

Package ‘ktspair’

December 14, 2011

Title k-Top Scoring Pairs for Microarray Classification

Version 1.0

Author Julien Damond <julien.damond@gmail.com>

Description These functions compute the k best pairs of genes used to classify samples based on the relative rank of the genes expression within each profile. A score based on the sensitivity and the specificity is calculated for every possible pair. The k pairs with the highest score will be selected with the restriction that a gene can appear in at most one pair. The value of k is either set as a parameter chosen by the user or computed through crossvalidation. Other functions related to the k-TSP are also available, for example the functions prediction, summary, plot, etc. can be found in the package.

Maintainer Julien Damond <julien.damond@gmail.com>

Depends R (>= 2.2.1), Biobase (>= 2.4.0)

License GPL-2

Archs i386

R topics documented:

ktspair-package	2
AUC.calc	3
bootstrap.graphic.ktsp	4
bootstrap.ktsp	5
cv	7
cv2	9
dat	10
eSet	11
grp	11
kts.pair	11
ktspcalc	13
ktspcalc2	16
ktspdata	18
ktspplot	18
make.consecutive.int	19

ordertsp	21
plot.ktsp	22
predict.ktsp	23
print.ktsp	24
rank_na	26
ROC.graphic.ktsp	27
ROC.offset	29
ROC.voting	30
summary.ktsp	32
Index	34

ktspair-package	<i>Computation of the k-TSP</i>
-----------------	---------------------------------

Description

The package computes the k-TSP. This method uses pairs of genes to perform a classification which compares the relative ordering of the gene expressions within each profile. It ranks pairs of genes with respect to a score based on the sensibility and the specificity achieved by each pair. It selects the k pairs that achieved the maximum score with the restriction that a gene can appear in at most one pair. The number of pairs of genes is computed through crossvalidation or can be chosen by the user. This package also contains function to display graphical properties of the k-TSP, to summarize the performance (accuracy, sensitivity, specificity, ROC curve) of the k-TSP, to predict new individual, etc. The functions deals also with "Expression Set" classes. The functions contained in this package are able to deal correctly with the presence of NAs in the dataset.

Details

Package:	ktspair
Type:	Package
Version:	1.0
Date:	2011-08-04
License:	GPL-2
LazyLoad:	yes

Author(s)

Julien Damond <julien.damond@gmail.com>
Maintainer: Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," Statist. Appl. in Genetics and Molecular Biology, 3, 2004.
A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," Bioinformatics, 21: 3896-3904, 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in Bioinformatics, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). `tspair`: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[kts.pair](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdta)
ktsp <- ktspcalc(dat, grp, 3)
ktsp <- ktspcalc(eSet, 1, 3)
ktsp <- ktspcalc(eSet, grp, 3)
ktsp
plot(ktsp)
summary(ktsp)
predict(ktsp)

## End(Not run)
```

AUC.calc

Compute the AUC

Description

Compute the AUC for a given curve (the area under the curve).

Usage

```
AUC.calc(at, values)
```

Arguments

<code>at</code>	The values of the curve on the x-axis
<code>values</code>	The values of the curve on the y-axis

Value

No value is returned

Note

This function is used in the function `ROC.graphic()` to compute the AUC for the ROC curves.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," Statist. Appl. in Genetics and Molecular Biology, 3, 2004.

A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," Bioinformatics, 21: 3896-3904, 2005.

Aik Choon Tan, Daniel Q. Naiman, Lei Xu, Raimond L. Winslow, and Donald Geman, "Simple decision rules for classifying human cancers from gene expression profiles," Bioinformatics, 21:3896-3904, October 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in Bioinformatics, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[kts.pair](#), [ktsplot](#), [predict.ktsp](#), [summary.ktsp](#), [ROC.voting](#), [ROC.offset](#), [ROC.graphic.ktsp](#)

bootstrap.graphic.ktsp

Graphical display of the bootstrap procedure

Description

This function displays the results of the bootstrap obtained with the function `bootstrap()`. It plots the frequency of appearance of the single genes as well as the pairs. If `k` was not set by the used an histogram of the values of `k` is also plotted.

Usage

```
bootstrap.graphic.ktsp(bootstrap, para1 = 0, para2 = 0, title = NULL, mtext = NU
```

Arguments

<code>bootstrap</code>	A bootstrap object obtained with the function <code>bootstrap()</code> .
<code>para1</code>	This parameter allows the user to control the lower bound of the frequency of appearance of the genes (the first plot).
<code>para2</code>	This parameter allows the user to control the lower bound of the frequency of appearance of pairs of the genes (the second plot).
<code>title</code>	This is used to determine the title of the graph.
<code>mtext</code>	Allow the user to specify a subtitle.

Details

The pair(s) of genes with the red dot represent(s) the pair(s) obtained on the original dataset.

Value

None, a graphic is plotted.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," Statist. Appl. in Genetics and Molecular Biology, 3, 2004.

A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," Bioinformatics, 21: 3896-3904, 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in Bioinformatics, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdta)
bootstrap1 <- bootstrap(dat, grp, k=3, n=20)
bootstrap2 <- bootstrap(dat, grp, n=30, seed=1)
bootstrap.graphic.ktsp(bootstrap1)
bootstrap.graphic.ktsp(bootstrap2)

## End(Not run)
```

bootstrap.ktsp

Bootstrap procedure for the k-TSP

Description

This function computes n bootstrap resamples of the original dataset, at each step it applies the k-TSP on the simulated dataset and records which pairs of genes were selected as well as the number k.

