

Package ‘hit’

October 13, 2022

Type Package

Title Hierarchical Inference Testing

Description Hierarchical inference testing (HIT) for (generalized) linear models with correlated covariates applicable to high-dimensional settings.

Version 0.4.0

URL <https://github.com/QTCAT/hit>

BugReports <https://github.com/QTCAT/hit/issues>

Depends R (>= 3.0.0)

Imports stats, parallel, Rcpp (>= 0.11.0), glmnet, speedglm

Suggests testthat

LinkingTo Rcpp

License GPL (>= 2)

RoxygenNote 5.0.1

NeedsCompilation yes

Author Jonas Klasen [aut, cre]

Maintainer Jonas Klasen <qtcat@gmx.de>

Repository CRAN

Date/Publication 2016-11-17 18:33:54

R topics documented:

hit-package	2
as.hierarchy	2
fast.anova	3
hit	4
names.hierarchy	6
reorder.hierarchy	6
summary.hit	7

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

hit-package	<i>Hierarchical Inference Testing</i>
-------------	---------------------------------------

Description

Hierarchical inference testing (HIT) for (generalized) linear models with correlated covariates applicable to high-dimensional settings.

Author(s)

Jonas Klasen

as.hierarchy	<i>Hierarchy Structure</i>
--------------	----------------------------

Description

Stores variable indexes of clustering hierarchies in a fast accessible manner.

Usage

```
as.hierarchy(x, max.height, height, names, ...)
```

Arguments

x	A S3 object e.g. from hclust or dendrogram.
max.height	Is the maximal height below the height of the global node which is considered.
height	A vector of heights at which nodes are grouped.
names	Variable names in the order in which the indexes should be assigned to the variables.
...	Further arguments.

Details

For the HIT algorithm it is important to have the hierarchical clustering structure in a fast accessible format. This is provided by the hierarchy object generated with this function.

Examples

```
##
set.seed(123)
n <- 80
p <- 90
# x with correlated columns
corMat <- toeplitz((p:1/p)^3)
corMatQ <- chol(corMat)
x <- matrix(rnorm(n * p), nrow = n) %*% corMatQ
colnames(x) <- paste0("x", 1:p)
# hierarchy
hc <- hclust(dist(t(x)))
hier <- as.hierarchy(hc)
```

fast.anova

Fast ANOVA

Description

A fast sequential analysis of variance (ANOVA). Mainly developed for internal use.

Usage

```
fast.anova(x, y, assign = NULL, family = gaussian(), test = c("LRT", "F"))
```

Arguments

<code>x</code>	Design matrix of dimension $n * p$.
<code>y</code>	Response vector of observations of length n .
<code>assign</code>	Integer vector assigning columns to terms can be also given as <code>x</code> attribute in which case the argument is ignored. If an intercept exist it is expected to be the first column in <code>x</code> and it has to be specified by a '0' in tis vector. For details about assign see model.matrix .
<code>family</code>	A description of the error distribution and link function to be used in the model. For GLMs this can be a character string naming a family function or the result of a call to a family function. (See family for details of family functions.)
<code>test</code>	The name of the test either "LRT" (default) for likelihood ratio test or "F" for F test.

See Also

[lm](#), [anova](#), and [aov](#).

Examples

```

y <- rnorm(n = 100)
x <- matrix(data = rnorm(1000), nrow = 100)
a <- 1:10
fast.anova(x = x, y = y, assign = a)

```

hit

*Hierarchical Inference Testing***Description**

Hierarchical inference testing for linear models with high-dimensional and/or correlated covariates by repeated sample splitting.

Usage

```

hit(x, y, hierarchy, family = "gaussian", B = 50, p.samp1 = 0.5,
    nfolds = 10, overall.lambda = FALSE, lambda.opt = "lambda.1se",
    alpha = 1, gamma = seq(0.05, 0.99, length.out = 100), max.p.esti = 1,
    mc.cores = 1L, trace = FALSE, ...)

