Package ‘FRCC’

February 19, 2015

Type Package
Title Fast Regularized Canonical Correlation Analysis
Version 1.0
Date 2012-06-25
Author Raul Cruz-Cano
Maintainer Raul Cruz-Cano <raulcruz@umd.edu>
Description This package implements the functions associated with Fast Regularized Canonical Correlation Analysis.
License GPL-2
LazyLoad yes
Depends CCP, MASS, calibrate, corpcor
Repository CRAN
Date/Publication 2012-10-04 19:01:31
NeedsCompilation no

R topics documented:

FRCC-package .................................................. 2
custom.draw.circle ............................................ 3
frcc ............................................................... 3
generate_multivariate_normal_sample ................. 5
microRNAs ..................................................... 6
off.diagonal.lambda ......................................... 7
plot_units ....................................................... 8
plot_variables ................................................. 9
rearrange.frcc ............................................... 10
soilspec ......................................................... 11
Topoisomerase_II_Inhibitors .............................. 12

Index 13
FRCC-package

Fast Regularized Canonical Correlation Analysis

Description

This package implements the Fast Regularized Canonical Correlation Analysis described in [Cruz-Cano et al., 2012]. The main idea of the algorithm is using the minimum risk estimators of the correlation matrices described in [Schafer and Strimmer, 2008] during the calculation of the Canonical correlation Structure. It can be considered an extension of the work for two set of variables (blocks) mentioned in [Tenenhaus and Tenenhaus, 2011].

Details

<table>
<thead>
<tr>
<th>Package:</th>
<th>FRCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type:</td>
<td>Package</td>
</tr>
<tr>
<td>Version:</td>
<td>1.0</td>
</tr>
<tr>
<td>Date:</td>
<td>2012-03-13</td>
</tr>
<tr>
<td>License:</td>
<td>GPL-2</td>
</tr>
<tr>
<td>LazyLoad:</td>
<td>yes</td>
</tr>
</tbody>
</table>

The function frcc provides the canonical structure for two sets of variables X and Y. The rest of the functions help to visualize and interpret the values of the canonical structure.

Author(s)

Raul Cruz-Cano
Maintainer: Raul Cruz-Cano <raulcruz@umd.edu>

References


Examples

# Examples of the functions included in this package are listed
# in the help file of each individual function.
**custom.draw.circle**

*Draws a circle.*

**Description**

Given a center, radius and color, this function draws a circle.

**Usage**

```r
custom.draw.circle(x, y, r, col)
```

**Arguments**

- `x`: X coordinate of the center
- `y`: Y coordinate of the center
- `r`: Radius of the circle
- `col`: Color of the circle

**Value**

This function does not return a value, it just draws a circle.

**Author(s)**

Michael Bedward

**References**

It comes from http://www.r-bloggers.com/circle-packing-with-r/

**Examples**

```r
# This is an internal function. No examples required.
```

---

**frcc**

*This function implements the Fast Regularized Canonical Correlation Analysis*

**Description**

This function implements the Fast Regularized Canonical Correlation algorithm described in [Cruz-Cano et al., 2012]. The main idea of the algorithm is using the minimum risk estimators of the correlation matrices described in [Schafer and Strimmer, 2008] during the calculation of the Canonical correlation Structure. It can be considered an extension of the work for two set of variables (blocks) mentioned in [Tenenhaus and Tenenhaus, 2011].
Usage

frcc(X, Y)

Arguments

X numeric matrix (n by p) which contains the observations on the X variables.
Y numeric matrix (n by q) which contains the observations on the Y variables.

Value

A list with the following components of the Canonical Structure:

cor Canonical correlations.
p_values The corresponding p-values for the each of the canonical correlations.
canonical_weights_X The canonical weights for the variables of the dataset X.
canonical_weights_Y The canonical weights for the variables of the dataset Y.
canonical_factor_loadings_X The interset canonical factor loadings for the variables of the dataset X.
canonical_factor_loadings_Y The interset canonical factor loadings for the variables of the dataset Y.

Author(s)

Raul Cruz-Cano

References


Examples

# Example # 1 Multivariate Normal Data
p<-10
q<-10
n<-50
res<-generate_multivariate_normal_sample(p,q,n)
X<-res$X
Y<-res$Y
rownames(X)<-c(1:n)
colnames(X)<-c(1:p)
colnames(Y)<-c(1:q)
**generate_multivariate_normal_sample**

It generates a sample from a multinormal distribution function.

**Description**

It generates a sample from a multinormal distribution function with the cross-covariance matrix described in [Cruz-Cano et al. 2012].

**Usage**

```
generate_multivariate_normal_sample(p, q, n)
```

**Arguments**

- `p` Number of desired variables in the dataset X.
- `q` Number of desired variables in the dataset Y.
- `n` sample size desired.

**Value**

A list of n sample units with the values for the variables of the datasets X and Y.

