

Package ‘stabiliser’

April 6, 2022

Title Stabilising Variable Selection

Version 1.0.1

Description

A stable approach to variable selection through stability selection and the use of a permutation-based objective stability threshold. 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

Config/testthat/edition 3

Depends R (>= 3.0.0)

Suggests rmarkdown, testthat (>= 3.0.0), markdown

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

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.hyde4@nottingham.ac.uk>

Repository CRAN

Date/Publication 2022-04-06 10:40:02 UTC

R topics documented:

simulate_data	2
stabilise	2
stabiliser_example	3
stab_plot	3
triangulate	4

Index**5**

simulate_data	<i>simulate_data</i>
---------------	----------------------

Description

Simulate a dataset. This can optionally include variables with a given associated with the outcome.

Usage

```
simulate_data(nrows, ncols, n_true = 0, amplitude = 0)
```

Arguments

nrows	The number of rows to simulate.
ncols	The number of columns to simulate.
n_true	The number of variables truly associated with the outcome.
amplitude	The strength of association between true variables and the outcome.

Value

A simulated dataset

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".
type	The type of model, either "linear" or "logistic"
quantile	The quantile of null stabilities to use as a threshold.

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_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`*triangulate*

Description

Triangulate multiple models using a stability object

Arguments

<code>object</code>	An object generated through the <code>stabilise()</code> function.
<code>quantile</code>	The quantile of null stabilities to use as a threshold.

Value

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

Index

* datasets

stabiliser_example, 3

simulate_data, 2

stab_plot, 3

stabilise, 2

stabiliser_example, 3

triangulate, 4