```

Arguments

x	Design matrix of dimension $n * p$, without intercept. Variables not part of the dendrogram are added to the HO-model, see Details below.
y	Quantitative response variable dimension n .
hierarchy	Object of class <code>as.hierarchy</code> . Must include all variables of x which should be tested.
family	Family of response variable distribution. Either y is "gaussian" or "poisson" in which case y must be a vector or it is "binomial" distributed and is either a vector of zeros and ones, factor with two levels, or a two-column matrix of counts or proportions. The second column is treated as the target class. For a factor, the last level in alphabetical order is the target class. For "binomial" if y is presented as a vector, it will be coerced into a factor.
B	Number of sample-splits.
p.samp1	Fraction of data used for the LASSO. The hierarchical ANOVA testing uses $1 - p.samp1$.
nfolds	Number of folds (default is 10). See <code>cv.glmnet</code> for more details.
overall.lambda	Logical, if true, lambda is estimated once, if false, lambda is estimated for each sample split.
lambda.opt	Criterion for optimum selection of cross-validated lasso. Either "lambda.1se" (default) or "lambda.min". See <code>cv.glmnet</code> for more details.
alpha	A single value in the range of 0 to 1 for the elastic net mixing parameter.

gamma	Vector of gamma-values.
max.p.esti	Maximum alpha level. All p-values above this value are set to one. Small max.p.esti values reduce computing time.
mc.cores	Number of cores for parallelising. Theoretical maximum is 'B'. For details see mclapply .
trace	If TRUE it prints current status of the program.
...	Additional arguments for cv.glmnet .

Details

The H0-model contains variables, with are not tested, like experimental-design variables. These variables are not penalised in the LASSO model selection and are always include in the reduced ANOVA model.

References

Mandozzi, J. and Buehlmann, P. (2013). *Hierarchical testing in the high-dimensional setting with correlated variables*. To appear in the Journal of the American Statistical Association. Preprint arXiv:1312.5556

Examples

```
# Simulation:
set.seed(123)
n <- 80
p <- 82
## x with correlated columns
corMat <- toeplitz((p:1/p)^2)
corMatQ <- chol(corMat)
x <- matrix(rnorm(n * p), nrow = n) %%% corMatQ
colnames(x) <- paste0("x", 1:p)
## y
mu <- x[, c(5, 10, 72)] %%% c(2, -2, 2)
y <- rnorm(n, mu)
## clustering of the clumns of x
hc <- hclust(dist(t(x)))

# HIT with AF
out <- hit(x, y, hc)
summary(out)
```

names.hierarchy	<i>Names of Hierarchy</i>
-----------------	---------------------------

Description

Names of variables of an hierarchy.

Usage

```
## S3 method for class 'hierarchy'  
names(x)
```

Arguments

x A [as.hierarchy](#).

reorder.hierarchy	<i>Reorder Hierarchy</i>
-------------------	--------------------------

Description

Reorder indexes according to a vector of names.

Usage

```
## S3 method for class 'hierarchy'  
reorder(x, names, ...)
```

Arguments

x A [as.hierarchy](#).

names Variable names in the order in which the indexes should be assigned to the variables.

... Further arguments passed to or from other methods (not used).

`summary.hit`*Summary of HIT*

Description

Significant clusters at alpha threshold.

Usage

```
## S3 method for class 'hit'  
summary(object, alpha = 0.05, max.height, ...)
```

Arguments

<code>object</code>	A <code>hit</code> object.
<code>alpha</code>	A alpha significance threshold.
<code>max.height</code>	max. Height to consider.
<code>...</code>	Further arguments passed to or from other methods (not used).

Index

anova, 3
aov, 3
as.hierarchy, 2, 4, 6

cv.glmnet, 4, 5

family, 3
fast.anova, 3

HIT (hit-package), 2
hit, 4, 7
hit-package, 2

lm, 3

mclapply, 5
model.matrix, 3

names.hierarchy, 6

reorder.hierarchy, 6

summary.hit, 7