Package ‘HDBRR’

April 20, 2021

Type Package

Title High Dimensional Bayesian Ridge Regression without MCMC

Version 0.1.8

Description The svd(singular value decomposition) or qr decomposition was using for the implementation, this avoid the recursion optimizing the time in the compute
<https://drive.google.com/drive/folders/1xJw7gM5_XiJipQ3grTZkfc6q4K0hzuCx?usp=sharing>.

License GPL (>= 2)

Encoding UTF-8

Imports numDeriv, parallel, bigstatsr, MASS

RoxygenNote 7.1.1

NeedsCompilation no

Author Sergio Perez-Elizalde Developer [aut],
Blanca Monroy-Castillo Developer [aut, cre],
Paulino Perez-Rodriguez User [ctb]

Maintainer Blanca Monroy-Castillo Developer <blancamonroy.96@gmail.com>

Repository CRAN

Date/Publication 2021-04-20 07:50:05 UTC

R topics documented:

HDBRR ............................................................... 2
matop ............................................................. 4

Index 6
Description
Ridge regression provides biased estimators of the regression parameters with lower variance. The HDBRR ("High Dimensional Bayesian Ridge Regression") function fits Bayesian Ridge regression without MCMC, this one uses the SVD or QR decomposition for the posterior computation.

Usage
\[
\text{HDBRR}(y, X, a1 = 1, a2 = 1, b1 = 1, b2 = 1/\text{ncol}(X), \\
\quad \text{intercept = TRUE, npts = NULL, c = NULL, corpred = NULL,} \\
\quad \text{method = c("svd", "qr"), bigmat = TRUE, ncores = 2)}
\]

Arguments
- **y**: The data vector (numeric, n) NAs allowed.
- **X**: Design Matrix of dimension \(n \times p\).
- **a1**: \(n\) a priori. \(a1/2\) is the shape parameter of the Gamma Inverse prior assigned to the residual variance. The default value for the \(a1\) parameter is 1.
- **a2**: \(p\) a priori. \(a2/2\) is the shape parameter of the Gamma Inverse prior assigned to the Beta's variance. The default value for the \(a2\) parameter is 1.
- **b1**: \(b1/2\) is the scale parameter of the Gamma Inverse prior assigned to the residual variance. The default value for the \(b1\) parameter is 1.
- **b2**: \(b2/2\) is the scale parameter of the Gamma Inverse prior assigned to the Beta's variance. The default value for the \(b2\) parameter is \(1/\text{ncol}(X)\). Since the beta's variance is small, in this case \(\text{ncol}(X) = p\), if \(p\) is large then \(b2\) is small.
- **intercept**: Logic value. The default value for the intercept is TRUE.
- **npts**: Number (integer) of points used to evaluate the \(u\)'s density for the numeric approach. The default value for the \(npts\) parameter is 200.
- **c**: Auxiliary constant for the prior density for Beta in the variable selection, in this case the prior density is a mixture of a normal distribution.
- **corpred**: The method for the compute of the correlation, there are two methods, Empirical Basis ("eb") and Bayesian ("b") method. The default value for the parameter corpred is NULL. If the values is NULL then the corr and edf values will be NULL.
- **method**: Options for the posterior computation. There are two methods available: "qr" and "svd" decomposition. The default value for the method is SVD decomposition.
- **bigmat**: Use of the bigstatsr package. The default value for bigmat is TRUE.
- **ncores**: Number of the cores for the auxiliary functions. The default value for the ncores is 2, you can detect your number of cores with detectCores() and use it.
Details

Ridge regression is a useful tool to deal with colinearity in the homocesastic linear regression model providing biased estimators of the regression parameters with lower variance than the least square estimators. The model

\[ y = X\beta + \epsilon \]

where \( \epsilon \) vector is assumed Normal with mean vector 0 and covariance matrix \( \sigma^2 I_n \). See more
https://drive.google.com/drive/folders/1xJw7gM5_xiJipQ3grTZkfc6q4K0hzuCx?usp=sharing

Value

List containing the following components:

- `betahat`: Vector (numeric, \( p \)) with the betas estimates.
- `yhat`: Vector (numeric, \( n \)) with the y's estimates.
- `varb`: Vector (numeric, \( p \)) with the beta's variance.
- `sigsqhat`: Value (numeric) of the residual variance estimate.
- `sigbsqhat`: Value (numeric) of the Beta's variance estimate.
- `u`: Vector (numeric, \( npts \)) with the u’s values.
- `postu`: Vector (numeric, \( npts \)) with the values of the u posterior.
- `uhat`: Value (numeric) of u estimated.
- `umode`: Value (numeric) of the posterior mode of u.
- `whichNa`: Value (integer) of NAs in the y vector.
- `phat`: Vector (numeric, \( p \)), selection probability of \( x_i \).
- `delta`: Used in the variable selection.
- `edf`: Value (numeric) of the effective degrees of freedom for regression.
- `corr`: Vector (numeric, \( n \)) of the correlation between \( y_{-i} \) estimates and \( y_{-i} \).

Author(s)

Sergio Perez-Elizalde, Blanca E. Monroy-Castillo, Paulino Perez-Rodriguez.

Examples

```r
n <- 30
p <- 100
X <- matrix(rnorm(n*(p-1),1,1/p),nrow = n,ncol = p-1)
Beta <- sample(1:p,p-1,rep = FALSE)
Beta <- c(1,Beta)
y <- cbind(rep(1,n),X) %*% Beta+rnorm(n,0,1)
fit <- HDBRR(y,X/sqrt(p), intercept = TRUE, c = 100, cor = "b")
```
Description

Compute the SVD or QR decomposition of the matrix X.

Usage

matop(y, X, method = c("svd", "qr"), bigmat = TRUE)

Arguments

y The data vector (numeric, n) NAs allowed.
X Design Matrix of dimension n x p.
method Options for the posterior computation. Two methods, "qr" and "svd" decomposition. The default value for the method is SVD decomposition.
bigmat Use of the bigstatsr package. The default value for bigmat is TRUE.

Details

Use the bigstatsr package when p >> n. Auxiliary in the HDBRR function.

Value

If the method used is svd then the list containing the following components:

y The data vector (numeric, n) NAs allowed.
X Design Matrix of dimension n x p.
D A vector containing the singular values of X, of length min(n, p).
L A matrix whose columns contain the left singular vectors of X.
R A matrix whose columns contain the right singular vectors of X.
ev A vector containing the square of D.
Ly The cross-product between the matrix L and vector y.
n Number of rows of X.
p Number of columns of X.

If the method used is qr then the list containing the following components:

y The data vector (numeric, n) NAs allowed.
X Design Matrix of dimension n x p.
R An upper triangular matrix of dimension n x p.
n Number of rows of X.
p Number of columns of X.
Author(s)
Sergio Perez-Elizalde, Blanca E. Monroy-Castillo, Paulino Perez-Rodriguez.

See Also
qr, svd

Examples
n <- 30
p <- 100
X <- matrix(rnorm(n*(p-1),1,1/p),nrow = n,ncol = p-1)
Beta <- sample(1:p,p-1,rep = FALSE)
Beta <- c(1,Beta)
y <- cbind(rep(1,n),X) %*% Beta+rnorm(n,0,1)
matop(y, X, bigmat = TRUE)
Index

* BRR without MCMC.
  HDBRR, 2
* BRR.
  HDBRR, 2

HDBRR, 2
matop, 4
qr, 5
svd, 5