Package ‘workboots’

March 9, 2022

Title Generate Bootstrap Prediction Intervals from a 'tidymodels' Workflow

Version 0.1.0

Description Provides functions for generating bootstrap prediction intervals from a 'tidymodels' workflow. 'tidymodels' <https://www.tidymodels.org/> is a collection of packages for modeling and machine learning using 'tidyverse' <https://www.tidyverse.org/> principles. This package is not affiliated with or maintained by 'RStudio' or the 'tidymodels' maintainers.

License MIT + file LICENSE

URL https://github.com/markjrieke/workboots

BugReports https://github.com/markjrieke/workboots/issues

Imports assertthat, dplyr, generics, purrr, rlang, rsample, stats, tibble, tidyr, vip, workflows

Encoding UTF-8

RoxygenNote 7.1.2

Suggests ggplot2, knitr, recipes, rmarkdown, testthat (>= 3.0.0), tidymodels, tune, xgboost

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Mark Rieke [aut, cre]

Maintainer Mark Rieke <markjrieke@gmail.com>

Repository CRAN

Date/Publication 2022-03-09 09:20:02 UTC

R topics documented:

predict_boots ........................................................................................................ 2
predict_single_boot ............................................................................................ 3
predict_boots

summarise_predictions ............................................... 3
vi_boots ........................................................................ 4
vi_single_boot ............................................................... 5

Index 7

predict_boots  Fit and predict from a workflow using many bootstrap resamples.

Description
Generate a prediction interval from arbitrary model types using bootstrap resampling. predict_boots() generates $n$ bootstrap resamples, fits a model to each resample (creating $n$ models), then creates $n$ predictions for each observation in new_data.

Usage
predict_boots(workflow, n = 100, training_data, new_data, ...)

Arguments
workflow An un-fitted workflow object.
n An integer for the number of bootstrap resampled models that will be created.
training_data A tibble or dataframe of data to be resampled and used for training.
new_data A tibble or dataframe used to make predictions.
... Additional params passed to rsample::bootstraps().

Details
Since predict_boots() fits a new model to each resample, the argument workflow must not yet be fit. Any tuned hyperparameters must be finalized prior to calling predict_boots().

Value
A tibble with a column indicating the row index of each observation in new_data and a nested list of the model predictions for each observation.

Examples
# Not run:
library(tidymodels)

# setup a workflow without fitting
wf <- workflow() %>%
  add_recipe(recipe(qsec ~ wt, data = mtcars)) %>%
  add_model(linear_reg())
# fit and predict 125 bootstrap resampled models to mtcars

```r
set.seed(123)
wtf %>%
predict_boots(n = 125, training_data = mtcars, new_data = mtcars)
## End(Not run)
```

---

**predict_single_boot**  
Fit a model and predict based on a single bootstrap resample

**Description**

Fit a model and predict based on a single bootstrap resample

**Usage**

```r
predict_single_boot(workflow, boot_splits, new_data, index)
```

**Arguments**

- **workflow**: An un-fitted workflow object.
- **boot_splits**: A bootstrap split object created by `rsample::bootstraps()`.
- **new_data**: New data to make predictions
- **index**: Index of `boot_splits` to use for training

---

**summarise_predictions**  
Append a tibble of predictions returned by `predict_boots()` with a prediction interval.

**Description**

Append a tibble of predictions returned by `predict_boots()` with a prediction interval.

**Usage**

```r
summarise_predictions(.data, conf = 0.95, summary_type = "quantile")
```

**Arguments**

- **.data**: a tibble of predictions returned by `predict_boots()`.
- **conf**: an integer between (0, 1) specifying the interval range.
- **summary_type**: currently fixed as "quantile". More options may be added in later releases.
Details

Generates a summary of predictions with a upper and lower interval range. Presently, the quantile() function from the {stats} package is used to determine the lower, 50th percentile, and upper interval ranges.

Value

Appends the tibble of predictions returned by predict_boots() with three new columns: .pred_lower, .pred, and .pred_upper.

Examples

```r
## Not run:
library(tidymodels)

# setup a workflow without fitting
wf <-
  workflow() %>%
  add_recipe(recipe(qsec ~ wt, data = mtcars)) %>%
  add_model(linear_reg())

# fit and predict 125 bootstrap resampled models to mtcars
set.seed(123)
preds <-
  wf %>%
  predict_boots(n = 125, training_data = mtcars, new_data = mtcars)

# append with prediction interval summary columns
preds %>%
  summarise_predictions(conf = 0.95)

## End(Not run)
```

vi_boots

Fit and estimate variable importance from a workflow using many bootstrap resamples.

Description

Generate a prediction interval from arbitrary model types using bootstrap resampling. predict_boots() generates n bootstrap resamples, fits a model to each resample (creating n models), then creates n estimates of variable importance for each variable in the model.

Usage

`vi_boots(workflow, n = 100, training_data, ...)`
vi_single_boot

Arguments

workflow An un-fitted workflow object.
n An integer for the number of bootstrap resampled models that will be created.
training_data A tibble or dataframe of data to be resampled and used for training.
... Additional params passed to rsample::bootstraps().

Details

Since vi_boots() fits a new model to each resample, the argument workflow must not yet be fit. Any tuned hyperparameters must be finalized prior to calling vi_boots().

Value

A tibble with a column indicating each variable in the model and a nested list of variable importances for each variable. The shape of the list may vary by model type. For example, linear models return two nested columns: the absolute value of each variable’s importance and the sign (POS/NEG), whereas XGboost models return a single nested column of variable importance. Similarly, the number of nested rows may vary by model type as some models may not utilize every possible predictor.

Examples

```r
## Not run:
library(tidymodels)

# setup a workflow without fitting
wf <-
  workflow() %>%
  add_recipe(recipe(qsec ~ wt, data = mtcars)) %>%
  add_model(linear_reg())

# fit and estimate variable importance from 125 bootstrap resampled models
set.seed(123)
wf %>%
  vi_boots(n = 125, training_data = mtcars)

## End(Not run)
```

vi_single_boot Fit a model and get the variable importance based on a single bootstrap resample

Description

Fit a model and get the variable importance based on a single bootstrap resample
Usage

   vi_single_boot(workflow, boot_splits, index)

Arguments

   workflow    An un-fitted workflow object.
   boot_splits A bootstrap split object created by rsample::bootstraps().
   index       Index of boot_splits to use for training
Index

predict_boots, 2
predict_single_boot, 3
summarise_predictions, 3
vi_boots, 4
vi_single_boot, 5