Package 'AVGAS'

January 20, 2025

Title A Variable Selection using Genetic Algorithms

Version 0.1.0

Description We provide a stage-wise selection method using genetic algorithm which can perform fast interaction selection in high-dimensional linear regression models with two-way interaction effects under strong, weak, or no heredity condition. Ye, C.,and Yang,Y. (2019) <doi:10.1109/TIT.2019.2913417>.

License GPL-2

Encoding UTF-8

Imports utils, Matrix, pracma, stats, dplyr, selectiveInference, VariableScreening, ggplot2

Language en-US

Author Leiyue Li [aut, cre],
Chenglong Ye [aut]

Maintainer Leiyue Li <lli289.git@gmail.com>

RoxygenNote 7.2.3

NeedsCompilation no

Repository CRAN

Date/Publication 2023-12-20 16:20:08 UTC

Contents

ABC   *Evaluating ABC for each fitted model*

Description

This function evaluates ABC score for fitted model, one model at a time. For a model I, the ABC is defined as

$$ABC(I) = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i^I \right)^2 + 2r_I \sigma^2 + \lambda \sigma^2 C_I.$$

When comparing ABC of fitted models to the same data set, the smaller the ABC, the better fit.

Usage

```
ABC(
  X,
  y,
  heredity = "Strong",
  nmain.p,
  sigma = NULL,
  extract = "No",
  varind = NULL,
  interaction.ind = NULL,
  pi1 = 0.32,
  pi2 = 0.32,
  pi3 = 0.32,
  lambda = 10
)
```

Arguments

X	Input data. An optional data frame, or numeric matrix of dimension n by nmain.p. Note that the two-way interaction effects should not be included in X because this function automatically generates the corresponding two-way interaction effects if needed.
y	Response variable. A n-dimensional vector, where n is the number of observations in X.
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
nmain.p	A numeric value that represents the total number of main effects in X.
sigma	The standard deviation of the noise term. In practice, sigma is usually unknown. In such case, this function automatically estimate sigma using root mean square error (RMSE). Default is NULL. Otherwise, users need to enter a numeric value.
extract	A either "Yes" or "No" logical vector that represents whether or not to extract specific columns from X. Default is "No".

varind	Only used when extract = "Yes". A numeric vector of class c() that specifies the indices of variables to be extracted from X. If varind contains indices of two-way interaction effects, then this function automatically generates corresponding two-way interaction effects from X.
interaction.ind	
	Only used when extract = "Yes". A two-column numeric matrix containing all possible two-way interaction effects. It must be generated outside of this function using t(utils::combn()) or indchunked(). See Example section for details.
pi1	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to the Details section.
pi2	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to the Details section.
pi3	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to the Details section.
lambda	A numeric value defined by users. Default is 10. For guidance on selecting an appropriate value, please refer to the Details section.

Details

- For inputs pi1, pi2, and pi3, the number needs to satisfy the condition: $\pi_1 + \pi_2 + \pi_3 = 1 - \pi_0$ where π_0 is a numeric value between 0 and 1, the smaller the better.
- For input lambda, the number needs to satisfy the condition: $\lambda \geq 5.1/log(2)$.

Value

A numeric value is returned. It represents the ABC score of the fitted model.

References

Ye, C. and Yang, Y., 2019. *High-dimensional adaptive minimax sparse estimation with interactions.*

See Also

[Extract](#), [initial](#).

Examples

```
# sigma is unknown
set.seed(0)
nmain.p <- 4
interaction.ind <- t(combn(4,2))
X <- matrix(rnorm(50*4,1,0.1), 50, 4)
epl <- rnorm(50,0,0.01)
y<- 1+X[,1]+X[,2]+X[,3]+X[,4]+epl
ABC(X, y, nmain.p = 4, interaction.ind = interaction.ind)
ABC(X, y, nmain.p = 4, extract = "Yes",
    varind = c(1,2,5), interaction.ind = interaction.ind)
#'
```

```
# users want to enter a suggested value for sigma

# model with only one predictor
try(ABC(X, y, nmain.p = 4, extract = "Yes",
  varind = 1, interaction.ind = interaction.ind)) # warning message
```

AVGAS   *A Variable selection using Genetic AlgorithmS*

Description

A Variable selection using Genetic AlgorithmS

Usage

```
AVGAS(
  X,
  y,
  heredity = "Strong",
  nmain.p,
  r1,
  r2,
  sigma = NULL,
  interaction.ind = NULL,
  lambda = 10,
  q = 40,
  allout = "No",
  interonly = "No",
  pi1 = 0.32,
  pi2 = 0.32,
  pi3 = 0.32,
  aprob = 0.9,
  dprob = 0.9,
  aprobm = 0.1,
  aprobi = 0.9,
  dprobm = 0.9,
  dprobi = 0.1,
  take = 3
)
```

