Package ‘BSGS’

June 24, 2015

Type Package

Title Bayesian Sparse Group Selection

Version 2.0

Date 2015-06-23

Author Kuo-Jung Lee and Ray-Bin Chen

Maintainer Kuo-Jung Lee <kuojunglee@mail.ncku.edu.tw>

Description The integration of Bayesian variable and sparse group variable selection approaches for regression models.

Depends R (>= 3.1.1), MASS, compiler, pscl, plyr, batchmeans

License GPL-2

LazyLoad yes

Repository CRAN

NeedsCompilation no

Date/Publication 2015-06-24 01:07:18

R topics documented:

BSGS.PE ........................................ 2
BSGS.Sample .................................. 2
BSGS.Simple ................................... 4
CGS.SMP.PE .................................. 5
CompWiseGibbsSimple ......................... 6
CompWiseGibbsSMP ......................... 7
Crisis2008 .................................. 8
Crisis2008BalancedData ................. 8
MSE.BSGS .................................. 9
MSE.CGS.SMP ................................. 9
TCR.TPR.FPR.BSGS ....................... 10
TCR.TPR.FPR.CGS.SMP ............... 11

Index 12
BSGS.PE

Posterior estimates of parameters.

Description

Provide the posterior estimates of parameters.

Usage

BSGS.PE(BSGS.Output)

Arguments

BSGS.Output A list of random samples generated from the posterior distribution by MCMC procedures.

Value

A list is returned with estimates of regression coefficients, $\beta$, the posterior probability of binary variable $\eta$ for group selection equal to 1, binary variable $\gamma$ for variable selection equal to 1, and variance, $\sigma^2$.

Examples

```r
## Not run:
BSGS.PE(output)
## End(Not run)
```

BSGS.Sample

Sample version of group-wise Gibbs sampler for sparse group selection.

Description

Generate the posterior samples by an approximation sampling method to perform Bayesian sparse group selection to identify the important groups of variables and variables within those.

Usage

BSGS.Sample(Y, X, Group.Index, r.value, eta.value, beta.value, tau2.value, rho.value, theta.value, sigma2.value, nu, lambda, Num.of.Iter.Inside.CompWise, Num.Of.Iteration, MCSE.Sigma2.Given)
ARGUMENTS

Y       vector of observations of length n.
X       design matrix of dimension n x p.
Group.Index  Specify the group label to which the variable belongs.
r.value  Initial values of indicator variables for individual variables.
etas.value  Initial values of indicator variables for the group variables.
betas.value  Initial values of regression coefficients, \( \beta \).
tau2.value  Variance in the prior distribution for regression coefficients.
rho.value  Prior inclusion probability for a variable.
theta.value  Prior inclusion probability for a group.
sigma2.value  Initial value of \( \sigma^2 \).
u  The hyperparameter in the prior distribution of \( \sigma^2 \).
lambdas  The hyperparameter in the prior distribution of \( \sigma^2 \).
Num.of.Iter.Inside.CompWise  Specify the number of iterations within component wise Gibbs sampler for variable selection within a group.
Num.Of.Iteration  Specify the number of iterations for sparse group variable selection.
MCSE.Sigma2.Given  Prespecified value which is used to stop simulating samples. When the MCSE of estimate of \( \sigma^2 \) less than the given value, the simulation is terminated.

VALUE

A list is returned with posterior random samples of regression coefficients, \( \beta \), binary variables for group selection, \( \eta \), binary variables for variable selection, \( \gamma \), variance, \( \sigma^2 \) and the number of iterations performed and the elapsed time in second required for the run.

EXAMPLES

```r
## Not run:
output = BSGS.Sample(Y, X, Group.Index, r.value, eta.value, beta.value, tau2.value,
rho.value, theta.value, sigma2.value, nu, lambda,

## End(Not run)
```
BSGS.Simple

The group-wise Gibbs sampler for sparse group selection.

Description

Generate the posterior samples to perform Bayesian sparse group selection to identify the important groups of variables and variables within those.

Usage


Arguments

Y vector of observations of length n.
X design matrix of dimension n × p.
Group.Index Specify the group label to which the variable belongs.
r.value Initial values of indicator variables for individual variables.
eta.value Initial values of indicator variables for the group variables.
beta.value Initial values of regression coefficients, β.
tau2.value Variance in the prior distribution for regression coefficients.
rho.value Prior inclusion probability for a variable.
theta.value Prior inclusion probability for a group.
sigma2.value Initial value of σ^2.
nu The hyperparameter in the prior distribution of σ^2.
lambda The hyperparameter in the prior distribution of σ^2.
Num.of.Iter.Inside.CompWise Specify the number of iterations within component wise Gibbs sampler for variable selection within a group.
Num.Of.Iteration Specify the number of iterations for sparse group variable selection.
MCSE.Sigma2.Given Prespecified value which is used to stop simulating samples. When the MCSE of estimate of σ^2 less than the given value, the simulation is terminated.

