

Package ‘xspliner’

December 21, 2018

Title Assisted Model Building, using Surrogate Black-Box Models to Train Interpretable Spline Based Additive Models

Version 0.0.2

Description Builds generalized linear model with automatic data transformation.

The 'xspliner' helps to build simple, interpretable models that inherits informations provided by more complicated ones.

The resulting model may be treated as explanation of provided black box, that was supplied prior to the algorithm.

Depends R (>= 3.0)

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Imports stats, pdp, dplyr, ggplot2, mgcv, magrittr, purrr, tidyr

Suggests ALEPlot, factorMerger, testthat, knitr, rmarkdown, ResourceSelection, randomForest, e1071, caret, breakDown, DALEX, xgboost, gridExtra, grid

VignetteBuilder knitr

URL <https://ModelOriented.github.io/xspliner/>

BugReports <https://github.com/ModelOriented/xspliner/issues>

NeedsCompilation no

Author Krystian Igras [aut, cre],
Przemyslaw Biecek [aut, ths]

Maintainer Krystian Igras <krystian8207@gmail.com>

Repository CRAN

Date/Publication 2018-12-21 15:20:03 UTC

R topics documented:

xspliner-package	2
approx_with_spline	2
build_xspliner	3
log_msg	4
plot.xspliner	4
predict.xspliner	5
print.xspliner	5
stats	6
summary.xspliner	6
transition	7
xf_opts_default	7
xspline	8

Index **10**

xspliner-package	<i>Easy way for approximating data with splines.</i>
------------------	--

Description

Easy way for approximating data with splines.

approx_with_spline	<i>Approximate spline on data</i>
--------------------	-----------------------------------

Description

It approximates data with spline function by fitting GAM model.

Usage

```
approx_with_spline(effect_data, response, predictor,
  env = parent.frame(), ...)
```

```
approx_with_monotonic_spline(effect_data, response, predictor,
  env = parent.frame(), monotonic, ...)
```

Arguments

effect_data	Black box response data, for example pdp curve.
response	Name of response value from effect_data.
predictor	Name of predictor value from effect_data.
env	Formula environment that should be used for fitting gam model.
...	Other arguments passed to <code>s</code> function.
monotonic	Possible options "up", "down" and "auto". If up the spline is increasing, when down decreasing.

Value

Object of class "gam". See [gamObject](#)

Examples

```
x <- sort(rnorm(20, 5, 5))
y <- rnorm(20, 2, 2)
env <- new.env()
approx_with_spline(data.frame(x = x, y = y), "y", "x", env)

approx_with_monotonic_spline(data.frame(x = x, y = y), "y", "x", env, "up")
```

build_xspliner	<i>Helper function for building GLM object with transformed variables.</i>
----------------	--

Description

Helper function for building GLM object with transformed variables.

Usage

```
build_xspliner(formula, model, data, xf_opts = xf_opts_default,
  xs_opts = xs_opts_default, link = "identity", family = "gaussian",
  env = parent.frame(), compare_stat = aic)
```

Arguments

formula	xspliner-specific formula object. Check vignette("xspliner") for more details.
model	Predictive model. Basic model used for extracting predictors transformation.
data	Training data of model.
xf_opts	Formula parameters used for factor variable transformations inherited from factorMerger package.
xs_opts	Predictive model response method and approximation parameters used for quantitative.
link	Link function that should be used in final model. The passed is used when cannot be extracted from model. By default 'identity'. See family for possibilities.
family	Family of response variable that should be used in final model. The passed is used when cannot be extracted from model. By default 'gaussian'. See family for possibilities.
env	Environment in which optional variables passed into parameters are stored.
compare_stat	Function of linear model (lm function output). Statistic that measures if linear model is better that transformed one. See stats .

log_msg	<i>Helper function to print out log messages</i>
---------	--

Description

Helper function to print out log messages

Usage

```
log_msg(message)
```

Arguments

message	Message that should be printed to R console
---------	---

plot.xspliner	<i>Plot method for 'xspliner' model</i>
---------------	---

Description

Plot method for 'xspliner' model

Usage

```
## S3 method for class 'xspliner'
plot(x, variable_names = NULL, model = NULL,
     plot_response = TRUE, plot_approx = TRUE, data = NULL,
     plot_data = FALSE, plot_deriv = FALSE, n_plots = 6,
     sort_by = NULL, compare_with = list(),
     prediction_funs = list(function(object, newdata) predict(object,
     newdata)), ...)
```

Arguments

x	Object of class 'xspliner'
variable_names	Names of predictors which transitions should be plotted.
model	Base model that xspliner is based on.
plot_response	If TRUE black box model response is drawn.
plot_approx	If TRUE black box model response approximation is drawn.
data	Training data used for building x model. Required for plot_data option and model comparing.
plot_data	If TRUE raw data is drawn.
plot_deriv	If TRUE derivative of approximation is showed on plot.

n_plots	Threshold for number of plots when plotting all variables.
sort_by	When comparing models determines according to which model should observations be ordered.
compare_with	Named list. Other models that should be compared with xspliner and model.
prediction_funs	Prediction functions that should be used in model comparison.
...	Another arguments passed into model specific method.

predict.xspliner *Predict xspliner method*

Description

Predict xspliner method

Usage

```
## S3 method for class 'xspliner'
predict(object, newdata, ...)
```

Arguments

object	Object of class 'xspliner'.
newdata	Data that should be prediction based on.
...	Another arguments passed into predict.glm method.

print.xspliner *Print method for xspliner object*

Description

Print method for xspliner object

Usage

```
## S3 method for class 'xspliner'
print(x, predictor, ...)
```

Arguments

x	xspliner object
predictor	predictor for xspliner model formula
...	Another arguments passed into model specific print method.

stats	<i>Statistics used for better linear model selection</i>
-------	--

Description

Used as compare_stat parameter in xspline method. Each function has attribute "higher-better". If "higher-better" is TRUE then model with higher statistic value is treated as better one.

