Package 'SHRED'

March 11, 2026

Type Package

Title Setwise Hierarchical Rate of Erroneous Discovery

Version 1.0.0

Maintainer Toby Kenney <tkenney@mathstat.dal.ca>

Description Setwise Hierarchical Rate of Erroneous Discovery (SHRED) methods for setwise variable selection with false discovery rate (FDR) control. Setwise variable selection means that sets of variables may be selected when the true variable cannot be identified. This allows us to maintain FDR control but increase power. Details of the SHRED methods are in Organ, Kenney & Gu (2026) <doi:10.48550/arXiv.2603.02160>.

License GPL-3

Encoding UTF-8

Imports graphics, stats, ClustOfVar

NeedsCompilation no

Author Sarah Organ [aut],
Toby Kenney [cre],
Hong Gu [aut]

Repository CRAN

Date/Publication 2026-03-11 20:00:03 UTC

Contents

---

convert.to.matrix	*Convert clustering to matrix form*

Description

Converts clustering from ClustOfVar package to membership matrix format.

Usage

```
convert.to.matrix(clust)
```

Arguments

clust   A clustering from the ClustOfVar package.

Details

For a clustering of p variables, this converts the clustering into a list of parents and

Value

A list containing the following components: "parent" The parent node of each node in the clustering "matrix" a 2p-1 by p matrix of 0 and 1 whose [i,j] entry is 1 if node j is in cluster i.

Author(s)

Sarah Organ, Toby Kenney, Hong Gu

Examples

```
set.seed(1)
X<-matrix(rnorm(200),20,10)
cl<-ClustOfVar::hclustvar(X)
convert.to.matrix(cl)
```

CumMinWeights	*Cumulative sum-of-minimal-weights sizing function*

Description

Calculates the sum-of-minimal-weights sizing function for the initial elements of a hierarchical tree.

Usage

```
CumMinWeights(weights,parents)
```

Arguments

weights	A vector of weights.
parents	A vector giving the index of the parent node for each node in the hierarchical tree or forest (NA for root nodes).

Details

For a subset of the hierarchical tree, minimal elements are elements that have no proper descendents in the subset. The sum-of-minimal-weights sizing function takes a subset A of the tree, and assigns the sum of weights corresponding to all minimal elements of A. Given a vector of weights and the hierarchical tree structure, this function calculates the sum-of-minimal-weights for every initial subset 1-to-k from this vector.

Value

A numerical vector of the sizing function.

Author(s)

Sarah Organ, Toby Kenney, Hong Gu

Examples

```
set.seed(1)
pv<-rbeta(31,1,5)
parents<-c(NA,rep(seq_len(15),each=2)) # perfect binary tree
weights<-2^-c(5,rep(4,2),rep(3,4),rep(2,8),rep(1,16)) #weighted by
                                                      #no. of leaves

permutation<-order(runif(31)) # random permutation
permutation.inv<-rep(0,31)
permutation.inv[permutation]<-seq_len(31) #inverse permutation

### change weights and parents to the new permutation
weights.ordered<-weights[permutation]
parents.ordered<-permutation.inv[parents[permutation]]

### Compute sum minimal weights
CumMinWeights(weights.ordered,parents.ordered)
```

---

cutoffs	*Compute cut-off slopes for SHRED*

Description

Computes the cut-off slope that controls FDR at the specified level under the assumptions

Usage

```
PRDS.cutoff(weights,parents,level)
SHRED.cutoff(weights,parents,level)
SHREDDER.cutoff(weights,parents,level)
```

Arguments

`weights`	A vector of weights for each hypothesis.
`parents`	A vector giving the index of the parent hypothesis for each hypothesis in the hierarchical tree or forest (NA for root hypotheses).
`level`	The desired FDR control level.

Details

The GLSUP with sum-of-minimal-weights sizing function is proven to control gFDR for certain choices of cut-off slope, under various assumptions. These functions calculate the appropriate slope to control gFDR under the corresponding assumptions - PRDS.cutoff computes the cut-off slope that controls gFDR under the PRDS assumption. SHRED.cutoff computes the cut-off slope that guarantees gFDR control, regardless of dependency between p-values. SHREDDER.cutoff computes the cutoff that controls gFDR for the SHREDDER method under the PPRDS assumption. This also often works in practice for the SHRED method.

