Package ‘tspair’

March 29, 2017

**Title**  Top Scoring Pairs for Microarray Classification

**Version**  1.32.0

**Author**  Jeffrey T. Leek <jtleek@jhu.edu>

**Description**  These functions calculate the pair of genes that show the maximum difference in ranking between two user specified groups. This "top scoring pair" maximizes the average of sensitivity and specificity over all rank based classifiers using a pair of genes in the data set. The advantage of classifying samples based on only the relative rank of a pair of genes is (a) the classifiers are much simpler and often more interpretable than more complicated classification schemes and (b) if arrays can be classified using only a pair of genes, PCR based tests could be used for classification of samples. See the references for the tspcalc() function for references regarding TSP classifiers.

**Maintainer**  Jeffrey T. Leek <jtleek@jhu.edu>

**Depends**  R (>= 2.10), Biobase (>= 2.4.0)

**License**  GPL-2

**biocViews**  Microarray

**NeedsCompilation**  yes

**R topics documented:**

- dat ................................................................. 2
- dat2 ............................................................... 2
- eSet1 ............................................................... 2
- eSet2 ............................................................... 3
- grp ................................................................. 3
- plot.tsp .......................................................... 3
- predict.tsp ....................................................... 4
- summary.tsp ...................................................... 5
- ts.pair ............................................................ 7
- tspcalc ........................................................... 8
- tspdata ........................................................... 9
- tspplot .......................................................... 10
- tspsig ............................................................ 11

**Index**  13
dat

Simulated gene expression data.

Description
This data set is a simulated gene expression matrix "dat" with 1000 genes and 50 arrays.

Usage
dat

Format
matrix

dat2

Simulated gene expression data.

Description
This data set is a simulated gene expression matrix "dat2" with 500 genes and 20 arrays used for testing predictions based on the data "dat".

Usage
dat2

Format
matrix

eSet1

A simulated expression set.

Description
This is a simulated expression set, where exprs(eSet1) = dat and pData(eSet1) = grp.

Usage
eSet1

Format
Expression Set
**eSet2**

A simulated expression set.

**Description**

This is a simulated test expression set, where exprs(eSet2) = dat2.

**Usage**

eSet2

**Format**

Expression Set

---

**grp**

A group indicator.

**Description**

This is a group indicator for the simulated gene expression data "dat" in the TSP package. The two groups are "healthy" and "diseased".

**Usage**

grp

**Format**

vector

---

**plot.tsp**

Graphical display of tsp objects

**Description**

Graphical display of tsp objects

**Usage**

```r
## S3 method for class 'tsp'
plot(x,y,...)
```

**Arguments**

- `x` A tsp object
- `y` Not used
- `...` Plotting arguments, not used
predict.tsp

**Details**

tspplot() creates a plot of each top scoring pair. The figure plots the expression for the first gene in the TSP pair versus the expression for the second gene in the TSP pair across arrays. The user defined groups are plotted in the colors red and blue. The score for the pair is shown across the top of each plot. If there is more than one TSP, hitting return will cycle from one TSP to the next.

**Value**

Nothing of interest.

**Author(s)**

Jeffrey T. Leek <jtleek@jhu.edu>

**References**


**See Also**

tspcalc, ts.pair, tspsig,predict.tsp, summary.tsp, tspplot

**Examples**

```r
## Not run:
## Load data
data(tspdata)
tsp1 <- tspcalc(dat, grp)
plot(tsp1)

## End(Not run)
```

---

**predict.tsp**

*Prediction based on a tsp object*

**Description**

This function can be used to predict outcome values for a data set based on a tsp object.

**Usage**

```r
## S3 method for class 'tsp'
predict(object, dat=NULL, select=NULL, ...)
```

**Arguments**

- `object`: A tsp object
- `dat`: Can take two values: (a) an m genes by n arrays matrix of expression data or (b) an eSet object
- `select`: An indicator of which TSP to use, defaults to the first TSP.
- `...`: Plotting arguments (ignored)
**summary.tsp**

**Details**

predict() accepts a tsp object calculated on an expression set or gene expression matrix. If no other data set is included, the tsp predictions for the original data set are produced. If a second gene expression matrix or expression set is included, predict() looks for the gene names of the TSP in tspobj and attempts to match them in the rownames or featureNames of the gene expression matrix. If rownames or featureNames are not available, the prediction is based on the row numbers. If a match is identified, predict() makes a prediction for each gene based on the output.

