Package 'MultiVarSel'

January 20, 2025

Type Package

Title Variable Selection in a Multivariate Linear Model

Version 1.1.3

Author Marie Perrot-Dockès, Céline Lévy-Leduc, Julien Chiquet

Maintainer Marie Perrot-Dockès <marie.perrocks@gmail.com>

Description

It performs variable selection in a multivariate linear model by estimating the covariance matrix of the residuals then use it to remove the dependence that may exist among the responses and eventually performs variable selection by using the Lasso criterion.
The method is described in the paper Perrot-Dockès et al. (2017) <arXiv:1704.00076>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Depends glmnet, Matrix (>= 1.2-11), parallel

Suggests R.rsp

VignetteBuilder R.rsp

NeedsCompilation no

Repository CRAN

Date/Publication 2019-03-21 10:23:33 UTC

Contents

MultiVarSel-package 2
copals_camera 3
group 3
metab 4
prot 4
variable_selection 4
whitening 5
whitening_choice 6

whitening_test 7
Y 8

Index **9**

`MultiVarSel-package` *Package*

Description

MultiVarSel consists of four functions: "whitening.R", "whitening_test.R", "whitening_choice.R" and "variable_selection.R". For further information on how to use these functions, we refer the reader to the vignette of the package.

Details

This package consists of four functions: "whitening.R", "whitening_test.R", "whitening_choice.R" and "variable_selection.R". For further information on how to use these functions, we refer the reader to the vignette of the package.

Author(s)

Marie Perrot-Dockes, Celine Levy-Leduc, Julien Chiquet

Maintainer: Marie Perrot-Dockes <marie.perrocks@gmail.com>

References

M. Perrot-Dockes et al. "A multivariate variable selection approach for analyzing LC-MS metabolomics data", arXiv:1704.00076

Examples

```
data("copals_camera")
Y <- scale(Y[, 1:50])
X <- model.matrix(~ group + 0)
residuals <- lm(as.matrix(Y) ~ X - 1)$residuals
S12_inv <- whitening(residuals, "AR1", pAR = 1, qMA = 0)
Frequencies <- variable_selection(
  Y = Y, X = X,
  square_root_inv_hat_Sigma = S12_inv,
  nb_repli = 10,  nb.cores = 1, parallel = FALSE
)
## Not run:
# Parallel computing
require(doMC)
registerDoMC(cores=4)
Freqs <- variable_selection(Y,X,square_root_inv_hat_Sigma,
                    nb_repli=10,parallel=TRUE,nb.cores=4)

## End(Not run)
```

copals_camera *Copals data*

Description

A Liquid Chromatography Mass Spectrometry dataset made of African copals samples.

Usage

```
data("copals_camera")
```

Format

It containes Y a data frame with 30 observations on 1019 variables and group a qualitative variable indicating the type of tree each row of Y is.

References

M. Perrot-Dockes et al. "A multivariate variable selection approach for analyzing LC-MS metabolomics data", arXiv:1704.00076 https://arxiv.org/pdf/1704.00076.pdf

Examples

```
data(copals_camera)
```

group *This is a qualitative variable indicating the type of tree each row of Y is.*

Description

This is a qualitative variable indicating the type of tree each row of Y is.

Author(s)

Marie Perrot-Dockes <marie.perrot-dockes@agroparistech.fr>

References

https://arxiv.org/pdf/1704.00076.pdf

`metab` *This is a dataset containing the abundance of 199 metabolites from 9 seeds samples just after germination. The temperature of seed maturation vary between the different seeds.*

Description

This is a dataset containing the abundance of 199 metabolites from 9 seeds samples just after germination. The temperature of seed maturation vary between the different seeds.

Author(s)

Marie Perrot-Dockes <marie.perrot-dockes@agroparistech.fr>

`prot` *This is a dataset containing the abundance of 724 proteins from 9 seeds samples just after germination. The temperature of seed maturation vary between the different seeds.*

Description

This is a dataset containing the abundance of 724 proteins from 9 seeds samples just after germination. The temperature of seed maturation vary between the different seeds.

Author(s)

Marie Perrot-Dockes <marie.perrot-dockes@agroparistech.fr>

`variable_selection` *This function allows the user to select the most relevant variables thanks to the estimation of their selection frequencies obtained by the stability selection approach.*

Description

This function allows the user to select the most relevant variables thanks to the estimation of their selection frequencies obtained by the stability selection approach.

Usage

```
variable_selection(Y, X, square_root_inv_hat_Sigma, nb_repli = 1000,
  parallel = FALSE, nb.cores = 1)
```

Arguments

Y	a response matrix
X	a matrix of covariables
square_root_inv_hat_Sigma	Estimation of the inverse of the square root of the covariance matrix of each row of the residuals matrix obtained by the whitening function.
nb_repli	numerical, number of replications in the stability selection
parallel	logical, if TRUE then a parallelized version of the code is used
nb.cores	numerical, number of cores used

Value

A data frame containing the selection frequencies of the different variables obtained by the stability selection, the corresponding level in the design matrix and the associated column of the observations matrix.

