Package 'SplitWise'

July 29, 2025

Type Package

Title Hybrid Stepwise Regression with Single-Split Dummy Encoding

Version 1.0.2

Description Implements 'SplitWise', a hybrid regression approach that transforms numeric variables into either single-split (0/1) dummy variables or retains them as continuous predictors. The transformation is followed by stepwise selection to identify the most relevant variables. The default 'iterative' mode adaptively explores partial synergies among variables to enhance model performance, while an alternative 'univariate' mode applies simpler transformations independently to each predictor. For details, see Kurbucz et al. (2025) <https://arxiv.org/abs/2505.15423>.

License GPL (>= 3)

Encoding UTF-8

Depends R (>= 3.5.0)

Imports rpart,
stats

LazyData true

RoxygenNote 7.3.2

Suggests knitr,
rmarkdown,
testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

R topics documented:

splitwise *SplitWise Regression*

Description

Transforms each numeric variable into either a single-split dummy or keeps it linear, then runs `stats::step()` for stepwise selection. The user can choose a simpler univariate transformation or an iterative approach.

Usage

```
splitwise(
  formula,
  data,
  transformation_mode = c("iterative", "univariate"),
  direction = c("backward", "forward", "both"),
  min_support = 0.1,
  min_improvement = 3,
  criterion = c("AIC", "BIC"),
  exclude_vars = NULL,
  verbose = FALSE,
  steps = 1000,
  k = 2,
  ...
)

## S3 method for class 'splitwise_lm'
print(x, ...)

## S3 method for class 'splitwise_lm'
summary(object, ...)

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

## S3 method for class 'splitwise_lm'
coef(object, ...)

## S3 method for class 'splitwise_lm'
fitted(object, ...)

## S3 method for class 'splitwise_lm'
residuals(object, ...)

## S3 method for class 'splitwise_lm'
model.matrix(object, ...)
```

Arguments

`formula`	A formula specifying the response and (initial) predictors, e.g. `mpg ~ .`.
`data`	A data frame containing the variables used in `formula`.

`transformation_mode`
: Either `"iterative"` or `"univariate"`. Default = `"iterative"`.

`direction`
: Stepwise direction: `"backward"`, `"forward"`, or `"both"`.

`min_support`
: Minimum fraction (between 0 and 0.5) of observations needed in either group when making a dummy split. Prevents over-fragmented or tiny dummy groups. Default = `0.1`.

`min_improvement`
: Minimum required improvement (in AIC/BIC units) for accepting a dummy split or variable transformation. Helps guard against overfitting from marginal improvements. Default = `2`.

`criterion`
: Either `"AIC"` or `"BIC"`. Default = `"AIC"`. **Note**: If you choose `"BIC"`, you typically want `k = log(nrow(data))` in stepwise.

`exclude_vars`
: A character vector naming variables that should be forced to remain linear (i.e., no dummy splits allowed). Default = `NULL`.

`verbose`
: Logical; if `TRUE`, prints debug info in transformation steps. If `FALSE`, the stepwise selection process is run quietly (trace = 0 in `step()`). Default = `FALSE`.

`steps`
: Maximum number of steps for `step()`. Default = 1000.

`k`
: Penalty multiple for the number of degrees of freedom (used by `step()`). E.g. 2 for AIC, `log(n)` for BIC. Default = 2.

`...`
: Additional arguments passed to `predict.lm`.

`x`
: A `"splitwise_lm"` object returned by `splitwise`.

`object`
: An object of class `splitwise_lm`, as returned by `splitwise`.

`newdata`
: A data frame of new data (with original predictors) to generate predictions for. The appropriate dummy variables will be generated using the transformation rules learned during model training. If omitted, predictions for the training data are returned.

Value

An S3 object of class `c("splitwise_lm", "lm")`, storing:

`splitwise_info`
: List containing transformation decisions, final data, and call.

Functions

- `print(splitwise_lm)`: Prints a summary of the splitwise_lm object.
- `summary(splitwise_lm)`: Provides a detailed summary, including how dummies were created.
- `predict(splitwise_lm)`: Generate predictions from a `splitwise_lm` object using learned transformation rules.
- `coef(splitwise_lm)`: Extract model coefficients from a SplitWise linear model.
- `fitted(splitwise_lm)`: Extract fitted values from a SplitWise linear model.
- `residuals(splitwise_lm)`: Extract residuals from a SplitWise linear model.
- `model.matrix(splitwise_lm)`: Extract the model matrix from a SplitWise linear model.

Examples

```
# Load the mtcars dataset
data(mtcars)

# Univariate transformations (AIC-based, backward stepwise)
model_uni <- splitwise(
  mpg ~ .,
  data               = mtcars,
  transformation_mode = "univariate",
  direction           = "backward"
)
summary(model_uni)

# Iterative approach (BIC-based, forward stepwise)
# Note: typically set k = log(nrow(mtcars)) for BIC in step().
model_iter <- splitwise(
  mpg ~ .,
  data               = mtcars,
  transformation_mode = "iterative",
  direction           = "forward",
  criterion           = "BIC",
  k                   = log(nrow(mtcars))
)
summary(model_iter)
```

Index