Arguments

X   Input data. An optional data frame, or numeric matrix of dimension n by nmain.p.
                Note that the two-way interaction effects should not be included in X because this
                function automatically generates the corresponding two-way interaction effects
                if needed.

y   Response variable. A n-dimensional vector, where n is the number of observa-
                tions in X.

heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
nmain.p	A numeric value that represents the total number of main effects in X.
r1	A numeric value indicating the maximum number of main effects. This number can be different from the r1 defined in [detect](#).
r2	A numeric value indicating the maximum number of interaction effects. This number can be different from the r1 defined in [detect](#).
sigma	The standard deviation of the noise term. In practice, sigma is usually unknown. In such case, this function automatically estimate sigma using root mean square error (RMSE). Default is NULL. Otherwise, users need to enter a numeric value.
interaction.ind	
	A two-column numeric matrix containing all possible two-way interaction effects. It must be generated outside of this function using t(utils::combn()). See Example section for details.
lambda	A numeric value defined by users. Default is 10. For guidance on selecting an appropriate value, please refer to the Details section.
q	A numeric value indicating the number of models in each generation (e.g., the population size). Default is 40.
allout	Whether to print all outputs from this function. A "Yes" or "No" logical vector. Default is "No". See Value section for details.
interonly	Whether or not to consider fitted models with only two-way interaction effects. A "Yes" or "No" logical vector. Default is "No".
pi1	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC](#).
pi2	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC](#).
pi3	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC](#).
aprob	A numeric value between 0 and 1, defined by users. The addition probability during mutation. Default is 0.9.
dprob	A numeric value between 0 and 1, defined by users. The deletion probability during mutation. Default is 0.9.
aprobm	A numeric value between 0 and 1, defined by users. The main effect addition probability during addition. Default is 0.1.
aprobi	A numeric value between 0 and 1, defined by users. The interaction effect addition probability during addition. Default is 0.9.
dprobm	A numeric value between 0 and 1, defined by users. The main effect deletion probability during deletion. Default is 0.9.
dprobi	A numeric value between 0 and 1, defined by users. The interaction effect deletion probability during deletion. Default is 0.1.
take	Only used when allout = "No". Number of top candidate models to display. Default is 3.

Value

A list of output. The components are:

`final_model`	The final selected model.
`cleaned_candidate_model`	
	All candidate models where each row corresponding to a fitted model; the first 1 to `r1 + r2` columns representing the predictor indices in that model, and the last column is a numeric value representing the ABC score of that fitted model. Duplicated models are not allowed.
`InterRank`	Rank of all candidate interaction effects. A two-column numeric matrix. The first column contains indices of ranked two-way interaction effects, and the second column contains its corresponding ABC score.

See Also

[initial](), [cross](), [mut](), [ABC](), [Genone](), and [Extract]().

Examples

```
# allout = "No"


# allout = "Yes"
```

---

cross   *Performing crossover*

Description

This function performs crossover which only stores all fitted models without making any comparison. The selected indices in each fitted model will be automatically re-ordered so that main effects comes first, followed by two-way interaction effects, and zero reservation spaces.

Usage

```
cross(parents, heredity = "Strong", nmain.p, r1, r2, interaction.ind = NULL)
```

Arguments

`parents`	A numeric matrix of dimension q by `r1+r2`, obtained from `initial` or previous generation where each row corresponding a fitted model and each column representing the predictor index in the fitted model.
`heredity`	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
`nmain.p`	A numeric value that represents the total number of main effects in `X`.
`r1`	A numeric value indicating the maximum number of main effects.

r2   A numeric value indicating the maximum number of interaction effects.

`interaction.ind`

> A two-column numeric matrix containing all possible two-way interaction effects. It must be generated outside of this function using `t(utils::combn())`. See Example section for details.

Value

A numeric matrix `single.child.bit` is returned. Each row representing a fitted model, and each column corresponding to the predictor index in the fitted model. Duplicated models are allowed.

See Also

initial.

Examples

```
# Under Strong heredity
set.seed(0)
nmain.p <- 4
interaction.ind <- t(combn(4,2))
X <- matrix(rnorm(50*4,1,0.1), 50, 4)
epl <- rnorm(50,0,0.01)
y<- 1+X[,1]+X[,2]+X[,1]*X[,2]+epl
p1 <- initial(X, y, nmain.p = 4, r1 = 3, r2 = 3,
    interaction.ind = interaction.ind, q = 5)
c1 <- cross(p1, nmain.p=4, r1 = 3, r2 = 3,
    interaction.ind = interaction.ind)
```

---

detect   *Suggesting values for* r2

Description

This function suggests the values for `r2`.

