Package ‘TreeDimensionTest’

March 12, 2022

Type Package
Title Trajectory Presence and Heterogeneity in Multivariate Data
Version 0.0.2
Date 2022-03-11
Author Lovemore Tenha [aut] (https://orcid.org/0000-0001-9705-2023),
Joe Song [aut, cre] (https://orcid.org/0000-0002-6883-6547)
Maintainer Joe Song <joemsong@cs.nmsu.edu>
Description Testing for trajectory presence and heterogeneity on
multivariate data. Two statistical methods (Tenha & Song 2022)
<doi:10.1371/journal.pcbi.1009829> are implemented. The tree dimension
test quantifies the statistical evidence for trajectory presence. The
subset specificity measure summarizes pattern heterogeneity using the
minimum subtree cover. There is no user tunable parameters for either
method. Examples are included to illustrate how to use the methods on
single-cell data for studying gene and pathway expression dynamics and
pathway expression specificity.
License LGPL (>= 3)
Imports fitdistrplus, igraph, nFactors, Rcpp (>= 1.0.2), RColorBrewer,
Rdpack
LinkingTo Rcpp
RoxygenNote 7.1.2
Encoding UTF-8
Suggests knitr, rmarkdown, testthat
VignetteBuilder knitr
NeedsCompilation yes
RdMacros Rdpack
Depends mlpack
Repository CRAN
Date/Publication 2022-03-12 10:30:07 UTC
R topics documented:

compute.stats .................................................. 2
empirical.distributions ........................................ 3
plot.treedim ..................................................... 3
separability ..................................................... 4
test.trajectory .................................................. 5

Index 7

---

compute.stats Tree Dimension Test Related Statistics

Description

Computes tree dimension measure, tree dimension test effect, number leaves and tree diameter from MST of a given dataset.

Usage

compute.stats(x, MST = c("boruvka", "exact"), dim.reduction = c("pca", "none"))

Arguments

x matrix of input data. Rows as observations and columns as features
MST name of MST to be used in test. There are 2 options; "exact" MST and "boruvka" which is faster for large samples
dim.reduction string parameter with value "pca" to perform dimensionality reduction or "none" to not perform dimensionality reduction

Value

A list with the following components:

- tdt_measure The tree dimension value for the given input data
- tdt_effect Effect size for tree dimension
- leaves Number of leaf/degree1 vertices in the MST of the data
- diameter The tree diameter of MST, where each edge is of unit length
- original_dimension If "pca" is selected, the number of dimensions in the original dataset
- pca_components If "pca" is selected, the number of pca components selected after dimensionality reduction
- mst A vector of edges of the mst computed on x. Length of vector is always even.
empirical.distributions

Empirical Null Distribution of Tree Dimension Test

Description
Computes empirical null distribution of S statistic and parameters for lognormal approximation for input of size rows * columns using multivariate normal randomization.

Usage
`empirical.distributions(rows, cols, perm = 100, MST = c("boruvka", "exact"))`

Arguments
- `rows` number of rows for data representing null case. Rows represent sample size.
- `cols` number of columns for data representing null case. Columns represent variables.
- `perm` number of simulations to compute null distribution. Default is 100.
- `MST` name of MST to be used in computing distribution. There are two options; "exact" MST and "boruvka" which is faster for large samples.

Value
A list with the following components:

- `dist` A vector with null distribution of s statistic
- `meanlog` The meanlog parameter estimation for the lognormal distribution on empirical null distribution S.
- `sdlog` The sdlog parameter estimation for lognormal distribution on empirical null distribution of S.

plot.treedim

Visualizing Euclidean Minimum Spanning Trees

Description
Plots an Euclidean minimum spanning tree from given input data.
separability

Usage

```r
## S3 method for class 'treedim'
plot(
  x,
  ..., 
  node.col = "orange",
  node.size = 5,
  main = "MST plot",
  legend.cord = c(-1.2, 1.1)
)
```

Arguments

- `x`: An object of type "treedim"; returned from test.trajectory, compute.stats or separability
- `...`: ignored
- `node.col`: vector of colors for the observations in `x` (vertices)
- `node.size`: numerical value to represent size of nodes in the plot
- `main`: title for the plot
- `legend.cord`: vector of the xy coordinates for the legend c(x,y)

Value

result plots a minimum spanning tree for input data `x`

---

separability  
*Separability of Labeled Data Points*

Description

Computes homogeneity of labeled observations with multiple label types.

Usage

`separability(x, labels)`

Arguments

- `x`: input data matrix, with rows as observations and columns as features
- `labels`: a vector of labels for the observations. A label could be a type of the observation e.g cell type in single-cell data
**Value**

A list with the following components:

- `label_separability` A vector of separability scores for each of the label types. A high score denotes high separability
- `overall_separability` Overall average separability score for all the labels

**Description**

Computes the statistical significance for the presence of trajectory in multivariate data.

**Usage**

```r
test.trajectory(x, perm = 100, MST = c("boruvka", "exact"), dim.reduction = c("pca", "none"))
```

**Arguments**

- `x` matrix of input data. Rows as observations and columns as features.
- `perm` number of simulations to compute null distribution parameters by maximum likelihood estimation.
- `MST` the MST algorithm to be used in test. There are two options: "exact" MST and "boruvka" which is approximate but faster for large samples.
- `dim.reduction` string parameter with value "pca" to perform dimensionality reduction or "none" to not perform dimensionality reduction before the test.

**Details**

If the input data is already after dimension reduction, use `dim.reduction="none"`. The method is described in (Tenha and Song 2022).

**Value**

A list with the following components:

- `tdt_measure` The tree dimension value for the given input data
- `statistic` The S statistic calculated on the input data. S statistic is derived from tree dimension
- `tdt_effect` Effect size for tree dimension
- `leaves` Number of leaf/degree1 vertices in the MST of the data
- diameter The tree diameter of MST, where each edge is of unit length
- p.value The pvalue for the S statistic. Pvalue measures presence of trajectory in input x.
- original_dimension If "pca" is selected, the number of dimensions in the original dataset
- pca_components If "pca" is selected, the number of pca components selected after dimensionality reduction
- mst A vector of edges of the mst computed on x. Length of vector is always even.

References

Index

compute.stats, 2
empirical.distributions, 3
plot.treedim, 3
separability, 4
test.trajectory, 5