

Package ‘hySAINT’

April 4, 2024

Title Hybrid Genetic and Simulated Annealing Algorithm for High Dimensional Linear Models with Interaction Effects

Version 1.2.1

Description

We provide a stage-wise selection method using genetic algorithms, designed to efficiently identify main and two-way interactions within high-dimensional linear regression models. Additionally, it implements simulated annealing algorithm during the mutation process. The relevant paper can be found at: Ye, C., and Yang, Y. (2019) <[doi:10.1109/TIT.2019.2913417](https://doi.org/10.1109/TIT.2019.2913417)>.

License GPL-2

Encoding UTF-8

Imports utils, Matrix, energy, pracma, stats, selectiveInference, VariableScreening, SIS

Language en-US

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

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

RoxygenNote 7.3.1

NeedsCompilation no

Repository CRAN

Date/Publication 2024-04-04 15:03:00 UTC

R topics documented:

ABC	2
Crossover	3
EVA	4
Extract	5
hySAINT	6
Initial	8
Mutation	9

Index	11
--------------	-----------

ABC

*ABC Evaluation***Description**

Gives ABC score for each fitted model. 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 dataset, the smaller the ABC, the better fit.

Usage

```
ABC(
  X,
  y,
  heredity = "Strong",
  sigma,
  varind = NULL,
  interaction.ind = NULL,
  lambda = 10
)
```

Arguments

<code>X</code>	Input data. An optional data frame, or numeric matrix of dimension n observations by p main effects.
<code>y</code>	Response variable. A n -dimensional vector.
<code>heredity</code>	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
<code>sigma</code>	The standard deviation of the noise term. In practice, <code>sigma</code> is usually unknown. Users can estimate <code>sigma</code> from function <code>selectiveInference::estimateSigma</code> , then use the output as the <code>sigma</code> value. See examples for details.
<code>varind</code>	A numeric vector that specifies the indices of variables to be extracted from <code>X</code> . Default is "No".
<code>interaction.ind</code>	A two-column numeric matrix. Each row represents a unique interaction pair, with the columns indicating the index numbers of the variables involved in each interaction. Note that <code>interaction.ind</code> must be generated outside of this function using <code>t(utils::combn(p, 2))</code> . See Example section for details.
<code>lambda</code>	A numeric value defined by users. The number needs to satisfy the condition: $\lambda \geq 5.1/\log(2)$. Default is 10.

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*.

Examples

```
# When sigma is known
set.seed(0)
interaction.ind <- t(combn(4,2))
X <- matrix(rnorm(50*4,1,0.1), 50, 4)
ep1 <- rnorm(50,0,0.01)
y <- 1+X[,1]+X[,2]+X[,1]*X[,2] + ep1
ABC(X, y, sigma = 0.01, varind = c(1,2,5), interaction.ind = interaction.ind)

# When sigma is not known
full <- Extract(X, varind = c(1:(dim(X)[2]+dim(interaction.ind)[1])), interaction.ind)
sigma <- selectiveInference::estimateSigma(full, y)$sigmahat # Estimate sigma
```

Crossover

Performing crossover

Description

This function gives offspring from parents. It performs crossover at a fixed probability of 0.6.

Usage

```
Crossover(X, myParent, EVAoutput, heredity = "Strong", r1, r2, numElite = 40)
```

Arguments

X	Input data. An optional data frame, or numeric matrix of dimension n observations by p main effects.
myParent	A numeric matrix with dimension numElite by r1 + r2.
EVAoutput	The output from function EVA.
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
r1	At most how many main effects do you want to include in your model?. For high-dimensional data, r1 cannot be larger than the number of screened main effects.
r2	At most how many interaction effects do you want to include in your model?
numElite	Number of elite parents. Default is 40.

Value

Offspring. If crossover occurred, it returns a numeric matrix with dimensions choose(numElite, 2) by r1+r2. Otherwise, numElite by r1 + r2.

See Also

[EVA, Initial.](#)

Examples

```
set.seed(0)
interaction.ind <- t(combn(10,2))
X <- matrix(rnorm(100*10,1,0.1), 100, 10)
ep1 <- rnorm(100,0,0.01)
y <- 1+X[,1]+X[,2]+X[,3]+X[,1]*X[,2]+X[,1]*X[,3]+ep1
EVAoutput <- EVA(X, y, r1 = 5, sigma = 0.01,
  interaction.ind = interaction.ind)
myParent <- Initial(X = X, y = y, EVAoutput, r1 = 5, r2 = 2)
Offsprings <- Crossover(X, myParent, EVAoutput, r1 = 5, r2 = 2)
```

 EVA

Evaluating main and interaction effects

Description

This function ranks each main and interaction effect. It also calculate the ABC score for each potential interactions across different heredity structures. If heredity = "No" and the the number of potential interactions exceed choose(1000, 2), distance correlation between each variable in X and y will be calculated so that it reduces the running time. This ensures a more efficient evaluation process.

