

# Package ‘RSDA’

October 12, 2022

**Type** Package

**Title** R to Symbolic Data Analysis

**Version** 3.0.13

**Date** 2022-07-15

**Description** Symbolic Data Analysis (SDA) was proposed by professor Edwin Diday in 1987, the main purpose of SDA is to substitute the set of rows (cases) in the data table for a concept (second order statistical unit). This package implements, to the symbolic case, certain techniques of automatic classification, as well as some linear models.

**License** GPL (>= 2)

**Encoding** UTF-8

**Depends** R (>= 3.4)

**URL** <http://www.oldemarrodriguez.com>

**Suggests** testthat (>= 2.1.0), knitr, rmarkdown

**RoxygenNote** 7.2.0

**Imports** vctrs (>= 0.2.4), dplyr (>= 0.8.5), forcats, scales, stringr, rlang (>= 0.4.5), purrr, magrittr, tidyselect, tibble (>= 3.0.0), stats, RJSONIO, XML, ggplot2, ggpolypath, reshape, glmnet, FactoMineR, princurve, nloptr, sqldf, randomcoloR, kknn, e1071, gbm, randomForest, rpart, neuralnet, umap, tsne

**VignetteBuilder** knitr

**LazyData** true

**NeedsCompilation** no

**Author** Oldemar Rodriguez [aut, cre],  
Jose Emmanuel Chacon [cph],  
Carlos Aguero [cph],  
Jorge Arce [cph]

**Maintainer** Oldemar Rodriguez <oldemar.rodriguez@ucr.ac.cr>

**Repository** CRAN

**Date/Publication** 2022-07-16 07:30:37 UTC

**R topics documented:**

abalone	4
as.data.frame.symbolic_histogram	5
as.data.frame.symbolic_interval	5
as.data.frame.symbolic_modal	6
as.data.frame.symbolic_set	6
calc.burt.sym	7
Cardiological	7
cardiologicalv2	8
centers.interval	8
classic.to.sym	9
cor	10
cov	11
deter.coefficient	12
dist.vect	13
dist.vect.matrix	14
ex1_db2so	15
example1	15
example2	17
example3	17
example4	18
example5	18
example6	19
example7	20
ex_cfa1	20
ex_cfa2	21
ex_mcfa1	21
ex_mcfa2	22
facedata	23
format.symbolic_histogram	24
format.symbolic_interval	24
format.symbolic_modal	25
format.symbolic_set	25
get_cats	26
get_props	26
interval.centers	26
interval.histogram.plot	27
interval.max	27
interval.min	28
interval.ranges	28
int_prost_test	28
int_prost_train	29
is.sym.histogram	29
is.sym.interval	30
is.sym.modal	30
is.sym.set	31
lynne1	31

mcfa.scatterplot . . . . .	32
mean.symbolic_interval . . . . .	32
median.symbolic_interval . . . . .	33
method_summary . . . . .	34
min.symbolic_interval . . . . .	34
neighbors.vertex . . . . .	35
norm.vect . . . . .	36
oils . . . . .	37
plot.symbolic_tbl . . . . .	37
plot.sym_tsne . . . . .	39
plot.sym_umap . . . . .	39
R2.L . . . . .	40
R2.U . . . . .	41
read.sym.table . . . . .	42
RMSE.L . . . . .	44
RMSE.U . . . . .	45
RSDA . . . . .	46
sd . . . . .	47
SDS.to.RSDA . . . . .	48
SODAS.to.RSDA . . . . .	49
sym.circle.plot . . . . .	50
sym.dist.interval . . . . .	51
sym.gbm . . . . .	52
sym.glm . . . . .	53
sym.histogram . . . . .	54
sym.interval . . . . .	54
sym.interval.pc . . . . .	55
sym.interval.pc.limits . . . . .	56
sym.kmeans . . . . .	57
sym.knn . . . . .	59
sym.lm . . . . .	60
sym.mcfa . . . . .	61
sym.modal . . . . .	62
sym.nnet . . . . .	62
sym.pca . . . . .	63
sym.predict . . . . .	65
sym.predict.symbolic_gbm_cm . . . . .	67
sym.predict.symbolic_gbm_crm . . . . .	68
sym.predict.symbolic_knn_cm . . . . .	68
sym.predict.symbolic_knn_crm . . . . .	69
sym.predict.symbolic_nnet_cm . . . . .	70
sym.predict.symbolic_nnet_crm . . . . .	70
sym.predict.symbolic_rf_cm . . . . .	71
sym.predict.symbolic_rf_crm . . . . .	72
sym.predict.symbolic_rt_cm . . . . .	72
sym.predict.symbolic_rt_crm . . . . .	73
sym.predict.symbolic_svm_cm . . . . .	74
sym.predict.symbolic_svm_crm . . . . .	74

sym.rf . . . . .	75
sym.rt . . . . .	76
sym.scatterplot . . . . .	77
sym.set . . . . .	78
sym.svm . . . . .	79
sym.tsne.symbolic_tbl . . . . .	80
sym.umap.symbolic_tbl . . . . .	81
sym.var . . . . .	81
USCrime . . . . .	83
uscrime_int . . . . .	84
uscrime_intv2 . . . . .	85
var . . . . .	85
variance.princ.curve . . . . .	86
vertex.interval . . . . .	87
VeterinaryData . . . . .	88
write.sym.table . . . . .	89
\$.symbolic_histogram . . . . .	90
\$.symbolic_modal . . . . .	90
\$.symbolic_set . . . . .	91

**Index** **92**

---

abalone	<i>SODAS XML data file.</i>
---------	-----------------------------

---

**Description**

Example of SODAS XML data file converted in a CSV file in RSDA format.

**Usage**

data(abalone)

**Format**

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 24 rows and 7 columns.

**Source**

<http://www.info.fundp.ac.be/asso/sodaslink.htm>

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

### Examples

```
data(abalone)
res <- sym.pca(abalone, 'centers')
plot(res, choix = "ind")
plot(res, choix = "var")
```

---

```
as.data.frame.symbolic_histogram
  a data.frame
```

---

### Description

a data.frame

### Usage

```
## S3 method for class 'symbolic_histogram'
as.data.frame(x, ...)
```

### Arguments

x	.....
...	...

---

```
as.data.frame.symbolic_interval
  convertir a data.frame
```

---

### Description

convertir a data.frame

### Usage

```
## S3 method for class 'symbolic_interval'
as.data.frame(x, ...)
```

### Arguments

x	a symbolic interval vector
...	further arguments passed to or from other methods.

---

```
as.data.frame.symbolic_modal
      Extract values
```

---

**Description**

Extract values

**Usage**

```
## S3 method for class 'symbolic_modal'
as.data.frame(x, ...)
```

**Arguments**

x                    An object to be converted  
...                   Further arguments to be passed from or to other methods.

---

```
as.data.frame.symbolic_set
      convertir a data.frame
```

---

**Description**

convertir a data.frame

**Usage**

```
## S3 method for class 'symbolic_set'
as.data.frame(x, ...)
```

**Arguments**

x                    a symbolic interval vector  
...                   further arguments passed to or from other methods.

---

calc.burt.sym	<i>Burt Matrix</i>
---------------	--------------------

---

**Description**

Burt Matrix

**Usage**

```
calc.burt.sym(sym.data, pos.var)
```

**Arguments**

sym.data	ddd
pos.var	ddd

---

Cardiological	<i>Cardiological data example</i>
---------------	-----------------------------------

---

**Description**

Cardiological interval data example.

**Usage**

```
data(Cardiological)
```

**Format**

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 11 rows and 3 columns.

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(Cardiological)
res.cm <- sym.lm(formula = Pulse~Syst+Diast, sym.data = Cardiological, method = 'cm')
pred.cm <- sym.predict(res.cm, Cardiological)
RMSE.L(Cardiological$Pulse, pred.cm$Fitted)
RMSE.U(Cardiological$Pulse, pred.cm$Fitted)
R2.L(Cardiological$Pulse, pred.cm$Fitted)
R2.U(Cardiological$Pulse, pred.cm$Fitted)
deter.coefficient(Cardiological$Pulse, pred.cm$Fitted)
```

---

cardiologicalv2      *Cardiological data example*

---

**Description**

Cardiological interval data example.

**Usage**

```
data(Cardiological)
```

**Format**

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 44 rows and 5 columns.

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

---

centers.interval      *Compute centers of the interval*

---

**Description**

Compute centers of the interval

**Usage**

```
centers.interval(sym.data)
```

**Arguments**

`sym.data`      Symbolic interval data table.

**Value**

Centers of teh intervals.

**Author(s)**

Jorge Arce.



## References

- Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.
- Hastie, T. (1984). Principal Curves and Surface. Ph.D Thesis Stanford University.
- Hastie, T. & Weingessel, A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension. R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.
- Hastie, T. & Stuetzle, W. (1989). Principal Curves. Journal of the American Statistical Association, Vol. 84-406, 502–516.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning; Data Mining, Inference and Prediction. Springer, New York.

## See Also

sym.interval.pc

---

classic.to.sym	<i>Generate a symbolic data frame</i>
----------------	---------------------------------------

---

## Description

Generate a symbolic data table from a classic data table.

## Usage

```
classic.to.sym(
  x = NULL,
  concept = NULL,
  variables = tidyselect::everything(),
  default.numeric = sym.interval,
  default.categorical = sym.modal,
  ...
)
```

## Arguments

x	A data.frame.
concept	These are the variable that we are going to use a concepts.
variables	These are the variables that we want to include in the symbolic data table.
default.numeric	function to use for numeric variables
default.categorical	function to use for categorical variables
...	A vector with names and the type of symbolic data to use, the available types are type_histogram (), type_continuous (), type.set (), type.modal (), by default type_histogram () is used for numeric variables and type_modal () for the categorical variables.

**Value**

a [tibble][[tibble::tibble-package]

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

---

cor

*Generic function for the correlation*

---

**Description**

This function compute the symbolic correlation

**Usage**

```
cor(x, ...)

## Default S3 method:
cor(
  x,
  y = NULL,
  use = "everything",
  method = c("pearson", "kendall", "spearman"),
  ...
)

## S3 method for class 'symbolic_interval'
cor(x, y, method = c("centers", "billard"), ...)

## S3 method for class 'symbolic_tbl'
cor(x, ...)
```

**Arguments**

x	A symbolic variable.
...	As in R cor function.
y	A symbolic variable.
use	An optional character string giving a method for computing covariances in the presence of missing values. This must be (an abbreviation of) one of the strings 'everything', 'all.obs', 'complete.obs', 'na.or.complete', or 'pairwise.complete.obs'.
method	The method to be use.

**Value**

Return a real number in [-1,1].

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

---

 cov

*Generic function for the covariance*


---

**Description**

This function compute the symbolic covariance.

**Usage**

```

cov(x, ...)

## Default S3 method:
cov(
  x,
  y = NULL,
  use = "everything",
  method = c("pearson", "kendall", "spearman"),
  ...
)

## S3 method for class 'symbolic_interval'
cov(x, y, method = c("centers", "billard"), na.rm = FALSE, ...)

## S3 method for class 'symbolic_tbl'
cov(x, ...)

```

**Arguments**

x	First symbolic variables.
...	As in R cov function.
y	Second symbolic variables.
use	an optional character string giving a method for computing covariances in the presence of missing values. This must be (an abbreviation of) one of the strings 'everything', 'all.obs', 'complete.obs', 'na.or.complete', or 'pairwise.complete.obs'.
method	The method to be use.
na.rm	As in R cov function.

**Value**

Return a real number.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

---

`deter.coefficient`      *Compute the determination coefficient*

---

**Description**

The determination coefficient represents a goodness-of-fit measure commonly used in regression analysis to capture the adjustment quality of a model.

**Usage**

```
deter.coefficient(ref, pred)
```

**Arguments**

<code>ref</code>	Variable that was predicted.
<code>pred</code>	The prediction given by the model.

**Value**

Return the determination coefficient.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. Computational Statistics and Data Analysis 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54, 333-347.

**See Also**

sym.glm

**Examples**

```
data(int_prost_test)
data(int_prost_train)
res.cm <- sym.lm(lpsa ~ ., sym.data = int_prost_train, method = "cm")
pred.cm <- sym.predict(res.cm, int_prost_test)
deter.coefficient(int_prost_test$lpsa, pred.cm$Fitted)
```

---

dist.vect

*Compute a distance vector*

---

**Description**

Compute a distance vector

**Usage**

