Package ‘sense’

September 6, 2021

**Type** Package

**Title** Automatic Stacked Ensemble for Regression Tasks

**Version** 1.0.0

**Author** Giancarlo Vercellino

**Maintainer** Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

**Description** Stacked ensemble for regression tasks based on ’mlr3’ framework with a pipeline for pre-processing numeric and factor features and hyper-parameter tuning using grid or random search.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**Depends** R (>= 4.1)

**Imports** mlr3 (>= 0.12.0), mlr3learners (>= 0.5.0), mlr3filters (>= 0.4.2), mlr3pipelines (>= 0.3.5-1), mlr3viz (>= 0.5.5), paradox (>= 0.7.1), mlr3tuning (>= 0.8.0), bbotk (>= 0.3.2), tictoc (>= 1.0.1), forcats (>= 0.5.1), readr (>= 2.0.1), lubridate (>= 1.7.10), purrr (>= 0.3.4), Metrics (>= 0.1.4), data.table (>= 1.14.0), visNetwork (>= 2.0.9)

**Suggests** xgboost (>= 1.4.1.1), rpart (>= 4.1-15), ranger (>= 0.13.1), kknn (>= 1.3.1), glmnet (>= 4.1-2), e1071 (>= 1.7-8), mlr3misc (>= 0.9.3), FSelectorRcpp (>= 0.3.8), care (>= 1.1.10), praznik (>= 8.0.0), lme4 (>= 1.1-27.1), nloptr (>= 1.2.2.2)

**URL** https://mlr3.mlr-org.com/

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-09-06 08:00:02 UTC

**R topics documented:**

- benchmark ................................................................. 2
- sense ................................................................. 2
Index

<table>
<thead>
<tr>
<th>benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>benchmark data set</td>
</tr>
</tbody>
</table>

Description

A data frame for regression task generated with mlbench friedman1.

Usage

benchmark

Format

A data frame with 11 columns and 150 rows.

Source

mlbench, friedman1

Description

Stacked ensamble for regression tasks based on 'mlr3' framework.

Usage

```
sense(
  df,
  target_feat,
  benchmarking = "all",
  super = "avg",
  algos = c("glmnet", "ranger", "xgboost", "rpart", "kknn", "svm"),
  sampling_rate = 1,
  metric = "mae",
  collapse_char_to = 10,
  num_preproc = "scale",
  fct_preproc = "one-hot",
  impute_num = "sample",
  missing_fusion = FALSE,
  inner = "holdout",
  outer = "holdout",
  folds = 3,
  repeats = 3,
)```
sense

```r
ratio = 0.5,
selected_filter = "information_gain",
selected_n_feats = NULL,
tuning = "random_search",
budget = 30,
resolution = 5,
n_evals = 30,
minute_time = 10,
patience = 0.3,
min_improve = 0.01,
java_mem = 64,
decimals = 2,
seed = 42
)
```

Arguments

- **df**: A data frame with features and target.
- **target_feat**: String. Name of the numeric feature for the regression task.
- **benchmarking**: Positive integer. Number of base learners to stack. Default: "all".
- **super**: String. Super learner of choice among the available learners. Default: "avg".
- **algos**: String vector. Available learners are: "glmnet", "ranger", "xgboost", "rpart", "kknn", "svm".
- **sampling_rate**: Positive numeric. Sampling rate before applying the stacked ensemble. Default: 1.
- **metric**: String. Evaluation metric for outer and inner cross-validation. Default: "mae".
- **collapse_char_to**: Positive integer. Conversion of characters to factors with predefined maximum number of levels. Default: 10.
- **num_preproc**: String. Options for scalar pre-processing: "scale" or "range". Default: "scale".
- **fct_preproc**: String. Options for factor pre-processing: "encodeimpact", "encodelmer", "one-hot", "treatment", "poly", "sum", "helmert". Default: "one-hot".
- **impute_num**: String. Options for missing imputation in case of numeric: "sample" or "hist". Default: "sample". For factor the default mode is Out-Of-Range.
- **missing_fusion**: String. Adding missing indicator features. Default: "FALSE".
- **inner**: String. Cross-validation inner cycle: "holdout", "cv", "repeated_cv", "subsampling". Default: "holdout".
- **outer**: String. Cross-validation outer cycle: "holdout", "cv", "repeated_cv", "subsampling". Default: "holdout".
- **folds**: Positive integer. Number of repetitions used in "cv" and "repeated_cv". Default: 3.
- **repeats**: Positive integer. Number of repetitions used in "subsampling" and "repeated_cv". Default: 3.
ratio Positive numeric. Percentage value for "holdout" and "subsampling". Default: 0.5.

selected_filter String. Filters available for regression tasks: "carscore", "cmim", "correlation", "find_correlation", "information_gain", "relief", "variance". Default: "information_gain".

selected_n_feats Positive integer. Number of features to select through the chosen filter. Default: NULL.

tuning String. Available options are "random_search" and "grid_search". Default: "random_search".


resolution Positive integer. Grid resolution for each hyper-parameter. Default: 5.


patience Positive numeric. Percentage of stagnating evaluations before termination. Default: 0.3.

min_improve Positive numeric. Minimum error improvement required before termination. Default: 0.01.

java_mem Positive integer. Memory allocated to Java. Default: 64.


seed Positive integer. Default: 42.

Value

This function returns a list including:

• benchmark_error: comparison between the base learners
• resampled_model: mlr3 standard description of the analytic pipeline.
• plot: mlr3 standard graph of the analytic pipeline.
• selected_n_feats: selected features and score according to the filtering method used.
• model_error: error measure for outer cycle of cross-validation.
• testing_frame: data set used for calculating the test metrics.
• test_metrics: metrics reported are mse, rmse, mae, mdae, rae, rse, rrse, smape.
• model_predict: prediction function to apply to new data on the same scheme.
• time_log: computation time.

Author(s)

Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

See Also

Useful links:

• https://mlr3.mlr-org.com/
Examples

```r
## Not run:
sense(benchmark, "y", algos = c("glmnet", "rpart"))
```

```r
## End(Not run)
```
Index

* datasets
  benchmark, 2

benchmark, 2

sense, 2

sense-package (sense), 2