

Package ‘BGGE’

May 10, 2018

Title Bayesian Genomic Linear Models Applied to GE Genome Selection

Version 0.6.2

Date 2018-05-10

Description Application of genome prediction for a continuous variable, focused on genotype by environment (GE) genomic selection models (GS). It consists a group of functions that help to create regression kernels for some GE genomic models proposed by Jarquín et al. (2014) <doi:10.1007/s00122-013-2243-1> and Lopez-Cruz et al. (2015) <doi:10.1534/g3.114.016097>. Also, it computes genomic predictions based on Bayesian approaches.

The prediction function uses an orthogonal transformation of the data and specific priors present by Cuevas et al. (2014) <doi:10.1534/g3.114.013094>.

Depends R (>= 3.1.1)

Imports stats,

Suggests BGLR, coda

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Author Italo Granato [aut, cre],
Luna-Vázquez Francisco J. [aut],
Cuevas Jaime [aut]

Maintainer Italo Granato <italo.granato@gmail.com>

Repository CRAN

Date/Publication 2018-05-10 16:55:01 UTC

R topics documented:

BGGE	2
getK	4
Index	7

 BGGE

Genotype x Environment models using regression kernel

Description

BGGE function fits Bayesian regression for continuous observations through regression kernels

Usage

```
BGGE(y, K, XF = NULL, ne, ite = 1000, burn = 200, thin = 3, verbose = FALSE, tol = 1e-10)
```

Arguments

y	Vector of data. Should be numeric and NAs are allowed.
K	list A two-level list Specify the regression kernels (co-variance matrix). The former is the Kernel, where is included the regression kernels. The later is the Type, specifying if the matrix is either D Dense or BD Block Diagonal. A number of regression kernels or random effects to be fitted are specified in this list.
XF	matrix Design matrix ($n \times p$) for fixed effects
ne	vector Number of genotypes by environment.
ite	numeric Number of iterations.
burn	numeric Number of iterations to be discarded as burn-in.
thin	numeric Thinin interval.
verbose	Should iteration history be printed on console? If TRUE or 1 then it is printed, otherwise, if another number \$n\$ is chosen the history is printed every \$n\$ times. The default is FALSE.
tol	a numeric tolerance level. Eigenvalues lower than tol are discarded. Default is 1e-10.

Details

The goal is to fit genomic prediction models for continuous outcomes through Gibbs sampler. BGGE uses a proposal for dimension reduction through an orthogonal transformation of observed data (y) as well as differential shrinkage because of the prior variance assigned to regression parameters. Further details on this approach can be found in Cuevas et al. (2014). The primary genetic model is

$$y = g + e$$

where y is the response, g is the unknown random effect and e is the residual effect. You can specify a number of random effects g , as many as desired, through a list of regression kernels related to each random effect in the argument K. The structure of K is a two level list, where the first element on the second level is the Kernel and the second element is a definition of type of matrix. There are two definitions, either matrix is D (dense) or BD (Block Diagonal). As we make the spectral decomposition on the kernels, for block diagonal matrices, we take advantage of its structure and make decomposition on the submatrices instead of one big matrix. For example, the

regression kernels should be a structure like $K = \text{list}(\text{list}(\text{Kernel} = G, \text{Type} = "D"), \text{list}(\text{Kernel} = G, \text{Type} = "BD"))$. The definition of one matrix as a block diagonal must be followed by the number of subjects in each submatrix in the block diagonal, present in the `ne`, which allows sub matrices to be drawn. Some genotype by environment models has the block diagonal matrix type or similar. The genotype x environment deviation matrix in MDs model (Sousa et al., 2017) has the structure of block diagonal. Also, the matrices for environment-specific variance in MDe models (Sousa et al., 2017) if summed, can form a structure of block diagonal, where is possible to extract sub matrices for each environment. In the case of all kernel be of the dense type, `ne` is ignored.

Value

A list with estimated posterior means of residual and genetic variance component for each term in the linear model and the genetic value predicted. Also the values along with the chains are released.

References

Cuevas, J., Perez-Elizalde, S., Soberanis, V., Perez-Rodriguez, P., Gianola, D., & Crossa, J. (2014). Bayesian genomic-enabled prediction as an inverse problem. *G3: Genes, Genomes, Genetics*, 4(10), 1991-2001.

Sousa, M. B., Cuevas, J., Oliveira, E. G. C., Perez-Rodriguez, P., Jarquin, D., Fritsche-Neto, R., Burgueno, J. & Crossa, J. (2017). Genomic-enabled prediction in maize using kernel models with genotype x environment interaction. *G3: Genes, Genomes, Genetics*, 7(6), 1995-2014.