```
dist.vect(vector1, vector2)
```

**Arguments**

vector1	First vector.
vector2	Second vector.

**Value**

Eclidean distance between the two vectors.

**Author(s)**

Jorge Arce

**References**

- Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.
- Hastie, T. (1984). Principal Curves and Surface. Ph.D. Thesis Stanford University.
- Hastie, T. & Weingessel, A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension. R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.
- Hastie, T. & Stuetzle, W. (1989). Principal Curves. Journal of the American Statistical Association, Vol. 84-406, 502–516.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, New York.

**See Also**

sym.interval.pc

---

dist.vect.matrix      *Compute the distance vector matrix*

---

**Description**

Compute the distance vector matrix.

**Usage**

```
dist.vect.matrix(vector, Matrix)
```

**Arguments**

vector	An n dimensional vector.
Matrix	An n x n matrix.

**Value**

The distance.

**Author(s)**

Jorge Arce.

**References**

- Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.
- Hastie, T. (1984). Principal Curves and Surface. Ph.D Thesis Stanford University.
- Hastie, T. & Weingessel, A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension. R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.
- Hastie, T. & Stuetzle, W. (1989). Principal Curves. Journal of the American Statistical Association, Vol. 84-406, 502–516.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning; Data Mining, Inference and Prediction. Springer, New York.

**See Also**

sym.interval.pc

---

`ex1_db2so`*Data example to generate symbolic objects*

---

**Description**

This is a small data example to generate symbolic objects.

**Usage**

```
data(ex1_db2so)
```

**Format**

An object of class `data.frame` with 19 rows and 5 columns.

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

**Examples**

```
data(ex1_db2so)
ex1 <- ex1_db2so
result <- classic.to.sym(
  x = ex1_db2so,
  concept = c(state, sex),
  variables = c(county, group, age),
  county = mean(county),
  age_hist = sym.histogram(age, breaks = pretty(ex1_db2so$age, 5))
)
result
```

---

`example1`*Data Example 1*

---

**Description**

This is a symbolic data table with variables of continuous, interval, histogram and set types.

**Usage**

```
data(example1)
```

**Format**

The labels \$C means that follows a continuous variable, \$I means an interval variable, \$H means a histogram variables and \$S means set variable. In the first row each labels should be follow of a name to variable and to the case of histogram a set variables types the names of the modalities (categories). In data rows for continuous variables we have just one value, for interval variables we have the minimum and the maximum of the interval, for histogram variables we have the number of modalities and then the probability of each modality and for set variables we have the cardinality of the set and next the elements of the set.

The format is the \*.csv file is:

```
$C F1 $I F2 F2 $M F3 M1 M2 M3 $S F4 e a 2 3 g b 1 4 i k c d
Case1 $C 2.8 $I 1 2 $M 3 0.1 0.7 0.2 $S 12 1 0 0 0 1 0 0 0 1 1 0 0
Case2 $C 1.4 $I 3 9 $M 3 0.6 0.3 0.1 $S 12 0 1 0 0 0 1 0 0 0 0 1 1
Case3 $C 3.2 $I -1 4 $M 3 0.2 0.2 0.6 $S 12 0 0 1 0 0 1 1 0 0 0 1 0
Case4 $C -2.1 $I 0 2 $M 3 0.9 0.0 0.1 $S 12 0 1 0 1 0 0 0 1 0 0 1 0
Case5 $C -3.0 $I -4 -2 $M 3 0.6 0.0 0.4 $S 12 1 0 0 0 1 0 0 0 1 1 0 0
```

The internal format is:

```
$N
[1] 5
$M
[1] 4
$sym.obj.names
[1] 'Case1' 'Case2' 'Case3' 'Case4' 'Case5'
$sym.var.names
[1] 'F1' 'F2' 'F3' 'F4'
$sym.var.types [1] '$C' '$I' '$H' '$S'
$sym.var.length
[1] 1 2 3 4
$sym.var.starts
[1] 2 4 8 13
$meta
```

```
$C F1 $I F2 F2 $M F3 M1 M2 M3 $S F4 e a 2 3 g b 1 4 i k c d Case1 $C 2.8 $I 1 2 $M 3 0.1 0.7
0.2 $S 12 1 0 0 0 1 0 0 0 1 1 0 0 Case2 $C 1.4 $I 3 9 $M 3 0.6 0.3 0.1 $S 12 0 1 0 0 0 1 0 0 0 0 1 1
Case3 $C 3.2 $I -1 4 $M 3 0.2 0.2 0.6 $S 12 0 0 1 0 0 1 1 0 0 0 1 0 Case4 $C -2.1 $I 0 2 $M 3 0.9
0.0 0.1 $S 12 0 1 0 1 0 0 0 1 0 0 1 0 Case5 $C -3.0 $I -4 -2 $M 3 0.6 0.0 0.4 $S 12 1 0 0 0 1 0 0 0 1
1 0 0 $data
F1 F2 F2.1 M1 M2 M3 e a 2 3 g b 1 4 i k c d Case1 2.8 1 2 0.1 0.7 0.2 1 0 0 0 1 0 0 0 1 1 0 0 Case2
1.4 3 9 0.6 0.3 0.1 0 1 0 0 0 1 0 0 0 0 1 1 Case3 3.2 -1 4 0.2 0.2 0.6 0 0 1 0 0 1 1 0 0 0 1 0 Case4
-2.1 0 2 0.9 0.0 0.1 0 1 0 1 0 0 0 1 0 0 1 0 Case5 -3.0 -4 -2 0.6 0.0 0.4 1 0 0 0 1 0 0 0 1 1 0 0
```

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

**Examples**

```
data(example1)
```



example1

---

example2

*Data Example 2*

---

### Description

This a symbolic data table with variables of continuous, interval, histogram and set types.

### Usage

data(example2)

### Format

```
$C F1 $I F2 F2 $M F3 M1 M2 M3 $C F4 $$ F5 e a 2 3 g b 1 4 i k c d
Case1 $C 2.8 $I 1 2 $M 3 0.1 0.7 0.2 $C 6.0 $$ 12 1 0 0 0 1 0 0 0 1 1 0 0
Case2 $C 1.4 $I 3 9 $M 3 0.6 0.3 0.1 $C 8.0 $$ 12 0 1 0 0 0 1 0 0 0 0 1 1
Case3 $C 3.2 $I -1 4 $M 3 0.2 0.2 0.6 $C -7.0 $$ 12 0 0 1 0 0 1 1 0 0 0 1 0
Case4 $C -2.1 $I 0 2 $M 3 0.9 0.0 0.1 $C 0.0 $$ 12 0 1 0 1 0 0 0 1 0 0 1 0
Case5 $C -3.0 $I -4 -2 $M 3 0.6 0.0 0.4 $C -9.5 $$ 12 1 0 0 0 1 0 0 0 1 1 0 0
```

### Examples

```
data(example2)
example2
```

---

example3

*Data Example 3*

---

### Description

This a symbolic data table with variables of continuous, interval, histogram and set types.

### Usage

data(example3)

### Format

```
$C F1 $I F2 F2 $M F3 M1 M2 M3 $C F4 $$ F5 e a 2 3 g b 1 4 i k c d $I F6 F6 $I F7 F7 Case1 $C
2.8 $I 1 2 $M 3 0.1 0.7 0.2 $C 6.0 $$ 12 1 0 0 0 1 0 0 0 1 1 0 0 $I 0.00 90.00 $I 9 24 Case2 $C 1.4
$I 3 9 $M 3 0.6 0.3 0.1 $C 8.0 $$ 12 0 1 0 0 0 1 0 0 0 0 1 1 $I -90.00 98.00 $I -9 9 Case3 $C 3.2 $I
-1 4 $M 3 0.2 0.2 0.6 $C -7.0 $$ 12 0 0 1 0 0 1 1 0 0 0 1 0 $I 65.00 90.00 $I 65 70 Case4 $C -2.1
$I 0 2 $M 3 0.9 0.0 0.1 $C 0.0 $$ 12 0 1 0 1 0 0 0 1 0 0 1 0 $I 45.00 89.00 $I 25 67 Case5 $C -3.0
$I -4 -2 $M 3 0.6 0.0 0.4 $C -9.5 $$ 12 1 0 0 0 1 0 0 0 1 1 0 0 $I 20.00 40.00 $I 9 40 Case6 $C 0.1
$I 10 21 $M 3 0.0 0.7 0.3 $C -1.0 $$ 12 1 0 0 0 0 0 1 0 1 0 0 0 $I 5.00 8.00 $I 5 8 Case7 $C 9.0 $I
4 21 $M 3 0.2 0.2 0.6 $C 0.5 $$ 12 1 1 1 0 0 0 0 0 0 0 0 0 $I 3.14 6.76 $I 4 6
```

**Examples**

```
data(example3)
example3
```

---

```
example4
```

---

*Data Example 4*

---

**Description**

```
data(example4) example4
```

**Usage**

```
data(example4)
```

**Format**

```
$C 2.8 $I 1 2 $M 3 0.1 0.7 0.2 $C 6 $$ F4 e a 2 3 g b 1 4 i k c d $I 0 90 Case2 $C 1.4 $I 3 9 $M 3
0.6 0.3 0.1 $C 8.0 $$ 12 1 0 0 0 1 0 0 0 1 1 0 0 $I -90.00 98.00 Case3 $C 3.2 $I -1 4 $M 3 0.2 0.2
0.6 $C -7.0 $$ 12 0 1 0 0 0 1 0 0 0 0 1 1 $I 65.00 90.00 Case4 $C -2.1 $I 0 2 $M 3 0.9 0.0 0.1 $C
0.0 $$ 12 0 0 1 0 0 1 1 0 0 0 1 0 $I 45.00 89.00 Case5 $C -3.0 $I -4 -2 $M 3 0.6 0.0 0.4 $C -9.5 $$
12 0 1 0 1 0 0 0 1 0 0 1 0 $I 90.00 990.00 Case6 $C 0.1 $I 10 21 $M 3 0.0 0.7 0.3 $C -1.0 $$ 12 1
0 0 0 1 0 0 0 1 1 0 0 $I 5.00 8.00 Case7 $C 9.0 $I 4 21 $M 3 0.2 0.2 0.6 $C 0.5 $$ 12 1 1 0 0 0 1
0 0 0 0 1 $I 3.14 6.76
```

**Examples**

```
data(example4)
example4
```

---

```
example5
```

---

*Data Example 5*

---

**Description**

This a symbolic data matrix wint continuos, interval, histograma a set data types.

**Usage**

```
data(example5)
```

**Format**

```
$H F0 M01 M02 $C F1 $I F2 F2 $H F3 M1 M2 M3 $$ F4 E1 E2 E3 E4
Case1 $H 2 0.1 0.9 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $$ 4 e g k i
Case2 $H 2 0.7 0.3 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $$ 4 a b c d
Case3 $H 2 0.0 1.0 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $$ 4 2 1 b c
Case4 $H 2 0.2 0.8 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $$ 4 3 4 c a
Case5 $H 2 0.6 0.4 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $$ 4 e i g k
```

**Examples**

```
data(example5)
example5
```

---

example6

*Data Example 6*

---

**Description**

This a symbolic data matrix with continuous, interval, histogram and set data types.

**Usage**

```
data(example6)
```

**Format**

```
$C F1 $M F2 M1 M2 M3 M4 M5 $I F3 F3 $M F4 M1 M2 M3 $C F5 $$ F4 e a 2 3 g b 1 4 i k c d
Case1 $C 2.8 $M 5 0.1 0.1 0.1 0.1 0.6 $I 1 2 $M 3 0.1 0.7 0.2 $C 6.0 $$ 12 1 0 0 0 1 0 0 0 1 1 0 0
Case2 $C 1.4 $M 5 0.1 0.1 0.1 0.1 0.6 $I 3 9 $M 3 0.6 0.3 0.1 $C 8.0 $$ 12 0 1 0 0 0 1 0 0 0 0 1 1
Case3 $C 3.2 $M 5 0.1 0.1 0.1 0.1 0.6 $I -1 4 $M 3 0.2 0.2 0.6 $C -7.0 $$ 12 0 0 1 0 0 1 1 0 0 0 1 0
Case4 $C -2.1 $M 5 0.1 0.1 0.1 0.1 0.6 $I 0 2 $M 3 0.9 0.0 0.1 $C 0.0 $$ 12 0 1 0 1 0 0 0 1 0 0 1 0
Case5 $C -3.0 $M 5 0.1 0.1 0.1 0.1 0.6 $I -4 -2 $M 3 0.6 0.0 0.4 $C -9.5 $$ 12 1 0 0 0 1 0 0 0 1 1 0
0
```

**Examples**

```
data(example6)
example6
```

---

 example7

*Data Example 7*


---

**Description**

This a symbolic data matrix wint continuos, interval, histograma a set data types.

**Usage**

