Package 'HDBRR'

October 12, 2022

Type Package

Title High Dimensional Bayesian Ridge Regression without MCMC

Version 1.1.4

Description Ridge regression provide 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.

License GPL (>= 2)

Depends R (>= 3.0.0)

Encoding UTF-8

Imports numDeriv, parallel, bigstatsr, MASS, graphics

RoxygenNote 7.2.1

NeedsCompilation no

LazyData true

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

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

Repository CRAN

Date/Publication 2022-10-05 23:10:08 UTC

R topics documented:

HDBRR	*High Dimensional Bayesian Ridge Regression without MCMC.*

Description

Ridge regression provide 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

```
HDBRR(y, X, n0 = 5, p0 = 5, s20 = NULL, d20 = NULL, h = 0.5,
    intercept = TRUE, vpapp = TRUE,npts = NULL,c = NULL,
    corpred = NULL, method = c("svd","qr"),bigmat = TRUE, ncores = 2, svdx = NULL)

## S3 method for class 'HDBRR'
summary(object, all.coef = FALSE, crit = log(4), ...)

## S3 method for class 'HDBRR'
plot(x, crit = log(4), var_select = FALSE, post = FALSE, ...)

## S3 method for class 'HDBRR'
predict(object,  ...)

## S3 method for class 'summary.HDBRR'
print(x, ...)

## S3 method for class 'HDBRR'
print(x, ...)

## S3 method for class 'HDBRR'
coef(object, all = FALSE, ...)
```

Arguments

y	The data vector (numeric, n) NAs allowed.
X	Design Matrix of dimension n x p.
n0,p0	n0/2 and p0/2 are the shape parameter of the Gamma Inverse prior assigned to the residual variance and the shape parameter of the Gamma Inverse prior assigned to the Beta's variance respectively. The default value for n0/2 and p0/2 parameter is 5.
s20,d20	(n0s20)/2 and (p0d20)/2 are the scale parameter of the Gamma Inverse prior assigned to the residual variance and the scale parameter of the Gamma Inverse prior assigned to the Beta's variance respectively. The default value for the s20 and d20 is NULL. If the scale is not specified a value is calculated with h and quantiles.

h	(numeric, 0<h<1) shrinkage factor. Only used if the hyper-parameters are not specified; If h -> 0 then we have greater shrinkage, this is, β -> 0. If h -> 1 then we have less shrinkage.
intercept	Logic value. The default value for the intercept is TRUE.
vpapp	Logic value. Compute an approximation of the predictive variance. The default value for the vpapp is TRUE.
npts	Number (integer) of points used to evaluate the u's density for the numeric aprroach. The default value for the npts parameter is 200.
c	ratio of Gaussian densities (Spike/Slab) in the prior mixture density of each Beta for variable selection.
corpred	The method for the compute of the correlation, there are two methods, Empirical Bayes ("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" decomposition of X*t(X) and "svd" decomposition of matrix X. 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 computation. The default value for the ncores is 2, you can detect your number of cores with detectCores() and use it (iOS and Linux).
object	A HDBRR object, typically generated by a call to HDBRR.
all.coef	Logical. Should results be returned for all ridge regression penalty parameters (all.coef = TRUE), or only those whose log(bayes factor)>crit.
crit	Numerical. The lower bound of the log Bayes factor in favour to include a variable in the model. The default value for crit is log(4).
...	Additional arguments to be passed to or from other methods.
x	A HDBRR object, typically generated by a call to HDBRR (for the print.HDBRR and plot.HDBRR functions) or an object of class summary.HDBRR (for the print.summary.HDBRR function).
var_select	Logical. If is TRUE a plot with variable selection is returned. The default value is FALSE.
post	Logical. If is TRUE a plot with marginal posterior of u is returned. The default value is FALSE.
all	Logical. All coefficients are returned. If is FALSE, then, if p > 250 only 250 coefficients are returned. The default value es FALSE.
svdx	It is possible to add the svd. The default value es NULL.

Details

Ridge regression is a useful tool to deal with colinerity 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 ϵ vector is assumed Normal with mean vector 0 and covariance matrix $\sigma^2 I_n$. For further details see the vignettes in the package.

Value

List containing the following components:

betahat	Vector (numeric, p) with the betas estimates.
yhat	Vector (numeric, n) with the y's estimates.
sdyhat	Vector (numeric, n) with the standard deviation of the predicts values.
sdpred	Vector (numeric, n) with the standard deviation of predict variances.
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.
svdx	The svd decomposition.

Author(s)

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

Examples

```
## Not run:

data("phenowheat")
mod <- lmer(pheno$HD~pheno$env+(1|pheno$Line))
y <- unlist(ranef(mod))
n <- length(y)
X <- scale(X, scale=F)
fitall <- HDBRR(y,X/sqrt(ncol(X)),intercept = FALSE, corpred = "eb", c = 100)
fitall
sumarry(fitall, crit = 0)
plot(fitall, crit = 0)
predict(fitall)


## End(Not run)
```

matop	*matop*

Description

Compute the SVD or QR decomposition of the matrix X.

Usage

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

Arguments

y	The data vector (numeric, n) NAs allowed. The default value is NULL, It is possible to compute the SVD or QR decomposition without y.
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 descomposition.
bigmat	Use of the bigstatsr package. The default value for bigmat is `TRUE`.

Details

Use the bigstartsr 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 lenght `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)
```

---

pheno   *Durum Wheat*

Description

The final number of SNPs included in the NCCR linkage map was 7594. The markers were centered and standardized. Phenotypic evaluation of the NCCR population was performed during two growing seasons (2010-2011 and 2011-2012) in locations in the Po Valley representative of the target environments where durum wheat is grown: Cadriano in the 2010-2011 growin season (Cad11) and the 2011-2012 growing season (Cad12); Poggio Renatico in the 2010-2011 growing season (Pr11) and Argelato in the 2011-2012 growing season (Arg12).

Source

International Maize and Wheat Improvement Center (CIMMYT), Mexico.

---

phenowheat   *Durum wheat dataset*

Description

This contain data from a multiparental durum wheat (Triticum turgidum L. spp. duram) trial consisting of 334 lines evaluated in a country-year combination. This population is characterized for Grain Yield (GY), grain volume weight (GVW), 1000-kernel weight (GWT) and heading date (HD) in the four environments. For further details see the vignettes in the package.

Usage

```
data(phenowheat)
```

Format

1. phenoMatrix phenowheat.pheno contains the phenotypic data.
2. XThe matrix phenowheat.X contains the Genotypic data.

X *Durum Wheat X*

Description

Is a matrix (338 x 7594) with A balanced, four-way multiparental cross population was developed from four elite durum wheat cultivars (Neodur, Claudio, Colosseo, and Rascon/Tarro) that were chosen as diverse contributors of different alleles of agronomic relevance.

Source

International Maize and Wheat Improvement Center (CIMMYT), Mexico.

Index