Package ‘poolVIM’

July 30, 2018

Type Package

Title Gene-Based Association Tests using the Actual Impurity Reduction (AIR) Variable Importance

Version 1.0.0

Date 2018-07-25

Author Stefano Nembrini <stefanonembrini@gmail.com>

Maintainer Stefano Nembrini <stefanonembrini@gmail.com>

Description Gene-based association tests using the actual impurity reduction (AIR) variable importance. The function aggregates AIR importance measures from a group of SNPs or probes and outputs a p-value for each gene. The procedures builds upon the method described in <doi:10.1093/Bioinformatics/Bty373> and will be published soon.

License GPL (>= 2)

LazyData TRUE

Imports stats, ranger

Depends R(>= 3.3.1), EmpiricalBrownsMethod(>= 1.6.0), Hmisc(>= 4.1)

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2018-07-30 17:30:08 UTC

R topics documented:

fisher .................................................. 2
gaussianize .............................................. 2
m_effective ............................................. 2
poolVIM ................................................ 3
tippett .................................................. 3

Index 5
fischer

**Description**
fischer

**Usage**
fischer(p, adjust, R)

**Arguments**
p vector of pvalues
adjust if correlation has to be taken into account
R correlation matrix

---

gaussianize

gaussianize null variable importances

**Description**
gaussianize null variable importances

**Usage**
gaussianize(x, a)

**Arguments**
x distr
a value to interpolate/extrapolate

---

m_effective

**Description**
meff.

**Usage**
m_effective(R)

**Arguments**
R R
poolVIM

after the Actual Impurity Reduction Importance is computed with a Random Forest, pvalues from different probes or SNPs belonging to the same gene can be aggregated in order to obtain a single pvalue for that gene. Correlation between probes can also be taken into account.

Description

after the Actual Impurity Reduction Importance is computed with a Random Forest, pvalues from different probes or SNPs belonging to the same gene can be aggregated in order to obtain a single pvalue for that gene. Correlation between probes can also be taken into account.

Usage

poolVIM(rf, genenames, x, method = "Tippett", adjust)

Arguments

rf a ranger object with "importance="impurity_corrected"

genenames a vector of the name of the gene to which each probe or SNP belongs, it has to be of size dim(x)[1]
x design matrix used by the random forest
method one of Tippett, Fisher, Kost, EBM
adjust "no" / "yes" depending if correlation has to be taken into account

Examples

n <- 250
x=replicate(50, runif(n))
dat <- data.frame(y = factor(rbinom(n, 1, .5)), x)
library(ranger)
rf <- ranger(y ~ ., dat, importance = "impurity_corrected", num.trees=100)
genenames=colnames(x)=rep(c("G1","G2"),50/2)
poolVIM(rf,genenames,x,method="Fisher",adjust="no")

tippett
tippett.

Description
tippett.

Usage
tippett(p, adjust, R)
Arguments

\begin{itemize}
  \item \texttt{p} \quad \text{vector of pvalues}
  \item \texttt{adjust} \quad \text{if correlation has to be taken into account}
  \item \texttt{R} \quad \text{correlation matrix}
\end{itemize}
Index

fisher, 2
gaussianize, 2
m_effective, 2
poolVIM, 3
tippett, 3