Usage

```
bootstrap.ktsp(dat, grp, k = NULL, seed = NULL, n = 500, length = 40, display =
```

Arguments

<code>dat</code>	Can either be (a) a matrix of m lines (the gene expressions) and n columns (the observations) or (b) an eSet object.
<code>grp</code>	Can either be (a) a character (or numeric) vector indicating the group of each observations or (b) an integer indicating the column of <code>pData(dat)</code> that represents the group of the observations.
<code>k</code>	If the number of pair of genes should be the same at each step or if it should be computed every time (by crossvalidation).
<code>seed</code>	If a seed should be set or not.
<code>n</code>	The number of bootstrap resamples.
<code>length</code>	To control the length of the list in the C code, see <code>kts.pair()</code> or <code>ktspcalc()</code> for more details.
<code>display</code>	If warnings should be displayed or not.
<code>med</code>	If the mean of the median between the two groups for each gene should be subtracted to the dataset or not.

Value

A bootstrap object with the following elements

<code>score</code>	A n by 9 matrix containing the scores of the k-TSP computed on the bootstrap. May contain NA (if $k < 9$).
<code>index</code>	A n by 18 matrix containing the index of the k-TSP computed on the bootstrap. May contain NA (if $k < 9$).
<code>k_value</code>	A vector of length n containing the number of pairs for each k-TSP (k).
<code>k</code>	A boolean indicating if the value of k was computed at every step or chosen as a parameter by the user.
<code>n</code>	The number of bootstrap resamples.
<code>genenames</code>	The names of the genes contained in the dataset. If none, a numerotation is used.
<code>ktsp</code>	A ktsp object computed on the original dataset.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in Bioinformatics, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
bootstrap1 <- bootstrap.ktsp(dat, grp, k=3, n=20)
bootstrap2 <- bootstrap.ktsp(dat, grp, n=20, seed=1)

## End(Not run)
```

cv

Crossvalidation for the parameter k

Description

This function computes the value of k through crossvalidation.

Usage

```
cv(dat, grp, cross = 5, display = FALSE, length = 40, seed = NULL, med = FALSE,
```

Arguments

dat	Can either be (a) a matrix of m lines (the gene expressions) and n columns (the observations) or (b) an eSet object.
grp	Can either be (a) a character (or numeric) vector indicating the group of each observations or (b) an integer indicating the column of pData(dat) that represents the group of the observations.
cross	The number of fold that should be used in the crossvalidation.
display	Allows the user to avoid the function ktspcalc() to print warning message over the loop.
length	This paramters allows the used to control the length of the list used in the C code.
seed	Allow the user to set a seed.
med	If the mean of the median between the two groups for each gene should be subtracted to the dataset or not.
healthy	This variable is used to determine which group will be considerer as the healthy group (specificity). Need to give the label of the group.

Details

This function computes the value of k through crossvalidation. The number of fold is given by the variable `cross` and by default is 5. It also computes the percentage of correct prediction based on the same partition as for the crossvalidation for the possible values of k .

Value

<code>k</code>	The selected value for k
<code>accuracy_k</code>	The estimated percentage of correct prediction achieved by the k -TSP with the selected k .
<code>accuracy</code>	A vector of the estimated percentage of correct prediction reached by the k -TSP with $k = 1, 3, 5, 7, 9$.
<code>sensitivity</code>	A vector of the estimated sensitivity reached by the k -TSP with $k = 1, 3, 5, 7, 9$.
<code>specificity</code>	A vector of the estimated specificity reached by the k -TSP with $k = 1, 3, 5, 7, 9$.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.
- J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k -TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.
- Jeffrey T. Leek <jtleek@jhu.edu> (). `tspair`: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdta)
cv <- cv(dat, grp, cross = 10)
ktsp <- ktspcalc(dat, grp, cv$k)
ktsp
cv

## End(Not run)
```


Description

This function computes the performance of the model where the pairs of genes has been selected by the user.

Usage

```
cv2(indice, dat, grp, cross = 5, display = FALSE, length = 40, seed = NULL, med
```

Arguments

<code>indice</code>	The pairs that should be used to construct the ktsp object
<code>dat</code>	Can either be (a) a matrix of m lines (the gene expressions) and n columns (the observations) or (b) an eSet object.
<code>grp</code>	Can either be (a) a character (or numeric) vector indicating the group of each observations or (b) an integer indicating the column of pData(dat) that represents the group of the observations.
<code>cross</code>	The number of fold that should be used in the crossvalidation.
<code>display</code>	Allows the user to avoid the function ktspcalc() to print warning message over the loop.
<code>length</code>	This paramters allows the used to control the length of the list used in the C code.
<code>seed</code>	Allow the user to set a seed.
<code>med</code>	If the mean of the median between the two groups for each gene should be subtracted to the dataset or not.
<code>healthy</code>	This variable is used to determine which group will be considerer as the healthy group (specificity). Need to give the label of the group.

