Package 'wsprv'

December 17, 2021

Type Package

Title Weighted Selection Probability for Rare Variant Analysis

Version 0.1.0

Author Xianglong Liang [aut, cre],
Hokeun Sun [ctb]

Maintainer Xianglong Liang <xianglongliang@gmail.com>

Description
A weighted selection probability to locate rare variants associated with multiple phenotypes.

License GPL-2

Depends R (>= 3.6.0)

Imports glmnet, mnormt, stats

Encoding UTF-8

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2021-12-17 19:40:02 UTC

R topics documented:

`wsprv` *A weighted selection probability is developed to locate individual rare variants associated with multiple phenotypes.*

Description

Recently, rare variant association studies with multiple phenotypes have drawn a lot of attentions because association signals can be boosted when rare variants are related with more than one phenotype. Most of existing statistical methods to identify rare variants associated with multiple phenotypes are based on a group test, where a gene or a genetic region is tested one at a time. However, these methods are not designed to locate individual rare variants within a gene or a genetic region. We propose a weighted selection probability to locate individual rare variants within a group after a multiple-phenotype based group test finds significance.

Usage

```
weight_sp(
  x,
  y,
  alpha = 1,
  penalty.factor = NULL,
  standardize = TRUE,
  type.multinomial = c("grouped", "ungrouped"),
  rep = 100,
  rate = 0.05,
  gamma = 0.01
)
```

Arguments

`x`	A $n \times (m + p)$ matrix with n samples, m covariates and p rare variants where m can be zero, i.e., there does not exist covariates.
`y`	A $n \times Q$ phenotype matrix with n samples and Q phenotypes where $Q > 1$.
`alpha`	The mixing parameter of elastic-net, `alpha=1` is the lasso, and `alpha=0` is the ridge. Default value is 1.
`penalty.factor`	Separate penalty factors factors can be applied to each coefficient. Can be `0` for some variables, which implies no shrinkage, and that variable is always included in the model.
`standardize`	Genotype standardization. Default is `TRUE`.
`type.multinomial`	A group lasso penalty is used on the multinomial coefficients for a variable when 'grouped'. It ensures the multinomial coefficents are all in or out. Default is 'grouped'.
`rep`	The number of bootstrap replications. We recommend to use 100 or more to compute weighted selection probability. Default value is 100.
`rate`	A tuning parameter represents rate of degree of freedom to the number of rare variants. Default value is 0.05.
`gamma`	The upper `gamma` quantile of selection frequencies of individual variants each phenotype to compute the threshold. Default value is 0.01.

Details

The penalty function of `elastic-net` is defined as

$$\lambda(\alpha||\beta||_1 + \frac{(1-\alpha)}{2}||\beta||_2^2),$$

where α is a mixing proportion of ridge and the lasso, and β is regression coefficients. This penalty is equivalent to the Lasso penalty if `alpha=1`.

Let η be the degree of freedom and it depends on the tuning parameter λ, and `rate` is computed as

$$rate = \frac{\eta}{p},$$

Note that $\eta \leq n$ is set up in `weight_sp` function.

Let δ_γ be a threshold of SF and it depends on the upper γ^{th} qunatile value of SF. Where $SF = \{SF_{11}(\eta), SF_{21}(\eta), \cdots, SF_{pQ}(\eta)\}$ is a set that contains selection frequencies of individual rare variants each phenotype.

Value

`res`	A matrix contains the order of weighted selection probabilities from the largest to the smallest and the corresponding weighted selection probabilities.
`eta`	eta used.
`bootstrap.rep`	The number of bootstrap replications used.
`rate`	The tuning parameter `rate` used.
`gamma`	The upper `gamma` quantile of selection frequencies of individual rare variants each phenotype used.

Examples

```
# Generate simulation data
 n <- 400
 p <- 100
 q <- 5
 MAF <- 0.01
 geno.prob <- rbind((1-MAF)^2,2*(1-MAF)*MAF,MAF^2)
 x <- matrix(NA,n,p)
 set.seed(1)
 for(i in 1:p) x[,i] <- sample(0:2,n,prob=geno.prob,replace=TRUE)
 beta <- c(rep(3.0,10),rep(0,(p-10)))
 cova <- matrix(0.75,q,q)
 diag(cova) <- 1
 require(mnormt)
 err.mat <- rmnorm(n,rep(0,q),cova)

 y1 <- x %*% beta+err.mat[,1]
 y2 <- x %*% beta+err.mat[,2]
```

```
y <- cbind(y1,y2,err.mat[,3:5])
# Weighted selection probabilities for individual rare variants without covariates.
#If rep=100, time consuming.
wsp.rv1 <- weight_sp(x,y,rep=5) # continuous phenotypes

# Weighted selection probabilities for individual rare variants with covariates.
#If rep=100, time consuming.
cx <- cbind(rnorm(n),sample(0:1,n,replace=TRUE))
x <- cbind(cx,x)
penalty.factor <- c(rep(0,2),rep(1,p))
colnames(x) <- c('Age','Gender',paste0('V',3:102))
wsp.rv2 <- weight_sp(x,y,penalty.factor=penalty.factor,rep=5) # continuous phenotypes
```

Index