

Package ‘lpda’

March 7, 2023

Type Package

Title Linear Programming Discriminant Analysis

Version 1.0.1

Author Maria Jose Nueda <mj.nueda@ua.es>

Maintainer Maria Jose Nueda <mj.nueda@ua.es>

Description Classification method obtained through linear programming.

It is advantageous with respect to the classical developments when the distribution of the variables involved is unknown or when the number of variables is much greater than the number of individuals.

LPDA method is published in Nueda, et al. (2022) “LPDA: A new classification method based on linear programming”.

<[doi:10.1371/journal.pone.0270403](https://doi.org/10.1371/journal.pone.0270403)>.

Depends R (>= 3.5.0)

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports Rglpk

Suggests rmarkdown, knitr, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2023-03-07 07:00:02 UTC

R topics documented:

bestPC	2
bestVariability	3
CVktest	4
CVloo	5

lpda	6
lpda.fit	8
lpda.pca	9
lpdaCV	10
palmdates	11
PCA	12
plot.lpda	13
predict.lpda	14
RNAseq	14
stand	15
stand2	15

Index	16
--------------	-----------

bestPC	<i>Choosing the best number of Principal Components (PCs) for lpda-pca model.</i>
--------	---

Description

bestPC computes the classification error for lpda.pca models applied with the number of components specified in PCs argument. The result is the average classification error rate from the R models computed for each number of PCs.

Usage

```
bestPC(data, group, ntest = 10, R = 10, PCs = c(10,15,20), f1 = NULL, f2 = NULL)
```

Arguments

data	Matrix containing data. Individuals in rows and variables in columns
group	Vector with the variable group
ntest	Number of samples to evaluate in the test-set.
R	Times the model is evaluated with each Variability indicated in Vars vector.
PCs	The PCs to check.
f1	Vector with weights for individuals of the first group. If NULL they are equally weighted.
f2	Vector with weights for individuals of the second group. If NULL they are equally weighted.

Value

bestPC returns a vector with the average prediction error rate obtained from the R models for each PC specified in PCs input.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

See Also

[lpda](#)

Examples

```
data(RNAseq)
group = as.factor(rep(c("G1", "G2"), each = 30))
bestPC(RNAseq, group, ntest = 10, R = 5, PCs = c(2, 10))
```

bestVariability

Choosing the best explained variability for lpda-pca model.

Description

bestVariability computes the classification error for lpda.pca models obtained with the number of components needed to reach the explained variability specified in 'Vars' argument. The result is the average classification error rate from the R models computed for each explained variability.

Usage

```
bestVariability(data, group, ntest = 10, R = 10, Vars = c(0.5, 0.7), f1 = NULL, f2 = NULL)
```

Arguments

data	Matrix containing data. Individuals in rows and variables in columns
group	Vector with the variable group
ntest	Number of samples to evaluate in the test-set.
R	Times the model is evaluated with each Variability indicated in Vars vector.
Vars	The different variabilities to check from which the best variability parameter will be chosen for lpdapca model.
f1	Vector with weights for individuals of the first group. If NULL they are equally weighted.
f2	Vector with weights for individuals of the second group. If NULL they are equally weighted.

Value

bestVar returns a vector with the average prediction error rate obtained from the R models for each variability specified in Vars input.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

See Also

[lpda](#)

Examples

```
data(RNAseq)
group = as.factor(rep(c("G1","G2"),each=30))
bestVariability(RNAseq, group, ntest = 10, R = 5, Vars = c(0.1,0.9))
```

CVktest

CVktest evaluates the error rate classification with crossvalidation

Description

CVktest evaluates the error rate classification in k samples that do not participate in the model

Usage

```
CVktest(data, group, scale = FALSE, pca = FALSE, PC = 2,
        Variability = NULL, ntest = 10, R = 10, f1 = NULL, f2 = NULL)
```

Arguments

data	Matrix containing data. Individuals in rows and variables in columns
group	Vector with the variable group
scale	Logical indicating if data is standarised.
pca	Logical indicating if a reduction of dimension is required
PC	Number of Principal Components (PC) for PCA. By default it is 2. When the number of PC is not decided, it can be determined choosing the desired proportion of explained variability (Variability parameter) or choosing the maximum number of errors allowed in the training set (Error.max).
Variability	Parameter for Principal Components (PC) selection. This is the desired proportion of variability explained for the PC of the variables.
ntest	Number of samples to evaluate in the test-set.
R	Number of times that the error is evaluated.
f1	Vector with weights for individuals of the first group. If NULL they are equally weighted.
f2	Vector with weights for individuals of the second group. If NULL they are equally weighted.

