

Package ‘oBGSelComp’

August 8, 2021

Type Package
Title oBGSelComp: overlapping Bayesian Group Selection for Compositional data
Version 1.0
Date 2020-10-01
Author Thierry Chekouo
Maintainer Thierry Chekouo <thierry.chekouotekou@ucalgary.ca>
Description The oBGSelComp package implements an MCMC algorithm for a Bayesian hierarchical model for variable selection with a group structure in the context of correlated multivariate compositional response variables.
SystemRequirements GSL (GNU Scientific Library)
License GPL (>= 2.0)
Imports sirt, gtools, AUC
RoxygenNote 7.1.1
Encoding UTF-8

R topics documented:

| | |
|-------------------------------|----------|
| GBMPathwayData | 1 |
| HallmarkPathwayData | 2 |
| KeggPathwayData | 3 |
| oBGSelCompAlgo | 3 |
| OncoPathwayData | 5 |
| SimulateData | 5 |
| Index | 7 |

| | |
|----------------|--|
| GBMPathwayData | <i>Data sets of patients with Glioblastoma and GBM pathway information</i> |
|----------------|--|

Description

A list of 4 data frames that contains tumor proportion, mRNA expression, and GBM pathway information and their names.

Usage

```
GBMPathwayData
```

Format

A list of 4 data frames:

TumorProportion Brain tumor proportions: edema, necrotic, non-enhancing and enhancing

mRNA mRNA expression of 718 genes.

GBMPathway GBM pathway information: it's a binary matrix that indicates the membership of a gene within an GBM pathway

GBMpathwaynames GBM pathway names

References

Thierry Chekouo et al (2021), *A Bayesian group selection with compositional responses for analysis of radiologic tumor proportions and their genomic determinants*, submitted.

Examples

```
data("GBMPathwayData")
```

| | |
|---------------------|---|
| HallmarkPathwayData | <i>Data sets of patients with Glioblastoma and Hallmark pathway information</i> |
|---------------------|---|

Description

A list of 4 data frames that contains tumor proportion, mRNA expression, and Hallmark pathway information and thier names.

Usage

```
HallmarkPathwayData
```

Format

A list of 4 data frames:

TumorProportion Brain tumor proportions: edema, necrotic, non-enhancing and enhancing

mRNA mRNA expression of 1172 genes.

HallmarkPathway Hallmark pathway information: it's a binary matrix that indicates the membership of a gene within a Hallmark pathway

Hallmarkpathnames Hallmark pathway names

References

Thierry Chekouo et al (2021), *A Bayesian group selection with compositional responses for analysis of radiologic tumor proportions and their genomic determinants*, submitted.

Examples

```
data("HallmarkPathwayData")
```

| | |
|-----------------|---|
| KeggPathwayData | <i>Data sets of patients with Glioblastoma and Kegg pathway information</i> |
|-----------------|---|

Description

A list of 4 data frames that contains tumor proportion, mRNA expression, and KEGG pathway information and their names.

Usage

```
KeggPathwayData
```

Format

A list of 4 data frames:

TumorProportion Brain tumor proportions: edema, necrotic, non-enhancing and enhancing

mRNA mRNA expression of 364 genes.

KeggPathway KEGG pathway information: it's a binary matrix that indicates the membership of a gene within a KEGG pathway.

Keggpathwaynames KEGG pathway names.

References

Thierry Chekouo et al (2021), *A Bayesian group selection with compositional responses for analysis of radiologic tumor proportions and their genomic determinants*, submitted.

Examples

```
data("KeggPathwayData")
```

| | |
|----------------|--|
| oBGSelCompAlgo | <i>An MCMC algorithm to perform an overlapping Bayesian group selection for compositional response data.</i> |
|----------------|--|

Description

The function implements an MCMC algorithm for a Bayesian model for compositional data using a Dirichlet model by allowing for straightforward incorporation of available high-dimensional covariate information within a log-linear regression framework. The method imposes prior distributions that account for the overlapping structure between groups.

Usage

```
oBGSelCompAlgo(outcome=outcome,X=X,Pathway=Pathway,sample=10000,
burnin=1000,hyperprob=c(1,1),hypersig=c(5,0.01),hyperlamb=c(5,2), seed=1)
```

Arguments

| | |
|-----------|--|
| outcome | A compositional multivariate response of dimension $n \times q$ |
| X | A set of covariate matrix of dimension $n \times p$ (e.g. gene expression) |
| Pathway | A binary pathway indicator matrix P of dimension $K \times p$ where cell (k, j) is 1 if the covariate j belongs to group (or pathway) k , and 0 otherwise. |
| sample | Total number of MCMC draws. It must be larger than burnin. |
| burnin | Number of draws to discard for burn-in |
| hyperprob | Hyperparameters of a beta(a,b) distribution, prior distribution of the probability of selecting groups |
| hypersig | Hyperparameters of an inverse gamma inv.gamma(a,b) distribution, prior distribution of the variance σ_l^2 |
| hyperlamb | Hyperparameters of a gamma(a,b) distribution, prior distribution of the shrinkage parameter λ_l |
| seed | Random number are generated with seed=1 during the MCMC algorithm |

Details

The function will return two R objects, which can be assigned to a variable. To see the results, use the "\$" operator.

