Package ‘BSM2bg’

November 2, 2020

Type Package

Title Implementing Bayesian Shrinkage Models Using Two Block Gibbs Samplers

Version 1.0

Date 2020-11-02

Author Rui Jin

Maintainer Rui Jin <rui-jin-1@uiowa.edu>

Description Two block Gibbs samplers for efficiently sampling from posterior distributions of Bayesian group lasso, Bayesian sparse group lasso and Bayesian fused lasso models. The package is an implementation of Jin and Tan (2019).

License GPL (>= 2)

Imports Rcpp (>= 1.0.3), gglasso, BSGS

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

R topics documented:

`bfl` *Bayesian fused Lasso*

Description

Inference for fused lasso model by two block Gibbs sampling from the Bayesian posterior distribution.

Usage

```
bfl(X, Y, beta, sigma2, lambda1 = 1, lambda2 = 1,
alpha = 0 ,xi = 0, K = 10000)
```

Arguments

X	a matrix of predictors, each row corresponds to one observation.
Y	a response vector of length equal to the leading dimension (rows) of X, i.e., length(Y) == nrow(X).
beta	initial value of the regression coefficients β, its length should equal to ncol(X).
sigma2	initial value of the variance parameter σ^2, which should be positive.
lambda1	initial value of the penalty parameter λ_1, which should be nonnegative.
lambda2	initial value of the penalty parameter λ_2, which should be nonnegative.
alpha	initial value of the shape parameter for the inverse gamma prior of σ^2.
xi	initial value of the rate parameter for the inverse gamma prior of σ^2.
K	total number of MCMC samples to be collected.

Details

The sampling algorithm implemented by bfl for the Bayesian fused lasso model is described in detail in Jin & Tan (2019). It samples from the joint posterior distribution of (β, σ^2) using a two block Gibbs sampler, which is defined in (3.7).

Value

bfl returns a list object, which contains the components listed below.

betas: a K*nrow(X) matrix of K samples from the (penalized) regression coefficients

sigma2s: a vector of K samples of the variance parameter

Author(s)

Rui Jin

References

Rui Jin and Aixin Tan. "Fast Markov chain Monte Carlo for high dimensional Bayesian regression models with shrinkage priors." (2019). https://arxiv.org/pdf/1903.06964.pdf

Examples

```
n <- 100
p <- 50
r <- 0.2
Xvarhalf <- diag(sqrt(1-r),p)+matrix((sqrt(1+(p-1)*r)-sqrt(1-r))/p,nrow=p,ncol=p)
X.raw <- matrix(rnorm(n*p),nrow=n,ncol=p)
X <- scale(X.raw)*sqrt(n/(n-1))
X <- matrix(as.vector(X),n,p)
beta_holder <- c(rnorm(p/10, 1, 0.1), rep(0, p/5), rnorm(p/10, 1, 0.1), rep(0, (3*p)/5))
Y.raw <- drop(X%*%beta_holder+rnorm(n))
Y <- Y.raw-mean(Y.raw)
beta <- rep(1, p)
sigma2 <- 1
bfl_2bg <- bfl(X, Y, beta, sigma2, lambda1 = 1, lambda2 = 1, K = 5000)
```

`bgl` *Bayesian group Lasso*

Description

Inference for group lasso model by two block Gibbs sampling from the Bayesian posterior distribution.

Usage

```
bgl(X, Y, group_size, beta, sigma2, lambda = 1, alpha = 0, xi = 0, K = 10000)
```

Arguments

X	a matrix of predictors, each row corresponds to one observation.
Y	a response vector of length equal to the leading dimension (rows) of X, i.e., length(Y) == nrow(X).
group_size	a vector that contains the number of predictors for each group (see example below).
beta	initial value of the regression coefficients β, its length should equal to ncol(X).
sigma2	initial value of the variance parameter σ^2, which should be positive.
lambda	initial value of the penalty parameter λ, which should be nonnegative.
alpha	initial value of the shape parameter for the inverse gamma prior of σ^2.
xi	initial value of the rate parameter for the inverse gamma prior of σ^2.
K	total number of MCMC samples to be collected.