Usage

```
detect(
  X,
  y,
  heredity = "Strong",
  nmain.p,
  sigma = NULL,
  r1,
  r2,
  interaction.ind = NULL,
  pi1 = 0.32,
```

```
  pi2 = 0.32,
  pi3 = 0.32,
  lambda = 10,
  q = 40
)
```

Arguments

X	Input data. An optional data frame, or numeric matrix of dimension n by nmain.p. Note that the two-way interaction effects should not be included in X because this function automatically generates the corresponding two-way interaction effects if needed.
y	Response variable. A n-dimensional vector, where n is the number of observations in X.
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
nmain.p	A numeric value that represents the total number of main effects in X.
sigma	The standard deviation of the noise term. In practice, sigma is usually unknown. In such case, this function automatically estimate sigma using root mean square error (RMSE). Default is NULL. Otherwise, users need to enter a numeric value.
r1	A numeric value indicating the maximum number of main effects.
r2	A numeric value indicating the maximum number of interaction effects.
interaction.ind	
	A two-column numeric matrix containing all possible two-way interaction effects. It must be generated outside of this function using t(utils::combn()). See Example section for details.
pi1	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC](#).
pi2	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC](#).
pi3	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC](#).
lambda	A numeric value defined by users. Default is 10. For guidance on selecting an appropriate value, please refer to [ABC](#).
q	A numeric value indicating the number of models in each generation (e.g., the population size). Default is 40.

Value

A list of output. The components are:

InterRank	Rank of all candidate interaction effects. A two-column numeric matrix. The first column contains indices of ranked two-way interaction effects, and the second column contains its corresponding ABC score.
mainind.sel	Selected main effects. A r1-dimensional vector.
mainpool	Ranked main effects in X.
plot	Plot of potential interaction effects and their corresponding ABC scores.

See Also

[initial](#).

Examples

```
# under Strong heredity

# under No heredity
set.seed(0)
nmain.p <- 4
interaction.ind <- t(combn(4,2))
X <- matrix(rnorm(50*4,1,0.1), 50, 4)
epl <- rnorm(50,0,0.01)
y<- 1+X[,1]+X[,2]+X[,1]*X[,2]+epl
d2 <- detect(X, y, heredity = "No", nmain.p = 4, r1 = 3, r2 = 3,
    interaction.ind = interaction.ind, q = 5)
```

Extract   *Extracting specific columns from a data*

Description

This function extracts specific columns from X based on varind. It provides an efficient procedure
for conducting ABC evaluation, especially when working with high-dimensional data.

Usage

```
Extract(X, varind, interaction.ind = NULL)
```

Arguments

X	Input data. An optional data frame, or numeric matrix of dimension n by nmain.p. Note that the two-way interaction effects should not be included in X because this function automatically generates the corresponding two-way interaction effects if needed.
varind	A numeric vector of class c() that specifies the indices of variables to be extracted from X. Duplicated values are not allowed. See Example section for details.
interaction.ind	
	A two-column numeric matrix containing all possible two-way interaction effects. It must be generated outside of this function using t(utils::combn()). See Example section for details.

Details

Please be aware that this function automatically renames column names into a designated format (e.g., X.1, X.2 for main effects, and X.1X.2 for interaction effect, etc), regardless of the original column names in X.

Under no heredity condition, this function can be applied in the context of interaction only linear regression models. See Example section for details.

Value

A numeric matrix is returned.

See Also

ABC, initial.

Examples

```
# Extract main effect X1 and X2 from X1,...X4
set.seed(0)
X1 <- matrix(rnorm(20), ncol = 4)
y1 <- X1[, 2] + rnorm(5)
interaction.ind <- t(combn(4,2))

# Extract main effect X1 and interaction effect X1X2 from X1,..X4
Extract(X1, varind = c(1,5), interaction.ind)

# Extract interaction effect X1X2 from X1,...X4
Extract(X1, varind = 5, interaction.ind)

# Extract using duplicated values in varind.
try(Extract(X1, varind = c(1,1), interaction.ind)) # this will not run
```

Genone   *Gathering useful information for first generation*

Description

This function automatically ranks all candidate interaction effects under Strong, Weak, or No heredity condition, compare and obtain first generation candidate models. The selected models will be re-ordered so that main effects come first, followed by interaction effects. Only two-way interaction effects will be considered.