Value

The cut-off slope to be used for the HGLSUP function.

Author(s)

Sarah Organ, Toby Kenney, Hong Gu

References

Setwise Hierarchical Variable Selection and the Generalized Linear Step-Up Procedure for False Discovery Rate Control

Sarah Organ, Toby Kenney, Hong Gu

http://arxiv.org/abs/2603.02160

Examples

```
parents<-c(NA,rep(seq_len(15),each=2)) # perfect binary tree
weights<-2^-c(5,rep(4,2),rep(3,4),rep(2,8),rep(1,16))
PRDS.cutoff(weights,parents,0.05)
SHRED.cutoff(weights,parents,0.05)
SHREDDER.cutoff(weights,parents,0.05)
```

---

get.p.vals   *Calculate p-values for set-wise variable selection*

Description

Performs hypothesis tests for the null hypothesis that a set of predictors contains no true predictor.

Usage

```
get.p.vals(x,y,clust,test)
```

Arguments

x	The matrix of predictors.
y	A vector containing the response variable.
clust	A matrix whose rows are indicator vectors for subsets of predictors of X.
test	either one of the character strings "gaussian", "binomial" or "poisson", or else a list with two components: model — a function for fitting the model based on a formula typically "lm" for gaussian regression, something based on "glm" for other glm models. p.val — a test function based on comparison of the two models typically based on the anova function

Details

The rows of the matrix clust define a collection of subsets of the predictor variables. This function, for each of these sets, computes the p-values for the hypotheses that the set contains no true predictors.

Value

A vector of p-values.

Author(s)

Sarah Organ, Toby Kenney, Hong Gu

Examples

```
set.seed(1)
X<-matrix(rnorm(200),20,10)%*%(diag(rep(1,10))-c(0.4,0.4,rep(0,8))%*%t(c(0.4,0.4,rep(0,8))))
Y<-rnorm(20)+X%*%c(0,2,0,2,2,0,0,0,0,0)
clusters<-ClustOfVar::hclustvar(X)
clust<-convert.to.matrix(clusters)
get.p.vals(X,Y,clust$matrix,"gaussian")
```

HGLSUP   *Hierarchical Generalised Linear Step-Up Procedure*

Description

Performs the Generalised Linear Step-up Procedure (GLSUP) on p-values arranged in a hierarchical tree.

Usage

```
HGLSUP(pvals,weights,parents,threshold)
```

Arguments

pvals	A vector of p-values.
weights	A vector of weights for each hypothesis.
parents	A vector giving the index of the parent hypothesis for each hypothesis in the hierarchical tree or forest (NA for root hypotheses).
threshold	The cut-off slope.

Details

The GLSUP with sizing function s(A) on subsets of the sets of hypotheses, and cut-off slope a, rejects all hypotheses with p-values less than a cut-off c, where c is the largest cut-off such that s($\{i \mid p_i < c\}$)>=ac. This function performs the GLSUP for a set of hypotheses arranged in a hierarchical tree or forest, with sizing function s(A) given by the sum of weights of minimal elements of the set A of hypotheses.

Value

A list containing the following components: "pv" The p-values in increasing order "ord" The order from the original vector of p-values "parent" The parent node of each p-value in the sorted list "weight" The weight assigned to each p-value in the sorted list "cum.weight" The cumulative sum of weights of minimal elements "selected" A vector of the indices of rejected hypotheses in the original order "pv.cut.off" The p-value cut-off below which hypotheses are rejected

Author(s)

Sarah Organ, Toby Kenney, Hong Gu

References

Setwise Hierarchical Variable Selection and the Generalized Linear Step-Up Procedure for False Discovery Rate Control

Sarah Organ, Toby Kenney, Hong Gu

http://arxiv.org/abs/2603.02160

Examples

```
set.seed(1)
pv<-rbeta(31,1,5)
parents<-c(NA,rep(seq_len(15),each=2)) # perfect binary tree
weights<-2^-c(5,rep(4,2),rep(3,4),rep(2,8),rep(1,16))
ans<-HGLSUP(pv,weights,parents,sum(weights)*20) # threshold under PRDS
ans$selected
```

SHRED	*SHRED setwise variable selection with FDR control*

Description

Performs variable selection, allowing the selection of sets of surrogates using the SHRED method for FDR control. This allows selection of sets of variables from a hierarchical clustering of the predictors.