Details

This function is used to compute the performance reached by the model where the pairs of genes have been chosen by the user.

Value

<code>k</code>	The selected value for k
<code>accuracy</code>	The estimated percentage of correct prediction reached by the k-TSP with the chosen pairs of genes.
<code>sensitivity</code>	The estimated sensitivity reached by the k-TSP with the chosen pairs of genes.
<code>specificity</code>	The estimated specificity reached by the k-TSP with the chosen pairs of genes.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.
- J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.
- Jeffrey T. Leek <jtleek@jhu.edu> (). *tspair*: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc2](#), [cv](#)

Examples

```
## Not run:
## Load data
data(ktspdta)
cv <- cv(dat, grp, cross = 10)
ktsp <- ktspcalc(dat, grp, cv$k)
ktsp
cv

## End(Not run)
```

dat

Simulated gene expressions of individuals.

Description

This is a 1000 by 50 simulated gene expression matrix, where the lines represent the genes and the column the individuals.

Usage

dat

Format

matrix

eSet

Simulated gene expressions of individuals.

Description

This dataset is a simulated gene expression set under the form of an eSet, where `exprs(eSet1) = dat` and `pData(eSet1) = grp`.

Usage

```
eSet
```

Format

Expression Set

grp

Group of the individuals.

Description

This is a vector of the group of the simulated profiles contained in the matrix "dat" in the k-TSP package. The groups are "healthy" and "diseased".

Usage

```
grp
```

Format

vector

kts.pair

Calculation of the k top scoring pairs.

Description

This function computes the k pairs of genes that achieved the maximum difference between the sensitivity and the specificity (in absolute value) between two specific groups based on the comparison of the expressions of the two genes present in the pairs. The function `ktspcalc()` is the general function and uses the function `kts.pair` once the dataset has been prepared for this function. The function `ktspcalc()` is also able to deal with eSets.

Usage

```
kts.pair(dat, grp, k, display = TRUE, length = 40, med = FALSE)
```

Arguments

<code>dat</code>	A matrix of <i>m</i> lines (the gene expressions) and <i>n</i> columns (the observations).
<code>grp</code>	A vector of 0 and 1 for the groups of the observations.
<code>k</code>	The number of pairs of genes that the function will select.
<code>display</code>	Allows the user to avoid the function <code>kts.pair()</code> to print warning message (mainly used in the function <code>crossvalidation</code>).
<code>length</code>	This paramters allows the used to control the length of the list used in the C code.
<code>med</code>	If the mean of the median between the two groups for each gene should be subtracted to the dataset or not.

Details

This function only works with matrices and vector of group containing only 0 and 1. For a more general use (eSets and labels for the groups) see the function `ktspcalc()`. This classifier can only be used for classifications with two groups. The k-TSP was introduced in Tan et al. (2005) and is an extension of the TSP, which was presented in Geman et al. (2004).

Value

A `ktsp` object with the following elements:

<code>index</code>	A <i>k</i> by 2 matrix composed of genes where the <i>i</i> th row stands for the <i>i</i> th best pair of genes with the restriction that a gene can appear in only one pair. The pairs are selected with respect to the score Delta and Gamma (in case of ties), see Tan et al. (2005) for more details about the k-TSP.
<code>ktspcore</code>	A vector of size <i>k</i> containing the scores Delta achieved by each selected pair of genes. The score Delta is based on the sensitivity and the specificity of a pair, see Geman et al. (2004) for more details.
<code>grp</code>	The group for each observation in a binary form
<code>ktspdat</code>	The row <i>i</i> and the row <i>i</i> + <i>k</i> represents the expressions of the genes present in the <i>i</i> th pair.
<code>k</code>	The number of pairs of genes.
<code>labels</code>	The name of the two groups that were present in the original variable <code>grp</code> .
<code>rankscore</code>	The score Gamma achieved by each pair of genes, for more details on this score see Geman et al. (2004).
<code>accuracy</code>	A vector of the estimated percentage of correct prediction for the k-TSP with <i>k</i> =1,3,5,7,9.
<code>accuracy_k</code>	The estimated percentage of correct prediction of the k-TSP with the selected <i>k</i> .
<code>sensitivity</code>	A vector of the estimated sensitivity for the k-TSP with <i>k</i> =1,3,5,7,9.
<code>specificity</code>	A vector of the estimated specificity for the k-TSP with <i>k</i> =1,3,5,7,9.
<code>med</code>	If the mean of the medians within each group has been subtracted to the dataset return the values of the mean of the median, return FALSE otherwise

note

The `length` sets to the list used in the C code (defined by the parameter `length`) has to be at least as big as *k*.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.