```
data(example6)
```

**Format**

```
$C F1 $H F2 M1 M2 M3 M4 M5 $I F3 F3 $H F4 M1 M2 M3 $C F5
Case1 $C 2.8 $H 5 0.1 0.2 0.3 0.4 0.0 $I 1 2 $H 3 0.1 0.7 0.2 $C 6.0
Case2 $C 1.4 $H 5 0.2 0.1 0.5 0.1 0.2 $I 3 9 $H 3 0.6 0.3 0.1 $C 8.0
Case3 $C 3.2 $H 5 0.1 0.1 0.2 0.1 0.5 $I -1 4 $H 3 0.2 0.2 0.6 $C -7.0
Case4 $C -2.1 $H 5 0.4 0.1 0.1 0.1 0.3 $I 0 2 $H 3 0.9 0.0 0.1 $C 0.0
Case5 $C -3.0 $H 5 0.6 0.1 0.1 0.1 0.1 $I -4 -2 $H 3 0.6 0.0 0.4 $C -9.5
```

**Examples**

```
data(example7)
example7
```

---

 ex\_cfa1

*Correspondence Analysis Example*


---

**Description**

Correspondence Analysis for Symbolic MultiValued Variables example.

**Usage**

```
data(ex_cfa1)
```

**Format**

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 4 rows and 4 columns.

**References**

Rodriguez, O. (2011). Correspondence Analysis for Symbolic MultiValued Variables. Workshop in Symbolic Data Analysis Namur, Belgium

---

`ex_cfa2`*Correspondence Analysis Example*

---

**Description**

Correspondence Analysis for Symbolic MultiValued Variables example.

**Usage**

```
data(ex_cfa2)
```

**Format**

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 6 rows and 5 columns.

**References**

Rodriguez, O. (2011). Correspondence Analysis for Symbolic MultiValued Variables. Workshop in Symbolic Data Analysis Namur, Belgium

---

`ex_mcfa1`*Multiple Correspondence Analysis Example*

---

**Description**

example for the `sym.mcfa` function.

example for the `sym.mcfa` function.

**Usage**

```
data(ex_mcfa1)
```

```
ex_mcfa1
```

**Format**

An object of class `data.frame` with 130 rows and 5 columns.

An object of class `data.frame` with 130 rows and 5 columns.

**Examples**

```

data("ex_mcfa1")
sym.table <- classic.to.sym(ex_mcfa1,
                           concept = suspect,
                           hair = sym.set(hair),
                           eyes = sym.set(eyes),
                           region = sym.set(region))

res <- sym.mcfa(sym.table, c(1,2))
mcfa.scatterplot(res[,1], res[,2], sym.data = sym.table, pos.var = c(1,2))

data("ex_mcfa1")
sym.table <- classic.to.sym(
  x = ex_mcfa1,
  concept = "suspect",
  variables = c(hair, eyes, region),
  hair = sym.set(hair),
  eyes = sym.set(eyes),
  region = sym.set(region)
)
sym.table

```

---

ex\_mcfa2

---

*Multiple Correspondence Analysis Example*


---

**Description**

example for the sym.mcfa function.

**Usage**

```
data(ex_mcfa2)
```

**Format**

An object of class data.frame with 130 rows and 7 columns.

**Examples**

```

data("ex_mcfa2")

ex <- classic.to.sym(ex_mcfa2,
                    concept = employee_id,
                    variables = c(employee_id, salary, region, evaluation, years_worked),
                    salary = sym.set(salary),
                    region = sym.set(region),
                    evaluation = sym.set(evaluation),
                    years_worked = sym.set(years_worked))

```

```
res <- sym.mcfa(ex, c(1,2,3,4))
mcfa.scatterplot(res[,1], res[,2], sym.data = ex, pos.var = c(1,2,3,4))
```

---

facedata

*Face Data Example*


---

### Description

Symbolic data matrix with all the variables of interval type.

### Usage

```
data('facedata')
```

### Format

```
$I;AD;AD;$I;BC;BC;.....
```

```
HUS1;$I;168.86;172.84;$I;58.55;63.39;.....
HUS2;$I;169.85;175.03;$I;60.21;64.38;.....
HUS3;$I;168.76;175.15;$I;61.4;63.51;.....
INC1;$I;155.26;160.45;$I;53.15;60.21;.....
INC2;$I;156.26;161.31;$I;51.09;60.07;.....
INC3;$I;154.47;160.31;$I;55.08;59.03;.....
ISA1;$I;164;168;$I;55.01;60.03;.....
ISA2;$I;163;170;$I;54.04;59;.....
ISA3;$I;164.01;169.01;$I;55;59.01;.....
JPL1;$I;167.11;171.19;$I;61.03;65.01;.....
JPL2;$I;169.14;173.18;$I;60.07;65.07;.....
JPL3;$I;169.03;170.11;$I;59.01;65.01;.....
KHA1;$I;149.34;155.54;$I;54.15;59.14;.....
KHA2;$I;149.34;155.32;$I;52.04;58.22;.....
KHA3;$I;150.33;157.26;$I;52.09;60.21;.....
LOT1;$I;152.64;157.62;$I;51.35;56.22;.....
LOT2;$I;154.64;157.62;$I;52.24;56.32;.....
LOT3;$I;154.83;157.81;$I;50.36;55.23;.....
PHI1;$I;163.08;167.07;$I;66.03;68.07;.....
PHI2;$I;164;168.03;$I;65.03;68.12;.....
PHI3;$I;161.01;167;$I;64.07;69.01;.....
ROM1;$I;167.15;171.24;$I;64.07;68.07;.....
ROM2;$I;168.15;172.14;$I;63.13;68.07;.....
ROM3;$I;167.11;171.19;$I;63.13;68.03;.....
```

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
## Not run:
data(facedata)
res.vertex.ps <- sym.interval.pc(facedata, 'vertex', 150, FALSE, FALSE, TRUE)
class(res.vertex.ps$sym.prin.curve) <- c('sym.data.table')
sym.scatterplot(res.vertex.ps$sym.prin.curve[,1], res.vertex.ps$sym.prin.curve[,2],
                labels=TRUE,col='red',main='PSC Face Data')

## End(Not run)
```

---

```
format.symbolic_histogram
```

*Symbolic modal conversion functions to and from Character*

---

**Description**

Symbolic modal conversion functions to and from Character

**Usage**

```
## S3 method for class 'symbolic_histogram'
format(x, ...)
```

**Arguments**

x	An object to be converted
...	Further arguments to be passed from or to other methods.

---

```
format.symbolic_interval
```

*Symbolic interval conversion functions to and from Character*

---

**Description**

Symbolic interval conversion functions to and from Character

**Usage**

```
## S3 method for class 'symbolic_interval'
format(x, ...)
```



**Arguments**

x                    An object to be converted  
...                   Further arguments to be passed from or to other methods.

---

format.symbolic\_modal    *Symbolic modal conversion functions to and from Character*

---

**Description**

Symbolic modal conversion functions to and from Character

**Usage**

```
## S3 method for class 'symbolic_modal'  
format(x, ...)
```

**Arguments**

x                    An object to be converted  
...                   Further arguments to be passed from or to other methods.

---

format.symbolic\_set    *Symbolic set conversion functions to and from Character*

---

**Description**

Symbolic set conversion functions to and from Character

**Usage**

```
## S3 method for class 'symbolic_set'  
format(x, ...)
```

**Arguments**

x                    An object to be converted  
...                   Further arguments to be passed from or to other methods.

---

get_cats	<i>Extract categories</i>
----------	---------------------------

---

**Description**

Extract categories

**Usage**

```
get_cats(x, ...)
```

**Arguments**

x	An object to be converted
...	Further arguments to be passed from or to other methods.

---

get_props	<i>Extract prop</i>
-----------	---------------------

---

**Description**

Extract prop

**Usage**

```
get_props(x, ...)
```

**Arguments**

x	An object to be converted
...	Further arguments to be passed from or to other methods.

---

interval.centers	<i>calcula centros</i>
------------------	------------------------

---

**Description**

calcula centros

**Usage**

```
interval.centers(x)
```

**Arguments**

x	tabla simbolica todos intervalos
---	----------------------------------

---

`interval.histogram.plot`*Histogram plot for an interval variable*

---

**Description**

Histogram plot for an interval variable

**Usage**

```
interval.histogram.plot(x, n.bins, ...)
```

**Arguments**

<code>x</code>	An symbolic data table.
<code>n.bins</code>	Numbers of breaks of the histogram.
<code>...</code>	Arguments to be passed to the barplot method.

**Value**

A list with componets : frequency and histogram

**Examples**

```
data(oils)
res <- interval.histogram.plot(x = oils[, 3], n.bins = 3)
res
```

---

`interval.max`*calcula maximos*

---

**Description**

calcula maximos

**Usage**

```
interval.max(x)
```

**Arguments**

<code>x</code>	tabla simbolica todos intervalos
----------------	----------------------------------

---

<code>interval.min</code>	<i>calcula minimos</i>
---------------------------	------------------------

---

**Description**

calcula minimos

**Usage**

`interval.min(x)`

**Arguments**

x	tabla simbolica todos intervalos
---	----------------------------------

---

<code>interval.ranges</code>	<i>calcula rangos</i>
------------------------------	-----------------------

---

**Description**

calcula rangos

**Usage**

`interval.ranges(x)`

**Arguments**

x	tabla simbolica todos intervalos
---	----------------------------------

---

<code>int_prost_test</code>	<i>Linear regression model data example.</i>
-----------------------------	--

---

**Description**

Linear regression model interval-valued data example.

**Usage**

`data(int_prost_test)`

**Format**

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 30 rows and 9 columns.

**References**

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer.

---

int_prost_train	<i>Linear regression model data example.</i>
-----------------	--

---

**Description**

Linear regression model interval-valued data example.

**Usage**

```
data(int_prost_train)
```

**Format**

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 67 rows and 9 columns.

**References**

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer.

---

is.sym.histogram	<i>Symbolic histogram</i>
------------------	---------------------------

---

**Description**

Symbolic histogram

**Usage**

```
is.sym.histogram(x)
```

**Arguments**

x                      an object to be tested

**Value**

returns TRUE if its argument's value is a `symbolic_histogram` and FALSE otherwise.

**Examples**

```
x <- sym.histogram(iris$Sepal.Length)
is.sym.histogram(x)
```

---

is.sym.interval	<i>Symbolic interval</i>
-----------------	--------------------------

---

**Description**

Symbolic interval

**Usage**

```
is.sym.interval(x)
```

**Arguments**

x                    an object to be tested

**Value**

returns TRUE if its argument's value is a symbolic\_vector and FALSE otherwise.

**Examples**

```
x <- sym.interval(1:10)
is.sym.interval(x)
is.sym.interval("d")
```

---

is.sym.modal	<i>Symbolic modal</i>
--------------	-----------------------

---

**Description**

Symbolic modal

**Usage**

```
is.sym.modal(x)
```

**Arguments**

x                    an object to be tested

**Value**

returns TRUE if its argument's value is a symbolic\_modal and FALSE otherwise.

**Examples**

```
x <- sym.modal(factor(c("a", "b", "b", "1")))
is.sym.modal(x)
```

---

is.sym.set	<i>Symbolic set</i>
------------	---------------------

---

**Description**

Symbolic set

**Usage**

```
is.sym.set(x)
```

**Arguments**

x                    an object to be tested

**Value**

returns TRUE if its argument's value is a symbolic\_set and FALSE otherwise.

**Examples**

```
x <- sym.set(factor(c("a", "b", "b", "1")))
is.sym.set(x)
```

---

lynne1	<i>Symbolic interval data example.</i>
--------	--

---

**Description**

Symbolic data matrix with all the variables of interval type.

**Usage**

```
data(lynne1)
```

**Format**

An object of class tbl\_df (inherits from tbl, data.frame) with 10 rows and 4 columns.

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(lynne1)
lynne1
```

---

mcfa.scatterplot      *Plot Interval Scatterplot*

---

### Description

Plot Interval Scatterplot

### Usage

```
mcfa.scatterplot(x, y, sym.data, pos.var)
```

### Arguments

x	symbolic table with only one column.
y	symbolic table with only one column.
sym.data	original symbolic table.
pos.var	column number of the variables to be plotted.

### Examples

```
data("ex_mcfa1")
sym.table <- classic.to.sym(ex_mcfa1,
  concept = suspect,
  hair = sym.set(hair),
  eyes = sym.set(eyes),
  region = sym.set(region)
)

res <- sym.mcfa(sym.table, c(1, 2))
mcfa.scatterplot(res[, 2], res[, 3], sym.data = sym.table, pos.var = c(1, 2))
```

---

mean.symbolic\_interval      *Symbolic mean for intervals*

---

### Description

This function compute the symbolic mean for intervals

### Usage

```
## S3 method for class 'symbolic_interval'
mean(x, method = c("centers", "interval"), trim = 0, na.rm = F, ...)