Usage

```
SHRED(x,y,test,method,level,weights=NULL)

## S3 method for class 'SHRED'
print(x,...)
## S3 method for class 'SHRED'
plot(x,...)
```

Arguments

x	A matrix of predictor variables
y	The response variable
test	Either one of the following character strings: "gaussian", "binomial" or "poisson", or else a list with two named components - the first component "model" is a function for modelling the data which takes as input a formula. For Gaussian regression, the "lm" function can be used for this. For GLM fitting, you will need to make a function that sets the parameters. The second component "test" is a function taking as input two nested models fitted using the "model" function specified in the first component, and producing a corresponding p.value for the null hypothesis that the sub-model fits the data as well as the complete model.
	The "make.test" function creates these named lists for the standard strings, and can be used as a template for creating more general lists.
method	The method used to calculate the cut-off in the SHRED method. Choices are: "PRDS", which uses a cut-off that is guaranteed to control FDR under the PRDS assumption for the p-values of the tests; "Arbitrary", which uses a stricter cut-off that is guaranteed to control FDR under arbitrary dependence between p-values; "Heuristic", which uses the cut-off for weighted BH, which is not guaranteed

to control FDR under hierarchically clustered hypotheses, but usually performs well in practice.

Alternatively, method can be a fixed value, which is used as the cut-off value.

level   The desired level of FDR control. If "method" is numeric, this is ignored, and the desired level of control should be incorporated into the value provided. If "method" is one of the options, then this the level at which FDR control is desired. For method="Arbitrary" or method="PRDS", the true FDR will usually be lower than this; for method="Heuristic", the true FDR could be higher than this, but in practice, will often be close to this value.

weights   If weights=NULL (the default), the weight of each set is the inverse of the number of elements in the set, as suggested by Organ et al.. Otherwise, weights should be a function that takes as input a vector of sizes of sets, and returns the corresponding weights.

...   Additional graphics or printing parameters. The graphics parameters are passed to other functions. For print.SHRED, any additional parameters are ignored.

Details

The SHRED method hierarchically clusters the predictors, then tests all clades of variables in the hierarchical clustering for significance. It then uses a BH or BY style method to control weighted FDR in the set of selected sets of variables, where the weight for each set is the inverse of the number of variables contained in it. For each set, the conclusion of a rejected hypothesis is that at least one of the variables in the set is a true variable. Thus, a selected set is a false positive if none of the variables contained in it is a true variable, and a true positive if any variable is a true variable. Because of the hierarchical structure of SHRED, the sets selected are always disjoint.

Value

An object of class "SHRED" which contains the following components "X" The matrix of predictor variables "Y" The vector of response variables "method" The method used to calculate the cut-off slope "test" The test used to obtain p-values "level" The desired FDR control level "cut.off.slope" The slope of the threshold "threshold" The threshold for final selection (NA if no sets selected) "cluster" The results of hierarchical clustering "cluster.matrix" A matrix giving the selected clades as indicator vectors "pv" The p-values for each test "ord" The order of the p-values "weight" The weights for each hypothesis "cum.weight" The sum of the weights for each selected number of hypotheses "selected" Logical vector indicating which sets were selected "selected.sets" Logical Matrix, whose rows correspond to selected sets of variables "selected.weight" The total weight of all rejected hypotheses

Author(s)

Sarah Organ, Toby Kenney, Hong Gu

References

Setwise Hierarchical Variable Selection and the Generalized Linear Step-Up Procedure for False Discovery Rate Control

