

Package ‘stabiliser’

August 10, 2021

Title Stabilising Variable Selection

Version 0.1.0

Description A stable approach to variable selection through stability selection and the use of an objective threshold based on a model from permuted data. Lima et al (2021) <[doi:10.1038/s41598-020-79317-8](https://doi.org/10.1038/s41598-020-79317-8)>, Meinshausen and Bühlmann (2010) <[doi:10.1111/j.1467-9868.2010.00740.x](https://doi.org/10.1111/j.1467-9868.2010.00740.x)>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Depends R (>= 2.10)

Suggests rmarkdown, knitr, testthat (>= 3.0.0), markdown

Config/testthat/edition 3

Imports dplyr, bigstep, rsample, tibble, purrr, tidyr, stringr, ggplot2, broom, caret, glmnet, ncvreg

VignetteBuilder knitr

NeedsCompilation no

Author Robert Hyde [aut, cre] (<<https://orcid.org/0000-0002-8705-9405>>),
Martin Green [aut],
Eliana Lima [aut]

Maintainer Robert Hyde <robert.hyde1@nottingham.ac.uk>

Repository CRAN

Date/Publication 2021-08-10 12:20:02 UTC

R topics documented:

boot_model	2
model_enet	2
model_lasso	3
model_mbic	3
model_mcp	3

model_selector	4
permute	4
perm_stab	4
rep_selector_boot	4
rep_selector_perm	5
sim_dat	5
stabilise	5
stabiliser	6
stabiliser_example	6
stab_plot	7
triangulate	7

Index	8
--------------	----------

boot_model	<i>boot_model</i>
------------	-------------------

Description

Function to calculate stability of variables' association with an outcome for a given model over a number of bootstrap repeats

Arguments

data	a dataframe containing an outcome variable to be permuted
outcome	the outcome as a string (i.e. "y")
boot_reps	the number of bootstrap samples
model	the model to be used (i.e. model_mbic)

model_enet	<i>model_enet</i>
------------	-------------------

Description

Function to model elastic net selection process on a given dataframe

Arguments

data	a dataframe containing an outcome variable to be permuted (usually coming from nested bootstrap data)
outcome	the outcome as a string (i.e. "y")

model_lasso	<i>model_lasso</i>
-------------	--------------------

Description

Function to model lasso selection process on a given dataframe

Arguments

data	a dataframe containing an outcome variable to be permuted (usually coming from nested bootstrap data)
outcome	the outcome as a string (i.e. "y")

model_mbic	<i>model_mbic</i>
------------	-------------------

Description

Function to model mbic selection process on a given dataframe

Arguments

data	a dataframe containing an outcome variable to be permuted (usually coming from nested bootstrap data)
outcome	the outcome as a string (i.e. "y")

model_mcp	<i>model_mcp</i>
-----------	------------------

Description

Function to model mcp selection process on a given dataframe

Arguments

data	a dataframe containing an outcome variable to be permuted (usually coming from nested bootstrap data)
outcome	the outcome as a string (i.e. "y")

model_selector	<i>model_selector</i>
----------------	-----------------------

Description

Determines which models to call.

permute	<i>permute</i>
---------	----------------

Description

Calculates permutation threshold for null model, where a specified model is run over multiple bootstrap resamples of multiple permuted version of the dataset.

Arguments

data	a dataframe containing an outcome variable to be permuted
outcome	the outcome to be permuted as a string (i.e. "y")
permutations	the number of times to be permuted per repeat
perm_boot_reps	the number of times to repeat each set of permutations

perm_stab	<i>perm_stab</i>
-----------	------------------

Description

Main function to call both permutation and bootstrapping functions; to be looped over multiple models selected by the user.

rep_selector_boot	<i>rep_selector_boot</i>
-------------------	--------------------------

Description

wrapper function to determine the number of bootstrap repeats

Usage

```
rep_selector_boot(data, boot_reps)
```

Arguments

data	the dataset to analyse.
boot_reps	the number of bootstrap samples

rep_selector_perm	<i>rep_selector_boot</i>
-------------------	--------------------------

Description

wrapper function to determine the number of permutations

Usage

```
rep_selector_perm(data, permutations)
```

Arguments

data	the dataset to analyse.
permutations	the number of times to be permuted per repeat

sim_dat	<i>Simulated data</i>
---------	-----------------------

Description

A simulated dataset

Usage

```
sim_dat
```

Format

A data frame with 100 rows and 1001 variables:

stabilise	<i>stabilise</i>
-----------	------------------

Description

Function to calculate stability of variables' association with an outcome for a given model over a number of bootstrap repeats

Arguments

data	A dataframe containing an outcome variable to be permuted.
outcome	The outcome as a string (i.e. "y").
boot_reps	The number of bootstrap samples. Default is "auto" which selects number based on dataframe size.
permutations	The number of times to be permuted per repeat. Default is "auto" which selects number based on dataframe size.
perm_boot_reps	The number of times to repeat each set of permutations. Default is 5.
models	The models to select for stabilising. Default is elastic net (models = c("enet")), other available models include "lasso", "mbic", "mcp".

Value

A list for each model selected. Each list contains a dataframe of variable stabilities, a numeric permutation threshold, and a dataframe of coefficients for both bootstrap and permutation.

stabiliser	<i>stabiliser</i>
------------	-------------------

Description

This package uses bootstrap resampling and an objective selection stability threshold to provide a robust method of selecting variables truly associated with an outcome.

Author(s)

Robert Hyde <robert.hyde1@nottingham.ac.uk>
 Martin Green
 Eliana Lima

stabiliser_example	<i>stabiliser_example</i>
--------------------	---------------------------

Description

A simulated dataset

Usage

stabiliser_example

Format

A data frame with 50 rows and 100 variables. The stabiliser_example dataset is a simulated example with the following properties: 1 simulated outcome variable: y 4 variables simulated to be associated with y: causal1, causal2... 95 variables simulated to have no association with y: junk1, junk2...

stab_plot	<i>stab_plot</i>
-----------	------------------

Description

Plot from stability object

Arguments

stabiliser_outcome
Outcome from stabilise() or triangulate() function.

Value

A ggplot object.

triangulate	<i>triangulate</i>
-------------	--------------------

Description

Triangulate multiple models using a stability object

Arguments

object An object generated through the stabilise() function.

Value

A combined list of model results including a dataframe of stability results for variables and a numeric permutation threshold.

Index

* datasets

sim_dat, 5

stabiliser_example, 6

boot_model, 2

model_enet, 2

model_lasso, 3

model_mbic, 3

model_mcp, 3

model_selector, 4

perm_stab, 4

permute, 4

rep_selector_boot, 4

rep_selector_perm, 5

sim_dat, 5

stab_plot, 7

stabilise, 5

stabiliser, 6

stabiliser-package (stabiliser), 6

stabiliser_example, 6

triangulate, 7