Package 'HomUHet'

February 22, 2021

Title Identifying and Separating Homogeneous and Heterogeneous
Predictors

Version 0.0.0.9000

Date 2020-01-22

Author Pei S. Yang [aut, cre]

Maintainer Pei S. Yang <yang.1736@osu.edu>

Depends glmnet, gglasso, dplyr, R (>= 2.10)

Imports mvtnorm, stats

Description This package contains functions to identify and separate predictors with homogeneous or heterogeneous effects across datasets.

License GPL (>=3)

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

R topics documented:

1

HomUHet   *fit a two-step penalized regression model*

Description

This function outputs the names of predictors with homogeneous or heterogeneous predictors across multiple data sets, the estimates of predictors, and solution plots

Usage

```
HomUHet(data, solution_path_plot = FALSE)
```

Arguments

data   The input dataframe containing observations from all studies where the first column is the study label, the second column is the response variable and the following columns are the predictors.

solution_path_plot
                    TRUE if outputting solution path plots is desired

Value

the names of identified predictors and their estimated effects

Homo   a character string of names of homogeneous predictors

Heter   a character string of N=names of heterogeneous predictors

coefficients   a data frame containing estimated coefficients of the homogeneous and heterogeneous predictors in K studies

---

HomUHet.sim   *simulate multiple data sets with both homogeneous and heterogeneous effects from the predictors*

Description

this function simulate data

Usage

```
HomUHet.sim(
  Pred_type = c("Gaussian", "SNP"),
  J,
  K,
  beta = NULL,
  rho = 0.5,
  sigma = 2,
  level = c("l", "m", "e"),
  nlower = 50,
  nupper = 300,
  allele_freq
)
```

Arguments

Pred_type	the predictor type; choose between Gaussian or SNP
J	the number of predictors.
K	the number of studies.
beta	the K x J coefficient matrix
rho	a number between 0 and 1. controlling the degree of correlation between predictors
sigma	a positive number. controlling the added noise to the simulated response variable
level	the level of heterogeneity. ignored if "beta" is supplied. "l" stands for low, "m" stands for medium, and "h" stands for high.
nlower	the lower bound of the K sample sizes
nupper	the upper bound of the K sample sizes
allele_freq	a J-length vector containing the allele frequencies for the J SNPs. ignored if Pred_type="Gaussian"

Value

the simulated data

HomUHet.sim.beta   *simulates homogeneous and heterogeneous coefficients of predictors*

Description

this function outputs matrix of coefficients of all predictors (homogeneous or heterogeneous, or 0 in the case when the corresponding predictor has no effect) and organize them as desired by the user

Usage

```
HomUHet.sim.beta(
  J,
  K,
  homo_coef,
  heter_distr = c("Gaussian", "Uniform"),
  heter_coef_param = rbind(c(1, 2, 3), c(2, 1, 2), c(2, 2, 3))
)
```

Arguments

J	The total number of predictors including the predictors with homogeneous and heterogeneous effects and the predictors without effects
K	The number of studies
homo_coef	the 2 x p1 input matrix to generate homogeneous coefficients of p1 number of predictors. The first row should be integers indicating which columns the user wants the homogeneous coefficients to be. The second row should be the homogeneous coefficients themselves

heter_distr   indicates which distribution the user wishes to use to generate the heterogeneous
                   coefficients

heter_coef_param

                   the 3 x p2 input matrix to generate heterogeneous coefficients of p2 number of
                   predictors. The first row should be integers indicating which columns the user
                   wants the heterogeneous coefficients to be. If heter_distr="Gaussian", the sec-
                   ond and third row for each column are, resepctively, the mean and standard devi-
                   ation of the gaussian distribution used as heter_distr. If heter_distr="Uniform",
                   the second and third row for each column are, respectively, the lower boundary
                   and upper boundary of the uniform distribution used as heter_distr.

Value

the simulated coefficient matrix and miscellenance information about it

beta   the K x J coefficient matrix

J   the number of predictors including both predictors which have effects and which
                   do not

K   the number of studies

homo_index   a vector containing the column numbers of homogeneous coefficients in the co-
                   efficient matrix

heter_index   a vector containing the column numbers of homogeneous coefficients in the co-
                   efficient matrix

HomUHet_data   *simulated data to be used as demonstration for HomUHet package*

Description

A dataset containing observations of 4 studies and 500 predictors

Usage

HomUHet_data

Format

A data frame with 444 rows and 502 variables:

Study_label integers indicating the study number

y the response variable ...

Index