Usage

```
Genone(
  X,
  y,
  heredity = "Strong",
```

```
    nmain.p,
    r1,
    r2,
    sigma = NULL,
    interaction.ind = NULL,
    lambda = 10,
    q = 40,
    allout = "No",
    interonly = "No",
    pi1 = 0.32,
    pi2 = 0.32,
    pi3 = 0.32,
    aprob = 0.9,
    dprob = 0.9,
    aprobm = 0.1,
    aprobi = 0.9,
    dprobm = 0.9,
    dprobi = 0.1
)
```

Arguments

X	Input data. An optional data frame, or numeric matrix of dimension n by `nmain.p`. Note that the two-way interaction effects should not be included in X because this function automatically generates the corresponding two-way interaction effects if needed.
y	Response variable. A n-dimensional vector, where n is the number of observations in X.
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
nmain.p	A numeric value that represents the total number of main effects in X.
r1	A numeric value indicating the maximum number of main effects.
r2	A numeric value indicating the maximum number of interaction effects.
sigma	The standard deviation of the noise term. In practice, sigma is usually unknown. In such case, this function automatically estimate sigma using root mean square error (RMSE). Default is NULL. Otherwise, users need to enter a numeric value.
interaction.ind	
	A two-column numeric matrix containing all possible two-way interaction effects. It must be generated outside of this function using `t(utils::combn())`. See Example section for details.
lambda	A numeric value defined by users. Default is 10. For guidance on selecting an appropriate value, please refer to the Details section.
q	A numeric value indicating the number of models in each generation (e.g., the population size). Default is 40.
allout	Whether to print all outputs from this function. A "Yes" or "No" logical vector. Default is "No". See Value section for details.

interonly	Whether or not to consider fitted models with only two-way interaction effects. A "Yes" or "No" logical vector. Default is "No".
pi1	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to ABC.
pi2	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to ABC.
pi3	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to ABC.
aprob	A numeric value between 0 and 1, defined by users. The addition probability during mutation. Default is 0.9.
dprob	A numeric value between 0 and 1, defined by users. The deletion probability during mutation. Default is 0.9.
aprobm	A numeric value between 0 and 1, defined by users. The main effect addition probability during addition. Default is 0.1.
aprobi	A numeric value between 0 and 1, defined by users. The interaction effect addition probability during addition. Default is 0.9.
dprobm	A numeric value between 0 and 1, defined by users. The main effect deletion probability during deletion. Default is 0.9.
dprobi	A numeric value between 0 and 1, defined by users. The interaction effect deletion probability during deletion. Default is 0.1.

Value

A list of output. The components are:

newparents	New parents models used for t+1-th generation. A numeric matrix of dimension q by r1+r2 where each row represents a fitted model. Duplicated models are allowed.
parents_models	A numeric matrix containing all fitted models from initial, cross, and mut where each row corresponding to a fitted model and each column representing the predictor index in that model. Duplicated models are allowed.
parents_models_cleaned	
	A numeric matrix containing fitted models from initial, cross, and mut with ABC scores. Each row corresponding to a fitted model; the first 1 to r1 + r2 columns representing the predictor indices in that model, and the last column is a numeric value representing the ABC score of that fitted model. Duplicated models are not allowed.
InterRank	Rank of all candidate interaction effects. A two-column numeric matrix. The first column contains indices of ranked two-way interaction effects, and the second column contains its corresponding ABC score.

See Also

initial, cross, mut, ABC, and Extract.

Examples

```
# allout = "No"
set.seed(0)
nmain.p <- 4
interaction.ind <- t(combn(4,2))
X <- matrix(rnorm(50*4,1,0.1), 50, 4)
epl <- rnorm(50,0,0.01)
y <- 1+X[,1]+X[,2]+X[,1]*X[,2]+epl
g1 <- Genone(X, y, nmain.p = 4, r1= 3, r2=3,
    interaction.ind = interaction.ind, q = 5)

# allout = "Yes"
g2 <- Genone(X, y, nmain.p = 4, r1= 3, r2=3,
    interaction.ind = interaction.ind, q = 5, allout = "Yes")
```

---

initial   *Setting up initial candidate models*

Description

This function automatically ranks all candidate interaction effects under Strong, Weak, or No heredity condition and obtains initial candidate models.

