-07-08
License:	GPL-2

Maintainers

Michail Tsagris <mtsagris@uoc.gr>

Author(s)

Michail Tsagris <mtsagris@uoc.gr>.

References

- Bouveyron C. Girard S. and Schmid C. (2007). High Dimensional Discriminant Analysis. Communications in Statistics: Theory and Methods, 36(14): 2607–2623.
- Bouveyron C. Celeux G. and Girard S. (2010). Intrinsic dimension estimation by maximum likelihood in probabilistic PCA. Technical Report 440372, Universite Paris 1 Pantheon-Sorbonne.
- Berge L. Bouveyron C. and Girard S. (2012). HDclassif: An R Package for Model-Based Clustering and Discriminant Analysis of High-Dimensional Data. Journal of Statistical Software, 46(6).
- Tsagris M.T., Preston S. and Wood A.T.A. (2011). A data-based power transformation for compositional data. In Proceedings of the 4th Compositional Data Analysis Workshop, Girona, Spain. <https://arxiv.org/pdf/1106.1451.pdf>

 alfa.hdda

HDDA for compositional data using the alpha-transformation

Description

HDDA for compositional data using the alpha-transformation.

Usage

```
alfa.hdda(xnew, ina, x, a = seq(-1, 1, by = 0.1), d_select = "Cattell", threshold = 0.2)
```

Arguments

xnew	A matrix with the new compositional data whose class is to be predicted.
ina	A group indicator variable for the compositional data.
x	The compositional data. Zero values are allowed.
a	Either a single value or a vector of α values.
d_select	Either "Cattell", "BIC" or "both". "Cattell": The Cattell's scree-test is used to gather the intrinsic dimension of each class. If the model is of common dimension (models 7 to 14), the scree-test is done on the covariance matrix of the whole dataset. "BIC": The intrinsic dimensions are selected with the BIC criterion. See Bouveyron et al. (2010) for a discussion of this topic. For common dimension models, the procedure is done on the covariance matrix of the whole dataset.
threshold	A float stricly within 0 and 1. It is the threshold used in the Cattell's Scree-Test.

Details

The compositional data are first using the α -transformation and then the HDDA algorithm is called. The function then will compute all the models, give their BIC and keep the model with the highest BIC value.

Value

A list with sub-lists, one for each value of α , where each sub-list includes:

mod	A list containing the output as returned by the function hdda from the package HDclassif .
class	The predicted class of each observation.
posterior	The posterior probabilities of each new observation.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

- Bouveyron C. Girard S. and Schmid C. (2007). High Dimensional Discriminant Analysis. Communications in Statistics: Theory and Methods, 36(14): 2607–2623.
- Bouveyron C. Celeux G. and Girard S. (2010). Intrinsic dimension estimation by maximum likelihood in probabilistic PCA. Technical Report 440372, Universite Paris 1 Pantheon-Sorbonne.
- Berge L. Bouveyron C. and Girard S. (2012). HDclassif: An R Package for Model-Based Clustering and Discriminant Analysis of High-Dimensional Data. Journal of Statistical Software, 46(6).
- Tsagris M.T., Preston S. and Wood A.T.A. (2011). A data-based power transformation for compositional data. In Proceedings of the 4th Compositional Data Analysis Workshop, Girona, Spain. <https://arxiv.org/pdf/1106.1451.pdf>

See Also

[cv.alfahdda](#)

Examples

```
x <- matrix( rgamma(60 * 100, runif(100, 4, 10), 1), ncol = 100, byrow = TRUE )
x <- x / rowSums(x) ## Dirichlet simulated values
xnew <- matrix( rgamma(20 * 100, runif(100, 4, 10), 1), ncol = 100, byrow = TRUE )
xnew <- xnew / rowSums(xnew) ## Dirichlet simulated values
ina <- rbinom(60, 1, 0.5)
alfa.hdda(xnew, ina, x, a = 0.5)
```

cv.alfahdda

Cross-Validation of the HDDA for compositional data using the alpha-transformation

Description

Cross-Validation of the HDDA for compositional data using the alpha-transformation.

Usage

```
cv.alfahdda(ina, x, a = seq(-1, 1, by = 0.1), d_select = "both",
threshold = c(0.001, 0.005, 0.05, 1:9 * 0.1), folds = NULL, stratified = TRUE,
nfolds = 10, seed = NULL)
```

Arguments

ina	A group indicator variable for the compositional data.
x	The compositional data. Zero values are allowed.
a	A vector of α values.
d_select	Either "Cattell", "BIC" or "both".
threshold	A vector with numbers strictly between 0 and 1. Each value corresponds to a threshold used in the Cattell's Scree-Test.

<code>folds</code>	If you have the list with the folds supply it here. You can also leave it NULL and it will create folds.
<code>stratified</code>	Do you want the folds to be created in a stratified way? The default value is TRUE.
<code>nfolds</code>	The number of folds in the cross validation.
<code>seed</code>	You can specify your own seed number here or leave it NULL.

Details

K-fold cross-validation for the high dimensional discriminant analysis with compositional data using the α -transformation is performed.

Value

A list including:

<code>k1</code>	A matrix with the configurations of hyper-parameters tested and the estimated Kullback-Leibler divergence, for each configuration.
<code>js</code>	A matrix with the configurations of hyper-parameters tested and the estimated Jensen-Shannon divergence, for each configuration.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

- Bouveyron C. Girard S. and Schmid C. (2007). High Dimensional Discriminant Analysis. Communications in Statistics: Theory and Methods, 36(14): 2607–2623.
- Bouveyron C. Celeux G. and Girard S. (2010). Intrinsic dimension estimation by maximum likelihood in probabilistic PCA. Technical Report 440372, Universite Paris 1 Pantheon-Sorbonne.
- Berge L. Bouveyron C. and Girard S. (2012). HDclassif: An R Package for Model-Based Clustering and Discriminant Analysis of High-Dimensional Data. Journal of Statistical Software, 46(6).
- Tsagris M.T., Preston S. and Wood A.T.A. (2011). A data-based power transformation for compositional data. In Proceedings of the 4th Compositional Data Analysis Workshop, Girona, Spain. <https://arxiv.org/pdf/1106.1451.pdf>

See Also

[alfa.hdda](#)

Examples

```
x <- matrix( rgamma(100 * 200, runif(200, 4, 10), 1), ncol = 200, byrow = TRUE )
x <- x / rowSums(x) ## Dirichlet simulated values
ina <- rbinom(100, 1, 0.5)
mod <- cv.alfahdda(ina, x, a = c(0.1, 0.5, 1), d_select = "both",
  threshold = seq(0.1, 0.5, by = 0.1), nfolds = 5)
```

Index

`alfa.hdda`, [3](#), [5](#)

`CompositionalHDDA-package`, [2](#)

`cv.alfahdda`, [4](#), [4](#)