Package ‘GSTiCluster’

January 6, 2015

Type Package

Title Integrative multi-omics clustering for disease subtype discovery by sparse overlapping group lasso and tight clustering

Version 1.0-6

Date 2015-01-06

Author SungHwan Kim, Yongseok Park and George C. Tseng

Maintainer SungHwan Kim <suk73@pitt.edu>

Description A group structured tight iCluster method with a sparse overlapping group lasso technique and a tight clustering concept via regularization to accommodate the information of inter-omics regulation flows, and to exclude outlier samples scattering away from the tight clusters.

License GPL-2

Depends pamr, doMC

R topics documented:

GSTiCluster-package .............................................. 1
GSTiCluster ......................................................... 2

Index 4

GSTiCluster-package Integrative multi-omics clustering for disease subtype discovery by sparse overlapping group lasso and tight clustering

Details

Package: GSTiCluster
Type: Package
Version: 1.0-6
Date: 2015-01-06
License: GPL (>= 2)
Description

GST-iCluster fits a latent variable model of integrative clustering to accommodate the information of inter-omics regulation flows, and to exclude outlier samples scattering away from the tight clusters.

Usage

GSTiCluster(datasets, k, Best_lam = NULL, max.iter = 100, is.scale = FALSE, f.module, is.tight = FALSE, lambda_1 = NULL)

Arguments

datasets        Input variable matrix (combined multi-omics datasets by samples)
k              # of integrative clusters
Best_lam      Regularization parameter for sparse overlapping group lasso
max.iter     # of maximum iteration of EM
is.scale    Logical value indicating whether data are scaled and centered
f.module     Input feature module matrix (Row: feature module / Col: feature)
is.tight    Logical value indicating whether tight clustering is performed
lambda_1    Regularization parameter for tight clustering

Value

Beta = Estimated sparse coefficient matrix
Z = Estimated average latent variable matrix
cluster = Integrative clustering label of samples
is.sample.penalized = Logical values whether samples are included
tight.cluster = Integrative tight clustering label of samples

Author(s)

SungWhan Kim <suk73@pitt.edu>
Examples

```r
set.seed(123)
library(doMC)
library(pamr)

# Generate two random omics datasets
mu <- c(-3,1,3)
Simul_mRNA <- rbind(cbind(matrix(rnorm(4, 0, mu[1]), 4, 5),
                      matrix(rnorm(4, 0, mu[2]), 4, 5),
                      matrix(rnorm(4, 0, mu[3]), 4, 5)),
                      matrix(rnorm(1, 0, 0.1), 1, 15))
mu <- c(1,3,-3)
Simul_methyl <- rbind(cbind(matrix(rnorm(4, 0, mu[1]), 4, 5),
                            matrix(rnorm(4, 0, mu[2]), 4, 5),
                            matrix(rnorm(4, 0, mu[3]), 4, 5)),
                        matrix(rnorm(1, 0, 0.1), 1, 15))

# feature modules across two datasets
null_gr <- rep(0, 10); null_gr[c(1:10, 51:60)] <- 1; Group_1 <- null_gr
null_gr <- rep(0, 10); null_gr[c(11:20, 61:70)] <- 1; Group_2 <- null_gr
null_gr <- rep(0, 10); null_gr[c(21:30, 71:80)] <- 1; Group_3 <- null_gr
null_gr <- rep(0, 10); null_gr[c(31:40, 81:90)] <- 1; Group_4 <- null_gr

# Singleton_group
Group_Single <- matrix(0, 2, 10)
Group_Single[1:10, 41:50] <- diag(10)

# Total feature module
Total_group <- rbind(Group_1, Group_2, Group_3, Group_4, Group_Single)

# Combined dataset
DList <- rbind(Simul_mRNA, Simul_methyl)

## Example of GS-iCluster
GSTiCluster(DList, k = 3, Best_lam = 5, f.module = Total_group, is.scale=TRUE)

# Combined dataset (including non-informative samples)
## Three non-informative samples
Dat.r.sam <- matrix(rnorm(dim(DList)[1] * 3, 0, 0.01), dim(DList)[1], 3)
DDList <- cbind(DList, Dat.r.sam) # dataset combined with three non-informative samples

## Example of GST-iCluster
GSTiCluster(DDList, k = 3, Best_lam = 10, f.module = Total_group, is.scale=TRUE,
            is.tight = TRUE, lambda_1 = 1)
```
Index

*Topic GST-iCluster
   GSTiCluster, 2

GSTiCluster, 2
GSTiCluster (GSTiCluster-package), 1
GSTiCluster-package, 1