Usage

```
initial(
  X,
  y,
  heredity = "Strong",
  nmain.p,
  sigma = NULL,
  r1,
  r2,
  interaction.ind = NULL,
  pi1 = 0.32,
  pi2 = 0.32,
  pi3 = 0.32,
  lambda = 10,
  q = 40
)
```

Arguments

X   Input data. An optional data frame, or numeric matrix of dimension n by nmain.p.
            Note that the two-way interaction effects should not be included in X because this
            function automatically generates the corresponding two-way interaction effects
            if needed.

y	Response variable. A n-dimensional vector, where n is the number of observations in X.
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
nmain.p	A numeric value that represents the total number of main effects in X.
sigma	The standard deviation of the noise term. In practice, sigma is usually unknown. In such case, this function automatically estimate sigma using root mean square error (RMSE). Default is NULL. Otherwise, users need to enter a numeric value.
r1	A numeric value indicating the maximum number of main effects. This number can be different from the r1 defined in [detect]().
r2	A numeric value indicating the maximum number of interaction effects. This number can be different from the r2 defined in [detect]().
interaction.ind	
	A two-column numeric matrix containing all possible two-way interaction effects. It must be generated outside of this function using t(utils::combn()). See Example section for details.
pi1	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC]().
pi2	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC]().
pi3	A numeric value between 0 and 1, defined by users. Default is 0.32. For guidance on selecting an appropriate value, please refer to [ABC]().
lambda	A numeric value defined by users. Default is 10. For guidance on selecting an appropriate value, please refer to [ABC]().
q	A numeric value indicating the number of models in each generation (e.g., the population size). Default is 40.

Value

A list of output. The components are:

initialize	Initial candidate models. A numeric matrix of dimension q by r1+r2 where each row represents a fitted model. Duplicated models are allowed.
InterRank	Rank of all candidate interaction effects. A two-column numeric matrix. The first column contains indices of ranked two-way interaction effects, and the second column contains its corresponding ABC score.
mainind.sel	Selected main effects. A r1-dimensional vector.
mainpool	Ranked main effects in X.

See Also

[ABC](), [Extract]().

Examples

```
# Under Strong heredity
```

---

mut   *Performing mutation*

Description

This function performs mutation which only stores all fitted models without making any comparison. The selected indices in each fitted model will be automatically re-ordered so that main effects comes first, followed by two-way interaction effects, and zero reservation spaces.

Usage

```
mut(
  parents,
  heredity = "Strong",
  nmain.p,
  r1,
  r2,
  interaction.ind = NULL,
  interonly = "No",
  aprob = 0.9,
  dprob = 0.9,
  aprobm = 0.1,
  aprobi = 0.9,
  dprobm = 0.9,
  dprobi = 0.1
)
```

Arguments

parents	A numeric matrix of dimension q by r1+r2, obtained from initial or previous generation where each row corresponding a fitted model and each column representing the predictor index in the fitted model.
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
nmain.p	A numeric value that represents the total number of main effects in X.
r1	A numeric value indicating the maximum number of main effects.
r2	A numeric value indicating the maximum number of interaction effects.
interaction.ind	
	A two-column numeric matrix containing all possible two-way interaction effects. It must be generated outside of this function using t(utils::combn()). See Example section for details.
interonly	Whether or not to consider fitted models with only two-way interaction effects. A "Yes" or "No" logical vector. Default is "No".
aprob	A numeric value between 0 and 1, defined by users. The addition probability during mutation. Default is 0.9.

dprob	A numeric value between 0 and 1, defined by users. The deletion probability during mutation. Default is 0.9.
aprobm	A numeric value between 0 and 1, defined by users. The main effect addition probability during addition. Default is 0.1.
aprobi	A numeric value between 0 and 1, defined by users. The interaction effect addition probability during addition. Default is 0.9.
dprobm	A numeric value between 0 and 1, defined by users. The main effect deletion probability during deletion. Default is 0.9.
dprobi	A numeric value between 0 and 1, defined by users. The interaction effect deletion probability during deletion. Default is 0.1.

Value

A numeric matrix `single.child.mutated` is returned. Each row representing a fitted model, and each column corresponding to the predictor index in the fitted model. Duplicated models are allowed.

See Also

[initial](#).

Examples

```
# Under Strong heredity, interonly = "No"
set.seed(0)
nmain.p <- 4
interaction.ind <- t(combn(4,2))
X <- matrix(rnorm(50*4,1,0.1), 50, 4)
epl <- rnorm(50,0,0.01)
y <- 1+X[,1]+X[,2]+X[,1]*X[,2]+epl
p1 <- initial(X, y, nmain.p = 4, r1 = 3, r2 = 3,
    interaction.ind = interaction.ind, q = 5)
m1 <- mut(p1, nmain.p = 4, r1 = 3, r2 = 3,
    interaction.ind =interaction.ind)
# Under Strong heredity, interonly = "Yes"
m2 <- mut(p1, heredity = "No", nmain.p = 4, r1 = 3, r2 = 3,
    interaction.ind =interaction.ind, interonly = "Yes")
```

Index