## S3 method for class 'symbolic_tbl'
mean(x, ...)
```



**Arguments**

x	A symbolic interval.
method	The method to be use.
trim	As in R mean function.
na.rm	As in R mean function.
...	As in R mean function.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

- Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.
- Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

---

median.symbolic\_interval  
*Symbolic Median*

---

**Description**

This function compute the median for symbolic intervals.

**Usage**

```
## S3 method for class 'symbolic_interval'  
median(x, na.rm = FALSE, method = c("centers", "interval"), ...)  
  
## S3 method for class 'symbolic_tbl'  
median(x, ...)
```

**Arguments**

x	A symbolic interval.
na.rm	As in R median function.
method	The method to be use.
...	As in R median function.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

- Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.
- Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

---

method_summary	<i>Summary method to CM and CRM regression model</i>
----------------	--

---

**Description**

Summary method to CM and CRM regression model

**Usage**

```
method_summary(ref, pred)
```

**Arguments**

ref	Real values
pred	Predicted values

---

min.symbolic_interval	<i>Maxima and Minima</i>
-----------------------	--------------------------

---

**Description**

Maxima and Minima

**Usage**

```
## S3 method for class 'symbolic_interval'
min(x, ...)

## S3 method for class 'symbolic_interval'
max(x, ...)

## S3 method for class 'symbolic_interval'
x$name = c("min", "max", "mean", "median")
```

**Arguments**

x	symbolic interval vector
...	further arguments passed to or from other methods.
name	...

**Value**

a new symbolic interval with the minimum of the minima and the minimum of the maxima

---

neighbors.vertex	<i>Compute neighbors vertex</i>
------------------	---------------------------------

---

**Description**

Compute neighbors vertex

**Usage**

```
neighbors.vertex(vertex, Matrix, num.neig)
```

**Arguments**

vertex	Vertes of the hipercube
Matrix	Interval Data Matrix.
num.neig	Number of vertices.

**Author(s)**

Jorge Arce

**References**

- Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.
- Hastie,T. (1984). Principal Curves and Surface. Ph.D Thesis Stanford University.
- Hastie,T. & Weingessel,A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension.R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.
- Hastie,T. & Stuetzle, W. (1989). Principal Curves. Journal of the American Statistical Association, Vol. 84-406, 502–516.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning; Data Mining, Inference and Prediction. Springer, New York.

**See Also**

sym.interval.pc

---

norm.vect	<i>Compute the norm of a vector.</i>
-----------	--------------------------------------

---

**Description**

Compute the norm of a vector.

**Usage**

```
norm.vect(vector1)
```

**Arguments**

vector1          An n dimensional vector.

**Value**

The L2 norm of the vector.

**Author(s)**

Jorge Arce

**References**

Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.

Hastie,T. (1984). Principal Curves and Surface. Ph.D Thesis Stanford University.

Hastie,T. & Weingessel,A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension.R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.

Hastie,T. & Stuetzle, W. (1989). Principal Curves. Journal of the American Statistical Association, Vol. 84-406, 502–516.

Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning; Data Mining, Inference and Prediction. Springer, New York.

**See Also**

sym.interval.pc

---

`oils`*Ichino Oils example data.*

---

**Description**

Symbolic data matrix with all the variables of interval type.

**Usage**

```
data(oils)
```

**Format**

```
$I GRA GRA $I FRE FRE $I IOD IOD $I SAP SAP  
L $I 0.930 0.935 $I -27 -18 $I 170 204 $I 118 196  
P $I 0.930 0.937 $I -5 -4 $I 192 208 $I 188 197  
Co $I 0.916 0.918 $I -6 -1 $I 99 113 $I 189 198  
S $I 0.920 0.926 $I -6 -4 $I 104 116 $I 187 193  
Ca $I 0.916 0.917 $I -25 -15 $I 80 82 $I 189 193  
O $I 0.914 0.919 $I 0 6 $I 79 90 $I 187 196  
B $I 0.860 0.870 $I 30 38 $I 40 48 $I 190 199  
H $I 0.858 0.864 $I 22 32 $I 53 77 $I 190 202
```

**References**

Cazes P., Chouakria A., Diday E. et Schektman Y. (1997). Extension de l'analyse en composantes principales a des donnees de type intervalle, Rev. Statistique Appliquee, Vol. XLV Num. 3 pag. 5-24, France.

**Examples**

```
data(oils)  
oils
```

---

`plot.symbolic_tbl`*Function for plotting a symbolic object*

---

**Description**

Function for plotting a symbolic object

**Usage**

```
## S3 method for class 'symbolic_tbl'
plot(
  x,
  col = NA,
  matrix.form = NA,
  border = FALSE,
  size = 1,
  title = TRUE,
  show.type = FALSE,
  font.size = 1,
  reduce = FALSE,
  hist.angle.x = 60,
  ...
)
```

**Arguments**

x	The symbolic object.
col	A specification for the default plotting color.
matrix.form	A vector of the form c(num.rows,num.columns).
border	A logical value indicating whether border should be plotted.
size	The magnification to be used for each graphic.
title	A logical value indicating whether title should be plotted.
show.type	A logical value indicating whether type should be plotted.
font.size	The font size of graphics.
reduce	A logical value indicating whether values different from zero should be plotted in modal and set graphics.
hist.angle.x	The angle of labels in y axis. Only for histogram plot
...	Arguments to be passed to methods.

**Value**

A plot of the symbolic data table.

**Author(s)**

Andres Navarro

**Examples**

```
## Not run:
data(oils)
plot(oils)
plot(oils, border = T, size = 1.3)

## End(Not run)
```

---

plot.sym_tsne	<i>Plot TSNE for symbolic data tables</i>
---------------	---

---

**Description**

Plot TSNE for symbolic data tables

**Usage**

```
## S3 method for class 'sym_tsne'  
plot(x, ...)
```

**Arguments**

x	data frame
...	other graphical parameters (see par and section 'Details' below).

---

plot.sym_umap	<i>Plot UMAP for symbolic data tables</i>
---------------	---

---

**Description**

Plot UMAP for symbolic data tables

**Usage**

```
## S3 method for class 'sym_umap'  
plot(x, ...)
```

**Arguments**

x	sym_umap object
...	other graphical parameters (see par and section 'Details' below).

---

R2.L

*Lower boundary correlation coefficient.*

---

**Description**

Compute the lower boundary correlation coefficient for two interval variables.

**Usage**

```
R2.L(ref, pred)
```

**Arguments**

ref	Variable that was predicted.
pred	The prediction given by the model.

**Value**

The lower boundary correlation coefficient.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347.

**See Also**

sym.glm

**Examples**

```
data(int_prost_train)
data(int_prost_test)
res.cm <- sym.lm(lpsa ~ ., sym.data = int_prost_train, method = "cm")
pred.cm <- sym.predict(res.cm, int_prost_test)
R2.L(int_prost_test$lpsa, pred.cm$Fitted)
```



---

R2.U *Upper boundary correlation coefficient.*

---

**Description**

Compute the upper boundary correlation coefficient for two interval variables.

**Usage**

```
R2.U(ref, pred)
```

**Arguments**

ref	Variable that was predicted.
pred	The prediction given by the model.

**Value**

The upper boundary correlation coefficient.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347.

**See Also**

sym.glm

**Examples**

```
data(int_prost_train)
data(int_prost_test)
res.cm <- sym.lm(lpsa ~ ., sym.data = int_prost_train, method = "cm")
pred.cm <- sym.predict(res.cm, int_prost_test)
R2.U(int_prost_test$lpsa, pred.cm$Fitted)
```

---

read.sym.table	<i>Read a Symbolic Table</i>
----------------	------------------------------

---

### Description

It reads a symbolic data table from a CSV file.

### Usage

```
read.sym.table(file, header = TRUE, sep, dec, row.names = NULL)
```

### Arguments

file	The name of the CSV file.
header	As in R function read.table
sep	As in R function read.table
dec	As in R function read.table
row.names	As in R function read.table

### Details

The labels \$C means that follows a continuous variable, \$I means an interval variable, \$H means a histogram variables and \$S means set variable. In the first row each labels should be follow of a name to variable and to the case of histogram a set variables types the names of the modalities (categories) . In data rows for continuous variables we have just one value, for interval variables we have the minimum and the maximum of the interval, for histogram variables we have the number of modalities and then the probability of each modality and for set variables we have the cardinality of the set and next the elements of the set.

The format is the CSV file should be like:

```
$C F1 $I F2 F2 $H F3 M1 M2 M3 $S F4 E1 E2 E3 E4
```

```
Case1 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $S 4 e g k i
```

```
Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $S 4 a b c d
```

```
Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $S 4 2 1 b c
```

```
Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $S 4 3 4 c a
```

```
Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $S 4 e i g k
```

The internal format is:

```
$N
```

```

[1] 5
$M
[1] 4
$sym.obj.names
[1] 'Case1' 'Case2' 'Case3' 'Case4' 'Case5'
$sym.var.names
[1] 'F1' 'F2' 'F3' 'F4'
$sym.var.types
[1] '$C' '$I' '$H' '$S'
$sym.var.length
[1] 1 2 3 4
$sym.var.starts
[1] 2 4 8 13
$meta
$C F1 $I F2 F2 $H F3 M1 M2 M3 $S F4 E1 E2 E3 E4
Case1 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $S 4 e g k i
Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $S 4 a b c d
Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $S 4 2 1 b c
Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $S 4 3 4 c a
Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $S 4 e i g k
$data
F1 F2 F2.1 M1 M2 M3 E1 E2 E3 E4
Case1 2.8 1 2 0.1 0.7 0.2 e g k i
Case2 1.4 3 9 0.6 0.3 0.1 a b c d
Case3 3.2 -1 4 0.2 0.2 0.6 2 1 b c
Case4 -2.1 0 2 0.9 0.0 0.1 3 4 c a
Case5 -3.0 -4 -2 0.6 0.0 0.4 e i g k

```

**Value**

Return a symbolic data table structure.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

**See Also**

display.sym.table

**Examples**

```
## Not run:
data(example1)
```

```
write.sym.table(example1,
  file = "temp4.csv", sep = "|", dec = ".", row.names = TRUE,
  col.names = TRUE
)
ex1 <- read.sym.table("temp4.csv", header = TRUE, sep = "|", dec = ".", row.names = 1)

## End(Not run)
```

---

RMSE.L

*Lower boundary root-mean-square error*

---

### Description

Compute the lower boundary root-mean-square error.

### Usage

```
RMSE.L(ref, pred)
```

### Arguments

ref	Variable that was predicted.
pred	The prediction given by the model.

### Value

The lower boundary root-mean-square error.

### Author(s)

Oldemar Rodriguez Rojas.

### References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347.

### See Also

sym.glm

---

RMSE.U	<i>Upper boundary root-mean-square error</i>
--------	--

---

**Description**

Compute the upper boundary root-mean-square error.

**Usage**

```
RMSE.U(ref, pred)
```

**Arguments**

ref	Variable that was predicted.
pred	The prediction given by the model.

**Value**

The upper boundary root-mean-square error.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347.

**See Also**

`sym.glm`

## Description

This work is framed inside the Symbolic Data Analysis (SDA). The objective of this work is to implement in R to the symbolic case certain techniques of the automatic classification, as well as some lineal models. These implementations will always be made following two fundamental principles in Symbolic Data Analysis like they are: Classic Data Analysis should always be a case particular case of the Symbolic Data Analysis and both, the exit as the input in an Symbolic Data Analysis should be symbolic. We implement for variables of type interval the mean, the median, the mean of the extreme values, the standard deviation, the deviation quartil, the dispersion boxes and the correlation also three new methods are also presented to carry out the lineal regression for variables of type interval. We also implement in this R package the method of Principal Components Analysis in two senses: First, we propose three ways to project the interval variables in the circle of correlations in such way that is reflected the variation or the inexactness of the variables. Second, we propose an algorithm to make the Principal Components Analysis for variables of type histogram. We implement a method for multidimensional scaling of interval data, denominated INTERSCAL.

## Details

Package: RSDA  
Type: Package  
Version: 3.0.13  
Date: 2022-07-15  
License: GPL (>=2)

Most of the function of the package starts from a symbolic data table that can be store in a CSV file withe follwing forma: In the first row the labels \$C means that follows a continuous variable, \$I means an interval variable, \$H means a histogram variables and \$S means set variable. In the first row each labels should be follow of a name to variable and to the case of histogram a set variables types the names of the modalities (categories) . In data rows for continuous variables we have just one value, for interval variables we have the minimum and the maximum of the interval, for histogram variables we have the number of modalities and then the probability of each modality and for set variables we have the cardinality of the set and next the elements of the set.

## Author(s)

Oldemar Rodriguez Rojas  
Maintainer: Oldemar Rodriguez Rojas <oldemar.rodriguez@ucr.ac.cr>

## References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Billard L., Douzal-Chouakria A. and Diday E. (2011) Symbolic Principal Components For Interval-Valued Observations, *Statistical Analysis and Data Mining*. 4 (2), 229-246. Wiley.

Bock H-H. and Diday E. (eds.) (2000). *Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data.* Springer, Germany.

Carvalho F., Souza R., Chavent M., and Lechevallier Y. (2006) Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. *Pattern Recognition Letters Volume 27, Issue 3, February 2006, Pages 167-179*

Cazes P., Chouakria A., Diday E. et Schektman Y. (1997). Extension de l'analyse en composantes principales a des donnees de type intervalle, *Rev. Statistique Appliquee*, Vol. XLV Num. 3 pag. 5-24, France.

Diday, E., Rodriguez O. and Winberg S. (2000). Generalization of the Principal Components Analysis to Histogram Data, 4th European Conference on Principles and Practice of Knowledge Discovery in Data Bases, September 12-16, 2000, Lyon, France.

Chouakria A. (1998) Extension des methodes d'analysis factorielle a des donnees de type intervalle, Ph.D. Thesis, Paris IX Dauphine University.

Makosso-Kallyth S. and Diday E. (2012). Adaptation of interval PCA to symbolic histogram variables, *Advances in Data Analysis and Classification July, Volume 6, Issue 2, pp 147-159.* Rodriguez, O. (2000). *Classification et Modeles Lineaires en Analyse des Donnees Symboliques.* Ph.D. Thesis, Paris IX-Dauphine University.

sd

*Generic function for the standard deviation***Description**

Compute the symbolic standard deviation.

**Usage**

