Package 'TreeDiagram'

July 21, 2025

Type Package

Title Tree Diagram

Version 0.1.1

Author Wen Tian (Wendy) Wang,Lloyd Elliott

Maintainer Wen Tian (Wendy) Wang <wtwang@sfu.ca>

Date 2021-03-25

Description Visualizing cuts for either axis-align or non axis-
align tree methods (e.g. decision tree, random tessellation process).

License BSD_2_clause + file LICENCE

Encoding UTF-8

Imports ggplot2,ape,cowplot,tree,stringr,stats

LazyData TRUE

Suggests testthat

RoxygenNote 7.1.1

NeedsCompilation no

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2021-03-27 05:10:03 UTC

Contents

---

cancer	*Breast cancer dataset*

Description

A dataset involves 9 quantitative predictors and a binary variable, indicating the present or absent of breast cancer.

Usage

```
data(cancer)
```

Format

A data frame with 116 rows and 10 variables:

Age Age (years)

BMI BMI (kg/m2)

Glucose Glucose (mg/dL)

Insulin Insulin (µU/mL)

HOMA Optimal cut-off value for homeostasis model assessment(HOMA) index of insulin

Leptin Leptin (ng/mL)

Adiponectin Adiponectin (µg/mL)

Resistin Resistin (ng/mL)

MCP.1 MCP-1(pg/dL)

Classification indicating the presence or absence of breast cancer. 1=Health controls;2=Patients

Source

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra#

References

Patrício, M., Pereira, J., Crisóstomo, J., Matafome, P., Gomes, M., Seiça, R., & Caramelo, F. (2018). Using Resistin, glucose, age and BMI to predict the presence of breast cancer. BMC Cancer, 18(1).

---

treeDiagram   *Draw a tree diagram for herarchical tree-based method*

Description

`treeDiagram` generates a tree diagram of any tree-based method and save automatically into user's current working directory. The classifying tree acts on the points provided in 'data' variable. The tree is specified by 'treedat' (for example, this could be a decision tree, or a character string generated from extended Newick's format, as described below). Both these variables must be provided. Output tree diagram will also provide tree information (e.g. node number, variable and value used at a split, minimum and maximum values of the split variable) for the first three nodes.

Usage

```
treeDiagram(data,treedat,cat_var,filename,pic_height=10,pic_width=10)
```

Arguments

`data`	An non-empty data frame.Variable names should not contain speical characters such as ">", "<", or "="
`treedat`	A character string (in extended Newick's format as described below) or the first object returned by `tree()` from **tree** package.
`cat_var`	A character string indicaing the predictor(classification) variable name in tree method.
`filename`	A character string. The name of output plot
`pic_height`	A real positive numeric value. This argument is optional, and the default argument is set as 10
`pic_width`	A real positive numeric value. This argument is optional, and the default argument is set as 10

Details

`treeDiagram` visualizes hierarchical clustering by projecting each tree split into an one dimensional density plot of the partitioned data. These projections are then arranged through rotation and translation to indicate the topology of the tree. As these projections are rotated and translated, a single plot with a fractal-like organisation is formed. The tree diagram allows users to access the quality of cut in terms of linearly separability, depth, balance of tree and distribution and classification of partitioned data in each cut.

Note

- Predictor variable must be a binary factor variable.
- The extended Newick's format is revised such that it follows the following rules:
    1. If a parent node has two children nodes, use the form of "variable = value"

2. If a parent node has only one child node, use the form of "var < value" if child node is on the left; otherwise use "var > value" for child node on the right
3. parent node is always at the end of brackets.
4. tip node is written in the form of "var = value"
5. tree is read from left to right

- Six required packages, "ggplot2","ape","cowplot","tree","stringr", will be automatically installed if user has not yet installed.

Author(s)

Wen Tian (Wendy) Wang (<wtwang@sfu.ca>), Lloyd T. Elliott

Examples

```
## Not run:
# read breast cancer data from UCI database website
cancer <- read.csv(
  url("https://archive.ics.uci.edu/ml/machine-learning-databases/00451/dataR2.csv"))
# make sure predictor variable is factored
cancer$Classification <- factor(cancer$Classification)
summary(cancer) #optional step giving an overview of the dataset

# set working directory (this step is optional)
# setwd("~/user location")

# e.g.1 : draw tree diagram according to the first object returned by 'tree()'
# create decision tree
library(tree)
t_cancer <- tree(Classification ~ ., data=cancer)

# plot tree diagram and save to your working directory
treeDiagram(cancer,t_cancer[[1]],"Classification","tree diagram for tree()")

# e.g.2 : draw tree diagram giving a newick format file of a tree
# newick format string of a decision tree for breast cancer
breast_cancer <- paste0(
  "(((BMI=25.745,BMI=29.722)Resistin=13.248)Age>44.5,",
  "(((((Age=70)Adiponectin<9.3482)BMI<32.275)Glucose<",
  "111)Leptin>7.93315)Age>48.5)Glucose=91.5;")


# plot tree diagram and save to your working directory
treeDiagram(cancer,breast_cancer,"Classification","tree diagram for newicks format file")

## End(Not run)
```

Index