A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). *tspair*: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
ktsp <- kts.pair(dat, grp, 3)
ktsp$index
ktsp$ktspscore

ktsp2 <- kts.pair(dat, grp, 9)
ktsp2
ktsp2 <- kts.pair(dat, grp, 9, length=40)
ktsp2

## End(Not run)
```

ktspcalc

Compute the k top scoring pairs based on a gene expression matrix and a group indicator vector.

Description

This function computes the k pairs of genes that achieved the maximum difference between the sensitivity and the specificity (in absolute value) between two specific groups based on the comparison of the expressions of the two genes present in the pairs.

Usage

```
ktspcalc(dat, grp, k = NULL, cross = 5, performance = FALSE, healthy = NULL, see
```

Arguments

<code>dat</code>	Can either be (a) a matrix of <i>m</i> lines (the gene expressions) and <i>n</i> columns (the observations) or (b) an <code>eSet</code> object.
<code>grp</code>	Can either be (a) a character (or numeric) vector indicating the group of each observations or (b) an integer indicating the column of <code>pData(dat)</code> that represents the group of the observations.
<code>k</code>	The number of pairs of genes used in the k-TSP (<i>k</i>). By default this parameter is computed through crossvalidation.
<code>cross</code>	The number of fold that should be used for the crossvalidation estimation of the parameter <i>k</i> . By default the number of fold is 5.
<code>performance</code>	An indicator if the performance of the model should be computed or not.
<code>healthy</code>	This variable is used to determine which group will be considered as the healthy group (specificity). Need to give the label of the group.
<code>seed</code>	The crossvalidation and the computation of the performance of the model are based on a random partition of the dataset. The variable <code>seed</code> allows the user to fix a seed.
<code>display</code>	Allows the user to avoid the function <code>ktspcalc()</code> to print warning message (mainly used in the function <code>crossvalidation</code>).
<code>length</code>	This parameter allows the user to control the length of the list used in the C code.
<code>med</code>	If the mean of the median between the two groups for each gene should be subtracted to the dataset or not.

Details

The original version of the k-TSP only works for two groups classification. It is possible to deal with multiclass classification by using trees or multiple steps methods that reduce the problem to a combination of several two classes classifications, see Aik et al. (2005) for more details. The function computes the score Delta (see Geman et al. (2004) for more details about this score) for every possible pair of genes. This makes the required computational time to grow rapidly in the number of genes. A pre-filtering step can be useful in some cases. This function is able to deal with NA present in the dataset. It considers only the patients for which the gene expressions were measured and adapts the computation of the score Delta to the number of measures without NA. This function also deals with "Expression Set" dataset. The group indicator can be replaced by the number of the column of `pData(eSet)` that contains the indicator vectors of the group. The user has the possibility to let the function compute a value for the parameter *k*. This value is computed through crossvalidation. The user can choose the number of fold. The special case of the Leave-One-Out crossvalidation is when the number of fold is equal to the number of observations. The user has the possibility to obtain the accuracy reached by the method via the variable `performance`. It will compute the accuracy using a partitioning of the dataset as in the crossvalidation. If the crossvalidation had to be computed, the performance will be computed at the same time and will be based on the same partition. If a reference for a healthy group is given (via the variable `healthy`), the sensitivity and the specificity will be computed (estimated by crossvalidation). The number of partition for the performance is also given by the variable `cross`. The user can also fix a seed via the variable `seed`. The parameter `length` has to be used carefully. Indeed, choosing a low value will speed up the computation but may produce results with less pairs of genes than expected (*k*). In order to obtain *k* pairs of genes, one has to set a bigger value for this parameter. We note that, on some datasets, it is not possible to obtain *k* pairs of genes, this may be due to the dataset itself and not the algorithm.

Value

A ktsp object with the following elements:

index	A k by 2 matrix composed of genes where the ith row stands for the ith best pair of genes with the restriction that a gene can appear in only one pair. The pairs are selected with respect to the score Delta and Gamma (in case of ties), see Tan et al. (2005) for more details about the k-TSP.
ktspcore	A vector of size k containing the scores Delta achieved by each selected pair of genes. The score Delta is based on the sensitivity and the specificity of a pair, see Geman et al. (2004) for more details.
grp	The group for each observation in a binary form
ktspdat	The row i and the row i+k represents the expressions of the genes present in the ith pair.
k	The number of pairs of genes.
labels	The name of the two groups that were present in the original variable grp.
rankscore	The score Gamma achieved by each pair of genes, for more details on this score see Geman et al. (2004).
accuracy	A vector of the estimated percentage of correct prediction for the k-TSP with k=1,3,5,7,9.
accuracy_k	The estimated percentage of correct prediction of the k-TSP with the selected k.
sensitivity	A vector of the estimated sensitivity for the k-TSP with k=1,3,5,7,9.
specificity	A vector of the estimated specificity for the k-TSP with k=1,3,5,7,9.
med	If the mean of the medians within each group has been subtracted to the dataset return the values of the mean of the median, return FALSE otherwise

note

The length sets to the list used in the C code (defined by the parameter length) has to be at least as big as k.

Warning

If NAs are present in the dataset, the computation of the score Delta will be based only on observations for which the measures of the two genes of the current comparison are not NA. This will reduce the number of observations used to compute the score Delta and can produce lower accuracy of the estimation compared to the scores for others pairs.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- Aik Choon Tan, Daniel Q. Naiman, Lei Xu, Raimond L. Winslow, and Donald Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21:3896-3904, October 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in Bioinformatics, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[kts.pair](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
ktsp <- ktscalcd(dat, grp, 3)
ktsp <- ktspcalc(eSet, grp, 3)
ktsp <- ktspcalc(eSet, 1, 3)
ktsp
ktsp$rankscore
ktsp$accuracy_k
ktsp$accuracy

ktsp2 <- ktspcalc(dat, grp, 9)
ktsp2
ktsp2 <- ktspcalc(dat, grp, 9, length=40)
ktsp2

## End(Not run)
```

