

# Package ‘ASGS’

August 30, 2021

**Type** Package

**Title** Algorithm Sets for Genomic Selection

**Version** 1.0.4

**Maintainer** Kuiqin Wang <wkq@stu.ouc.edu.cn>

**Description** Sets of algorithms that can be used for GS (Genomic Selection), including machine learning models and tradition models like Bayes.

For machine learning methods, ANN (Artificial Neural Network), RKHS (Reproducing kernel Hilbert space), RF(Random Forest),

GBM(Generalized Boosted Regression Models), SVM(support vector machine) is included. For traditional methods, BayesA, BayesB,

BayesC, BayesLasso, BayesRR (Bayesian Ridge Regression), GBLUP(Genomic Best Linear Unbiased Prediction),

rrBLUP(ridge regression best linear unbiased prediction) is included. Simple validation is implemented to get the prediction accuracy,

In order to get a better performance for machine learning methods, an auto hyperparameter adjusting is implemented.

**License** GPL-3

**Encoding** UTF-8

**Imports** brnn, kernlab, BGLR, randomForest, gbm, rrBLUP

**LazyData** true

**RoxygenNote** 7.1.1

**Depends** R (>= 3.5.0)

## R topics documented:

ANN . . . . .	2
BA . . . . .	3
BB . . . . .	3
BC . . . . .	4
BL . . . . .	5
BRR . . . . .	6
GBLUP . . . . .	7
GBM . . . . .	7
RF . . . . .	8
RKHS . . . . .	9
RRBLUP . . . . .	10
SVM . . . . .	11
trout . . . . .	12

**Index****13**


---

ANN	ANN
-----	-----

---

**Description**

Calculates the Genomic Estimated Breeding Value based on ANN method.

**Usage**

```
ANN(X, Y, r)
```

**Arguments**

X	X is a matrix of marker genotype of size $n \times p$ where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size $n \times 1$ .
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define $r=0.1$ ).

**Details**

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 10 times and accuracy is averaged. For machine learning methods, an auto hyperparameter adjusting through grid search is implemented to achieve better prediction accuracy. In each split, the 0.9 of the training population is used to adjust hyperparameters, of which 0.9 selected to be training sets and 0.1 to be testing sets, 10 times of simple validation is implemented.

**Value**

\$fit ANN model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

**Examples**

```
library(ASGS);rm(list=ls());data(trout);
X=trout[1:50,2:11];Y=as.data.frame(trout[1:50,1]);r=0.1;ASGS::ANN(X,Y,r)
```

BA

*BayesA***Description**

Calculates the Genomic Estimated Breeding Value based on BayesA method.

**Usage**

```
BA(X, Y, r)
```

**Arguments**

X	X is a matrix of marker genotype of size n×p where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size n×1.
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define r=0.1).

**Details**

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged.

**Value**

\$fit BayesA model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

**Examples**

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:20,2:11];Y=as.data.frame(trout[1:20,1]);r=0.1;ASGS::BA(X,Y,r)
```

BB

*BayesB***Description**

Calculates the Genomic Estimated Breeding Value based on BayesB method.

**Usage**

```
BB(X, Y, r)
```

### Arguments

X	X is a matrix of marker genotype of size $n \times p$ where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size $n \times 1$ .
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define $r=0.1$ ).

### Details

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged.

### Value

\$fit BayesB model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

### Examples

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:20,2:11];Y=as.data.frame(trout[1:20,1]);r=0.1;ASGS::BB(X,Y,r)
```

---

BC	<i>BayesC</i>
----	---------------

---

### Description

Calculates the Genomic Estimated Breeding Value based on BayesC method.

### Usage

```
BC(X, Y, r)
```

### Arguments

X	X is a matrix of marker genotype of size $n \times p$ where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size $n \times 1$ .
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define $r=0.1$ ).

## Details

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged.

