Package 'imputeMDR'

February 20, 2015

Type Package

Title The Multifactor Dimensionality Reduction (MDR) Analysis for Incomplete Data

Version 1.1.2

Date 2011-06-28

Author Junghyun Namkung, Taeyoung Hwang, MinSeok Kwon, Sunggon Yi and Wonil Chung

Maintainer Junghyun Namkung <`jh.namkung@gmail.com`>

Description This package provides various approaches to handling missing values for the MDR analysis to identify gene-gene interactions using biallelic marker data in genetic association studies

License GPL (>= 2)

LazyLoad yes

Repository CRAN

Date/Publication 2012-12-21 09:01:16

NeedsCompilation yes

R topics documented:

`imputeMDR-package` *The Multifactor Dimensionality Reduction (MDR) Analysis for Incomplete Data*

Description

This provides various approaches to handling missing values for the MDR analysis of incomplete data to identify gene-gene interactions using biallelic marker data in genetic association studies

Details

Package:	imputeMDR
Type:	Package
Version:	1.1.1
Date:	2011-06-28
License:	GPL (>2)
LazyLoad:	yes

impute.mdr()

Author(s)

Junghyun Namkung, Taeyoung Hwang, MinSeok Kwon, Sunggon Yi and Wonil Chung Maintainer: Junghyun Namkung <jh.namkung@gmail.com>

References

Namkung J, Elston RC, Yang JM, Park T. "Identification of gene-gene interactions in the presence of missing data using the multifactor dimensionality reduction method" Genet Epidemiol. 2009 Nov;33(7):646-56.

See Also

`impute.mdr`

`impute.mdr` *The Multifactor Dimensionality Reduction (MDR) Analysis for Incomplete Data*

Description

This function provides various approaches to handling missing values for the MDR analysis of incomplete data to identify gene-gene interactions using biallelic marker data in genetic association studies

Usage

```
impute.mdr(dataset, colresp, cs, combi, cv.fold = 10, na.method = 0, max_iter = 30, randomize = FALSE)
```

Arguments

`dataset`	A matrix of SNP data with class variable (response; phenotype; disease status). Genotypes must be coded as allele counts (0,1,2). Missing genotypes should be coded as 3
`colresp`	Column number of class variables in the dataset. No missing value is allowed for the class variable
`cs`	The value used to indicate "case (affected)" for class variable
`combi`	The number of SNPs considered simultaneously as predictor variables (An order of interactions to analyze)
`cv.fold`	The number of folds k for k-fold cross-validation
`na.method`	Options for missing handling approaches. na.method = 0 for complete data, na.method = 1 for treating missing genotypes as another genotype category, na.method=2 for using available data for given number of SNPs under consideration as a model, na.method=3 for using method of imputing missing information by using EM (expectation-maximization) algorithm
`max_iter`	The number of maximum iteration in EM impute approach (na.method=3). In order to apply one-step EM approach, set this argument as 1
`randomize`	Logical. If 'TRUE' the cross validation sets are randomized

Value

`min.comb`	Marker combinations with the minimum error rate in each cross validation
`train.erate`	Training errors for selected marker combination
`test.erate`	Test error of the selected marker combination
`best.combi`	The best combination that was selected most frequently across k-fold cross validation

Author(s)

Junghyun Namkung, Taeyoung Hwang, MinSeok Kwon, Sunggon Yi and Wonil Chung

Maintainer: Junghyun Namkung <jh.namkung@gmail.com>

References

Namkung J, Elston RC, Yang JM, Park T. "Identification of gene-gene interactions in the presence of missing data using the multifactor dimensionality reduction method" Genet Epidemiol. 2009 Nov;33(7):646-56.

Examples

```
## sample data with missing values
data(incomplete)
## analysis example of 2nd order gene-gene interaction test
impute.mdr(incomplete, colresp=1, cs=1, combi=2, cv.fold = 10,na.method=2)
```

`incomplete` *A simulated example data containing missing values*

Description

This data is an example of a simulated case-control study data with 10 10 biallelic marker genotypes (ex : single nucleotide polymorphisms; SNPs) for 200 cases and 200 controls are included. Class variable (response; phenotype; disease status) is in the first column.

Usage

```
data(incomplete)
```

Format

A data frame with 400 observations on the following 11 variables.

`class` a numeric vector

`snp1` a numeric vector

`snp2` a numeric vector

`snp3` a numeric vector

`snp4` a numeric vector

`snp5` a numeric vector

`snp6` a numeric vector

`snp7` a numeric vector

`snp8` a numeric vector

`snp9` a numeric vector

`snp10` a numeric vector

References

Namkung J, Elston RC, Yang JM, Park T. "Identification of gene-gene interactions in the presence of missing data using the multifactor dimensionality reduction method" Genet Epidemiol. 2009 Nov;33(7):646-56.

Examples

```
data(incomplete)
```

Index