Details

The sampling algorithm implemented by bgl for the Bayesian group lasso model is described in detail in Jin & Tan (2019). It samples from the joint posterior distribution of (β, σ^2) using a two block Gibbs sampler, which is defined in (3.3).

Value

bgl returns a list object, which contains the components listed below.

betas: a K*nrow(X) matrix of K samples from the (penalized) regression coefficients

sigma2s: a vector of K samples of the variance parameter

Author(s)

Rui Jin

Maintainer: Rui Jin <rui-jin-1@uiowa.edu>

References

Rui Jin and Aixin Tan. "Fast Markov chain Monte Carlo for high dimensional Bayesian regression models with shrinkage priors." (2019). https://arxiv.org/pdf/1903.06964.pdf

Examples

```
require(gglasso)
data(bardet)
group_size <- rep(5, 20)
X.raw <- bardet$x
Y.raw <- bardet$y
n <- dim(X.raw)[1]
p <- dim(X.raw)[2]
X <- scale(X.raw)*sqrt(n/(n-1))
X <- matrix(as.vector(X),n,p)
Y <- Y.raw-mean(Y.raw)
bgl_2bg<- bgl(X = X, Y = Y, group_size = group_size, beta = rep(1,p),
sigma2= 1, lambda = 0.0601 ,K = 5000)
```

`bsgl` *Bayesian sparse group Lasso*

Description

This function perform efficient Bayesian sparse group Lasso baseded on the two block Gibbs sampler.

Usage

```
bsgl(X, Y, group_size, beta, sigma2, lambda1 = 1, lambda2 = 1,
alpha = 0, xi = 0, K = 10000)
```

Arguments

`X`	a matrix of predictors, each row corresponds to one observation.
`Y`	a response vector of length equal to the leading dimension (rows) of `X`, i.e., `length(Y) == nrow(X)`.
`group_size`	a vector that contains the number of predictors for each group (see example below).
`beta`	initial value of the regression coefficients β, its length should equal to `ncol(X)`.
`sigma2`	initial value of the variance parameter σ^2, which should be positive.
`lambda1`	initial value of the penalty parameter λ_1, which should be nonnegative.
`lambda2`	initial value of the penalty parameter λ_2, which should be nonnegative.
`alpha`	initial value of the shape parameter for the inverse gamma prior of σ^2.
`xi`	initial value of the rate parameter for the inverse gamma prior of σ^2.
`K`	total number of MCMC samples to be collected.

Details

The sampling algorithm implemented by `bsgl` for the Bayesian sparse group lasso model is described in detail in Jin & Tan (2019). It samples from the joint posterior distribution of (β, σ^2) using a two block Gibbs sampler, which is defined in (3.5).

Value

`bsgl` returns a list object, which contains the components listed below.

`betas`: a `K*nrow(X) matrix` of `K` samples from the (penalized) regression coefficients

`sigma2s`: a vector of `K` samples of the variance parameter

Author(s)

Rui Jin

Maintainer: Rui Jin <rui-jin-1@uiowa.edu>

References

Rui Jin and Aixin Tan. "Fast Markov chain Monte Carlo for high dimensional Bayesian regression models with shrinkage priors." (2019). https://arxiv.org/pdf/1903.06964.pdf

Examples

```
require(BSGS)
data(Crisis2008BalancedData)
var.names <- colnames(Crisis2008BalancedData)[-1]
country.all <- rownames(Crisis2008BalancedData)
cov.of.interest <- colnames(Crisis2008BalancedData)[-1]
Y <- Crisis2008BalancedData[, 1]
Y <- Y - mean(Y)
X <- Crisis2008BalancedData[, -1]
dummy.variable <- cov.of.interest[lapply(apply(X, 2, unique), length) == 2]
non.dummy.X <- X[, !(colnames(X) %in% dummy.variable)]
X.normalized <- apply(non.dummy.X, 2, function(XX) (XX - mean(XX))/sd(XX))
X[, !(colnames(X) %in% dummy.variable)] <- X.normalized
n <- dim(X)[1]
p <- dim(X)[2]
group_size <- c(10, 3, 4, 2, 4, 11, 4, 1, 12)
bsgl_2bg <- bsgl(X, Y, group_size, rep(1, dim(X)[2]), 1, lambda1 = .104,
lambda2 = .082, K = 5000)
```

Index