Package 'HMVD'

May 13, 2016

Type Package

Title Group Association Test using a Hidden Markov Model

Version 1.0

Date 2016-05-12

Author Yichen Cheng

Maintainer Yichen Cheng <ycheng@fredhutch.org>

Description Perform association test between a group of variable and the outcome.

License GPL-3

Depends MASS

NeedsCompilation no

Repository CRAN

Date/Publication 2016-05-13 01:18:58

R topics documented:

HMVD-package   *Group Association Test using a Hidden Markov Model*

Description

HMVD performs an association test between a group of variables and the outcome. Posterior probabilities are provided for each variable indicating how likely each variable is associated with the outcome.

Details

Package:	HMVD
Type:	Package
Version:	1.0
Date:	2016-05-12
License:	GPL-3

~~ An overview of how to use the package, including the most important functions ~~

Author(s)

Maintainer: Yichen Cheng<ycheng@fredhutch.org>

References

Cheng, Y., Dai, J. and Kooperberg, C. (2015). Group association test using hidden Markov model. Biostatistics, in pres.

Examples

```
############################################################################
#### compute the p-value and do parameter estimation for continuous outcome
n = 4000; m = 20
X = matrix(rnorm(n*m),n)
Y = rowMeans(X[,1:4])*.2 + rnorm(n)
HMVD(Y,X)$p.value #### approximate p-value

HMVD(Y,X,nperm.max = 20)$p.value.perm #### p-value based on permutations
#### in practice we would use more permutations

out = HMVD(Y,X,method='estimation')
round(out$prob,2) ###posterior probability
out$theta ### common effect size


#### compute the p-value and do parameter estimation for binary outcome
n = 4000; m = 20
X = matrix(rnorm(n*m),n)
p = rowMeans(X[,1:4])*.4
Y = rbinom(n,1,p = exp(p)/(1+exp(p)))
HMVD(Y,X,model.type='D')$p.value #### approximate p-value

HMVD(Y,X,nperm.max = 20)$p.value.perm #### p-value based on permutations
#### in practice we would use more permutations

out = HMVD(Y,X,model.type='D',method='estimation')
round(out$prob,2) ###posterior probability
out$theta ### common effect size
```

HMVD   *Group Association Test using a Hidden Markov Model*

Description

Perform group association test between outcome Y and a group of features X. The function HMVD can be used in two ways. If the 'method' parameter is set to 'test', then HMVD performs a hypothesis test of association of X and Y. If the 'method' parameter is set to 'estimation', then HMVD performs a model fitting using a generalized hidden Markov model. If the 'method' is 'estimation', a list of posterior probabilities will also be provided to indicate how likely each feature is to be associated with the outcome.

Usage

```
HMVD(Y,X,XX=NULL,weight=FALSE,C=1,nperm.max=0,model.type="C",method="test",adj=TRUE)
```

Arguments

Y	a numeric vector for the outcome variable. No missing data is allowed.
X	a numeric matrix. Each row represents one subject and each column represents one feature. No missing data is allowed.
XX	optional. A numeric matrix of covariates with each row for one subject and each column for one covariate. No missing data is allowed.
weight	a logical value indicate whether each feature will be multiplied by a weight before entering the model. When 'weight' is set to be true, every feature will be divided by its standard deviation.
C	a positive number (default=1). It specifies the level of penalty added to the likelihood function to avoid the composite null issue.
nperm.max	a number providing the maximum number of permutations to run. When nperm.max is non positive, no permutation test is run. When nperm.max is a positive number, a permutation test is run with maximum number of permutations set to be nperm.max.
model.type	can be either 'C' or 'D'. 'C' corresponds to continuous outcomes and 'D' corresponds to binary outcomes.
method	can be either 'test' or 'estimation'. When method is set to 'test', a hypothesis test for whether there is association between X and Y is performed. When method is set to 'estimation', HMVD performs a model fitting and estimate the effect size of associated features as well as the transition matrix. Also, a list of posterior probabilities will be provided to indicate how likely each feature is to be associated with the outcome.
adj	a logical value. If true, an adjustment is applied to avoid non-convergence in glm.fit. This adjustment can yield a small bias, so it is only recommended when there is non-convergence. See 'details'.

Details

When no permutation test is run, the p-value for association test is calculated based using a Chi-square approximation/ When nperm.max is positive, the p-value is calculated using a permutation test, where the number of permutations is defined according to a function based on nperm.max and approximate p-value. The actual number of permutations that is run is given in the output. See nperm in the value section for detail.

For binary outcomes, it is possible that the algorithm will not converge. A warning message will be printed. It is suggested to use adj=TRUE option and re-run the HMVD function. If adj=TRUE, two extra data points will be added to the original data. For each feature G, min(G) with Y=.5 and max(G) with Y=.5 are added to avoid the non-convergence in glm.fit.

Value

A	estimated transition matrix of the Markov chain. This is only provided when 'method' is set to 'estimation'.
keep1	a series of indices for the variables that are kept after removing constant features (features with 0 variance).
keep2	a series of indices for the variables that are kept after removing highly correlated features as well as constant features.
nperm	number of permutations run to get p-value based on the permutation test. It cannot be greater than nperm.max. This is only provided when 'method' is set to 'test' and 'nperm.max' is greater than 0.
p.value	p-value for the association test calculated using a Chi-square approximation. This is only provided when 'method' is set to 'test'.
p.value.perm	p-value for association test calculated using a permutation test with the number of permutations equal to nperm. This is only provided when 'method' is set to 'test' and 'nperm.max' is greater than 0.
prob	a list of posterior probabilities for each feature. Higher value indicates higher probability of being associated with the outcome. This is only provided when 'method' is set to 'estimation'.
theta	estimated common effect size for associated features. This is only provided when 'method' is set to 'estimation'.

Author(s)

Yichen Cheng

References

Cheng, Y., Dai, J.Y. and Kooperberg, C. (2015) Group association test using a hidden Markov model. Biostatistics, in press.

Examples

```
###########################################################################
#### compute the p-value and do parameter estimation for continuous outcome
n = 4000; m = 20
```

```
X = matrix(rnorm(n*m),n)
Y = rowMeans(X[,1:4])*.2 + rnorm(n)
HMVD(Y,X)$p.value #### approximate p-value

HMVD(Y,X,nperm.max = 20)$p.value.perm #### p-value based on permutations
### In practice we would increase the number of permutations

out = HMVD(Y,X,method='estimation')
round(out$prob,2) ###posterior probability
out$theta ### common effect size

#### compute the p-value and do parameter estimation for binary outcome
n = 4000; m = 20
X = matrix(rnorm(n*m),n)
p = rowMeans(X[,1:4])*.4
Y = rbinom(n,1,p = exp(p)/(1+exp(p)))
HMVD(Y,X,model.type='D')$p.value #### approximate p-value

HMVD(Y,X,nperm.max = 20)$p.value.perm #### p-value based on permutations
### In practice we would increase the number of permutations

out = HMVD(Y,X,model.type='D',method='estimation')
round(out$prob,2) ###posterior probability
out$theta ### common effect size
```

Index