Value

A list is returned with posterior random samples of regression coefficients, β, binary variables for group selection, η, binary variables for variable selection, γ, variance, σ^2 and the number of iterations performed and the elapsed time in second required for the run.
Examples

```r
## Not run:

output = BSGS.Simple(Y, X, Group.Index, r.value, eta.value, beta.value, tau2.value,
                     rho.value, theta.value, sigma2.value, nu, lambda, Num.of.Iter.Inside.CompWise,
                     Num.Of.Iteration, MCSE.Sigma2.Given)

## End(Not run)
```

**CGS.SMP.PE**

*Posterior estimates of parameters.*

**Description**

Calculate the posterior estimates of parameters based on the samples generated from the posterior distribution by the stochastic matching pursuit (SMP).

**Usage**

```r
CGS.SMP.PE(CGS.SMP.Output)
```

**Arguments**

- `CGS.SMP.Output` A list of random samples for parameters

**Value**

A list is returned with estimates of regression coefficients, $\beta$, binary variables for variable selection, $\gamma$, and variance, $\sigma^2$.

**Examples**

```r
## Not run:

output = CompWiseGibbsSMP(Y, X, beta.value, r, tau2, rho, sigma2, nu, lambda,
                          num.of.inner.iter, num.of.iteration, MCSE.Sigma2.Given)
CGS.SMP.PE(output)

## End(Not run)
```
CompWiseGibbsSimple Generate the posterior samples from the posterior distribution using the component-wise Gibbs sampler (CWGS).

Description

Generate the posterior samples using MCMC procedures.

Usage

CompWiseGibbsSimple(Y, X, beta.value, r, tau2, rho, sigma2, nu, lambda, num.of.inner.iter, num.of.iteration, MCSE.Sigma2.Given)

Arguments

Y vector of observations of length n.
X design matrix of dimension n \times p.
beta.value Initial values of regression coefficients, β.
r Initial values of indicator variables for individual regressors.
tau2 Variance in the prior distribution for regression coefficients.
rho Prior probability including a variable.
sigma2 Initial value of σ².
nu The hyperparameter in the prior distribution of σ².
lambda The hyperparameter in the prior distribution of σ².
num.of.inner.iter The number of iterations before sampling σ².
num.of.iteration The number of iterations to be runned for sparse group variable selection.
MCSE.Sigma2.Given Prespecified value which is used to stop simulating samples when the MCSE of estimate of σ² less then given values.

Value

A list is returned with posterior samples of regression coefficients, β, variance σ², binary variables, γ, the number of iterations performed, and the time in second required for the run.

Examples

```r
## Not run:
CompWiseGibbsSimple(Y, X, beta.value, r, tau2, rho, sigma2, nu, lambda, 
num.of.inner.iter.default, num.of.iteration, MCSE.Sigma2.Given)
## End(Not run)
```
Stochastic matching pursuit for variable selection.

Description

Perform MCMC procedure to generate the posterior samples to estimate posterior quantities of interest in Bayesian variable selection using stochastic matching pursuit approach (SMP).

Usage

CompWiseGibbsSMP(Y, X, beta.value, r, tau2, rho, sigma2, nu, lambda, num.of.inner.iter, num.of.iteration, MCSE.Sigma2.Given)

Arguments

Y vector of observations of length n.
X design matrix of dimension n × p.
beta.value Initial values of regression coefficients, β.
r Initial values of indicator variables for individual regressors.
tau2 Variance in the prior distribution for regression coefficients.
rho Prior probability including a variable.
sigma2 Initial value of σ^2.
nu Given value in the prior distribution of σ^2.
lambda Given value in the prior distribution of σ^2.
um.of.inner.iter The number of iterations before sampling σ^2.
um.of.iteration The number of iterations to be runned for sparse group variable selection.
MCSE.Sigma2.Given Prespecified value which is used to stop simulating samples when the MCSE of estimate of σ^2 less then given values.

Value

A list is returned with posterior samples of regression coefficients, β, variance σ^2, binary variables, γ, the number of iterations performed, and the time in second required for the run.

Examples

## Not run:
CompWiseGibbsSMP(Y, X, beta.value, r, tau2, rho, sigma2, nu0, lambda0, num.of.inner.iter, num.of.iteration, MCSE.Sigma2.Given)
## End(Not run)
Crisis2008

A cross-sectional data set from Rose and Spiegel.