Usage

```
EVA(
  X,
  y,
  heredity = "Strong",
  r1,
  sigma,
  varind = NULL,
  interaction.ind = NULL,
  lambda = 10
)
```

Arguments

X	Input data. An optional data frame, or numeric matrix of dimension n observations by p main effects.
y	Response variable. A n-dimensional vector.
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".

<code>r1</code>	At most how many main effects do you want to include in your model?. For high-dimensional data, <code>r1</code> cannot be larger than the number of screened main effects.
<code>sigma</code>	The standard deviation of the noise term. In practice, <code>sigma</code> is usually unknown. Users can estimate <code>sigma</code> from function <code>selectiveInference::estimateSigma</code> , then use the output as the <code>sigma</code> value.
<code>varind</code>	A numeric vector that specifies the indices of variables to be extracted from <code>X</code> . Default is "No".
<code>interaction.ind</code>	A two-column numeric matrix. Each row represents a unique interaction pair, with the columns indicating the index numbers of the variables involved in each interaction. Note that <code>interaction.ind</code> must be generated outside of this function using <code>t(utils::combn(p, 2))</code> . See Example section for details.
<code>lambda</code>	A numeric value defined by users. The number needs to satisfy the condition: $\lambda \geq 5.1/\log(2)$. Default is 10.

Value

A list of output. The components are: ranked main effect, `ranked.mainpool`; and a 4-column matrix contains potential interactions ranked by ABC score, `ranked.intermat`.

See Also

[ABC](#), [Extract](#)

Examples

```
# Strong heredity
set.seed(0)
interaction.ind <- t(combn(10,2))
X <- matrix(rnorm(100*10,1,0.1), 100, 10)
ep1 <- rnorm(100,0,0.01)
y <- 1+X[,1]+X[,2]+X[,3]+X[,1]*X[,2]+X[,1]*X[,3]+ep1
EVAoutput <- EVA(X, y, r1 = 5, sigma = 0.01, interaction.ind = interaction.ind)
```

Extract	<i>Extracting columns and generating required interaction effects from data</i>
---------	---

Description

This function simplifies the data preparation process by enabling users to extract specific columns from their dataset `X`, and automatically generating any necessary interaction effects based on `varind`.

Usage

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

Arguments

<code>X</code>	Input data. An optional data frame, or numeric matrix of dimension n observations by p main effects. Note that the interaction effects should not be included in X because this function automatically generates the corresponding interaction effects if needed.
<code>varind</code>	A numeric vector that specifies the indices of variables to be extracted from X . Duplicated values are not allowed. See Example for details.
<code>interaction.ind</code>	A two-column numeric matrix. Each row represents a unique interaction pair, with the columns indicating the index numbers of the variables involved in each interaction. Note that <code>interaction.ind</code> must be generated outside of this function using <code>t(utils::combn(p,2))</code> . See Example section for details.

Value

A numeric matrix is returned.

Examples

```
# Generate interaction.ind
interaction.ind <- t(combn(4,2))

# Generate data
set.seed(0)
X <- matrix(rnorm(20), ncol = 4)
y <- X[, 2] + rnorm(5)

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

# Extract X5 from X1, ..., X4
Extract(X, varind = 5, interaction.ind)

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

hySAINT

Hybrid Genetic and Simulated Annealing Algorithm

Description

This is the main function of package hySAINT. It implements both genetic algorithm and simulated annealing. The simulated annealing technique is used within mutation operator.

Usage

```
hySAINT(
  X,
  y,
  heredity = "Strong",
  r1,
  r2,
  sigma,
  interaction.ind = NULL,
  varind = NULL,
  numElite = 40,
  max.iter = 500,
  initial.temp = 1000,
  cooling.rate = 0.95,
  lambda = 10
)
```

Arguments

X	Input data. An optional data frame, or numeric matrix of dimension n observations by p main effects.
y	Response variable. A n-dimensional vector.
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
r1	At most how many main effects do you want to include in your model?. For high-dimensional data, r1 cannot be larger than the number of screened main effects.
r2	At most how many interaction effects do you want to include in your model?
sigma	The standard deviation of the noise term. In practice, sigma is usually unknown. Users can estimate sigma from function <code>selectiveInference::estimateSigma</code> , then use the output as the sigma value.
interaction.ind	A two-column numeric matrix. Each row represents a unique interaction pair, with the columns indicating the index numbers of the variables involved in each interaction. Note that <code>interaction.ind</code> must be generated outside of this function using <code>t(utils::combn(p, 2))</code> . See Example section for details.
varind	A numeric vector that specifies the indices of variables to be extracted from X.
numElite	Number of elite parents. Default is 40.
max.iter	Maximum number of iterations. Default is 500.
initial.temp	Initial temperature. Default is 1000.
cooling.rate	A numeric value represents the speed at which the temperature decreases. Default is 0.95.
lambda	A numeric value defined by users. The number needs to satisfy the condition: $\lambda \geq 5.1/\log(2)$. Default is 10.

Value

An object with S3 class "hySAINT".

Final.variable.names Name of the selected effects.

Final.variable.idx Index of the selected effects.

Final.model.score Final Model ABC.

All.iter.score Best ABC scores from initial parents and all iterations.

See Also

[ABC](#), [EVA](#), [Initial](#), [Crossover](#), [Mutation](#)

Examples

