Package ‘rfPermute’

February 15, 2013

Type Package

Title Estimate permutation p-values for Random Forest importance metrics.

Version 1.5.1

Date 2012-11-30

Author Eric Archer

Maintainer Eric Archer <eric.archer@noaa.gov>

Depends R (>= 2.15.1), randomForest, parallel

Description Estimate significance of importance metrics for a Random Forest model by permuting the response variable. Produces null distribution of importance metrics for each predictor variable and p-value of observed.

License GPL (>= 2)

LazyLoad yes

Collate


Repository CRAN

Date/Publication 2012-11-30 19:47:24

NeedsCompilation no

R topics documented:

  plot.rfPermute ............................................................. 2
  rfPermute ................................................................. 2
  rp.importance ............................................................. 4

Index 5
plot.rfPermute \hspace{1cm} \textit{Plot Random Forest importance null distributions.}

\textbf{Description}

Plot the Random Forest null distributions importance metrics, observed values, and p-values for each predictor variable from the object produced by a call to \texttt{rfPermute}.

\textbf{Usage}

\begin{verbatim}
## S3 method for class 'rfPermute'
plot(x, imp.type, ...)
\end{verbatim}

\textbf{Arguments}

\begin{itemize}
  \item \texttt{x} \hspace{1cm} An object produced by a call to \texttt{rfPermute}.
  \item \texttt{imp.type} \hspace{1cm} Either a numeric or character vector giving the importance metric(s) to plot.
  \item \texttt{...} \hspace{1cm} Optional graphical arguments to be sent to \texttt{par}.
\end{itemize}

\textbf{Details}

The function will generate an individual plot for each variable and importance metric on the default graphics device.

\textbf{Author(s)}

Eric Archer <eric.archer@noaa.gov>

\texttt{rfPermute} \hspace{1cm} \textit{Estimate permutation p-values for importance metrics.}

\textbf{Description}

Estimate significance of importance metrics for a Random Forest model by permuting the response variable. Produces null distribution of importance metrics for each predictor variable and p-value of observed.

\textbf{Usage}

\begin{verbatim}
rfPermute(x, ...) \hspace{1cm}
\end{verbatim}

\begin{verbatim}
# Default S3 method:
rfPermute(x, y, ..., nrep = 100)
\end{verbatim}

\begin{verbatim}
# S3 method for class 'formula'
rfPermute(formula, data = NULL, ...,
          subset, na.action = na.fail, nrep = 100)
\end{verbatim}
rfPermute

Arguments

x, y, formula, data, subset, na.action, ...
   See randomForest for definitions.

nrep   Number of permutation replicates to run to construct null distribution and calculate p-values (default = 100).

Details

All other parameters are as defined in randomForest.formula. A Random Forest model is first created as normal to calculate the observed values of variable importance. rfPermute then permutes the response variable nrep times, with a new Random Forest model built for each permutation step. Permutations are done using the mclapply function in the package parallel. Set options("mc.cores") to the number of cores desired beforehand.

Value

An rfPermute object which contains all of the components of a randomForest object plus:

null.dist   A list containing three matrices. The first two matrices are null distributions for the importance metrics (%IncMSE and IncNodePurity for regression models, and MeanDecreaseAccuracy and MeanDecreaseGini for classification models) and have nrep rows and one column for each predictor variable. The third matrix (pval) has one row for each predictor variable and one column for each importance metric. The values are the permutation p-values for the respective importance metrics calculated as: (N(rep >= obs) + 1)/(nrep + 1).

Author(s)

Eric Archer <eric.archer@noaa.gov>

See Also

plot.rfPermute for plotting null distributions from the rfPermute object
randomForest, mclapply in package parallel

Examples

# A regression model using the ozone example
data(airquality)
ozone.rfP <- rfPermute(Ozone ~ ., data = airquality, ntree = 500, na.action = na.omit, nrep = 100)
print(ozone.rfP$importance) # The original importance metrics.
print(ozone.rfP$null.dist$pval) # The p-values for each variable.
plot(ozone.rfP, imp.type = 1) # Plot the null distributions and observed values.
rp.importance

Extract rfPermute importance scores and p-values.

Description

Extract a matrix of the observed importance scores and p-values from the object produced by a call to rfPermute.

Usage

rp.importance(x, sort.by = "MeanDecreaseAccuracy", decreasing = T)

Arguments

x  An object produced by a call to rfPermute.
sort.by  character vector giving the importance metric(s) or p-values to sort by.
decreasing  logical. Should the sort order be increasing or decreasing?

Details

p-values can be given to the sort.by argument by adding ".pval" to the column name of the desired column from the importance element of the rfPermute object

Author(s)

Eric Archer <eric.archer@noaa.gov>
Index

*Topic **classif**
  rfPermute, 2
*Topic **regression**
  rfPermute, 2
*Topic **tree**
  rfPermute, 2

mclapply, 3
parallel, 3
plot.rfPermute, 2, 3

randomForest, 3
rfPermute, 2
rp.importance, 4