Value

lpdaktest The prediction error rate.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

See Also

[lpdaCV](#)

CVloo	<i>CVloo evaluates the error rate classification with leave one out procedure</i>
-------	---

Description

CVloo evaluates the error rate classification with leave one out procedure.

Usage

```
CVloo(data, group, scale = FALSE, pca = FALSE, PC = 2,
       Variability = NULL, f1 = NULL, f2 = NULL)
```

Arguments

data	Matrix containing data. Individuals in rows and variables in columns.
group	Vector with the variable group.
scale	Logical indicating if data is standarised.
pca	Logical indicating if a reduction of dimension is required.
PC	Number of Principal Components (PC) for PCA. By default it is 2. When the number of PC is not decided, it can be determined choosing the desired proportion of explained variability (Variability parameter) or choosing the maximum number of errors allowed in the training set (Error.max).
Variability	Parameter for Principal Components (PC) selection. This is the desired proportion of variability explained for the PC of the variables.
f1	Vector with weights for individuals of the first group. If NULL they are equally weighted.
f2	Vector with weights for individuals of the second group. If NULL they are equally weighted.

Value

CVloo The prediction error rate.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

See Also

[lpdaCV](#)

lpda

Computing discriminating hyperplane for two groups

Description

This function computes a discriminating hyperplane for two groups with original data (calling `lpda.fit`) or with principal components (calling `lpda.pca`)

Usage

```
lpda(data, group, scale = FALSE, pca = FALSE, PC = 2, Variability = NULL,
      f1 = NULL, f2 = NULL)
```

Arguments

<code>data</code>	Matrix containing data. Individuals in rows and variables in columns
<code>group</code>	Vector with the variable group
<code>scale</code>	Logical indicating if data is standarised. When <code>pca=TRUE</code> data is always scaled.
<code>pca</code>	Logical indicating if Principal Components Analysis is required
<code>PC</code>	Number of Principal Components (PC) for PCA. By default it is 2. When the number of PC is not decided, it can be determined choosing the desired proportion of explained variability (Variability parameter).
<code>Variability</code>	Parameter for Principal Components (PC) selection. This is the minimum desired proportion of variability explained for the PC of the variables. The analysis is always done with a minimum of 2 PCs. If it is NULL the PCA will be computed with PC parameter.
<code>f1</code>	Vector with weights for individuals of the first group. If NULL they are equally weighted.
<code>f2</code>	Vector with weights for individuals of the second group. If NULL they are equally weighted.

Value

`lpda` returns an object of class "lpda".

The functions `predict` and `plot` can be used to obtain the predicted classes and a plot in two dimensions with the distances to the computed hyperplane for the two classes.

`coef` Hyperplane coefficients

data	Input data matrix when <code>pca = FALSE</code> and scores when <code>pca = TRUE</code>
group	Input group vector
scale	Input scale argument
pca	Input <code>pca</code> argument
loadings	Principal Components loadings. Shown when <code>pca = TRUE</code>
scores	Principal Components scores. Shown when <code>pca = TRUE</code>
var.exp	A matrix containing the explained variance for each component and the cumulative variance. Shown when <code>pca = TRUE</code>
PCs	Number of Principal Components in the analysis. Shown when <code>pca = TRUE</code>

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

References

Nueda MJ, Gandía C, Molina MD (2022) LPDA: A new classification method based on linear programming. PLoS ONE 17(7): e0270403. <<https://doi.org/10.1371/journal.pone.0270403>>

See Also

[lpda.pca](#), [lpda.fit](#)

Examples

```
##### palmdates example in lpda package:
data(palmdates)
group = as.factor( c(rep("Spanish",11),rep("Foreign",10)) )

# with concentration data:
model = lpda(data = palmdates$conc, group = group )
pred = predict(model)
table(pred$fitted, group)
plot(model, main = "Palmdates example")

model.pca = lpda(data = palmdates$conc, group = group, pca=TRUE, PC = 2)
plot(model.pca, PCscores = TRUE, main = "Palmdates example")

# with spectra data
model.pca = lpda(data = palmdates$spectra, group = group, pca=TRUE, Variability = 0.9)
model.pca$PCs # 4 PCs to explain 90% of the variability
plot(model.pca, PCscores = TRUE, main = "Spectra palmdates")
```

`lpda.fit`*lpda.fit computes the discriminating hyperplane for two groups*

Description

`lpda.fit` computes the discriminating hyperplane for two groups, giving as a result the coefficients of the hyperplane.

