Package ‘mob’

July 6, 2020

Title Monotonic Optimal Binning

Version 0.2.1

Description Generate the monotonic binning and perform the woe (weight of evidence) transformation for the logistic regression used in the consumer credit scorecard development. The woe transformation is a piecewise transformation that is linear to the log odds. For a numeric variable, all of its monotonic functional transformations will converge to the same woe transformation.

License GPL (>= 2)

URL https://github.com/statcompute/mob

Author WenSui Liu

Maintainer WenSui Liu <liuwensui@gmail.com>

Depends R (>= 3.3.3)

Imports stats, gbm, Rborist

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2020-07-06 09:20:20 UTC

R topics documented:

arb_bin ................................................................. 2
bad_bin ............................................................... 2
cal_woe ............................................................... 3
gbm_bin ............................................................... 4
hmeq ................................................................. 4
iso_bin ............................................................... 5
knn_bin .............................................................. 6
qcut ................................................................. 6
qtl_bin .............................................................. 7
rng_bin .............................................................. 7
arb_bin

Monotonic binning based on decision tree model

Description

The function arb_bin implements the monotonic binning based on the decision tree.

Usage

arb_bin(x, y)

Arguments

x A numeric vector
y A numeric vector with 0/1 binary values

Value

A list of binning outcomes, including a numeric vector with cut points and a dataframe with binning summary

Examples

data(hmeq)
arb_bin(hmeq$DEROG, hmeq$BAD)

bad_bin

Monotonic binning by quantile with cases Y = 1

Description

The function bad_bin implements the quantile-based monotonic binning by the iterative discretization based on cases with Y = 1.

Usage

bad_bin(x, y)

Arguments

x A numeric vector
y A numeric vector with 0/1 binary values
Value

A list of binning outcomes, including a numeric vector with cut points and a dataframe with binning summary.

Examples

data(hmeq)
bad_bin(hmeq$DEROG, hmeq$BAD)

Description

The function `cal_woe` applies the WoE transformation to a numeric vector based on the binning outcome from a binning function, e.g. `qtl_bin()` or `iso_bin()`.

Usage

`cal_woe(x, bin)`

Arguments

`x`  
A numeric vector that will be transformed to WoE values.

`bin`  
A list with the binning outcome from the binning function, e.g. `qtl_bin()` or `iso_bin()`.

Value

A numeric vector with WoE transformed values.

Examples

```r
data(hmeq)
bin_out <- qtl_bin(hmeq$DEROG, hmeq$BAD)
cal_woe(hmeq$DEROG[1:10], bin_out)
```
Monotonic binning based on generalized boosted model

**Description**

The function `gbm_bin` implements the monotonic binning based on the generalized boosted model (GBM).

**Usage**

```r
gbm_bin(x, y)
```

**Arguments**

- `x` A numeric vector
- `y` A numeric vector with 0/1 binary values

**Value**

A list of binning outcomes, including a numeric vector with cut points and a dataframe with binning summary

**Examples**

```r
data(hmeq)
gbm_bin(hmeq$DEROG, hmeq$BAD)
```

**Description**

A dataset containing characteristics and delinquency information for 5,960 home equity loans.
Format

A data frame with 5960 rows and 13 variables:

- **BAD** indicator of applicant defaulted on loan or seriously delinquent
- **LOAN** Amount of the loan request, in dollar
- **MORTDUE** Amount due on existing mortgage, in dollar
- **VALUE** Value of current property, in dollar
- **REASON** DebtCon = debt consolidation; HomeImp = home improvement
- **JOB** Occupational categories
- **YOJ** Years at present job
- **DEROG** Number of major derogatory reports
- **DELINQ** Number of delinquent credit lines
- **CLAGE** Age of oldest credit line in months
- **NINQ** Number of recent credit inquiries
- **CLNO** Number of credit lines
- **DEBTINC** Debt-to-income ratio

Source

http://www.creditriskanalytics.net/datasets-private2.html

---

iso_bin

*Monotonic binning based on isotonic regression*

Description

The function `iso_bin` implements the monotonic binning based on the isotonic regression.

Usage

```r
iso_bin(x, y)
```

Arguments

- **x** A numeric vector
- **y** A numeric vector with 0/1 binary values

Value

A list of binning outcomes, including a numeric vector with cut points and a dataframe with binning summary

Examples

```r
data(hmeq)
iso_bin(hmeq$DEROG, hmeq$BAD)
```
**kmn_bin**  
*Monotonic binning based on k-means clustering*

**Description**

The function `kmn_bin` implements the monotonic binning based on the k-means clustering.

**Usage**

```r
kmn_bin(x, y)
```

**Arguments**

- `x` A numeric vector
- `y` A numeric vector with 0/1 binary values

**Value**

A list of binning outcomes, including a numeric vector with cut points and a dataframe with binning summary.

**Examples**

```r
data(hmeq)
kmn_bin(hmeq$DEROG, hmeq$BAD)
```

---

**qcut**  
*Discretizing a numeric vector*

**Description**

The function `qcut` discretizes a numeric vector into N pieces based on quantiles.

**Usage**

```r
qcut(x, n)
```

**Arguments**

- `x` A numeric vector.
- `n` An integer indicating the number of categories to discretize.

**Value**

A numeric vector to divide the vector `x` into `n` categories.
```r
Examples
x <- 1:10
# [1] 1 2 3 4 5 6 7 8 9 10
v <- qcut(1:10, 4)
# [1] 3 5 8
findInterval(x, sort(c(v, -Inf, Inf)), left.open = TRUE)
# [1] 1 1 1 2 2 3 3 3 4 4
```

---

**qtl_bin**

*Monotonic binning by quantile*

**Description**

The function `qtl_bin` implements the quantile-based monotonic binning by the iterative discretization.

**Usage**

```r
qtl_bin(x, y)
```

**Arguments**

- `x` A numeric vector
- `y` A numeric vector with 0/1 binary values

**Value**

A list of binning outcomes, including a numeric vector with cut points and a dataframe with binning summary.

**Examples**

```r
data(hmeq)
qtl_bin(hmeq$DEROG, hmeq$BAD)
```

---

**rng_bin**

*Monotonic binning by quantile based on value range*

**Description**

The function `rng_bin` implements the quantile-based monotonic binning by the iterative discretization based on the equal-width range of values.

**Usage**

```r
rng_bin(x, y)
```
**rng_bin**

**Arguments**

- **x**  A numeric vector
- **y**  A numeric vector with 0/1 binary values

**Value**

A list of binning outcomes, including a numeric vector with cut points and a dataframe with binning summary

**Examples**

```r
data(hmeq)
rng_bin(hmeq$DEROG, hmeq$BAD)
```
Index

* datasets
  hmeq, 4
  arb_bin, 2
  bad_bin, 2
  cal_woe, 3
  gbm_bin, 4
  hmeq, 4
  iso_bin, 5
  kmm_bin, 6
  qcut, 6
  qtl_bin, 7
  rng_bin, 7