ktspcalc2

The k-TSP with chosen pairs of genes

Description

This function creates a ktsp object where the pairs of genes used have been selected by the user.

Usage

```
ktspcalc2(indice, dat, grp, length = 40, cross = 5, display = display, med = FALSE)
```

Arguments

indice	The pairs that should be used to construct the ktsp object
dat	Can either be (a) a matrix of m lines (the gene expressions) and n columns (the observations) or (b) an eSet object.
grp	Can either be (a) a character (or numeric) vector indicating the group of each observations or (b) an integer indicating the column of pData(dat) that represents the group of the observations.

length	This paramters allows the used to control the length of the list used in the C code.
cross	The number of fold that should be used in the crossvalidation.
display	Allows the user to avoid the function ktspcalc() to print waring message over the loop.
med	If the mean of the median between the two groups for each gene should be subtracted to the dataset or not.
performance	An indicator if the performance of the model should be computed or not.
seed	The crossvalidation and the computation of the performance of the model are based on a random partition of the dataset. The variable seed allows the user to fix a seed.
healthy	This variable is used to determine which group will be considerer as the healthy group (specificity). Need to give the label of the group in the same form as given in the variable grp.

Details

This function would be useful when the user wants to select pairs of genes based on the results of the bootstrap. The pairs of genes will be then reordered based on the score Delta and Gamma from the best one to the worst one. When the genes are inserted (through the variable indice), the two first genes will represent the first pair, the two next genes the second pair, etc.

Value

A ktsp object, see the function ktspcalc() for more details.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.
- J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.
- Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
kts2 <- ktspcalc2(c(100,200,300,400),dat, grp)
kts2

## End(Not run)
```

ktspdata

Simulated dataset of gene expressions in a matrix form.

Description

This dataset can be used to illustrate the functions contained in the package "ktspair". The gene expressions are simulated from poisson random variables, where the means come from a normal distribution with different parameters for the group of healthy and diseased people. The first 25 measures are healthy people (represented by 0 in the group vector) and the 25 last measures are diseased people (represented by 1 in the group vector).

Usage

```
dat
```

ktspplot

Graphical representation of kts objects

Description

Show a graphical representation of kts objects based on the expressions of the genes present in the pairs constituting the k-TSP.

Usage

```
ktspplot(ktsobj, select = NULL)
```

Arguments

ktsobj	A kts object (computed through the functions ktspcalc or kts.pair).
select	If the number of TSPs in the k-TSP is greater than 1, one can choose the representation of a single TSP (select = 1,2,...,k). By default the graphical representation of every TSPs in succession is plotted.

Details

ktspplot() creates a two dimensional graph for a single TSP present in the k-TSP. The graph plots the expression of the first gene in the pair versus the expression of the second gene in the pair of interest. The group of the observations are plotted in red and blue, the score of the current pair is shown on the top of the graph. The black line of slope 1 and intercept 0 represents the decision rule of the current TSP. If not only one TSP is selected, hitting return will plot the next TSP in the k-TSP.

Value

No value is returned. Only a graph is plotted.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.

A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). *tspair*: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [kts.pair.predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdta)
ktsp <- ktspcalc(dat,grp,3)
ktspplot(ktsp,select=1)
ktspplot(ktsp)

## End(Not run)
```

make.consecutive.int

Transform the group vector into a binary vector

Description

This function transforms the vector corresponding to the groups of the observations into a binary vector.

Usage

```
make.consecutive.int(y)
```

Arguments

`y` A numeric or character vector.

Value

Return a vector containing only 0 and 1.

Note

This function is useful to link the functions `kts.pair` and `ktspcalc` (`kts.pair` accepts only a vector `grp` containing 0 and 1).

Author(s)

Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.