```
set.seed(0)
interaction.ind <- t(combn(10,2))
X <- matrix(rnorm(100*10,1,0.1), 100, 10)
ep1 <- rnorm(100,0,0.01)
y <- 1+X[,1]+X[,2]+X[,3]+X[,1]*X[,2]+X[,1]*X[,3]+ep1
hySAINT(X, y, r1 = 5, r2 = 2, sigma = 0.01, interaction.ind = interaction.ind, max.iter = 5)
```

Initial

Creating initial parents

Description

This function gives initial parents.

Usage

```
Initial(X, y, EVAoutput, heredity = "Strong", r1, r2, numElite = 40)
```

Arguments

X	Input data. An optional data frame, or numeric matrix of dimension n observations by p main effects.
y	Response variable. A n-dimensional vector.
EVAoutput	The output from function EVA
heredity	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
r1	At most how many main effects do you want to include in your model?. For high-dimensional data, r1 cannot be larger than the number of screened main effects.
r2	At most how many interaction effects do you want to include in your model?
numElite	Number of elite parents. Default is 40.

Value

Initial parents. A numeric matrix with dimensions numElite by r1+r2.

See Also

[EVA](#)

Examples

```
set.seed(0)
interaction.ind <- t(combn(10,2))
X <- matrix(rnorm(100*10,1,0.1), 100, 10)
ep1 <- rnorm(100,0,0.01)
y <- 1+X[,1]+X[,2]+X[,3]+X[,1]*X[,2]+X[,1]*X[,3]+ep1
EVAoutput <- EVA(X, y, r1 = 5, sigma = 0.01,
  interaction.ind = interaction.ind)
myParent <- Initial(X = X, y = y, EVAoutput, r1 = 5, r2 = 2)
```

Mutation

Performing mutation

Description

This function gives mutant from parents.

Usage

```
Mutation(
  myParent,
  EVAoutput,
  r1,
  r2,
  initial.temp = 1000,
  cooling.rate = 0.95,
  X,
  y,
  heredity = "Strong",
  sigma,
  varind = NULL,
  interaction.ind = NULL,
  lambda = 10
)
```

Arguments

<code>myParent</code>	A numeric matrix with dimension <code>numElite</code> by <code>r1 + r2</code> .
<code>EVAoutput</code>	The output from function <code>EVA</code> .
<code>r1</code>	At most how many main effects do you want to include in your model?. For high-dimensional data, <code>r1</code> cannot be larger than the number of screened main effects.
<code>r2</code>	At most how many interaction effects do you want to include in your model?
<code>initial.temp</code>	Initial temperature. Default is 1000.
<code>cooling.rate</code>	A numeric value represents the speed at which the temperature decreases. Default is 0.95.
<code>X</code>	Input data. An optional data frame, or numeric matrix of dimension <code>n</code> observations by <code>p</code> main effects.
<code>y</code>	Response variable. A <code>n</code> -dimensional vector.
<code>heredity</code>	Whether to enforce Strong, Weak, or No heredity. Default is "Strong".
<code>sigma</code>	The standard deviation of the noise term. In practice, <code>sigma</code> is usually unknown. Users can estimate <code>sigma</code> from function <code>selectiveInference::estimateSigma</code> , then use the output as the <code>sigma</code> value.
<code>varind</code>	A numeric vector that specifies the indices of variables to be extracted from <code>X</code> .
<code>interaction.ind</code>	A two-column numeric matrix. Each row represents a unique interaction pair, with the columns indicating the index numbers of the variables involved in each interaction. Note that <code>interaction.ind</code> must be generated outside of this function using <code>t(utils::combn(p, 2))</code> .
<code>lambda</code>	A numeric value defined by users. The number needs to satisfy the condition: $\lambda \geq 5.1/\log(2)$. Default is 10.

Value

`Mutant`. A numeric matrix with dimensions `numElite` by `r1+r2`.

See Also

[EVA](#), [Initial](#).

Examples

```
set.seed(0)
interaction.ind <- t(combn(10,2))
X <- matrix(rnorm(100*10,1,0.1), 100, 10)
ep1 <- rnorm(100,0,0.01)
y <- 1+X[,1]+X[,2]+X[,3]+X[,1]*X[,2]+X[,1]*X[,3]+ep1
EVAoutput <- EVA(X, y, r1 = 5, sigma = 0.01,
  interaction.ind = interaction.ind)
myParent <- Initial(X = X, y = y, EVAoutput, r1 = 5, r2 = 2)
Mutation(myParent, EVAoutput, r1 = 5, r2 = 2, X = X, y = y,
  sigma = 0.1, interaction.ind = interaction.ind)
```

Index

ABC, [2](#), [5](#), [8](#)

Crossover, [3](#), [8](#)

EVA, [4](#), [4](#), [8–10](#)

Extract, [5](#), [5](#)

hySAINT, [6](#)

Initial, [4](#), [8](#), [8](#), [10](#)

Mutation, [8](#), [9](#)