**Author(s)**

Raul Cruz-Cano
References


Examples

```r
p<-10
cq<-10
n<-50
res<-generate_multivariate_normal_sample(p,q,n)
X<-res$X
Y<-res$Y
rownames(X)<-c(1:n)
colnames(X)<-c(1:p)
colnames(Y)<-c(1:q)
my_res<-frcc(X,Y)
```

---

**microRNAs**

*NCI-60 microRNA data*

---

Description

Contains the expression level of 365 microRNA data in the NCI-60 cell lines.

Usage

`data(microRNA)`

Details

The NCI-60 is a set of cell cultures grown under controlled conditions by the National Cancer Institute. The NCI-60 cell lines include experimental units from the breast (8), central nervous system (6), colorectal (7), lung (9), prostate (2), ovarian (6) and renal (8) cancers. It also includes leukemia (6) and melanoma (8) cell lines. MicroRNAs are a type of RNA molecules found in eukaryotic cells. Each microRNA is a short RNA sequence (around 22 nucleotides) which is involved in the regulation of multiple target genes. A large number of published papers deal with the problem of finding the microRNA expression signature of different cancers with the goal of designing early detection methods and providing therapeutic targets.

Value

A matrix with the expression level of 365 microRNAs for the 60 cell lines in the NCI-60 dataset as described in [Cruz-Cano et al., 2012]. The original source of the dataset is [DTP, 2009]

Author(s)

Raul Cruz-Cano
**off.diagonal.lambda**

**References**

**Examples**
```r
# Example #3 NCI-60 micrRNA Data
data("Topoisomerase_II_Inhibitors")
data("micrRNA")
my_res <- frcc(t(micrRNA), -1*t(Topoisomerase_II_Inhibitors))
```

---

**Description**
Calculates the value of the shrinkage coefficient for the off-diagonal matrices as described in Cruz-Cano et al., 2012.

**Usage**
```r
off.diagonal.lambda(xs, p, q)
```

**Arguments**
- `xs`: Matrix with the values for the datasets X and Y.
- `p`: Number of variables in the dataset X.
- `q`: Number of variables in the dataset Y.

**Value**
Shrinkage coefficient for the off-diagonal matrices used to calculate the FRCC canonical structure.

**Author(s)**
Raul Cruz-Cano

**References**

**Examples**
```r
## This is an internal function. No need for examples.
```
plot_units

Plots the experimental units in the Canonical Variates Space

Description

This function plots the experimental units used in the FRCCA as points in a two-dimensional plane in which the axis are the canonical variates selected by the user.

Usage

plot_units(X, Y, res.mrcc, i, text_size = 0.8, point_size = 2)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>numeric matrix (n by p) which contains the observations on the X variables.</td>
</tr>
<tr>
<td>Y</td>
<td>numeric matrix (n by p) which contains the observations on the Y variables.</td>
</tr>
<tr>
<td>res.mrcc</td>
<td>List containing a canonical structure provided by the function frcc for the dataset X and Y.</td>
</tr>
<tr>
<td>i</td>
<td>Canonical Variate which will be used for the axes (X for horizontal and Y for vertical).</td>
</tr>
<tr>
<td>text_size</td>
<td>Character expansion factor for the labels of the experimental units.</td>
</tr>
<tr>
<td>point_size</td>
<td>Character expansion factor for the point representing the experimental units.</td>
</tr>
</tbody>
</table>

Value

This function just creates the units plot. It does not return a value.

Author(s)

Raul Cruz-Cano

References


Examples

```R
#Example: NCI-60 micrRNA Data
data("Topoisomerase_II_Inhibitors")
data("microRNA")
my_res <- frcc(t(microRNA), -1* t(Topoisomerase_II_Inhibitors))
for (i in 1:dim(microRNA)[2]) {
  colnames(microRNA)[i]<-substr(colnames(microRNA)[i], 1, 2)
}#end for i
dev.new()
plot_units(t(microRNA), -1* t(Topoisomerase_II_Inhibitors), my_res, 1, 1, text_size=0.01)
```
plot_variables

Plot variables in the Canonical Factor Loadings Space

Description

This function plots the variables used in the FRCCA as points in a two-dimensional plane in which the axis are the canonical factor loadings selected by the user.

Usage

plot_variables(res.mrcc, i, j, inner_circle_radius = 0.5, text_size = 0.8)

Arguments

res.mrcc List containing a canonical structure provided by the function frcc.
i Canonical Factor Loadings which will be used as the horizontal axis.
j Canonical Factor Loadings which will be used as the vertical axis.
inner_circle_radius Radius of the circle which is used to determine which variables are significant. Only the significant variables will be labeled.
text_size Character expansion factor for the labels of the variables.

Value

This function just creates the variables plot. It does not return a value.

Author(s)

Raul Cruz-Cano

References


Examples

# Example: Multivariate Normal Data
p<-10
t<-10
n<-50
res<-generate_multivariate_normal_sample(p,t,n)
X<-res$X
Y<-res$Y
rownames(X)<-c(1:n)
colnames(X)<-c(1:p)
colnames(Y)<- c(1:q)
my_res<-frcc(X,Y)
dev.new()
plot_variables(my_res,1,2,text_size=1.0)
**Description**

By using the minimum risk estimators of the correlation matrices instead of the sample correlation matrices the FRCC algorithm might disrupt the order of the canonical correlations and hence of the canonical structure. This is unacceptable for the algorithm used to calculate the p-values which requires the canonical correlations to be ordered in a descending order. This function rearranges the canonical structure according to the canonical correlations from largest to smallest.