```
sd(x, ...)
```

```
## Default S3 method:
sd(x, na.rm = FALSE, ...)
```

```
## S3 method for class 'symbolic_interval'
sd(x, method = c("centers", "interval", "billard"), na.rm = FALSE, ...)
```

```
## S3 method for class 'symbolic_tbl'
sd(x, ...)
```

**Arguments**

x	A symbolic variable.
...	As in R sd function.
na.rm	As in R sd function.
method	The method to be use.

**Value**

return a real number.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

---

SDS.to.RSDA

*SDS SODAS files to RSDA files.*

---

**Description**

To convert SDS SODAS files to RSDA files.

**Usage**

```
SDS.to.RSDA(file.path, labels = FALSE)
```

**Arguments**

file.path	Disk path where the SODAS *.SDA file is.
labels	If we want to include SODAS SDA files labels in RSDA file.

**Value**

A RSDA symbolic data file.

**Author(s)**

Olger Calderon and Roberto Zuniga.

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

**See Also**

SODAS.to.RSDA



## Examples

```
## Not run:  
# We can read the file directly from the SODAS SDA file as follows:  
# We can save the file in CSV to RSDA format as follows:  
setwd('C:/Program Files (x86)/DECISIA/SODAS version 2.0/bases/')  
result <- SDS.to.RSDA(file.path='hani3101.sds')  
# We can save the file in CSV to RSDA format as follows:  
write.sym.table(result, file='hani3101.csv', sep=';', dec='.', row.names=TRUE,  
  
## End(Not run)
```

---

SODAS.to.RSDA

*XML SODAS files to RSDA files.*

---

## Description

To convert XML SODAS files to RSDA files.

## Usage

```
SODAS.to.RSDA(XMLPath, labels = T)
```

## Arguments

XMLPath	Disk path where the SODAS *.XML file is.
labels	If we want to include SODAS XML files labels in RSDA file.

## Value

A RSDA symbolic data file.

## Author(s)

Olger Calderon and Roberto Zuniga.

## References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

## See Also

SDS.to.RSDA

**Examples**

```

## Not run:
# We can read the file directly from the SODAS XML file as follows:
# abalone<-SODAS.to.RSDA('C:/Program Files (x86)/DECISIA/SODAS version 2.0/bases/abalone.xml)
# We can save the file in CSV to RSDA format as follows:
# write.sym.table(sodas.ex1, file='abalone.csv', sep=';',dec='.', row.names=TRUE,
#               col.names=TRUE)
# We read the file from the CSV file,
# this is not necessary if the file is read directly from
# XML using SODAS.to.RSDA as in the first statement in this example.
data(abalone)
res <- sym.interval.pca(abalone, "centers")
sym.scatterplot(sym.var(res$Sym.Components, 1), sym.var(res$Sym.Components, 2),
  labels = TRUE, col = "red", main = "PCA Oils Data"
)
sym.scatterplot3d(sym.var(res$Sym.Components, 1), sym.var(res$Sym.Components, 2),
  sym.var(res$Sym.Components, 3),
  color = "blue", main = "PCA Oils Data"
)
sym.scatterplot.ggplot(sym.var(res$Sym.Components, 1), sym.var(res$Sym.Components, 2),
  labels = TRUE
)
sym.circle.plot(res$Sym.Prin.Correlations)

## End(Not run)

```

---

sym.circle.plot

*Symbolic Circle of Correlations*


---

**Description**

Plot the symbolic circle of correlations.

**Usage**

```
sym.circle.plot(prin.corre)
```

**Arguments**

prin.corre      A symbolic interval data matrix with correlations between the variables and the principals componets, both of interval type.

**Value**

Plot the symbolic circle

**Author(s)**

Oldemar Rodriguez Rojas

## References

Rodriguez O. (2012). The Duality Problem in Interval Principal Components Analysis. The 3rd Workshop in Symbolic Data Analysis, Madrid.

## Examples

```
data(oils)
res <- sym.pca(oils, "centers")
sym.circle.plot(res$Sym.Prin.Correlations)
```

---

sym.dist.interval      *Distance for Symbolic Interval Variables.*

---

## Description

This function computes and returns the distance matrix by using the specified distance measure to compute distance between symbolic interval variables.

## Usage

```
sym.dist.interval(
  sym.data,
  gamma = 0.5,
  method = "Minkowski",
  normalize = TRUE,
  SpanNormalize = FALSE,
  q = 1,
  euclidean = TRUE,
  pond = rep(1, length(variables))
)
```

## Arguments

sym.data	A symbolic object
gamma	gamma value for the methods ichino and minkowski.
method	Method to use (Gowda.Diday, Ichino, Minkowski, Hausdorff)
normalize	A logical value indicating whether normalize the data in the ichino or hausdorff method.
SpanNormalize	A logical value indicating whether
q	q value for the hausdorff method.
euclidean	A logical value indicating whether use the euclidean distance.
pond	A numeric vector
variables	Numeric vector with the number of the variables to use.

## Value

An object of class 'dist'

sym.gbm

*Generalized Boosted Symbolic Regression***Description**

Generalized Boosted Symbolic Regression

**Usage**

```

sym.gbm(
  formula,
  sym.data,
  method = c("cm", "crm"),
  distribution = "gaussian",
  interaction.depth = 1,
  n.trees = 500,
  shrinkage = 0.1
)

```

**Arguments**

formula	A symbolic description of the model to be fit. The formula may include an offset term (e.g. $y \sim \text{offset}(n)+x$ ). If <code>keep.data = FALSE</code> in the initial call to <code>gbm</code> then it is the user's responsibility to resupply the offset to <code>gbm.more</code> .
sym.data	symbolic data table
method	cm crm
distribution	distribution
interaction.depth	Integer specifying the maximum depth of each tree (i.e., the highest level of variable interactions allowed). A value of 1 implies an additive model, a value of 2 implies a model with up to 2-way interactions, etc. Default is 1.
n.trees	Integer specifying the total number of trees to fit. This is equivalent to the number of iterations and the number of basis functions in the additive expansion. Default is 100.
shrinkage	A shrinkage parameter applied to each tree in the expansion. Also known as the learning rate or step-size reduction; 0.001 to 0.1 usually work, but a smaller learning rate typically requires more trees. Default is 0.1.

**References**

Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515

Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347

Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>

Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

sym.glm	<i>Lasso, Ridge and and Elastic Net Linear regression model to interval variables</i>
---------	---

---

### Description

Execute Lasso, Ridge and and Elastic Net Linear regression model to interval variables.

### Usage

```
sym.glm(sym.data, response = 1, method = c('cm', 'crm'),
alpha = 1, nfolds = 10, grouped = TRUE)
```

### Arguments

sym.data	Should be a symbolic data table read with the function <code>read.sym.table(...)</code> .
response	The number of the column where is the response variable in the interval data table.
method	'cm' to generalized Center Method and 'crm' to generalized Center and Range Method.
alpha	alpha=1 is the lasso penalty, and alpha=0 the ridge penalty. $0 < \alpha < 1$ is the elastic net method.
nfolds	Number of folds - default is 10. Although nfolds can be as large as the sample size (leave-one-out CV), it is not recommended for large datasets. Smallest value allowable is nfolds=3
grouped	This is an experimental argument, with default TRUE, and can be ignored by most users.

### Value

An object of class 'cv.glmnet' is returned, which is a list with the ingredients of the cross-validation fit.

### Author(s)

Oldemar Rodriguez Rojas

### References

Rodriguez O. (2013). A generalization of Centre and Range method for fitting a linear regression model to symbolic interval data using Ridge Regression, Lasso and Elastic Net methods. The IFCS2013 conference of the International Federation of Classification Societies, Tilburg University Holland.

**See Also**

sym.lm

---

sym.histogram      *Create an symbolic\_histogram type object*

---

**Description**

Create an symbolic\_histogram type object

**Usage**

```
sym.histogram(x = double(), breaks = NA_real_)
```

**Arguments**

x                      character vector  
breaks                 a vector giving the breakpoints between histogram cells

**Value**

a symbolic histogram

**Examples**

```
sym.histogram(iris$Sepal.Length)
```

---

sym.interval      *Create an symbolic\_interval type object*

---

**Description**

Create an symbolic\_interval type object

**Usage**

```
sym.interval(x = numeric(), .min = min, .max = max)
```

**Arguments**

x                      numeric vector  
.min                    function that will be used to calculate the minimum interval  
.max                    function that will be used to calculate the maximum interval

**Value**

a symbolic interval

**Examples**

```
sym.interval(c(1, 2, 4, 5))
sym.interval(1:10)
```

---

sym.interval.pc	<i>Compute a symbolic interval principal components curves</i>
-----------------	--

---

**Description**

Compute a symbolic interval principal components curves

**Usage**

```
sym.interval.pc(sym.data, method = c('vertex', 'centers'), maxit, plot, scale, center)
```

**Arguments**

sym.data	Should be a symbolic data table read with the function read.sym.table(...)
method	It should be 'vertex' or 'centers'.
maxit	Maximum number of iterations.
plot	TRUE to plot immediately, FALSE if you do not want to plot.
scale	TRUE to standardize the data.
center	TRUE to center the data.

**Value**

prin.curve: This a symbolic data table with the interval principal components. As this is a symbolic data table we can apply over this table any other symbolic data analysis method (symbolic propagation).

cor.ps: This is the interval correlations between the original interval variables and the interval principal components, it can be use to plot the symbolic circle of correlations.

**Author(s)**

Jorge Arce.

## References

- Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.
- Hastie, T. (1984). Principal Curves and Surface. Ph.D Thesis Stanford University.
- Hastie, T. & Weingessel, A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension. R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.
- Hastie, T. & Stuetzle, W. (1989). Principal Curves. Journal of the American Statistical Association, Vol. 84-406, 502–516.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning; Data Mining, Inference and Prediction. Springer, New York.

## See Also

sym.interval.pca

## Examples

```
## Not run:
data(oils)
res.vertex.ps <- sym.interval.pc(oils, "vertex", 150, FALSE, FALSE, TRUE)
class(res.vertex.ps$sym.prin.curve) <- c("sym.data.table")
sym.scatterplot(res.vertex.ps$sym.prin.curve[, 1], res.vertex.ps$sym.prin.curve[, 2],
  labels = TRUE, col = "red", main = "PSC Oils Data"
)

data(facedata)
res.vertex.ps <- sym.interval.pc(facedata, "vertex", 150, FALSE, FALSE, TRUE)
class(res.vertex.ps$sym.prin.curve) <- c("sym.data.table")
sym.scatterplot(res.vertex.ps$sym.prin.curve[, 1], res.vertex.ps$sym.prin.curve[, 2],
  labels = TRUE, col = "red", main = "PSC Face Data"
)

## End(Not run)
```

---

sym.interval.pc.limits

*Symbolic interval principal curves limits*

---

## Description

Symbolic interval principal curves limits.

## Usage