**Value**

predict A class prediction for each array of dat based on the TSP from tspobj

**Author(s)**

Jeffrey T. Leek <jtleek@jhu.edu>

**References**


**See Also**

tspplot.ts.pair, tspcalc.tspsig.summary.tsp

**Examples**

```r
## Not run:
## Load data
data(tspdata)

## Run tspcalc() on a data matrix and grp vector
tsp1 <- tspcalc(dat,grp)

## Get predictions for a new eSet or data matrix
predict.tsp(tsp1,dat2,1)
predict(tsp1,eSet2,1)

## End(Not run)
```

**summary.tsp**

**Summary of a tsp object**

**Description**

This function can be used to summarize a TSP object.

**Usage**

```r
## S3 method for class 'tsp'
summary(object,select=NULL,printall=FALSE,...)
```
summary.tsp

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>A tsp object</td>
</tr>
<tr>
<td>select</td>
<td>A numerical indicator of which TSP to summarize, if NULL, all TSPs are summarized one at a time</td>
</tr>
<tr>
<td>printall</td>
<td>If printall=TRUE all TSPs are summarized at once</td>
</tr>
</tbody>
</table>

Details

summary() accepts a tsp object calculated on an expression set or gene expression matrix. The result is a table of frequencies for the TSP indicator versus the group indicator.

Value

A table.

Author(s)

Jeffrey T. Leek <jtleek@jhu.edu>

References


See Also

tspplot, ts.pair, tspcalc, tspsig, predict.tsp

Examples

```r
## Not run:
## Load data
data(tspdata)

## Run tspcalc() on a data matrix and grp vector
tsp1 <- tspcalc(dat,grp)

## Get the summary for the tsp object.
summary(tsp1)
summary.tsp(tsp1)

## End(Not run)
```
ts.pair Calculation of top scoring pairs.

**Description**

This function calculates the pair of genes that show the maximum difference in ranking between two user specified groups. The function tspcalc() is the general version of this function that accepts eSets.

**Usage**

```r
ts.pair(dat, grp)
```

**Arguments**

- `dat`: An m genes by n arrays matrix of expression data.
- `grp`: A group indicator with values 0 and 1 for the two groups.

**Details**

ts.pair() only works for two group classification. All the caveats for tspcalc() apply here, but this function only works for matrix and vector arguments, for a more general function see tspcalc(). The top scoring pairs methodology was originally described in Geman et al. (2004). and the unique TSP idea was described in Tan et al. (2005).

**Value**

A tsp object with elements:

- `index`: A two-column matrix where each row is a pair of indices from the data matrix achieving the top score.
- `tspscore`: The rank based score described in Geman et al. (2004), essentially the empirical average of sensitivity and specificity for the pair.
- `score`: The tie-breaking score described in Tan et al. (2005).
- `grp`: The binary group indicator.
- `tspdat`: Row i and row (i + Number of total TSPs) of this data matrix represent the expression data for the ith TSP pair.
- `labels`: The group labels for the two groups as defined by the grp variable.

**Author(s)**

Jeffrey T. Leek <jtleek@jhu.edu>

**References**


tspcalc

Estimate top scoring pairs from a gene expression matrix

Description

This function calculates the pair of genes that show the maximum difference in ranking between two user specified groups. The "top scoring pair" (TSP) maximizes the average of sensitivity and specificity over all rank based classifiers using a pair of genes in the data set.

Usage

tspcalc(dat, grp)

Arguments

dat
Can take two values: (a) an m genes by n arrays matrix of expression data or (b) an eSet object

grp
Can take one of two values: (a) A group indicator in character or numeric form, (b) an integer indicating the column of pData(dat) to use as the group indicator.

Details

tspcalc only works for two group classification. The computation time grows rapidly in the number of genes, so for large gene expression matrices one should be prepared to wait or do a pre-filtering step. The top scoring pairs methodology was originally described in Geman et al. (2004) and the unique TSP idea was described in Tan et al. (2005).

Value

A tsp object with elements:

- index: A two-column matrix where each row is a pair of indices from the data matrix achieving the top score
- tspscore: The rank based score described in Geman et al. (2004), essentially the empirical average of sensitivity and specificity for the pair.
- score: The tie-breaking score described in Tan et al. (2005).
- grp: The binary group indicator.
- tspdat: Row i and row (i + Number of total TSPs) of this data matrix represent the expression data for the ith TSP pair.
- labels: The group labels for the two groups as defined by the grp variable.

