Package 'iForm'

April 28, 2018

Type Package

Title Forward Selection Under Marginality Principle

Version 1.0

Date 2016-03-29

Author Kirk Gosik <kgosik@broadinstitute.org>

Maintainer Kirk Gosik <kgosik@broadinstitute.org>

Description Extended variable selection approaches to jointly model main and
interaction effects from high-dimensional data.

License GPL-3

LazyData TRUE

RoxygenNote 6.0.1.9000

Suggests knitr,rmarkdown

VignetteBuilder knitr

NeedsCompilation no

R topics documented:

`iForm`   *Interaction Screening for Ultra-High Dimensional Data*

Description

Extended variable selection approaches to jointly model main and interaction effects from high-dimensional data orignally proposed by Hao and Zhang (2014) and extended by Gosik and Wu (2016). Based on a greedy forward approach, their model can identify all possible interaction effects through two algorithms, iFORT and iFORM, which have been proved to possess sure screening property in an ultrahigh-dimensional setting.

Usage

```
iForm(formula, data, heredity = "strong", higher_order = FALSE)
```

Arguments

`formula`	an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
`data`	data.frame of your data with the response and all p predictors
`heredity`	a string specifying the heredity to be considered. NULL, weak, strong
`higher_order`	logical TRUE indicating to include order-3 interactions in the search (default FALSE)

Details

Runs the iFORM selection procedure on the dataset and returns a linear model of the final selected model. The model is of an R object of class "lm"

Value

a summary of the linear model returned after the selection procedure

Author(s)

Kirk Gosik

See Also

`lm`
`model.frame`

Examples

```
iForm(formula = hp ~ ., data = mtcars, heredity = "strong", higher_order = FALSE)
```

`iformselect`	*Selection for iForm procedure*

Description

This is a helper function to run the selection procedure under differen heredity principles and different levels of interactions included in the selection.

Usage

```
iformselect(x, y, p, n, C, S, bic, heredity, higher_order)
```

Arguments

`x`	data.frame of your data with all p predictors
`y`	vector of observed responses
`p`	number of predictors in the dataset
`n`	number of observations in the dataset
`C`	vector of candidate predictors to consider in this step of the procedure
`S`	vector of solution predictor selected from previous steps of the procedure
`bic`	vector of bic values calculated for each step of the procedure
`heredity`	a string specifying the heredity to be considered. NULL, weak, strong
`higher_order`	logical TRUE indicating to include order-3 interactions in the search (default FALSE)

Details

Runs the iFORM selection procedure for specifed heredity and level of interactions. It returns the solution to be fit from `iForm`

Value

the response vector, the solution set of predictors and the calculated bic values

Author(s)

Kirk Gosik

`rss_map_func` *Finding minimum RSS*

Description

Helper function to take in the candidate set and solution set along with the observations and previous data to calculate the residual sum of sqaures for each of the candidate predictors given what has already been selected.

Usage

```
rss_map_func(C, S, y, data)
```

Arguments

`C`	vector of candidate predictors to consider in this step of the procedure
`S`	vector of solution predictor selected from previous steps of the procedure
`y`	vector of observed responses
`data`	data.frame of your data with the response and all p predictors

Details

Mapping function to calculcate the residual sum of squares for each of the candidate predictors

Value

A vector of the RSS values for each candidate predictor

Author(s)

Kirk Gosik

`strong_order2` *Creating interactions based off of strong heredity principle*

Description

Helper function to give all possible order-2 interactions following the strong heredity principle.

Usage

```
strong_order2(S, data)
```

Arguments

`S`	vector of solution predictor selected from previous steps of the procedure
`data`	data.frame of your data with the response and all p predictors

Details

Finds all p choose 2 combinations of predicotrs in the solution set

Value

A vector of the RSS values for each candidate predictor

Author(s)

Kirk Gosik

See Also

`model.matrix`

`strong_order3`	*Creating interactions based off of strong heredity principle*

Description

Helper function to give all possible order-3 interactions following the strong heredity principle.

Usage

```
strong_order3(S, data)
```

Arguments

`S`	vector of solution predictor selected from previous steps of the procedure
`data`	data.frame of your data with the response and all p predictors

Details

Finds all p choose 3 combinations between the predicotrs in the solution set.

Value

A vector of the RSS values for each candidate predictor

Author(s)

Kirk Gosik

---

weak_order2   *Creating interactions based off of weak heredity principle*

Description

Helper function to give all possible order-2 interactions following the weak heredity principle.

Usage

```
weak_order2(S, C, data)
```

Arguments

S	vector of solution predictor selected from previous steps of the procedure
C	vector of candidate predictors to consider in this step of the procedure
data	data.frame of your data with the response and all p predictors

Details

Finds all p choose 3 combinations between the predicotrs in the solution set and the predictors in the candidate set.

Value

A vector of the RSS values for each candidate predictor

Author(s)

Kirk Gosik

---

weak_order3   *Creating interactions based off of weak heredity principle*

Description

Helper function to give all possible order-3 interactions following the strong heredity principle.

Usage

```
weak_order3(S, C, data)
```

Arguments

S	vector of solution predictor selected from previous steps of the procedure
C	vector of candidate predictors to consider in this step of the procedure
data	data.frame of your data with the response and all p predictors

Details

Finds all p choose 3 combinations between the predicotrs in the solution set and the predictors in the candidate set.

Value

A vector of the RSS values for each candidate predictor

Author(s)

Kirk Gosik