## Value

\$fit BayesC model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

## Examples

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:20,2:11];Y=as.data.frame(trout[1:20,1]);r=0.1;ASGS::BC(X,Y,r)
```

---

BL	<i>BayesLasso</i>
----	-------------------

---

## Description

Calculates the Genomic Estimated Breeding Value based on BayesLasso method.

## Usage

```
BL(X, Y, r)
```

## Arguments

X	X is a matrix of marker genotype of size n×p where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size n×1.
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define r=0.1).

## Details

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged.

**Value**

\$fit BayesL model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

**Examples**

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:30,2:11];Y=as.data.frame(trout[1:30,1]);r=0.1;ASGS::BL(X,Y,r)
```

---

BRR	<i>BayesRR</i>
-----	----------------

---

**Description**

Calculates the Genomic Estimated Breeding Value based on BayesRR method.

**Usage**

```
BRR(X, Y, r)
```

**Arguments**

X	X is a matrix of marker genotype of size n×p where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size n×1.
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define r=0.1).

**Details**

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged.

**Value**

\$fit BayesRR model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

**Examples**

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:20,2:11];Y=as.data.frame(trout[1:20,1]);r=0.1;ASGS::BRR(X,Y,r)
```

GBLUP

*GBLUP***Description**

Calculates the Genomic Estimated Breeding Value based on GBLUP method.

**Usage**

```
GBLUP (X, Y, r)
```

**Arguments**

X	X is a matrix of marker genotype of size n×p where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size n×1.
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define r=0.1).

**Details**

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged.

**Value**

\$fit GBLUP model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

**Examples**

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:20,2:9];Y=as.data.frame(trout[1:20,1]);r=0.1;ASGS::GBLUP(X,Y,r)
```

GBM

*GBM***Description**

Calculates the Genomic Estimated Breeding Value based on GBM method.

**Usage**

```
GBM(X, Y, r)
```

**Arguments**

X	X is a matrix of marker genotype of size $n \times p$ where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size $n \times 1$ .
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define $r=0.1$ ).

**Details**

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 10 times and accuracy is averaged. For machine learning methods, an auto hyperparameter adjusting through grid search is implemented to achieve better prediction accuracy. In each split, the 0.9 of the training population is used to adjust hyperparameters, of which 0.9 selected to be training sets and 0.1 to be testing sets, 10 times of simple validation is implemented.

**Value**

\$fit GBM model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

**Examples**

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[,2:6];Y=as.data.frame(trout[,1]);r=0.1;ASGS::GBM(X,Y,r)
```

---

RF	<i>RF</i>
----	-----------

---

**Description**

Calculates the Genomic Estimated Breeding Value based on RF method.

**Usage**

```
RF(X, Y, r)
```

**Arguments**

X	X is a matrix of marker genotype of size $n \times p$ where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size $n \times 1$ .
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define $r=0.1$ ).



## Details

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged. For machine learning methods, an auto hyperparameter adjusting through grid search is implemented to achieve better prediction accuracy. In each split, the 0.9 of the training population is used to adjust hyperparameters, of which 0.9 selected to be training sets and 0.1 to be testing sets, 10 times of simple validation is implemented.

## Value

\$fit RF model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

## Examples

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:50,2:11];Y=as.data.frame(trout[1:50,1]);r=0.1;ASGS::RF(X,Y,r)
```

---

RKHS	<i>RKHS</i>
------	-------------

---

## Description

Calculates the Genomic Estimated Breeding Value based on RKHS method.

## Usage

```
RKHS(X, Y, r)
```

## Arguments

X	X is a matrix of marker genotype of size n×p where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size n×1.
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define r=0.1).

## Details

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged. For machine learning methods, an auto hyperparameter adjusting through grid search is implemented to achieve better prediction accuracy. In each split, the 0.9 of the training population is used to adjust hyperparameters, of which 0.9 selected to be training sets and 0.1 to be testing sets, 10 times of simple validation is implemented.

