Package ‘stabiliser’

August 10, 2021

Title Stabilising Variable Selection

Version 0.1.0


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
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)

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**

**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**

**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**

**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

**Description**

Determines which models to call.

permute

**Description**

Calculates permutation threshold for null model, where a specified model is run over multiple boot-strap 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

**Description**

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

rep_selector_boot

**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

---

**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**  
**Simulated data**

---

**Description**

A simulated dataset

**Usage**

```
sim_dat
```

**Format**

A data frame with 100 rows and 1001 variables:

---

**stabilise**  
**stabilise**

---

**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.

---

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

---

Description

A simulated dataset

Usage

stabiliser_example
**stab_plot**

**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...

**Description**
Plot from stability object

**Arguments**
- **stabiliser_outcome**
  Outcome from stabilise() or triangulate() function.

**Value**
A ggplot object.

**triangulate**

**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