

Package ‘rfviz’

February 23, 2022

Type Package

Title Interactive Visualization Tool for Random Forests

Version 1.0.1

Description An interactive data visualization and exploration toolkit that implements Breiman and Cutler's original random forest Java based visualization tools in R, for supervised and unsupervised classification and regression within the algorithm random forest.

Depends R (>= 3.5.1), randomForest, loon, tcltk

License GPL (>= 2)

URL https://www.stat.berkeley.edu/~breiman/RandomForests/cc_graphics.htm

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Imports stats, utils

NeedsCompilation no

Author Chris Kuchar [aut, cre]

Maintainer Chris Kuchar <chrisjkuchar@gmail.com>

Repository CRAN

Date/Publication 2022-02-23 13:00:02 UTC

R topics documented:

rfviz-package	2
glass	3
rf_prep	4
rf_viz	5

Index	8
--------------	----------

rfviz-package

Rfviz: An Interactive Visualization Package for Random Forests in R

Description

Rfviz is an interactive package and toolkit in R, using Tcl/Tk code on the backend, to help in viewing and interpreting the results Random Forests for both Supervised Classification and Regression in a user-friendly way.

Details

Currently, rfviz implements the following statistical graphs, with functions to view any combination of the plots:

The three plots are:

1. The classic multidimensionally scaled proximities are plotted as a 3-D XYZ scatterplot.
2. The raw input data is plotted in a parallel coordinate plot.
3. The local importance scores of each observation are plotted in a parallel coordinate plot.

rfviz is built using the package Loon on the backend, and implements the random forests algorithm.

For detailed instructions in the use of these plots in this package, visit <https://github.com/chriskuchar/rfviz/blob/master/Rfviz.md>

Note

For instructions on how to use randomForests, use `?randomForest`. For more information on loon, use `?loon`.

Author(s)

Chris Kuchar <chrisjkuchar@gmail.com>, based on original Java graphics by Leo Breiman and Adele Cutler.

References

Liaw A, Wiener M (2002). "Classification and Regression by randomForest." *_R News_*, *2*(3), 18-22. <https://CRAN.R-project.org/doc/Rnews/>

Waddell A, Oldford R. Wayne (2018). "loon: Interactive Statistical Data Visualization" <https://github.com/waddella/loon>

Breiman, L. (2001), Random Forests, *Machine Learning* 45(1), 5-32.

Breiman, L (2002), "Manual On Setting Up, Using, And Understanding Random Forests V3.1", https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf

Breiman, L., Cutler, A., Random Forests Graphics. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_graphics.htm

See Also

[randomForest](#), [rf_prep](#), [rf_viz](#), [l_plot3D](#), [l_serialaxes](#)

glass

Glass Identification Data Set

Description

A dataset containing 6 types of glass; defined in terms of their oxide content

Usage

glass

Format

A data frame with 214 rows and 10 variables:

RI Refractive Index

Na Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)

Mg Magnesium

Al Aluminum

Si Silicon

K Potassium

Ca Calcium

Ba Barium

Fe Iron

Type Class Attribute

Source

<https://archive.ics.uci.edu/ml/datasets/glass+identification>

rf_prep	<i>A function to create Random Forests output in preparation for visualization with rf_viz</i>
---------	--

Description

A function using Random Forests which outputs a list of the Random Forests output, the predictor variables data, and response variable data.

Usage

```
rf_prep(x, y = NULL, ...)
```

Arguments

x	A data frame or a matrix of predictors.
y	A response vector. If a factor, classification is assumed, otherwise regression is assumed. If omitted, randomForest will run in unsupervised mode.
...	Optional parameters to be passed down to the randomForest function. Use ?randomForest to see the optional parameters.

Value

The parallel coordinate plots of the input data, the local importance scores, and the 3-D XYZ classic multidimensional scaling proximities from the output of the random forest algorithm.

Note

For instructions on how to use randomForests, use ?randomForest. For more information on loon, use ?loon.

For detailed instructions in the use of these plots in this package, visit <https://github.com/chriskuchar/rfviz/blob/master/Rfviz.md>

Author(s)

Chris Kuchar <chrisjkuchar@gmail.com>, based on original Java graphics by Leo Breiman and Adele Cutler.

References

Liaw A, Wiener M (2002). "Classification and Regression by randomForest." *_R News_*, *2*(3), 18-22. <https://CRAN.R-project.org/doc/Rnews/>

Waddell A, Oldford R. Wayne (2018). "loon: Interactive Statistical Data Visualization" <https://github.com/waddella/loon>

Breiman, L. (2001), Random Forests, *Machine Learning* 45(1), 5-32.