**Value**

\$fit RKHS model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

**Examples**

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:20,2:9];Y=as.data.frame(trout[1:20,1]);r=0.1;ASGS::RKHS(X,Y,r)
```

---

RRBLUP	<i>rrBLUP</i>
--------	---------------

---

**Description**

Calculates the Genomic Estimated Breeding Value based on rrBLUP method.

**Usage**

```
RRBLUP(X, Y, r)
```

**Arguments**

X	X is a matrix of marker genotype of size n×p where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size n×1.
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define r=0.1).

**Details**

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged.

**Value**

\$Vu variance of random effect \$Ve error variance \$beta estimate of fixed effects \$u estimate of random effects \$LL maximized log likelihood \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

**Examples**

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[,2:101];Y=as.data.frame(trout[,1]);r=0.1;ASGS::RRBLUP(X,Y,r)
```

---

SVM	<i>SVM</i>
-----	------------

---

### Description

Calculates the Genomic Estimated Breeding Value based on SVM method.

### Usage

```
SVM(X, Y, r)
```

### Arguments

X	X is a matrix of marker genotype of size n×p where n are no of Individuals under study (i.e. genotype, lines) and p are number of markers.
Y	Y is a vector of individuals of size n×1.
r	fraction of testing data (ranges from (0 to 1)) used during model fitting (suppose if one want to use 0.9 of data for model training and remaining 0.1 for model testing so one has to define r=0.1).

### Details

This function fits model by dividing data into two part i.e. training population and testing population. Former one is used to build the models and later one for performance evaluation. The performance of model is evaluated by calculating model's prediction accuracy of both training and testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value). Whole procedures is repeated 50 times and accuracy is averaged. For machine learning methods, an auto hyperparameter adjusting through grid search is implemented to achieve better prediction accuracy. In each split, the 0.9 of the training population is used to adjust hyperparameters, of which 0.9 selected to be training sets and 0.1 to be testing sets, 10 times of simple validation is implemented.

### Value

\$fit SVM model fitting \$Accuracy prediction accuracy of the testing population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value) \$Train accuracy prediction accuracy of the training population (pearson correlation coefficient between actual phenotypic value and predicted phenotypic value)

### Examples

```
library(ASGS);rm (list=ls ());data(trout);
X=trout[1:50,2:11];Y=as.data.frame(trout[1:50,1]);r=0.1;ASGS::SVM(X,Y,r)
```

---

trout	<i>rainbow trout dataset</i>
-------	------------------------------

---

**Description**

Dataset used in this study is subset from rainbow trout. In our sample dataset it has one response for 100 lines genotyped for 500 markers.

**Usage**

```
data(trout)
```

**Format**

A data frame with 100 rows as genotypes with 501 columns (First column contains information of phenotypic trait under study and the rest columns represent information of markers).

**Details**

This dataset come from a genome Wide association study for resistance to *Piscirickettsia salmonis* in rainbow trout, with 2047 individual genotyped and 26068 markers that can be used for GS. We selected the endweight trait as the response.

**References**

Agustin Barria, Rodrigo Marín Nahuelpi, Pablo Cáceres, María E López, Liane N Bassini, Jean P Lhorente, José M Yáñez, Single Step Genome Wide Association Study for Resistance to *Piscirickettsia salmonis* in Rainbow Trout (*Oncorhynchus mykiss*), G3 Genes/Genomes/Genetics, Volume 9, Issue 11, 1 November 2019, Pages 3833 to 3841, <https://doi.org/10.1534/g3.119.400204>

**Examples**

```
library(ASGS);rm (list=ls ());data(trout);X=trout[,2:101];Y=as.data.frame(trout[,1]);
```

# Index

ANN, [2](#)

BA, [3](#)

BB, [3](#)

BC, [4](#)

BL, [5](#)

BRR, [6](#)

GBLUP, [7](#)

GBM, [7](#)

RF, [8](#)

RKHS, [9](#)

RRBLUP, [10](#)

SVM, [11](#)

trout, [12](#)