```
sym.interval.pc.limits(sym.data, prin.curve, num.vertex, lambda, var.ord)
```



**Arguments**

sym.data	Symbolic interval data table.
prin.curve	Principal curves.
num.vertex	Number of vertices of the hipercube.
lambda	Lambda.
var.ord	Order of the variables.

**Author(s)**

Jorge Arce.

**References**

- Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.
- Hastie, T. (1984). Principal Curves and Surface. Ph.D Thesis Stanford University.
- Hastie, T. & Weingessel, A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension. R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.
- Hastie, T. & Stuetzle, W. (1989). Principal Curves. Journal of the American Statistical Association, Vol. 84-406, 502–516.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning; Data Mining, Inference and Prediction. Springer, New York.

**See Also**

sym.interval.pc

---

sym.kmeans	<i>Symbolic k-Means</i>
------------	-------------------------

---

**Description**

This is a function is to carry out a k-means overs a interval symbolic data matrix.

**Usage**

```
sym.kmeans(sym.data, k = 3, iter.max = 10, nstart = 1,
algorithm = c('Hartigan-Wong', 'Lloyd', 'Forgy', 'MacQueen'))
```

**Arguments**

sym.data	Symbolic data table.
k	The number of clusters.
iter.max	Maximun number of iterations.
nstart	As in R kmeans function.
algorithm	The method to be use, as in kmeans R function.

**Value**

This function return the following information:

K-means clustering with 3 clusters of sizes 2, 2, 4

Cluster means:

GRA FRE IOD SAP

1 0.93300 -13.500 193.500 174.75

2 0.86300 30.500 54.500 195.25

3 0.91825 -6.375 95.375 191.50

Clustering vector:

L P Co S Ca O B H

1 1 3 3 3 3 2 2

Within cluster sum of squares by cluster:

[1] 876.625 246.125 941.875

(between\_SS / total\_SS = 92.0

Available components:

[1] 'cluster' 'centers' 'totss' 'withinss' 'tot.withinss' 'betweenss'

[7] 'size'

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Carvalho F., Souza R., Chavent M., and Lechevallier Y. (2006) Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. *Pattern Recognition Letters* Volume 27, Issue 3, February 2006, Pages 167-179

**See Also**

sym.hclust

**Examples**

```
data(oils)
sk <- sym.kmeans(oils, k = 3)
sk$cluster
```

sym.knn

*Symbolic k-Nearest Neighbor Regression***Description**

Symbolic k-Nearest Neighbor Regression

**Usage**

```
sym.knn(
  formula,
  sym.data,
  method = c("cm", "crm"),
  scale = TRUE,
  kmax = 20,
  kernel = "triangular"
)
```

**Arguments**

formula	a formula object.
sym.data	symbolc data.table
method	cm or crm
scale	logical, scale variable to have equal sd.
kmax	maximum number of k, if ks is not specified.
kernel	kernel to use. Possible choices are "rectangular" (which is standard unweighted knn), "triangular", "epanechnikov" (or beta(2,2)), "biweight" (or beta(3,3)), "triweight" (or beta(4,4)), "cos", "inv", "gaussian" and "optimal".

**References**

Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515

Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347

Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>

Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

sym.lm

*CM and CRM Linear regression model.*

### Description

To execute the Center Method (CR) and Center and Range Method (CRM) to Linear regression.

### Usage

```
sym.lm(formula, sym.data, method = c('cm', 'crm'))
```

### Arguments

formula	An object of class 'formula' (or one that can be coerced to that class): a symbolic description of the model to be fitted.
sym.data	Should be a symbolic data table read with the function <code>read.sym.table(...)</code> .
method	'cm' to Center Method and 'crm' to Center and Range Method.

### Details

Models for `lm` are specified symbolically. A typical model has the form `response ~ terms` where `response` is the (numeric) response vector and `terms` is a series of terms which specifies a linear predictor for response. A terms specification of the form `first + second` indicates all the terms in `first` together with all the terms in `second` with duplicates removed. A specification of the form `first:second` indicates the set of terms obtained by taking the interactions of all terms in `first` with all terms in `second`. The specification `first*second` indicates the cross of `first` and `second`. This is the same as `first + second + first:second`.

### Value

`sym.lm` returns an object of class 'lm' or for multiple responses of class `c('mlm', 'lm')`

### Author(s)

Oldemar Rodriguez Rojas

### References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347.

**Examples**

```
data(int_prost_train)
data(int_prost_test)
res.cm <- sym.lm(lpsa ~ ., sym.data = int_prost_train, method = "cm")
res.cm
```

---

sym.mcfa

*sym.mcfa*

---

**Description**

This function executes a Multiple Correspondence Factor Analysis for variables of set type.

**Usage**

```
sym.mcfa(sym.data, pos.var)
```

**Arguments**

sym.data      A symbolic data table containing at least two set type variables.  
pos.var      Column numbers in the symbolic data table that contain the set type variables.

**Author(s)**

Jorge Arce

**References**

- Arce J. and Rodriguez, O. (2018). Multiple Correspondence Analysis for Symbolic Multi-Valued Variables. On the Symbolic Data Analysis Workshop SDA 2018.
- Benzecri, J.P. (1973). L'Analyse des Données. Tomo 2: L'Analyse des Correspondances. Dunod, Paris.
- Castillo, W. and Rodriguez O. (1997). Algoritmo e implementacion del analisis factorial de correspondencias. Revista de Matematicas: Teoria y Aplicaciones, 24-31.
- Takagi I. and Yadosiha H. (2011). Correspondence Analysis for symbolic contingency tables base on interval algebra. Elsevier Procedia Computer Science, 6, 352-357.
- Rodriguez, O. (2007). Correspondence Analysis for Symbolic Multi-Valued Variables. CARME 2007 (Rotterdam, The Netherlands), <http://www.carme-n.org/carme2007>.

**Examples**

```
data("ex_mcf1")
sym.table <- classic.to.sym(ex_mcf1,
  concept = suspect,
  hair = sym.set(hair),
  eyes = sym.set(eyes),
  region = sym.set(region)
)
sym.table
```

---

sym.modal	<i>Create an symbolic_modal type object</i>
-----------	---

---

**Description**

Create an symbolic\_modal type object

**Usage**

```
sym.modal(x = character())
```

**Arguments**

x	character vector
---	------------------

**Value**

a symbolic modal

**Examples**

```
sym.modal(factor(c("a", "b", "b", "1")))
```

---

sym.nnet	<i>Symbolic neural networks regression</i>
----------	--

---

**Description**

Symbolic neural networks regression

**Usage**

```

sym.nnet(
  formula,
  sym.data,
  method = c("cm", "crm"),
  hidden = c(10),
  threshold = 0.05,
  stepmax = 1e+05
)

```

**Arguments**

formula	a symbolic description of the model to be fitted.
sym.data	symbolic data.table
method	cm crm
hidden	a vector of integers specifying the number of hidden neurons (vertices) in each layer.
threshold	a numeric value specifying the threshold for the partial derivatives of the error function as stopping criteria.
stepmax	the maximum steps for the training of the neural network. Reaching this maximum leads to a stop of the neural network's training process.

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

sym.pca

---

*Interval Principal Components Analysis.*


---

**Description**

Cazes, Chouakria, Diday and Schektman (1997) proposed the Centers and the Tops Methods to extend the well known principal components analysis method to a particular kind of symbolic objects characterized by multi-values variables of interval type.

**Usage**

```

sym.pca(sym.data, ...)

## S3 method for class 'symbolic_tbl'
sym.pca(
  sym.data,
  method = c("classic", "tops", "centers", "principal.curves", "optimized.distance",
            "optimized.variance"),
  ...
)

```

**Arguments**

sym.data	Should be a symbolic data table
...	further arguments passed to or from other methods.
method	It is use so select the method, 'classic' execute a classical principal component analysis over the centers of the intervals, 'tops' to use the vertices algorithm and 'centers' to use the centers algorithm.

**Value**

Sym.Components: This a symbolic data table with the interval principal components. As this is a symbolic data table we can apply over this table any other symbolic data analysis method (symbolic propagation).

Sym.Prin.Correlations: This is the interval correlations between the original interval variables and the interval principal components, it can be use to plot the symbolic circle of correlations.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Cazes P., Chouakria A., Diday E. et Schektman Y. (1997). Extension de l'analyse en composantes principales a des donnees de type intervalle, Rev. Statistique Appliquee, Vol. XLV Num. 3 pag. 5-24, France.

Chouakria A. (1998) Extension des methodes d'analysis factorielle a des donnees de type intervalle, Ph.D. Thesis, Paris IX Dauphine University.

Makosso-Kallyth S. and Diday E. (2012). Adaptation of interval PCA to symbolic histogram variables, Advances in Data Analysis and Classification July, Volume 6, Issue 2, pp 147-159.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.



**See Also**

sym.histogram.pca

**Examples**

```
## Not run:
data(oils)
res <- sym.pca(oils, "centers")

sym.scatterplot(res$Sym.Components[, 1], res$Sym.Components[, 1],
  labels = TRUE, col = "red", main = "PCA Oils Data"
)
sym.scatterplot3d(res$Sym.Components[, 1], res$Sym.Components[, 2],
  res$Sym.Components[, 3],
  color = "blue", main = "PCA Oils Data"
)
sym.scatterplot.ggplot(res$Sym.Components[, 1], res$Sym.Components[, 2],
  labels = TRUE
)
sym.circle.plot(res$Sym.Prin.Correlations)

res <- sym.pca(oils, "classic")
plot(res, choix = "ind")
plot(res, choix = "var")

data(lynne2)
res <- sym.pca(lynne2, "centers")

sym.scatterplot(res$Sym.Components[, 1], res$Sym.Components[, 2],
  labels = TRUE, col = "red", main = "PCA Lynne Data"
)
sym.scatterplot3d(res$Sym.Components[, 1], res$Sym.Components[, 2],
  res$Sym.Components[, 3],
  color = "blue", main = "PCA Lynne Data"
)
sym.scatterplot.ggplot(res$Sym.Components[, 1], res$Sym.Components[, 2],
  labels = TRUE
)
sym.circle.plot(res$Sym.Prin.Correlations)

data(StudentsGrades)
st <- StudentsGrades
s.pca <- sym.pca(st)
plot(s.pca, choix = "ind")
plot(s.pca, choix = "var")

## End(Not run)
```

**Description**

To execute predict method the Center Method (CR) and Center and Range Method (CRM) to Linear regression.

**Usage**

```
sym.predict(model, ...)

## S3 method for class 'symbolic_lm_cm'
sym.predict(model, new.sym.data, ...)

## S3 method for class 'symbolic_lm_crm'
sym.predict(model, new.sym.data, ...)

## S3 method for class 'symbolic_glm_cm'
sym.predict(model, new.sym.data, response, ...)

## S3 method for class 'symbolic_glm_crm'
sym.predict(model, new.sym.data, response, ...)
```

**Arguments**

model	The output of lm method.
...	additional arguments affecting the predictions produced.
new.sym.data	Should be a symbolic data table read with the function read.sym.table(...).
response	The number of the column where is the response variable in the interval data table.

**Value**

sym.predict produces a vector of predictions or a matrix of predictions and bounds with column names fit, lwr, and upr if interval is set. For type = 'terms' this is a matrix with a column per term and may have an attribute 'constant'

**Author(s)**

Oldemar Rodriguez Rojas

**References**

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347.

**See Also**

sym.glm

**Examples**

```

data(int_prost_train)
data(int_prost_test)
model <- sym.lm(lpsa ~ ., sym.data = int_prost_train, method = "cm")
pred.cm <- sym.predict(model, int_prost_test)
pred.cm

```

---

```

sym.predict.symbolic_gbm_cm
      Predict model_gbm_cm model

```

---

**Description**

Predict model\_gbm\_cm model

**Usage**

```

## S3 method for class 'symbolic_gbm_cm'
sym.predict(model, new.sym.data, n.trees = 500, ...)

```

**Arguments**

model	model
new.sym.data	new data
n.trees	Integer specifying the total number of trees to fit. This is equivalent to the number of iterations and the number of basis functions in the additive expansion. Default is 100.
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

```
sym.predict.symbolic_gbm_crm
```

*Predict model\_gbm\_crm model*

---

### Description

Predict model\_gbm\_crm model

### Usage

```
## S3 method for class 'symbolic_gbm_crm'
sym.predict(model, new.sym.data, n.trees = 500, ...)
```

### Arguments

model	model
new.sym.data	new data
n.trees	Integer specifying the total number of trees to fit. This is equivalent to the number of iterations and the number of basis functions in the additive expansion. Default is 100.
...	optional parameters

### References

Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515

Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347

Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>

Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

```
sym.predict.symbolic_knn_cm
```

*Predict model\_knn\_cm model*

---

### Description

Predict model\_knn\_cm model

### Usage