Examples

```r
## Not run:
## Load data
data(tspdata)
tsp1 <- ts.pair(dat, grp)
tsp1$index

## End(Not run)
```

See Also
tspplot, tspcalc, tspsig, predict.tsp, summary.tsp
tspdata

Author(s)

Jeffrey T. Leek <jtleek@jhu.edu>

References


See Also
tspplot, ts.pair, tspsig, predict.tsp, summary.tsp

Examples

## Not run:
## Load data
data(tspdata)

## Run tspcalc() on a data matrix and grp vector
tsp1 <- tspcalc(dat,grp)
tsp1$index

## Run tspcalc() on an expression set and a column of the pData matrix
tsp2 <- tspcalc(eSet1,1)
tsp2$index

## End(Not run)

---

tspdata  

Simulated gene expression data in both matrix and expression set format.

Description

These data sets can be used to illustrate the TSP functions. The data consist of two simulated data sets "dat", "dat2", a group indicator "grp" for "dat", and two expression sets "eSet1" and "eSet2" where the expression of eSet1 is "dat", the expression of eSet2 is dat2, and the pData for eSet1 consists of "grp".

Usage
dat
tsspplot

Graphical display of tsp objects

Description

Graphical display of tsp objects

Usage

```
tsspplot(tspobj)
```

Arguments

- `tspobj` A tsp object.

Details

tsspplot() creates a plot of each top scoring pair. The figure plots the expression for the first gene in the TSP pair versus the expression for the second gene in the TSP pair across arrays. The user defined groups are plotted in the colors red and blue. The score for the pair is shown across the top of each plot. If there is more than one TSP, hitting return will cycle from one TSP to the next.

Value

Nothing of interest.

Author(s)

Jeffrey T. Leek <jtleek@jhu.edu>

References


See Also

```
tspcalc, ts.pair, tspsig, predict.tsp, summary.tsp
```

Examples

```
## Not run:
## Load data
data(tsspdata)
tsp1 <- tspcalc(dat, grp)
tsspplot(tsp1)
## End(Not run)
```
tspsig

Significance calculation for top scoring pairs

Description

This function calculates the significance of a top-scoring pair. It can be run after tspcalc() to calculate how strong a TSP is.

Usage

```r
tspsig(dat, grp, B=50, seed=NULL)
```

Arguments

dat Can take two values: (a) an m genes by n arrays matrix of expression data or (b) an eSet object

grp Can take one of two values: (a) A group indicator in character or numeric form, (b) an integer indicating the column of pData(dat) to use as the group indicator

B The number of permutations to perform in calculation of the p-value, default is 50.

seed If this is a numeric argument, the seed will be set for reproducible p-values.

Details

tpsig() only works for two group classification. The computation time grows rapidly in the number of genes, so for large gene expression matrices one should be prepared to wait or do a pre-filtering step. A progress bar is shown which gives some indication of the time until the calculation is complete. The top scoring pairs methodology was originally described in Geman et al. (2004).

Value

p A p-value for testing the null hypothesis that there is no TSP for the data set dat.

nullscores The null TSP scores from the permutation test.

Author(s)

Jeffrey T. Leek <jtleek@jhu.edu>

References


See Also

tspplot, ts.pair, tspcalc, predict.tsp, summary.tsp
Examples

## Not run:
## Load data
data(tspdata)

## Run tspcalc() on a data matrix and grp vector
tsp1 <- tspcalc(dat,grp)

## Run tspsig() to get a p-value
p <- tspsig(dat,grp)
p

## End(Not run)
Index

∗Topic **datasets**
  dat, 2
  dat2, 2
  eSet1, 2
  eSet2, 3
  grp, 3
  tspdata, 9

∗Topic **misc**
  plot.tsp, 3
  predict.tsp, 4
  summary.tsp, 5
  ts.pair, 7
  tspcalc, 8
  tspplot, 10
  tspsig, 11

dat, 2
dat2, 2

eSet1, 2
eSet2, 3

grp, 3

plot.tsp, 3
predict(predict.tsp), 4
predict.tsp, 4, 6, 8–11

summary(summary.tsp), 5
summary.tsp, 4, 5, 5, 8–11

ts.pair, 4–6, 7, 9–11
tspcalc, 4–6, 8, 8, 10, 11
tspdata, 9
tspplot, 4–6, 8, 9, 10, 11
tpsig, 4–6, 8–10, 11