Description

A cross-sectional data set from Rose and Spiegel (2011), which is available at http://faculty.haas.berkeley.edu/arose. The response variable is 2008-2009 growth rate for the crisis measure. This dataset consists of 119 explanatory factors for the crisis for as many as 107 countries, but there are data missing for a number of countries.

Usage

data(Crisis2008)

Crisis2008BalancedData

A cross-sectional data set from Rose and Spiegel with the removal of missing values.

Description

A cross-sectional data set from Rose and Spiegel (2011), which is available at http://faculty.haas.berkeley.edu/arose. The response variable is 2008-2009 growth rate for the crisis measure. Rose and Spiegel originally consider 119 explanatory factors for the crisis for as many as 107 countries, but there are data missing for a number of countries. To maintain a balanced data set, we use 51 regressors for a sample of 72 countries. These regressors can be classified into the nine theoretical groups of the crisis’ origin (the number in parentheses indicates the number of variables considered in the group): principal factors (10), financial policies (three), financial conditions (four), asset price appreciation (two), macroeconomic policies (four), institutions (11), geography (four), financial linkages (one), and trade linkages (12).

Usage

data(Crisis2008BalancedData)
MSE.BSGS

Mean square error (MSE).

Description
Calculate the mean square error for the sparse group selection problems.

Usage
MSE.BSGS(Output, Y, X)

Arguments
- **Output**: A list of random samples generated by either “BSGS.Simple” or “BSGS.Sample” for parameters from spare group selection problems.
- **Y**: Observations.
- **X**: Design matrix.

Value
Return the mean square error.

Examples
```r
## Not run:
MSE.BSGS(output, Y, X)

## End(Not run)
```

MSE.CGS.SMP

Mean square error (MSE).

Description
Calculate the mean square error.

Usage
MSE.CGS.SMP(Output, Y, X)
Arguments

Output  A list of random samples for parameters generated by CGS or SMP algorithm.
Y     Observations.
X     Design matrix.

Value

Return the mean square error.

Examples

## Not run:
MSE.CGS.SMP(output, Y, X)

## End(Not run)

---

TCR.TPR.FPR.BSGS  Evaluate TCR, TPR and FPR for sparse group variable selection problems.

Description

Calculate the true classification rate (TCR), the true positive rate (TPR), and the false positive rate (FPR).

Usage

TCR.TPR.FPR.BSGS(Output, True.r, Critical.Point)

Arguments

Output  A list of random samples for parameters.
True.r  The true value of indicator variable.
Critical.Point  When the posterior estimate of \( r = 1 \) greater than this critical point, then it would be assign to 1, and otherwise 0.

Value

A list is returned with TCP, TPR and FPR.
Examples

## Not run:
```r
output = BSGS.Simple.SaveAllSimulatedSamples(Y, X, Group.Index, r.value, eta.value,
    beta.value, tau2.value, rho.value, theta.value, sigma2.value, nu, lambda,
TCR.TPR.FPR.BSGS(output, r.true, critical.value)
```
## End(Not run)

### TCR.TPR.FPR.CGS.SMP

Evaluate TCR, TPR and FPR for variable selection problems.

Description

Calculate the true classification rate (TCR), the true positive rate (TPR), and the false positive rate (FPR).

Usage

```r
TCR.TPR.FPR.CGS.SMP(Output, True.r, Critical.Point)
```

Arguments

- **Output**: A list of random samples for parameters.
- **True.r**: The true value of indicator variable.
- **Critical.Point**: When the posterior estimate of \( r = 1 \) greater than this critical point, then it would be assign to 1, and otherwise 0.

Value

A list is returned with TCP, TPR and FPR.

Examples

## Not run:
```r
output = BSGS.Simple.SaveAllSimulatedSamples(Y, X, Group.Index, r.value, eta.value,
    beta.value, tau2.value, rho.value, theta.value, sigma2.value, nu, lambda,
TCR.TPR.FPR.BSGS(output, r.true, critical.value)
```
## End(Not run)
Index

BSGS.PE, 2
BSGS.Sample, 2
BSGS.Simple, 4

CGS.SMP.PE, 5
CompWiseGibbsSimple, 6
CompWiseGibbsSMP, 7
Crisis2008, 8
Crisis2008BalancedData, 8

MSE.BSGS, 9
MSE.CGS.SMP, 9

TCR.TPR.FPR.BSGS, 10
TCR.TPR.FPR.CGS.SMP, 11