**Usage**

```r
rearrange.frcc(res.frcc)
```

**Arguments**

- `res.frcc` List containing a canonical structure produced by the function `frcc`.

**Value**

- `res.frcc` List containing the sorted canonical structure.

**Author(s)**

Raul Cruz-Cano

**References**


**Examples**

```r
## This is an internal function. No need for examples.
```
soilspec

**Soil Specification Data**

**Description**
Contains the Soil Specification Data.

**Usage**
data(soilspec)

**Details**
The original purpose of the experiment was to determine the relationships between several soil characteristics of the limestone grassland in Wales and the abundance of certain plant species. These variables were measured in a random sample of 10 x 10 square meters in the community of Anglesey, North Wales. The dataset comprises data from 45 samples on 8 species of plants (H. pubescens, P. bertolonii, T. pretense, P. sanguisorba, R. squarrosus, H. pilosella, B. media and T. drucei) and 3 soil characteristics (d=depth, P=extractable phosphate and K=exchangeable potassium) and their interactions (d x P, d x K and P x K). Previous work shows that these soil characteristics are influential in determining how much each of the existing plant species can flourish. This set of plants was selected because they have a diverse response to variation of the soil variables.

**Value**
A matrix with the information corresponding to the 8 types of soils (columns 2-9) and the soil characteristics and their interactions (columns 10-16) for the 45 soil samples as described in [Cruz-Cano et al., 2012]. The first columns keeps track of the number of the site of origin. The original source of the data is [Gittins, 2005]. It was first used as an R dataset in [De’ath and Walsh, 2001].

**Author(s)**
Raul Cruz-Cano

**References**

**Examples**

```r
#Example #2 Soil Specification Data
data(soilspec)
list_of_units_to_be_used<-sample(1:nrow(soilspec),14)#We will only 14 soil samples
X<- soilspec[list_of_units_to_be_used,2:9]
```
Y <- soilspec[1:10]
colnames(X) <- c("H. pubescens", "P. bertolonii", "T. pretense", 
"P. sanguisorba", "R. squarrosus", "H. pilosella", "B. media", "T. drucei")
colnames(Y) <- c("d", "P", "K", "d x P", "d x K", "P x K")
my.res <- frcc(X, Y)

Topoisomerase_II_Inhibitors

NCI-60 Topoisomerase II Inhibitor Data.

Description

Load a matrix with Topoisomerase II Inhibitor Drugs dataset.

Usage

data(Topoisomerase_II_Inhibitors)

Details

The NCI-60 is a set of cell cultures grown under controlled conditions by the National Cancer 
Institute. The NCI-60 cell lines include experimental units from the breast (8), central nervous 
system (6), colorectal (7), lung (9), prostate (2), ovarian (6) and renal (8) cancers. It also includes 
leukemia (6) and melanoma (8) cell lines.

Value

A matrix with the growth inhibitory responses of the 15 Topoisomerase II Inhibitor Drugs for the 
60 cell lines in the NCI-60 dataset as described in [Cruz-Cano et al., 2012]. This is a subset of the 
A118 drugs dataset original described in [DTP, 2009].

Author(s)

Raul Cruz-Cano

References

DTP (2009). DTP Human Tumor Cell Line Screen.. Standard mechanism. Available from: 
http://dtp.nci.nih.gov/

Examples

#Example #3 NCI-60 miR-10a Data
data("Topoisomerase_II_Inhibitors")
data("miRNA")
my_res <- frcc(t(miRNA),-1*t(Topoisomerase_II_Inhibitors))
Index

*Topic Fast Regularized Canonical Correlation Analysis
  FRCC-package, 2
*Topic NCI-60
  microRNAs, 6
  Topoisomerase_II_Inhibitors, 12
*Topic Regularized Canonical Correlation Analysis
  FRCC-package, 2
*Topic Topoisomerase_II_Inhibitor
  Topoisomerase_II_Inhibitors, 12
*Topic \textasciitildekwd1
  custom.draw.circle, 3
  frcc, 3
  generate_multivariate_normal_sample, 5
  off.diagonal.lambda, 7
  plot_units, 8
  plot_variables, 9
  rearrange.frcc, 10
*Topic \textasciitildekwd2
  custom.draw.circle, 3
  frcc, 3
  generate_multivariate_normal_sample, 5
  off.diagonal.lambda, 7
  plot_units, 8
  plot_variables, 9
  rearrange.frcc, 10
*Topic microRNA
  microRNAs, 6
*Topic soil
  soilspec, 11

custom.draw.circle, 3

FRCC (FRCC-package), 2
frcc, 3
FRCC-package, 2

generate_multivariate_normal_sample, 5
microRNA (microRNAs), 6
microRNAs, 6
off.diagonal.lambda, 7
plot_units, 8
plot_variables, 9
rearrange.frcc, 10
soilspec, 11
Topoisomerase_II_Inhibitors, 12