R documentation

of 'discretize.Rd'

August 8, 2017

---

discretize *Discretize continuous data.*

Description

Converts continuous columns to discrete.

Usage

```
discretize(dataset, input, ndigs=0, nlevels=10)
```

Arguments

dataset	The dataset to discretize, data frame/table.
input	Optional specification for partitioning, giving the number of partitions and labels for each partition. List of lists, one list per column to be converted. The outermost list states the columns to be converted, and each inner liss holdt the name of the column, the number of partitions, and a list of labels for each partition.
ndigs	Number of digits to retain in forming labels/values for the discretized data, if input is not supplied. E.g. if ndigs is 2 and the original datum is 38.12, it becomes 38.
nlevels	The number of partitions to form, if input is NULL.

Details

If input is not specified, each numeric column in the data will be discretized, with one exception: If a column is numeric but has fewer distinct values than nlevels, it is presumed to be an informal R factor and is not converted. However, it is best to use makeFactor() on such variables.

Author(s)

Norm Matloff <matloff@cs.ucdavis.edu>, Vincent Yang <vinyang@ucdavis.edu>, and Harrison Nguyen <hhnguy@ucdavis.edu>

Examples

```
data(prgeng)
pe <- prgeng[,c(1,3,5,7:9)]  # extract vars of interest
pe25 <- pe[pe$wageinc < 250000,]  # delete extreme values
pe25disc <- discretize(pe25)  # age, wageinc and wkswrkd discretized

data(mlb)
# extract the height, weight, age, and position of players
m <- mlb[,4:7]

inp1 <- list("name" = "Height",
             "partitions"=4,
             "labels"=c("short", "shortmid", "tallmid", "tall"))

inp2 <- list("name" = "Weight",
             "partitions"=3,
             "labels"=c("light", "med", "heavy"))

inp3 <- list("name" = "Age",
             "partitions"=3,
             "labels"=c("young", "med", "old"))

# create one list to pass everything to discretize()
discreteinput <- list(inp1, inp2, inp3)

# at this point, all of the data has been discretized
discretizedmlb <- discretize(m, discreteinput)
```

Index