A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). `tspair`: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdta)
make.consecutive.int(grp)

## End(Not run)
```

ordertsp	<i>Ordering of the pairs of genes</i>
----------	---------------------------------------

Description

Determine the order (decreasing) of the pairs of genes represented by their scores.

Usage

```
ordertsp(delta, gamma)
```

Arguments

delta	The score Delta of the pairs of genes
gamma	The score Gamma of the pairs of genes

Value

The position in which each pair should appear to be ordered (based on delta and gamma).

Note

This function is used in the function `ktspcalc2()` to order the selected pairs of genes.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- Aik Choon Tan, Daniel Q. Naiman, Lei Xu, Raimond L. Winslow, and Donald Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21:3896NAK3904, October 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.
- J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.
- Jeffrey T. Leek <jtleek@jhu.edu> (). `tspair`: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[kts.pair](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#), [ktspcalc2](#)

plot.ktsp

Graphical representation of ktsp objects

Description

Show a graphical representation of ktsp objects based on the expressions of the genes present in the pairs constituting the k-TSP.

Usage

```
## S3 method for class 'ktsp'
plot(x, ...)
```

Arguments

x	A ktsp object.
...	Plotting arguments, not used

Details

This is the generic function for ktspplot.

Value

No value is returned. Only a graph is plotted.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.
- J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.
- Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [kts.pair](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdta)
ktsp <- kts.pair(dat,grp,3)
plot(ktsp,select=1)
plot(ktsp)

## End(Not run)
```

predict.ktsp	<i>Prediction using a ktsp object</i>
--------------	---------------------------------------

Description

This function is used to predict the group of an individual based on its gene expressions via a ktsp object.

Usage

```
predict.ktsp(object, dat = NULL, select = NULL, display = TRUE)
```

Arguments

object	A ktsp object
dat	A dataset from which predictions should be computed. It can be either under the form of a matrix or under the form of an eSet.
select	An integer to perform the prediction based on a single TSP present in the k-TSP.
display	Allows the user to avoid the function ktspcalc() to print warning message over the loop.

Details

By default, this function computes predictions of the original dataset (the one used to compute the k-TSP) based on the k-TSP. It is possible to predict new observations via the variable "dat". It can either be under the form of a matrix or of an eSet. The function searches for the gene names of the ktsp and try to identify them in the rownames or featuresNames of the matrix of gene expressions "dat". If rownames or featureNames are not available, it uses the number of the row as name to perform the predictions. By default, the prediction is based on all the pairs present in the k-TSP and by using a voting system. It is also possible to have predictions based on a single pair by using the variable "select". In order to obtain a well defined prediction, the number of pairs of genes used has to be an odd number. If the number of pairs is even, the function will not consider the pair with the lowest score Delta in the calculation of the prediction.

Value

A vector containing the class predictions of each individual based on the k-TSP is returned.

warning

If rownames or featureNames of the gene expression matrix "dat" are not available, the number of the row will be used as name. To be correct the order of the genes in the matrix "dat" must be the same as in the ktsp object (or equivalently the same as in the original dataset).

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," Statist. Appl. in Genetics and Molecular Biology, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," Bioinformatics, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.
- J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in Bioinformatics, December 2011.
- Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [kts.pair](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdta)
ktsp <- kts.pair(dat, grp, 3)
predict(ktsp)
predict(ktsp, select=1)

## End(Not run)
```

```
print.ktsp
```

Print the results of the k-TSP

Description

This function is used to print the results of a ktsp object.

Usage

```
print.ktsp(x)
```


Arguments

x A ktsp object.

Details

It prints the results of the k-TSP. More precisely, the number of pairs used in the k-TSP, the gene present in the pairs and the score of each pair.

Value

No value is returned. A table is plotted.

Note

Once the function ktsp.print is defined, writing "x" has the same effect as writing "print.ktsp(x)", where x is a ktsp object.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," Statist. Appl. in Genetics and Molecular Biology, 3, 2004.

A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," Bioinformatics, 21: 3896-3904, 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in Bioinformatics, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [kts.pair](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
ktsp <- ktspcalc(dat, grp, 3)
ktsp.print(ktsp)
ktsp

## End(Not run)
```

rank_na

*Rank the gene expression and Replace NA***Description**

This function computes the ranks of the gene expression within each row (within one profile) and replaces the NAs present in the dataset by a given value.

Usage

```
rank_na(dat, na = -1e+05)
```

Arguments

dat	A gene expression dataset under the form of a matrix.
na	The value by which the NAs should be replaced, by default -1e+05.

Value

A matrix of the same size as the matrix dat. The columns of this matrix contain the ranks of the gene expressions within each profile. In other words the ith row of the new matrix contains the ranks of the gene expression of the ith column of the matrix dat. If NA are present in the matrix dat, they won't be taken into account as computing the ranks and will be replaced by the value na, which is, by default, -1e+05.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.

A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). *tspair*: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [kts.pair](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
rank_na(dat, 2000)

## End(Not run)
```

ROC.graphic.ktsp *Graphical display of the ROC curve*

Description

This functions displays the results of the ROC curve obtained through the function ROC().

Usage