See Also

[BGLR](#)

Examples

```
# multi-environment main genotypic model
library(BGLR)
data(wheat)
X<-wheat.X[1:200,1:600] # Subset of 200 subjects and 600 markers
rownames(X) <- 1:200
Y<-wheat.Y[1:200,]
A<-wheat.A[1:200,1:200] # Pedigree

GB<-tcrossprod(X)/ncol(X)
K<-list(G = list(Kernel = GB, Type = "D"))
y<-Y[,1]
fit<-BGGE(y = y, K = K, ne = length(y), ite = 300, burn = 100, thin = 2)

# multi-environment main genotypic model
Env <- as.factor(c(2,3)) #subset of 2 environments
pheno_gen0 <- data.frame(env = gl(n = 2, k = nrow(Y), labels = Env),
                        GID = gl(n = nrow(Y), k = 1, length = nrow(Y) * length(Env)),
                        value = as.vector(Y[,2:3]))

K <- getK(Y = pheno_gen0, X = X, kernel = "GB", model = "MM")
y <- pheno_gen0[,3]
fit <- BGGE(y = y, K = K, ne = rep(nrow(Y), length(Env)), ite = 300, burn = 100, thin = 1)
```

getK

Kernel matrix for GE genomic selection models

Description

Create kernel matrix for GE genomic prediction models

Usage

```
getK(Y, X, kernel = c("GK", "GB"), setKernel = NULL, bandwidth = 1,
      model = c("SM", "MM", "MDs", "MDe"), quantil = 0.5,
      intercept.random = FALSE)
```

Arguments

Y	data.frame Phenotypic data with three columns. The first column is a factor for environments, the second column is a factor identifying genotypes, and the third column contains the trait of interest
X	Marker matrix with individuals in rows and markers in columns. Missing markers are not allowed.
kernel	Kernel to be created internally. Methods currently implemented are the Gaussian GK and the linear GBLUP kernel
setKernel	matrix Single kernel matrix in case it is necessary to use a different kernel from GK or GBLUP
bandwidth	vector Bandwidth parameter to create the Gaussian Kernel (GK) matrix. The default for the bandwidth is 1. Estimation of this parameter can be made using a Bayesian approach as presented in Perez-Elizalde et al. (2015)
model	Specifies the genotype \times environment model to be fitted. It currently supported the models SM, MM, MDs and MDe. See Details
quantil	Specifies the quantile to create the Gaussian kernel.
intercept.random	if TRUE, kernel related to random intercept of genotype is included.

Details

The aim is to create kernels to fit GE interaction models applied to genomic prediction. Two standard genomic kernels are currently supported: GB creates a linear kernel resulted from the cross-product of centered and standardized marker genotypes divide by the number of markers p :

$$GB = \frac{XX^T}{p}$$

Another alternative is the Gaussian Kernel GK, resulted from:

$$GK(x_i, x_{i'}) = \exp\left(\frac{-hd_{ii'}^2}{q(d)}\right)$$

where $d_{ii'}^2$ is the genetic distance between individuals based on markers scaled by some percentile $q(d)$ and $bandwidth$ is the bandwidth parameter. However, other kernels can be provided through `setKernel`. In this case, arguments λ , `kernel` and `h` are ignored.

Currently, the supported models for GE kernels are:

- SM: is the single-environment main genotypic effect model - It fits the data for a single environment, and only one kernel is produced.
- MM: is the multi-environment main genotypic effect model - It considers the main random genetic effects across environments. Thus, just one kernel is produced, of order $n \times n$, related to the main effect across environments.
- MDs: is the multi-environment single variance genotype x environment deviation model - It is an extension of MM by adding the random interaction effect of environments with genotype information. Thus, two kernels are created, one related to the main effect across environment, and the second is associated with single genotype by environment effect.
- MDe: is the multi-environment, environment-specific variance genotype x environment deviation model - It separates the genetic effects into the main genetic effects and the specific genetic effects (for each environment). Thus, one kernel for across environments effect and j kernels are created, one for each environment.

These GE genomic models were compared and named by Sousa et al. (2017) and can be increased by using the kernel related to random intercept of genotype through `intercept.random`.

Value

This function returns a two-level list, which specifies the kernel and the type of matrix. The latter is a classification according to its structure, i. e., if the matrix is dense or a block diagonal. For the main effect (G), the matrix is classified as dense (D). On the other hand, matrices for environment-specific and genotype by environment effect (GE) are considered diagonal block (BD). This classification is used as part of the prediction through the BGGE function.

References

- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño, and G. de los Campos. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595-607.
- Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.-L. Jannink, R.P. Singh, E. Autrique, and G. de los Campos. 2015. Increased prediction accuracy in wheat breeding trials using a marker x environment interaction genomic selection model. *G3: Genes, Genomes, Genetics.* 5(4): 569-82.
- Perez- Elizalde, S. J. Cuevas, P. Perez-Rodriguez, and J. Crossa. 2015. Selection of the Bandwidth Parameter in a Bayesian Kernel Regression Model for Genomic-Enabled Prediction. *Journal of Agricultural, Biological, and Environmental Statistics (JABES)*, 20(4):512-532.

Sousa, M. B., Cuevas, J., Oliveira, E. G. C., Perez-Rodriguez, P., Jarquin, D., Fritsche-Neto, R., Burgueno, J. & Crossa, J. (2017). Genomic-enabled prediction in maize using kernel models with genotype x environment interaction. *G3: Genes, Genomes, Genetics*, 7(6), 1995-2014.

Examples

```
# create kernel matrix for model MDs using wheat dataset
library(BGLR)

data(wheat)
X <- scale(wheat.X, scale = TRUE, center = TRUE)
rownames(X) <- 1:599
pheno_genos <- data.frame(env = gl(n = 4, k = 599),
                          GID = gl(n=599, k=1, length = 599*4),
                          value = as.vector(wheat.Y))

K <- getK(Y = pheno_genos, X = X, kernel = "GB", model = "MDs")
```

Index

BGGE, 2

BGLR, 3

getK, 4