Examples

```
data("copals_camera")
Y <- scale(Y[, 1:50])
X <- model.matrix(~ group + 0)
residuals <- lm(as.matrix(Y) ~ X - 1)$residuals
S12_inv <- whitening(residuals, "AR1", pAR = 1, qMA = 0)
Frequencies <- variable_selection(
  Y = Y, X = X,
  square_root_inv_hat_Sigma = S12_inv,
  nb_repli = 10, nb.cores = 1, parallel = FALSE
)
```

---

whitening *This function provides an estimation of the inverse of the square root of the covariance matrix of each row of the residuals matrix.*

Description

This function provides an estimation of the inverse of the square root of the covariance matrix of each row of the residuals matrix.

Usage

```
whitening(residuals, typeDep, pAR = 1, qMA = 0)
```

Arguments

residuals	the residuals matrix obtained by fitting a linear model to each column of the response matrix as if they were independent
typeDep	character in c("AR1", "ARMA", "nonparam") defining which type of dependence to use
pAR	numerical, only use if typeDep = "ARMA", the parameter p for the ARMA(p, q) process
qMA	numerical, only use if typeDep = "ARMA", the parameter q for the ARMA(p, q) process

Value

It returns the estimation of the inverse of the square root of the covariance matrix of each row of the residuals matrix.

Examples

```
data(copals_camera)
Y <- scale(Y[, 1:100])
X <- model.matrix(~ group + 0)
residuals <- lm(as.matrix(Y) ~ X - 1)$residuals
whitening(residuals, "AR1")
```

---

whitening_choice *This function helps to choose the best whitening strategy among the following types of dependence modellings: AR1, ARMA, non parametric and without any whitening.*

Description

This function helps to choose the best whitening strategy among the following types of dependence modellings: AR1, ARMA, non parametric and without any whitening.

Usage

```
whitening_choice(residuals, typeDeps = "AR1", pAR = 1, qMA = 0,
  threshold = 0.05)
```

Arguments

residuals	the residuals matrix obtained by fitting a linear model to each column of the response matrix as if they were independent
typeDeps	character in c("AR1", "ARMA", "nonparam", "no_whitening") defining which dependence structure to use to whiten the residuals.
pAR	numerical, only use if typeDep = "ARMA", the parameter p for the ARMA(p, q) process

qMA	numerical, only use if typeDep = "ARMA", the parameter q for the ARMA(p, q) process
threshold	significance level of the test

Value

It provides a table giving the p-values for the different whitening tests applied to the residuals multiplied on the right by the inverse of the square root of the estimated covariance matrix. If the p-value is small (in general smaller than 0.05) it means that the hypothesis that each row of the residuals "whitened" matrix is a white noise, is rejected.

Examples

```
data(copals_camera)
Y <- scale(Y[, 1:100])
X <- model.matrix(~ group + 0)
residuals <- lm(as.matrix(Y) ~ X - 1)$residuals
whitening_choice(residuals, c("AR1", "nonparam", "ARMA", "no_whitening"),
  pAR = 1, qMA = 1 )
```

---

whitening_test

This function provides the p-value of an adaptation of the Portmanteau statistic to test if there is some dependence in the rows of the residuals matrix given as an argument of the function.

Description

This function provides the p-value of an adaptation of the Portmanteau statistic to test if there is some dependence in the rows of the residuals matrix given as an argument of the function.

Usage

```
whitening_test(residuals)
```

Arguments

residuals	the residuals matrix obtained by fitting a linear model to each column of the response matrix as if they were independent

Value

the p-value of a whitening test. If the p-value is small (generally smaller than 0.05) it means that the hypothesis that each row of the residuals matrix is a white noise, is rejected.

Examples

```
data(copals_camera)
Y <- scale(Y[, 1:100])
X <- model.matrix(~ group + 0)
residuals <- lm(as.matrix(Y) ~ X - 1)$residuals
whitening_test(residuals)
```

Y *This is a metabolomic dataset from 30 copals samples of trees coming from Africa*

Description

This is a metabolomic dataset from 30 copals samples of trees coming from Africa

Author(s)

Marie Perrot-Dockes <marie.perrot-dockes@agroparistech.fr>

References

https://arxiv.org/pdf/1704.00076.pdf

Index

∗ **datasets**
    copals_camera, 3
∗ **data**
    group, 3
    metab, 4
    prot, 4
    Y, 8
∗ **package**
    MultiVarSel-package, 2

copals_camera, 3

group, 3

metab, 4
MultiVarSel-package, 2
MultiVarSelect *(*MultiVarSel-package*)*, 2

prot, 4

variable_selection, 4

whitening, 5
whitening_choice, 6
whitening_test, 7

Y, 8