Sarah Organ, Toby Kenney, Hong Gu

http://arxiv.org/abs/2603.02160

Examples

```
set.seed(1)
X<-matrix(rnorm(200),20,10)%*%(diag(rep(1,10))-c(0.4,0.4,rep(0,8))%*%t(c(0.4,0.4,rep(0,8))))
Y<-rnorm(20)+X%*%c(0,3,0,3,3,0,0,0,0,0)

selection<-SHRED(X,Y,"gaussian","PRDS",0.05)
### This fits a linear model of Y on subsets of X and uses a cut-off
### that controls FDR at the 0.05 level under the PRDS assumption for
### the p-values of all null hypotheses.

print(selection)
plot(selection)
```

SHREDDER	*SHREDDER*

Description

Perform variable selection with the SHREDDER method for FDR control

Usage

```
SHREDDER(x,y,test,level,weights)
```

Arguments

x	A matrix of predictor variables
y	The response variable
test	Either one of the following character strings: "gaussian", "binomial" or "poisson", or else a list with two named components - the first component "model" is a function for modelling the data which takes as input a formula. For Gaussian regression, the "lm" function can be used for this. For GLM fitting, you will need to make a function that sets the parameters. The second component "test" is a function taking as input two nested models fitted using the "model" function specified in the first component, and producing a corresponding p.value for the null hypothesis that the sub-model fits the data as well as the complete model.
	The "make.test" function creates these named lists for the standard strings, and can be used as a template for creating more general lists.
level	The desired level of FDR control. If "method" is numeric, this is ignored, and the desired level of control should be incorporated into the value provided. If "method" is one of the options, then this the level at which FDR control is desired. For method="Arbitrary" or method="PRDS", the true FDR will usually be lower than this; for method="Heuristic", the true FDR could be higher than this, but in practice, will often be close to this value.

weights   If weights=NULL (the default), the weight of each set is the inverse of the num-
                   ber of elements in the set, as suggested by Organ et al.. Otherwise, weights
                   should be a function that takes as input a vector of sizes of sets, and returns the
                   corresponding weights.

Details

The SHREDDER method hierarchically clusters the predictors, then tests all clades of variables
in the hierarchical clustering for significance. It then uses a BH or BY style method to control
weighted FDR in the set of selected sets of variables, where the weight for each set is the inverse of
the number of variables contained in it. For each set, the conclusion of a rejected hypothesis is that
at least one of the variables in the set is a true variable. Thus, a selected set is a false positive if none
of the variables contained in it is a true variable, and a true positive if any variable is a true variable.
SHREDDER only selects sets of variables when the p-values for all larger sets are below the cut-off.
Because of the hierarchical structure of SHREDDER, the sets selected are always disjoint.

Value

An object of class "SHREDDER" which contains the following components "X" The matrix of pre-
dictor variables "Y" The vector of response variables "method" "SHREDDER" "test" The test used
to obtain p-values "level" The desired FDR control level "cut.off.slope" The slope of the threshold
"threshold" The threshold for final selection (NA if no sets selected) "cluster" The results of hierar-
chical clustering "cluster.matrix" A matrix giving the selected clades as indicator vectors "pv" The
p-values for each test "ord" The order of the p-values "weight" The weights for each hypothesis
"cum.weight" The sum of the weights for each selected number of hypotheses "selected" Logical
vector indicating which sets were selected "selected.sets" Logical Matrix, whose rows correspond
to selected sets of variables "selected.weight" The total weight of all rejected hypotheses

Author(s)

Sarah Organ, Toby Kenney, Hong Gu

References

Setwise Hierarchical Variable Selection and the Generalized Linear Step-Up Procedure for False
Discovery Rate Control

Sarah Organ, Toby Kenney, Hong Gu

http://arxiv.org/abs/2603.02160

Examples

```
set.seed(1)
X<-matrix(rnorm(200),20,10)%*%(diag(rep(1,10))-c(0.4,0.4,rep(0,8))%*%t(c(0.4,0.4,rep(0,8))))
Y<-rnorm(20)+X%*%c(0,2,0,2,2,0,0,0,0,0)

selection<-SHREDDER(X,Y,"gaussian",0.05)
### This fits a linear model of Y on subsets of X and uses a cut-off
### that controls FDR at the 0.05 level under the PRDS assumption for
### the p-values of all null hypotheses.
```

```
selection
plot(selection)
```

Index