```
## S3 method for class 'symbolic_knn_cm'
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	model
new.sym.data	new data
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

sym.predict.symbolic\_knn\_crm

*Predict model\_knn\_crm model*

---

**Description**

Predict model\_knn\_crm model

**Usage**

```
## S3 method for class 'symbolic_knn_crm'
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	model
new.sym.data	new data
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

```
sym.predict.symbolic_nnet_cm  
Predict nnet_cm model
```

---

**Description**

Predict nnet\_cm model

**Usage**

```
## S3 method for class 'symbolic_nnet_cm'  
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	model
new.sym.data	new data
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38
- 

```
sym.predict.symbolic_nnet_crm  
Predict nnet_crm model
```

---

**Description**

Predict nnet\_crm model

**Usage**

```
## S3 method for class 'symbolic_nnet_crm'  
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	model
new.sym.data	new data
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

sym.predict.symbolic\_rf\_cm

*Predict rf\_cm model*

---

**Description**

Predict rf\_cm model

**Usage**

```
## S3 method for class 'symbolic_rf_cm'
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	model
new.sym.data	new data
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

```
sym.predict.symbolic_rf_crm  
    Predict rf_crm model
```

---

**Description**

Predict rf\_crm model

**Usage**

```
## S3 method for class 'symbolic_rf_crm'  
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	model
new.sym.data	new data
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

```
sym.predict.symbolic_rt_cm  
    Predict rt_cm model
```

---

**Description**

Predict rt\_cm model

**Usage**

```
## S3 method for class 'symbolic_rt_cm'  
sym.predict(model, new.sym.data, ...)
```



**Arguments**

model	a model_rt_crm object
new.sym.data	new data
...	arguments to predict.rpart

**References**

Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515

Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347

Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>

Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

sym.predict.symbolic\_rt\_crm

*Predict rt\_crm model*

---

**Description**

Predict rt\_crm model

**Usage**

```
## S3 method for class 'symbolic_rt_crm'
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	a model_rt_crm object
new.sym.data	new data
...	optional parameters

**References**

Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515

Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347

Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>

Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

```
sym.predict.symbolic_svm_cm  
Predict model_svm_cm model
```

---

**Description**

Predict model\_svm\_cm model

**Usage**

```
## S3 method for class 'symbolic_svm_cm'  
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	model
new.sym.data	new data
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38
- 

```
sym.predict.symbolic_svm_crm  
Predict model_svm_crm model
```

---

**Description**

Predict model\_svm\_crm model

**Usage**

```
## S3 method for class 'symbolic_svm_crm'  
sym.predict(model, new.sym.data, ...)
```

**Arguments**

model	model
new.sym.data	new data
...	optional parameters

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

sym.rf

*Symbolic Regression with Random Forest***Description**

Symbolic Regression with Random Forest

**Usage**

```
sym.rf(formula, sym.data, method = c("cm", "crm"), ntree = 500)
```

**Arguments**

formula	a formula, with a response but no interaction terms. If this a a data frame, that is taken as the model frame (see model.frame).
sym.data	symbolic data table
method	cm crm
ntree	Number of trees to grow. This should not be set to too small a number, to ensure that every input row gets predicted at least a few times.

**References**

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

 sym.rt

*Symbolic Regression Trees*


---

## Description

Symbolic Regression Trees

## Usage

```
sym.rt(
  formula,
  sym.data,
  method = c("cm", "crm"),
  minsplit = 20,
  maxdepth = 10
)
```

## Arguments

formula	a formula, with a response but no interaction terms. If this a a data frame, that is taken as the model frame (see model.frame).
sym.data	a symbolic data table
method	cm crm
minsplit	the minimum number of observations that must exist in a node in order for a split to be attempted.
maxdepth	Set the maximum depth of any node of the final tree, with the root node counted as depth 0. Values greater than 30 rpart will give nonsense results on 32-bit machines.

## References

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

sym.scatterplot	<i>Symbolic Scatter Plot</i>
-----------------	------------------------------

---

**Description**

This function could be use to plot two symbolic variables in a X-Y plane.

**Usage**

```
sym.scatterplot(sym.var.x, sym.var.y, labels = FALSE, ...)
```

**Arguments**

sym.var.x	First symbolic variable
sym.var.y	Second symbolic variable.
labels	As in R plot function.
...	As in R plot function.

**Value**

Return a graphics.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

**See Also**

sym.scatterplot3d

**Examples**

```
## Not run:  
data(example3)  
sym.data <- example3  
sym.scatterplot(sym.data[, 3], sym.data[, 7], col = "blue", main = "Main Title")  
sym.scatterplot(sym.data[, 1], sym.data[, 4],  
  labels = TRUE, col = "blue",  
  main = "Main Title"  
)
```

```
sym.scatterplot(sym.data[, 2], sym.data[, 6],
  labels = TRUE,
  col = "red", main = "Main Title", lwd = 3
)

data(oils)
sym.scatterplot(oils[, 2], oils[, 3],
  labels = TRUE,
  col = "red", main = "Oils Data"
)
data(lynne1)

sym.scatterplot(lynne1[, 2], lynne1[, 1],
  labels = TRUE,
  col = "red", main = "Lynne Data"
)

## End(Not run)
```

---

sym.set

*Create an symbolic\_set type object*

---

## Description

Create an symbolic\_set type object

## Usage

```
sym.set(x = NA)
```

## Arguments

x                    character vector

## Value

a symbolic set

## Examples

```
sym.set(factor(c("a", "b", "b", "1")))
```

---

 sym.svm

*Symbolic Support Vector Machines Regression*


---

## Description

Symbolic Support Vector Machines Regression

## Usage

```
sym.svm(
  formula,
  sym.data,
  method = c("cm", "crm"),
  scale = TRUE,
  kernel = "radial"
)
```

## Arguments

formula	a symbolic description of the model to be fit.
sym.data	symbolic data.table
method	method
scale	A logical vector indicating the variables to be scaled. If scale is of length 1, the value is recycled as many times as needed. Per default, data are scaled internally (both x and y variables) to zero mean and unit variance. The center and scale values are returned and used for later predictions.
kernel	the kernel used in training and predicting. You might consider changing some of the following parameters, depending on the kernel type.

## References

- Lima-Neto, E.A., De Carvalho, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis* 52, 1500-1515
- Lima-Neto, E.A., De Carvalho, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis* 54, 333-347
- Lima Neto, E.d.A., de Carvalho, F.d.A.T. Nonlinear regression applied to interval-valued data. *Pattern Anal Applic* 20, 809–824 (2017). <https://doi.org/10.1007/s10044-016-0538-y>
- Rodriguez, O. (2018). Shrinkage linear regression for symbolic interval-valued variables. *Journal MODULAD* 2018, vol. Modulad 45, pp.19-38

---

sym.tsne.symbolic\_tbl *TSNE for symbolic data tables*

---

### Description

TSNE for symbolic data tables

### Usage

```
sym.tsne(sym.data, initial_config = NULL, k = 2, initial_dims = 30,  
  perplexity = 30,max_iter = 1000, min_cost = 0, epoch_callback = NULL,  
  whiten = TRUE,epoch=100)
```

### Arguments

sym.data	symbolic data table
initial_config	an argument providing a matrix specifying the initial embedding for X.
k	the dimension of the resulting embedding.
initial_dims	The number of dimensions to use in reduction method.
perplexity	Perplexity parameter. (optimal number of neighbors)
max_iter	Maximum number of iterations to perform.
min_cost	The minimum cost value (error) to halt iteration.
epoch_callback	A callback function used after each epoch (an epoch here means a set number of iterations)
whiten	A boolean value indicating whether the matrix data should be whitened.
epoch	The number of iterations in between update messages.

### Examples

```
## Not run:  
res <- sym.tsne(oils)  
res  
plot(res)  
  
## End(Not run)
```



---

sym.umap.symbolic\_tbl *UMAP for symbolic data tables*

---

### Description

UMAP for symbolic data tables

### Usage

```
sym.umap(sym.data, config = umap::umap.defaults,
method = c("naive", "umap-learn"),preserve.seed = TRUE,...)
```

### Arguments

sym.data	symbolic data table
config	object of class umap.config
method	character, implementation. Available methods are 'naive' (an implementation written in pure R) and 'umap-learn' (requires python package 'umap-learn')
preserve.seed	logical, leave TRUE to insulate external code from randomness within the umap algorithms; set FALSE to allow randomness used in umap algorithms to alter the external random-number generator
...	list of settings; values overwrite defaults from config; see documentation of umap.default for details about available settings

### Examples

```
## Not run:
res <- sym.umap(oils)
res
plot(res)

## End(Not run)
```

---

sym.var *Symbolic Variable*

---

### Description

This function get a symbolic variable from a symbolic data table.

### Usage

```
sym.var(sym.data, number.sym.var)
```

**Arguments**

sym.data            The symbolic data table  
number.sym.var    The number of the column for the variable (feature) that we want to get.

**Value**

Return a symbolic data variable with the following structure:

\$N

[1] 7

\$var.name

[1] 'F6'

\$var.type

[1] '\$F'

\$obj.names

[1] 'Case1' 'Case2' 'Case3' 'Case4' 'Case5' 'Case6' 'Case7'

\$var.data.vector

F6 F6.1

Case1 0.00 90.00

Case2 -90.00 98.00

Case3 65.00 90.00

Case4 45.00 89.00

Case5 20.00 40.00

Case6 5.00 8.00

Case7 3.14 6.76

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

**See Also**

sym.obj

---

USCrime

*Us crime classic data table*

---

**Description**

Us crime classic data table that can be used to generate symbolic data tables.

**Usage**

```
data(USCrime)
```

**Format**

An object of class `data.frame` with 1994 rows and 103 columns.

**Source**

<http://archive.ics.uci.edu/ml/>

**References**

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer.

**Examples**

```
## Not run:  
data(USCrime)  
us.crime <- USCrime  
dim(us.crime)  
head(us.crime)  
summary(us.crime)  
names(us.crime)  
nrow(us.crime)
```

```

result <- classic.to.sym(us.crime,
  concept = "state",
  variables = c(NumInShelters, NumImmig),
  variables.types = c(
    NumInShelters = type.histogram(),
    NumImmig = type.histogram()
  )
)
result

## End(Not run)

```

---

uscrime\_int

*Us crime interval data table.*


---

### Description

Us crime classic data table genetated from uscrime data.

### Usage

```
data(uscrime_int)
```

### Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 46 rows and 102 columns.

### References

Rodriguez O. (2013). A generalization of Centre and Range method for fitting a linear regression model to symbolic interval data using Ridge Regression, Lasso and Elastic Net methods. The IFCS2013 conference of the International Federation of Classification Societies, Tilburg University Holland.

### Examples

```

data(uscrime_int)
car.data <- uscrime_int
res.cm.lasso <- sym.glm(
  sym.data = car.data, response = 102, method = "cm", alpha = 1,
  nfolds = 10, grouped = TRUE
)
plot(res.cm.lasso)
plot(res.cm.lasso$glmnet.fit, "norm", label = TRUE)
plot(res.cm.lasso$glmnet.fit, "lambda", label = TRUE)

pred.cm.lasso <- sym.predict(res.cm.lasso, response = 102, car.data)
RMSE.L(car.data$ViolentCrimesPerPop, pred.cm.lasso)
RMSE.U(car.data$ViolentCrimesPerPop, pred.cm.lasso)

```

```
R2.L(car.data$ViolentCrimesPerPop, pred.cm.lasso)
R2.U(car.data$ViolentCrimesPerPop, pred.cm.lasso)
deter.coefficient(car.data$ViolentCrimesPerPop, pred.cm.lasso)
```

---

uscrime_intv2	<i>Us crime interval data table.</i>
---------------	--------------------------------------

---

### Description

Us crime classic data table genetated from uscrime data.

### Usage

```
data(uscrime_int)
```

### Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 46 rows and 102 columns.

### References

Rodriguez O. (2013). A generalization of Centre and Range method for fitting a linear regression model to symbolic interval data using Ridge Regression, Lasso and Elastic Net methods. The IFCS2013 conference of the International Federation of Classification Societies, Tilburg University Holland.

---

var	<i>Symbolic Variance</i>
-----	--------------------------

---

### Description

Compute the symbolic variance.

### Usage

```
var(x, ...)