Usage

```
lpda.fit(data, group, f1 = NULL, f2 = NULL)
```

Arguments

<code>data</code>	Matrix containing data. Individuals in rows and variables in columns
<code>group</code>	Vector with the variable group
<code>f1</code>	Vector with weights for individuals of the first group
<code>f2</code>	Vector with weights for individuals of the second group

Value

<code>coef</code>	Hyperplane coefficients
-------------------	-------------------------

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

References

Nueda MJ, Gandía C, Molina MD (2022) LPDA: A new classification method based on linear programming. PLoS ONE 17(7): e0270403. <<https://doi.org/10.1371/journal.pone.0270403>>

See Also

[lpda](#)

lpda.pca	<i>lpda.pca computes a PCA to the original data and selects the desired PCs when Variability is supplied</i>
----------	--

Description

lpda.pca computes the discriminating hyperplane for two groups with Principal Components (PC)

Usage

```
lpda.pca(data, group, PC = 2, Variability = NULL)
```

Arguments

data	Matrix containing data. Individuals in rows and variables in columns
group	Vector with the variable group
PC	Number of Principal Components (PC) for PCA. By default it is 2. When the number of PC is not decided, it can be determined choosing the desired proportion of explained variability (Variability parameter).
Variability	Parameter for Principal Components (PC) selection. This is the minimum desired proportion of variability explained for the PC of the variables. The analysis is always done with a minimum of 2 PCs. If it is NULL the PCA will be computed with PC parameter.

Value

loadings	Principal Components loadings.
scores	Principal Components scores.
var.exp	A matrix containing the explained variance for each component and the cumulative variance.
PCs	Number of Principal Components in the analysis.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

References

Nueda MJ, Gandía C, Molina MD (2022) LPDA: A new classification method based on linear programming. PLoS ONE 17(7): e0270403. <<https://doi.org/10.1371/journal.pone.0270403>>

See Also

[lpda](#)

lpdaCV	<i>lpdaCV evaluates the error rate classification with a crossvalidation procedure</i>
--------	--

Description

lpdaCV evaluates the error rate classification with a crossvalidation procedure

Usage

```
lpdaCV(data, group, scale = FALSE, pca = FALSE, PC = 2, Variability = NULL,
        CV = "loo", ntest = 10, R = 10, f1 = NULL, f2 = NULL)
```

Arguments

data	Matrix containing data. Individuals in rows and variables in columns
group	Vector with the variable group
scale	Logical indicating if data is standarised.
pca	Logical indicating if a reduction of dimension is required
PC	Number of Principal Components (PC) for PCA. By default it is 2. When the number of PC is not decided, it can be determined choosing the desired proportion of explained variability (Variability parameter) or choosing the maximum number of errors allowed in the training set (Error.max).
Variability	Parameter for Principal Components (PC) selection. This is the desired proportion of variability explained for the PC of the variables.
CV	Crossvalidation mode: loo "leave one out" or ktest: that leaves k in the test set.
ntest	Number of samples to evaluate in the test-set.
R	Number of times that the error is evaluated.
f1	Vector with weights for individuals of the first group. If NULL they are equally weighted.
f2	Vector with weights for individuals of the second group. If NULL they are equally weighted.

Value

lpdaCV The prediction error rate.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

See Also

[lpda](#)

Examples

```
data(RNAseq)
group = as.factor(rep(c("G1","G2"), each = 30))
lpdaCV(RNAseq, group, pca = TRUE, CV = "ktest", ntest = 2)
```

palmdates	<i>Spectrometry and composition chemical of Spanish and Arabian palm dates</i>
-----------	--

Description

A data set with scores of 21 dates on spectrometry and concentration measurements of the substances that better define the quality of the dates: fibre, sorbitol, fructose, glucose and myo-inositol. The first 11 dates are Spanish (from Elche, Alicante) and the last 10 are from other countries, mainly Arabian.

Usage

```
palmdates
```

Format

A data frame with 2 elements:

conc a data frame with 5 columns: fibre, sorbitol, fructose, glucose and myo-inositol.

spectra a data frame with 2050 columns.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

References

Abdrabo, S.S., Gras, L., Grindlay, G. and Mora, J. (2021) Evaluation of Fourier Transform-Raman Spectroscopy for palm dates characterization. Journal of food composition and analysis. Submitted.