Breiman, L (2002), "Manual On Setting Up, Using, And Understanding Random Forests V3.1", https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf

Breiman, L., Cutler, A., Random Forests Graphics. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_graphics.htm

See Also

[randomForest](#), [rf_viz](#), [l_plot3D](#), [l_serialaxes](#)

Examples

```
#Preparation for classification with Iris data set
rfprep <- rf_prep(x=iris[,1:4], y=iris$Species)

#Preparation for regression with mtcars data set
rfprep <- rf_prep(x=mtcars[,-1], y=mtcars$mpg)

#Preparation for the unsupervised case with Iris data set
rfprep <- rf_prep(x=iris[,1:4], y=NULL)
```

rf_viz

Random Forest Plots for interpreting Random Forests output

Description

The Input Data, Local Importance Scores, and Classic Multidimensional Scaling Plots

Usage

```
rf_viz(rfprep, input = TRUE, imp = TRUE, cmd = TRUE, hl_color = "orange")
```

Arguments

rfprep	A list of prepared Random Forests input data to be used in visualization, created using the function <code>rf_prep</code> .
input	Should the Input Data Parallel Coordinate Plot be included in the visualization?
imp	Should the Local Importance Scores Parallel Coordinate Plot be included in the visualization?
cmd	Should the Classic Multidimensional Scaling Proximites 3-D XYZ Scatter Plot be included in the visualization?
hl_color	The highlight color when you select points on the plot(s).

Value

Any combination of the parallel coordinate plots of the input data, the local importance scores, and the 3-D XYZ classic multidimensional scaling proximities from the output of the random forest algorithm.

Note

For instructions on how to use randomForests, use ?randomForest. For more information on loon, use ?loon.

For detailed instructions in the use of these plots in this package, visit <https://github.com/chriskuchar/rfviz/blob/master/Rfviz.md>

Author(s)

Chris Kuchar <chrisjkuchar@gmail.com>, based on original Java graphics by Leo Breiman and Adele Cutler.

References

Liaw A, Wiener M (2002). "Classification and Regression by randomForest." *_R News_*, *2*(3), 18-22. <https://CRAN.R-project.org/doc/Rnews/>

Waddell A, Oldford R. Wayne (2018). "loon: Interactive Statistical Data Visualization" <https://github.com/waddella/loon>

Breiman, L. (2001), Random Forests, *Machine Learning* 45(1), 5-32.

Breiman, L (2002), "Manual On Setting Up, Using, And Understanding Random Forests V3.1", https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf

Breiman, L., Cutler, A., Random Forests Graphics. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_graphics.htm

See Also

[randomForest](#), [rf_prep](#), [l_plot3D](#), [l_serialaxes](#)

Examples

```
#Classification with iris data set
rfprep <- rf_prep(x = iris[,1:4], y = iris$Species)

#View all three plots
Myrfplots <- rf_viz(rfprep, input = TRUE, imp = TRUE, cmd = TRUE, hl_color = 'orange')

#Select data on any of the plots then run:
iris[Myrfplots$input['selected'], ]
iris[Myrfplots$imp['selected'], ]
iris[Myrfplots$cmd['selected'], ]

#Rotate 3-D XYZ Scatterplot
#1. Click on 3-D XYZ Scatterplot
#2. Press 'r' on keyboard to enter rotation mode
#3. Click and drag mouse to rotate plot
#4. Press 'r' to leave rotation mode

#View only the Input Data and CMD Scaling Proximities Plots
Myrfplots <- rf_viz(rfprep, input = TRUE, imp = FALSE, cmd = TRUE, hl_color = 'orange')
```

```
#Regression with mtcars data set
rfprep2 <- rf_prep(x = mtcars[,-1], y = mtcars$mpg)

#View all three plots
Myrfplots <- rf_viz(rfprep2, input = TRUE, imp = TRUE, cmd = TRUE, hl_color = 'orange')

#Unsupervised clustering with iris data set
rfprep <- rf_prep(x = iris[,1:4], y = NULL)

#View the Input Data and CMD Scaling Proximities Plots for the unsupervised case.
#(Importance Scores Plot not valid here)
Myrfplots <- rf_viz(rfprep, input = TRUE, imp = FALSE, cmd = TRUE, hl_color = 'orange')
```

Index

*** datasets**

glass, 3

glass, 3

l_plot3D, 3, 5, 6

l_serialaxes, 3, 5, 6

randomForest, 3, 5, 6

rf_prep, 3, 4, 6

rf_viz, 3, 5, 5

rfviz (rfviz-package), 2

rfviz-package, 2