## Default S3 method:
var(x, y = NULL, na.rm = FALSE, use, ...)

## S3 method for class 'symbolic_interval'
var(x, method = c("centers", "interval", "billard"), na.rm = FALSE, ...)

## S3 method for class 'symbolic_tbl'
var(x, ...)
```

**Arguments**

x	A symbolic interval.
...	As in R median function.
y	NULL (default) or a vector, matrix or data frame with compatible dimensions to x. The default is equivalent to $y = x$ (but more efficient).
na.rm	logical. Should missing values be removed?
use	an optional character string giving a method for computing covariances in the presence of missing values. This must be (an abbreviation of) one of the strings 'everything', 'all.obs', 'complete.obs', 'na.or.complete', or 'pairwise.complete.obs'.
method	The method to be use.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

- Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.
- Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

---

variance.princ.curve    *Variance of the principal curve*

---

**Description**

Variance of the principal curve

**Usage**

```
variance.princ.curve(data, curve)
```

**Arguments**

data	Classic data table.
curve	The principal curve.

**Value**

The variance of the principal curve.

**Author(s)**

Jorge Arce.

**References**

- Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.
- Hastie, T. (1984). Principal Curves and Surface. Ph.D Thesis Stanford University.
- Hastie, T. & Weingessel, A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension. R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.
- Hastie, T. & Stuetzle, W. (1989). Principal Curves. Journal of the American Statistical Association, Vol. 84-406, 502–516.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning; Data Mining, Inference and Prediction. Springer, New York.

**See Also**

sym.interval.pc

---

vertex.interval	<i>Vertex of the intervals</i>
-----------------	--------------------------------

---

**Description**

Vertex of the intervals

**Usage**

```
vertex.interval(sym.data)
```

**Arguments**

sym.data      Symbolic interval data table.

**Value**

Vertices of the intervals.

**Author(s)**

Jorge Arce.

**References**

- Arce J. and Rodriguez O. (2015) 'Principal Curves and Surfaces to Interval Valued Variables'. The 5th Workshop on Symbolic Data Analysis, SDA2015, Orleans, France, November.
- Hastie, T. (1984). Principal Curves and Surface. Ph.D Thesis Stanford University.
- Hastie, T. & Weingessel, A. (2014). princurve - Fits a Principal Curve in Arbitrary Dimension. R package version 1.1–12 <http://cran.r-project.org/web/packages/princurve/index.html>.

Hastie, T. & Stuetzle, W. (1989). Principal Curves. *Journal of the American Statistical Association*, Vol. 84-406, 502–516.

Hastie, T., Tibshirani, R. & Friedman, J. (2008). *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. Springer, New York.

### See Also

sym.interval.pc

---

VeterinaryData

*Symbolic interval data example*

---

### Description

Symbolic data matrix with all the variables of interval type.

### Usage

```
data(VeterinaryData)
```

### Format

\$I Height Height \$I Weight Weight

1 \$I 120.0 180.0 \$I 222.2 354.0

2 \$I 158.0 160.0 \$I 322.0 355.0

3 \$I 175.0 185.0 \$I 117.2 152.0

4 \$I 37.9 62.9 \$I 22.2 35.0

5 \$I 25.8 39.6 \$I 15.0 36.2

6 \$I 22.8 58.6 \$I 15.0 51.8

7 \$I 22.0 45.0 \$I 0.8 11.0

8 \$I 18.0 53.0 \$I 0.4 2.5

9 \$I 40.3 55.8 \$I 2.1 4.5

10 \$I 38.4 72.4 \$I 2.5 6.1



**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(VeterinaryData)
VeterinaryData
```

---

write.sym.table	<i>Write Symbolic Data Table</i>
-----------------	----------------------------------

---

**Description**

This function write (save) a symbolic data table from a CSV data file.

**Usage**

```
write.sym.table(sym.data, file, sep, dec, row.names = NULL, col.names = NULL)
```

**Arguments**

sym.data	Symbolic data table
file	The name of the CSV file.
sep	As in R function read.table
dec	As in R function read.table
row.names	As in R function read.table
col.names	As in R function read.table

**Value**

Write in CSV file the symbolic data table.

**Author(s)**

Oldemar Rodriguez Rojas

**References**

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

**See Also**

read.sym.table

**Examples**

```
## Not run:
data(example1)
write.sym.table(example1, file = "temp4.csv", sep = "|",
                dec = ".", row.names = TRUE, col.names = TRUE)
ex1 <- read.sym.table("temp4.csv", header = TRUE,
                    sep = "|", dec = ".", row.names = 1)

## End(Not run)
```

---

\$.symbolic\_histogram    *\$ operator for histograms*

---

**Description**

\$ operator for histograms

**Usage**

```
## S3 method for class 'symbolic_histogram'
x$name
```

**Arguments**

x	.....
name	...

---

\$.symbolic\_modal        *\$ operator for modals*

---

**Description**

\$ operator for modals

**Usage**

```
## S3 method for class 'symbolic_modal'
x$name = c("cats", "props", "counts")
```

**Arguments**

x	.....
name	...

---

\$.symbolic_set	<i>\$ operator for set</i>
-----------------	----------------------------

---

**Description**

\$ operator for set

**Usage**

```
## S3 method for class 'symbolic_set'  
x$name = c("levels", "values")
```

**Arguments**

x	.....
name	...

# Index

- \* **Circle**
  - sym.circle.plot, 50
- \* **Correlation**
  - cor, 10
- \* **Covariance**
  - cov, 11
- \* **Curve**
  - centers.interval, 8
  - dist.vect, 13
  - dist.vect.matrix, 14
  - neighbors.vertex, 35
  - norm.vect, 36
  - sym.interval.pc, 55
  - sym.interval.pc.limits, 56
  - variance.princ.curve, 86
  - vertex.interval, 87
- \* **Elastic**
  - deter.coefficient, 12
- \* **Intervals**
  - sym.pca, 63
- \* **Kmeans**
  - sym.kmeans, 57
- \* **Lasso**
  - deter.coefficient, 12
  - sym.glm, 53
- \* **Mean**
  - mean.symbolic\_interval, 32
- \* **Median**
  - median.symbolic\_interval, 33
- \* **Net**
  - deter.coefficient, 12
- \* **PCA**
  - sym.pca, 63
- \* **Plot**
  - plot.symbolic\_tbl, 37
  - sym.scatterplot, 77
- \* **Principal**
  - centers.interval, 8
  - dist.vect, 13
  - dist.vect.matrix, 14
  - neighbors.vertex, 35
  - norm.vect, 36
  - sym.interval.pc, 55
  - sym.interval.pc.limits, 56
  - variance.princ.curve, 86
  - vertex.interval, 87
- \* **Regression**
  - sym.glm, 53
- \* **Ridge**
  - deter.coefficient, 12
  - sym.glm, 53
- \* **Symbolic**
  - cor, 10
  - cov, 11
  - deter.coefficient, 12
  - mean.symbolic\_interval, 32
  - median.symbolic\_interval, 33
  - plot.symbolic\_tbl, 37
  - read.sym.table, 42
  - sd, 47
  - SDS.to.RSDA, 48
  - SODAS.to.RSDA, 49
  - sym.circle.plot, 50
  - sym.glm, 53
  - sym.kmeans, 57
  - sym.lm, 60
  - sym.predict, 65
  - sym.scatterplot, 77
  - sym.var, 81
  - var, 85
- \* **Table**
  - read.sym.table, 42
- \* **Variable**
  - sym.var, 81
- \* **Variance**
  - var, 85
- \* **correlation**
  - R2.L, 40

- R2.U, 41
- \* **datasets**
  - abalone, 4
  - Cardiological, 7
  - cardiologicalv2, 8
  - ex1\_db2so, 15
  - ex\_cfa1, 20
  - ex\_cfa2, 21
  - ex\_mcfa1, 21
  - ex\_mcfa2, 22
  - example1, 15
  - example2, 17
  - example3, 17
  - example4, 18
  - example5, 18
  - example6, 19
  - example7, 20
  - facedata, 23
  - int\_prost\_test, 28
  - int\_prost\_train, 29
  - lynne1, 31
  - oils, 37
  - USCrime, 83
  - uscrime\_int, 84
  - uscrime\_intv2, 85
  - VeterinaryData, 88
- \* **data**
  - plot.symbolic\_tbl, 37
  - SDS.to.RSDA, 48
  - SODAS.to.RSDA, 49
- \* **lm**
  - sym.lm, 60
  - sym.predict, 65
- \* **lower**
  - R2.L, 40
  - RMSE.L, 44
- \* **package**
  - RSDA, 46
- \* **root-mean-square**
  - RMSE.L, 44
  - RMSE.U, 45
- \* **sd**
  - sd, 47
- \* **table**
  - plot.symbolic\_tbl, 37
  - SDS.to.RSDA, 48
  - SODAS.to.RSDA, 49
- \* **upper**
  - R2.U, 41
  - RMSE.U, 45
  - \$.symbolic\_histogram, 90
  - \$.symbolic\_interval
    - (min.symbolic\_interval), 34
  - \$.symbolic\_modal, 90
  - \$.symbolic\_set, 91
  - abalone, 4
  - as.data.frame.symbolic\_histogram, 5
  - as.data.frame.symbolic\_interval, 5
  - as.data.frame.symbolic\_modal, 6
  - as.data.frame.symbolic\_set, 6
  - calc.burt.sym, 7
  - Cardiological, 7
  - cardiologicalv2, 8
  - centers.interval, 8
  - classic.to.sym, 9
  - cor, 10
  - cov, 11
  - deter.coefficient, 12
  - dist.vect, 13
  - dist.vect.matrix, 14
  - ex1\_db2so, 15
  - ex\_cfa1, 20
  - ex\_cfa2, 21
  - ex\_mcfa1, 21
  - ex\_mcfa2, 22
  - example1, 15
  - example2, 17
  - example3, 17
  - example4, 18
  - example5, 18
  - example6, 19
  - example7, 20
  - facedata, 23
  - format.symbolic\_histogram, 24
  - format.symbolic\_interval, 24
  - format.symbolic\_modal, 25
  - format.symbolic\_set, 25
  - get\_cats, 26
  - get\_props, 26
  - int\_prost\_test, 28
  - int\_prost\_train, 29

interval.centers, 26  
interval.histogram.plot, 27  
interval.max, 27  
interval.min, 28  
interval.ranges, 28  
is.sym.histogram, 29  
is.sym.interval, 30  
is.sym.modal, 30  
is.sym.set, 31  
  
lynne1, 31  
  
max.symbolic\_interval  
    (min.symbolic\_interval), 34  
mcfa.scatterplot, 32  
mean.symbolic\_interval, 32  
mean.symbolic\_tbl  
    (mean.symbolic\_interval), 32  
median.symbolic\_interval, 33  
median.symbolic\_tbl  
    (median.symbolic\_interval), 33  
method\_summary, 34  
min.symbolic\_interval, 34  
  
neighbors.vertex, 35  
norm.vect, 36  
  
oils, 37  
oilils(oils), 37  
  
plot.sym\_tsne, 39  
plot.sym\_umap, 39  
plot.symbolic\_tbl, 37  
  
R2.L, 40  
R2.U, 41  
read.sym.table, 42  
RMSE.L, 44  
RMSE.U, 45  
RSDA, 46  
  
sd, 47  
SDS.to.RSDA, 48  
SODAS.to.RSDA, 49  
sym.circle.plot, 50  
sym.dist.interval, 51  
sym.gbm, 52  
sym.glm, 53  
sym.histogram, 54  
sym.interval, 54  
sym.interval.pc, 55  
sym.interval.pc.limits, 56  
sym.interval.pca(sym.pca), 63  
sym.kmeans, 57  
sym.knn, 59  
sym.lm, 60  
sym.mcfa, 61  
sym.modal, 62  
sym.nnet, 62  
sym.pca, 63  
sym.predict, 65  
sym.predict.symbolic\_gbm\_cm, 67  
sym.predict.symbolic\_gbm\_crm, 68  
sym.predict.symbolic\_knn\_cm, 68  
sym.predict.symbolic\_knn\_crm, 69  
sym.predict.symbolic\_nnet\_cm, 70  
sym.predict.symbolic\_nnet\_crm, 70  
sym.predict.symbolic\_rf\_cm, 71  
sym.predict.symbolic\_rf\_crm, 72  
sym.predict.symbolic\_rt\_cm, 72  
sym.predict.symbolic\_rt\_crm, 73  
sym.predict.symbolic\_svm\_cm, 74  
sym.predict.symbolic\_svm\_crm, 74  
sym.rf, 75  
sym.rt, 76  
sym.scatterplot, 77  
sym.set, 78  
sym.svm, 79  
sym.tsne(sym.tsne.symbolic\_tbl), 80  
sym.tsne.symbolic\_tbl, 80  
sym.umap(sym.umap.symbolic\_tbl), 81  
sym.umap.symbolic\_tbl, 81  
sym.var, 81  
  
USCrime, 83  
uscrime\_int, 84  
uscrime\_intv2, 85  
  
var, 85  
variance.princ.curve, 86  
vertex.interval, 87  
VeterinaryData, 88  
  
write.sym.table, 89