Usage

```
aic(glm_model)
hoslem(glm_model)
```

Arguments

glm_model Linear model - glm function output.

summary.xspliner	<i>Summary method for xspliner object</i>
------------------	---

Description

Summary method for xspliner object

Usage

```
## S3 method for class 'xspliner'
summary(object, predictor, ...)
```

Arguments

object xspliner object
predictor predictor for xspliner model formula
... Another arguments passed into model specific method.

transition	<i>Extract variable transformation from xspliner</i>
------------	--

Description

Extract variable transformation from xspliner

Usage

```
transition(model, ...)

## S3 method for class 'xspliner'
transition(model, predictor, type = "function", ...)
```

Arguments

model	xspliner model
...	Other parameters passed to method. Currently not used.
predictor	variable name for which transformation should be extracted
type	If 'function' then transformation function is extracted. For 'base' there is sourced object on which transformation was built - in case of quantitative variable GAM model, for qualitative factorMerger.

xf_opts_default	<i>Default parameters for transition methods</i>
-----------------	--

Description

While constructing formula interpreted by xspliner package, some parameters may be specified within xs(..) or xf(..) symbols. Below are default parameters. See details in vignette("xspliner")

Usage

```
xf_opts_default
xs_opts_default
```

Format

An object of class list of length 2.

xspline

*Builds predictive model based GLM.***Description**

The method provides main functionality on building GLM models with automatic variables transformation. The transformations are based on specified single variable responses for selected black-box model. See details in vignette("xspliner").

Usage

```
xspline(object, ...)

model_surrogate_xspliner(object, ...)

## Default S3 method:
xspline(object, lhs = NULL, response = NULL,
        predictors = NULL, data = NULL, form = "additive", bare = NULL,
        env = parent.frame(), ...)

## S3 method for class 'formula'
xspline(object, model, data = NULL,
        consider = "specials", env = parent.frame(), ...)

## S3 method for class 'explainer'
xspline(object, env = parent.frame(), ...)
```

Arguments

object	Predictive model, formula or explainer (see DALEX) object.
...	Other arguments passed to xspline methods or build_xspliner .
lhs	Left-hand side of model formula. Can be transformed response.
response	Name of response variable of model.
predictors	Predictor values that should be used in final model.
data	Training data of model.
form	Can be 'additive' (default) or 'multiplicative'. Specifies formula form in final model.
bare	Variable names that mustn't be transformed in final model.
env	Environment in which optional variables passed into parameters are stored. variables transformation. See vignette("xspliner") for details.
model	When object is formula - predictive model. Basic model used for extracting predictors transformation.
consider	One of c("specials", "all"). If "specials", only components with xs or xf call are considered in transition.

Details

`model_surrogate_xspliner` is a wrapper of `xspline` method to assure consistency with <https://github.com/ModelOriented/DrWH> tools

Value

GLM object of class `'xspliner'`.

Examples

```
# preparing blackbox model
library(randomForest)
rf_iris <- randomForest(
  Petal.Width ~ Sepal.Length + Petal.Length + Species,
  data = iris)

# formula based xspliner
xs_iris <- xspline(
  Petal.Width ~ xs(Sepal.Length) + xs(Petal.Length) + xf(Species),
  model = rf_iris)
summary(xs_iris)
plot(xs_iris, "Sepal.Length")

# passing just the model
xs_iris <- xspline(rf_iris)
summary(xs_iris)
plot(xs_iris, "Sepal.Length")

# using DALEX
library(DALEX)
xs_iris_explainer <- explain(rf_iris)
xs_iris <- xspline(rf_iris)
summary(xs_iris)
plot(xs_iris, "Sepal.Length")
```

Index

*Topic **datasets**

- [xf_opts_default](#), 7
- [aic \(stats\)](#), 6
- [approx_with_monotonic_spline](#)
 - [\(approx_with_spline\)](#), 2
- [approx_with_spline](#), 2
- [build_xspliner](#), 3, 8
- [family](#), 3
- [gamObject](#), 3
- [hoslem \(stats\)](#), 6
- [log_msg](#), 4
- [model_surrogate_xspliner \(xspline\)](#), 8
- [plot.xspliner](#), 4
- [predict.glm](#), 5
- [predict.xspliner](#), 5
- [print.xspliner](#), 5
- [s](#), 2
- [stats](#), 3, 6
- [summary.xspliner](#), 6
- [transition](#), 7
- [xf_opts_default](#), 7
- [xs_opts_default \(xf_opts_default\)](#), 7
- [xspline](#), 8
- [xspliner-package](#), 2