PCA

Principal Component Analysis

Description

Computes a Principal Component Analysis when both when $p > n$ and when $p \leq n$.

Usage

PCA(X)

Arguments

X Matrix or data.frame with variables in columns and observations in rows.

Value

eigen A eigen class object with eigenvalues and eigenvectors of the analysis.
var.exp A matrix containing the explained variance for each component and the cumulative variance.
scores Scores of the PCA analysis.
loadings Loadings of the PCA analysis.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

Examples

```
## Simulate data matrix with 500 variables and 10 observations
datasim = matrix(sample(0:100, 5000, replace = TRUE), nrow = 10)
## PCA
myPCA = PCA(datasim)
## Extracting the variance explained by each principal component
myPCA$var.exp
```

plot.lpda	<i>Plot method for lpda classification</i>
-----------	--

Description

plot.lpda is applied to an lpda class object. It shows a plot in two dimensions with the distances to the computed hyperplane of each individual coloring each case with the real class.

Usage

```
## S3 method for class 'lpda'  
plot(x, PCscores = FALSE, xlim = NULL, main = NULL,  
     legend.pos = "topright", ...)
```

Arguments

x	Object of class inheriting from "lpda"
PCscores	Logical to show the first 2 PCscores. Only possible when PCA is applied.
xlim	An optional vector with two values with the x-axis range. If omitted, it will be computed.
main	An optional title for the plot.
legend.pos	The position for the legend. By default it is topright. NULL when no legend is required.
...	Other arguments passed.

Value

Two dimensional plot representing the distances to the computed hyperplane of each individual colored with the real class.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

See Also

[lpda](#)

predict.lpda	<i>Predict method for lpda classification</i>
--------------	---

Description

Predict method for lpda classification

Usage

```
## S3 method for class 'lpda'
predict(object, datatest = object$data,...)
```

Arguments

object	Object of class inheriting from "lpda"
datatest	An optional data to predict their class. If omitted, the original data is used.
...	Other arguments passed.

Value

fitted	Predicted class
eval	Evaluation of each individual in the fitted model

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

See Also

[lpda](#)

RNAseq	<i>Simulated RNA-Seq dataset example</i>
--------	--

Description

A simulated RNA-Seq dataset example.

Usage

```
RNAseq
```

Format

A data frame with 600 variables (in columns) and 60 samples (rows).

Details

This dataset is a RNA-Seq simulated example. It has been simulated as Negative Binomial distributed and transformed to rpkm (Reads per kilo base per million mapped reads). It contains 600 genes (in columns) and 60 samples (rows), 30 of each one of the experimental groups. First 30 samples are from first group and the remaining samples from the second one.

Author(s)

Maria Jose Nueda, <mj.nueda@ua.es>

stand	<i>stand center and scale a data matrix</i>
-------	---

Description

stand center and scale a data matrix

Usage

stand(X)

Arguments

X a data matrix with individuals in rows and variables in columns

Value

Scaled data matrix

stand2	<i>stand2 center and scale a data matrix with the parameters of another one</i>
--------	---

Description

stand2 center and scale a data matrix with the parameters of another one

Usage

stand2(X, X2)

Arguments

X the data matrix from which mean and standard deviation is computed
 X2 the data matrix to center and scale

Value

Scaled X2 data matrix

Index

* **classification**

- bestPC, [2](#)
- bestVariability, [3](#)
- lpda, [6](#)
- lpda.fit, [8](#)
- lpda.pca, [9](#)
- lpdaCV, [10](#)
- plot.lpda, [13](#)
- predict.lpda, [14](#)

* **datasets**

- palmdates, [11](#)
- RNAseq, [14](#)

bestPC, [2](#)

bestVariability, [3](#)

CVktest, [4](#)

CVloo, [5](#)

lpda, [3](#), [4](#), [6](#), [8–10](#), [13](#), [14](#)

lpda.fit, [7](#), [8](#)

lpda.pca, [7](#), [9](#)

lpdaCV, [5](#), [6](#), [10](#)

palmdates, [11](#)

PCA, [12](#)

plot.lpda, [13](#)

predict.lpda, [14](#)

RNAseq, [14](#)

stand, [15](#)

stand2, [15](#)