Package 'oBGSelComp'

October 1, 2020

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 6.1.0

Encoding UTF-8

R topics documented:

`oBGSelCompAlgo` *An MCMC algorithm to perform an overlapping Bayesian group selection for compostional response data.*

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 varianceσ_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 (2020), *A Bayesian Group Selection approach for the analysis of volumetric images of brain cancers 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
```

SimulateData *Generation of simulated data as explained in the reference manuscript.*

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)
```

Arguments

n	The sumber of subjects/samples
betaAbs	Effect values in absolute values of impaortant features
sig	Value of the variance of latent values alpha'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 takes three values;0,10 and 50
seed	Seed to generate random numbers

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 (2020), *A Bayesian Group Selection approach for the analysis of volumetric images of brain cancers and their genomic determinants*, *submitted*.

Examples

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

Index