Value

| | |
|-------------|---|
| PostProbGrp | Marginal posterior probabilities of the group selection indicators, $Prob(\gamma_{lk} = 1)$ |
| BetaSample | An MCMC sample of the regression effects β_{lk} |

References

Thierry Chekouo et al (2021), *A Bayesian group selection with compositional responses for analysis of radiologic tumor proportions and their genomic determinants*, submitted.

See Also

[SimulateData](#)

Examples

```
library(oBGSelComp)
## overlap groups (10 percent overlap)
dat=SimulateData(n=100,betaAbs=1,sig=.1,r2=0.5,propOverl=10,seed=1)
Res=oBGSelCompAlgo(outcome=dat$y,X=dat$X,Pathway=dat$Pathway,sample=10000,
burnin=1000,hyperprob=c(1,1),hypersig=c(4,0.01),hyperlamb=c(5,2), seed=1)
str(Res)
library(AUC)
q=4
AUC1=rep(0,q)
```

```

for (l in 1:q){
  AUC1[l]= auc(roc(Res$PostProbGrp[l,],as.factor(dat$GamKnown[l,])))
}
AUC1

```

| | |
|-----------------|--|
| OncoPathwayData | <i>Data sets of patients with Glioblastoma and Oncogenetic pathway information</i> |
|-----------------|--|

Description

A list of 4 data frames that contains tumor proportion, mRNA expression, and Oncogenetic pathway information and their names.

Usage

```
OncoPathwayData
```

Format

A list of 4 data frames:

TumorProportion Brain tumor proportions: edema, necrotic, non-enhancing and enhancing

mRNA mRNA expression of 1032 genes.

OncoPathway Oncogenetic pathway information: it's a binary matrix that indicates the membership of a gene within an Oncogenetic pathway

Oncopathnames Oncogenetic signature names

References

Thierry Chekouo et al (2021), *A Bayesian group selection with compositional responses for analysis of radiologic tumor proportions and their genomic determinants*, submitted.

Examples

```
data("OncoPathwayData")
```

| | |
|--------------|---|
| SimulateData | <i>Generation of simulated data as explained in the reference manuscript.</i> |
|--------------|---|

Description

This function generates compositional response and covariate data as described in the manuscript (see Section 5.2 in the manuscript).

Usage

```
SimulateData(n=100,betaAbs=1,sig=.1,r2=0.5,propOverl=0,seed=1,K=20,p=200,q=4,nbrpath=4)
```

Arguments

| | |
|-----------|---|
| n | The number of subjects/samples |
| betaAbs | Effect values in absolute values of important features |
| sig | Value of the variance of latent values α 's given the feature set |
| r2 | Variance value that represents correlation of features within pathways |
| propOverl | Proportion of overlap between "consecutive" groups. It can only take three values; 0, 10 and 50 |
| seed | Seed to generate random numbers |
| K | K is the number of pathways/groups |
| p | p is the number of covariates (e.g., genes) |
| q | q is the number of (compositional or proportion) outcomes |
| nbrpath | It's the number of important pathways; It's the nbrpath's first pathways |

Details

The function will return three data matrices y, X, Pathway and GamKnown which are respectively the compositional response data, the covariate matrix, the Pathway matrix indicator and the true gamma ($\gamma(l,k)=1$ if pathway k is important for outcome l, and 0 otherwise)

References

Thierry Chekouo et al (2021), *A Bayesian group selection with compositional responses for analysis of radiologic tumor proportions and their genomic determinants*, submitted.

Examples

```
dat=SimulateData(n=100,betaAbs=1,sig=.1,r2=0.5,propOverl=0,seed=1)
str(dat)
```

Index

- * **HallmarkPathwayData**
 - HallmarkPathwayData, [2](#)
- * **datasets**
 - GBMPathwayData, [1](#)
 - KeggPathwayData, [3](#)
 - OncoPathwayData, [5](#)
- GBMPathwayData, [1](#)
- HallmarkPathwayData, [2](#)
- KeggPathwayData, [3](#)
- oBGSelCompAlgo, [3](#)
- OncoPathwayData, [5](#)
- SimulateData, [4](#), [5](#)