```
ROC.graphic.ktsp(roc, m = 1, boxplot = TRUE, AUC = TRUE, auc.x = 0.55, auc.y = 0)
```

Arguments

roc	A roc object obtained with the function ROC().
m	Allows the user to determine the proportion (1/m) of boxplot that should be plotted.
boxplot	A logical integer that determines if the boxplot should be plotted or not (if FALSE, the values are plotted as points).
AUC	If the AUC should be computed or not.
auc.x	If AUC=TRUE, the x coordinate where the AUC value will be plotted.
auc.y	If AUC=TRUE, the y coordinate where the AUC value will be plotted.
box.col	The color of the boxplots.
line.col	The color of the line that represent the median of the sensitivity for several values of the specificity.
multiple.col	If multiple cutoff were used to construct the roc object (in the function ROC()), the colors that should be used on the summary graph.
maintitle	Allow the user to specify a title.
mtext	Allow the user to specify a subtitle.
undertitle	Allow the user to control the title display in the single graph display more accurately.
graphic	If a graphic should be displayed or not.

Details

In the function ROC() a bootstrap procedure was used to construct several representations of ROC curves. In the function ROC.graphic(), these curves are used to create boxplot around the values of the sensitivity for several values of the specificity. If, for a given value of the specificity, several sensitivities were available, the median was used to represent this point and the different values of the sensitivity were used to draw a boxplot.

Value

A ROC.graphic object with the following elements

at	Either a vector or a 3 rows matrix (if mult.cutoff=FALSE or TRUE resp.) representing the x-axis of the curve for median of the sensitivity (the red line).
median	Either a vector or a 3 rows matrix (if mult.cutoff=FALSE or TRUE resp.) representing the y-axis of the curve for median of the sensitivity (the red line).
auc	Either a single value or a vector of length 3 (if mult.cutoff=FALSE or TRUE resp.) standing for the AUC of the model(s).

Author(s)

Julien Damond <julien.damond@gmail.com>

References

D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.

A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.

J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). *tspair*: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
roc1 <- ROC(dat, grp, n=200, healthy="healthy", mult.cutoff=FALSE)
roc2 <- ROC(dat, grp, n=200, healthy="healthy", mult.cutoff=TRUE)
ROC.graphic.ktsp(roc1)
ROC.graphic.ktsp(roc2)

## End(Not run)
```

ROC.offset

*ROC curve based on the offset method for k-TSP method***Description**

This function creates a ROC object used by the function ROC.graphic.ktsp() to plot a ROC curve based on the offset method.

Usage

```
ROC.offset(dat, grp, n = 200, healthy = NULL, seed = NULL, para1 = 200, para2 =
```

Arguments

dat	Can either be (a) a matrix of m lines (the gene expressions) and n columns (the observations) or (b) an eSet object.
grp	Can either be (a) a character (or numeric) vector indicating the group of each observations or (b) an integer indicating the column of pData(dat) that represents the group of the observations.
n	The number of bootstrap resample that should be used to compute the ROC curve.
healthy	This variable is used to determine which group will be considerer as the healthy group (specificity). Need to give the label of the group.
seed	If a seed should be set or not.
para1	This paramater is used to control the smoothness of the ROC curves computed on the bootstrap.
para2	This paramater is also used to control the smoothness of the ROC curves computed on the bootstrap.
mult.cutoff	If multiple value of the cutoff in the majority system procedure should be used or not, the cutoff will be 0.25, 0.5 and 0.75.c
length	This paramters allows the used to control the length of the list used in the C code.
display	If warnings should be displayed or not.
med	If the mean of the median between the two groups for each gene should be subtracted to the dataset or not.

Details

The parameters para1 and para2 are used to make the ROC curve smoother. The way of computing the ROC curve is based on the offset method. This procedure is described in Damond et al. (paper is submitted in Bioinformatic, December 2011).

Value

A ROC object with the following elements

spec	A n by 2(para1+para2)+1 matrix representing the sensitivities achieved among the bootstraps.
------	--

<code>sens</code>	A n by $2(\text{para1}+\text{para2})+1$ matrix representing the specificities achieved among the bootstraps.
<code>mult.cutoff</code>	A boolean indicating if several cutoffs in the majority procedure have been used or not.
<code>n</code>	The number of bootstrap resample used.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.
- J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.
- Jeffrey T. Leek <jtleek@jhu.edu> (). `tspair`: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#), [ROC.voting](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
roc1 <- ROC.offset(dat, grp, n=20, healthy="healthy", mult.cutoff=FALSE)
roc2 <- ROC.offset(dat, grp, n=20, healthy="healthy", mult.cutoff=TRUE)

## End(Not run)
```

ROC.voting

ROC curve based on the voting system for k-TSP method

Description

This function creates a ROC object used by the function `ROC.graphic.ktsp()` to plot a ROC curve based on a cutoff on the majority of voting system.

Usage

```
ROC.voting(dat, grp, n = 200, healthy = NULL, seed = NULL, para1 = 200, para2 =
```

Arguments

<code>dat</code>	Can either be (a) a matrix of m lines (the gene expressions) and n columns (the observations) or (b) an eSet object.
<code>grp</code>	Can either be (a) a character (or numeric) vector indicating the group of each observations or (b) an integer indicating the column of pData(dat) that represents the group of the observations.
<code>n</code>	The number of bootstrap resample that should be used to compute the ROC curve.
<code>healthy</code>	This variable is used to determine which group will be considerer as the healthy group (specificity). Need to give the label of the group.
<code>seed</code>	If a seed should be set or not.
<code>para1</code>	This paramater is used to control the smoothness of the ROC curves computed on the bootstrap.
<code>para2</code>	This paramater is also used to control the smoothness of the ROC curves computed on the bootstrap.
<code>length</code>	This paramters allows the used to control the length of the list used in the C code.
<code>display</code>	If warnings should be displayed or not.
<code>med</code>	If the mean of the median between the two groups for each gene should be subtracted to the dataset or not.

Details

The parameters `para1` and `para2` are used to make the ROC curve smoother. The way of computing the ROC curve is based on a cutoff on the voting system. This procedure is described in Damond et al. (paper is submitted in Bioinformatic, December 2011).

Value

A ROC object with the following elements

<code>spec</code>	A n by 2(para1+para2)+1 matrix representing the sensitivities achieved among the boostraps.
<code>sens</code>	A n by 2(para1+para2)+1 matrix representing the specificities achieved among the boostraps.
<code>n</code>	The number of bootstrap resample used.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.

J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in Bioinformatics, December 2011.

Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#), [ROC.offset](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
roc <- ROC.voting(dat, grp, n=20, healthy="healthy")

## End(Not run)
```

summary.ktsp

Summary of ktsp object

Description

This function summarizes a ktsp object.

Usage

```
summary.ktsp(object, select = NULL, printall = FALSE)
```

Arguments

object	A ktsp object.
select	To obtain the summary for only a single TSP in the k-TSP.
printall	If the summary for the whole method k-TSP or if all the summary of the TSPs should be plotted in succession.

Details

It is possible to obtain the summary of one TSP present in the k-TSP via the variable select. To obtain the summary of all the TSPs in succession one should use the variable printall with the value TRUE. By default, a summary of the whole method k-TSP will be plotted.

Value

No value is returned. Only a table is plotted.

Note

It has the same effect to write "summary.ktsp(x)" or "summary(x)" for a ktsp object x.

Author(s)

Julien Damond <julien.damond@gmail.com>

References

- D. Geman, C. d'Avignon, D. Naiman and R. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Statist. Appl. in Genetics and Molecular Biology*, 3, 2004.
- A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, 21: 3896-3904, 2005.
- J. Damond, supervised by S. Morgenthaler and S. Hosseinian, "Presentation and study of robustness for several methods to classify individuals based on their gene expressions", Master thesis, Swiss Federal Institute of Technology Lausanne (Switzerland), 2011.
- J. Damond, S. Morgenthaler, S. Hosseinian, "The robustness of the TSP and the k-TSP and the computation of ROC curves", paper is submitted in *Bioinformatics*, December 2011.
- Jeffrey T. Leek <jtleek@jhu.edu> (). tspair: Top Scoring Pairs for Microarray Classification. R package version 1.10.0.

See Also

[ktspcalc](#), [ktspplot](#), [predict.ktsp](#), [summary.ktsp](#)

Examples

```
## Not run:
## Load data
data(ktspdata)
ktsp <- kts.pair(dat,grp,3)
summary.ktsp(ktsp, select=1)
summary(ktsp, select=1)
summary(ktsp, printall=FALSE)
summary(ktsp, printall=TRUE)

## End(Not run)
```

Index

*Topic **datasets**

`dat`, 10
`eSet`, 11
`grp`, 11
`ktspdta`, 18

*Topic **misc**

`AUC.calc`, 3
`bootstrap.graphic.ktsp`, 4
`bootstrap.ktsp`, 5
`cv`, 7
`cv2`, 9
`kts.pair`, 11
`ktspair-package`, 2
`ktspcalc`, 13
`ktspcalc2`, 16
`ktspplot`, 18
`make.consecutive.int`, 19
`ordertsp`, 21
`plot.ktsp`, 22
`predict.ktsp`, 23
`print.ktsp`, 24
`rank_na`, 26
`ROC.graphic.ktsp`, 27
`ROC.offset`, 29
`ROC.voting`, 30
`summary.ktsp`, 32

`AUC.calc`, 3

`bootstrap.graphic.ktsp`, 4
`bootstrap.ktsp`, 5

`cv`, 7, 10
`cv2`, 9

`dat`, 10

`eSet`, 11

`grp`, 11

`kts.pair`, 3, 4, 11, 16, 19, 21, 22, 24–26
`ktspair` (*ktspair-package*), 2
`ktspair-package`, 2

`ktspcalc`, 5, 7, 8, 13, 13, 17, 19, 20, 22,
24–26, 28, 30, 32, 33
`ktspcalc2`, 10, 16, 21
`ktspdta`, 18
`ktspplot`, 3–5, 7, 8, 13, 16, 17, 18, 20, 21,
24, 28, 30, 32, 33

`make.consecutive.int`, 19

`ordertsp`, 21

`plot.ktsp`, 22
`predict.ktsp`, 3–5, 7, 8, 13, 16, 17, 19–22,
23, 25, 26, 28, 30, 32, 33
`print.ktsp`, 24

`rank_na`, 26
`ROC.graphic.ktsp`, 4, 27
`ROC.offset`, 4, 29, 32
`ROC.voting`, 4, 30, 30

`summary.ktsp`, 3–5, 7, 8, 13, 16, 17, 19–22,
24–26, 28, 30, 32, 32, 33