

Package: onehot (via r-universe)

October 27, 2024

Type Package

Title Fast Onehot Encoding for Data.frames

Version 0.1.3

Author Eric E. Graves [aut, cre]

Maintainer Eric E. Graves <gravcon5@gmail.com>

Description Quickly create numeric matrices for machine learning algorithms that require them. It converts factor columns into onehot vectors.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

Depends Matrix

Suggests testthat

Repository <https://gravesee.r-universe.dev>

RemoteUrl <https://github.com/gravesee/onehot>

RemoteRef HEAD

RemoteSha 26918e160d30a834b192d95e9bd1c7587b994f2c

Contents

make_names	2
onehot	2
predict.onehot	3
print.onehot	4
sas	4
summary.onehot	5

Index	6
--------------	----------

make_names	<i>Make column names for a onehot object</i>
------------	--

Description

Make column names for a onehot object

Usage

```
make_names(info, sep)
```

Arguments

x a [onehot](#) object

Examples

```
data(iris)
encoder <- onehot(iris)
make_names(encoder$Species)
```

onehot	<i>Onehot Encode a data.frame</i>
--------	-----------------------------------

Description

Onehot Encode a data.frame

Usage

```
onehot(data, sentinel = -999, max_levels = 10, add_NA_factors = TRUE)
```

Arguments

data data.frame to convert factors into onehot encoded columns

sentinel Numeric value with which to replace NAs. Applies to numeric columns only.

max_levels maximum number of levels to onehot encode per factor variable. Factors with levels exceeding this number will be skipped.

add_NA_factors if TRUE, adds NA indicator column for factors.

Details

By default, with `addNA=FALSE`, no NAs are returned for non-factor columns. Indicator columns are created for factor levels and NA factors are ignored. The exception is when NA is an explicit factor level.

`stringsAsFactors=TRUE` will convert character columns to factors first. Other wise characters are ignored. Only factor, numeric, integer, and logical vectors are valid for onehot. Other classes will be skipped entirely.

`addNA=TRUE` will create indicator columns for every field. This will add `ncols` columns to the output matrix. A sparse matrix may be better in such cases.

Value

a onehot object describing how to transform the data

Examples

```
data(iris)
encoder <- onehot(iris)

## add NA indicator columns
encoder <- onehot(iris, add_NA_factors=TRUE)

## limit which factors are onehot encoded
encoder <- onehot(iris, max_levels=5)

## Impute numeric NA values with sentinel value
encoder <- onehot(iris, sentinel=-1)
```

predict.onehot	<i>Predict onehot objects</i>
----------------	-------------------------------

Description

Predict onehot objects

Usage

```
## S3 method for class 'onehot'
predict(object, data, sparse = FALSE, sep = "_", ...)
```

Arguments

object	an object of class <code>onehot</code>
data	a data.frame to onehot encode using object
sparse	if TRUE, returns a <code>dgCMatrix-class</code>
...	further arguments passed to or from other methods

Value

a matrix with factor variable onehot encoded

Examples

```
data(iris)
encoder <- onehot(iris)
x <- predict(encoder, iris)
x_sparse <- predict(encoder, iris, sparse=TRUE)
```

<code>print.onehot</code>	<i>Print information about a onehot object</i>
---------------------------	--

Description

Print information about a onehot object

Usage

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

Arguments

<code>x</code>	onehot object to print
<code>...</code>	other arguments pass to or from other functions

<code>sas</code>	<i>Generate SAS code for onehot object</i>
------------------	--

Description

Generate SAS code for onehot object

Usage

```
sas(x, sep = "_", ...)
```

Arguments

<code>x</code>	a onehot object
<code>sep</code>	a character vector used to separate the name of a factor from the value.

Value

Returns a character vector of SAS code that can be written to file using `writeln`

summary.onehot	<i>Summarize onehot object</i>
----------------	--------------------------------

Description

Summarize onehot object

Usage

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

Arguments

object	a onehot object
...	other arguments pass to or from other functions

Examples

```
## Create some dummy data with different column types  
x <- data.frame(HairEyeColor)  
x$Hair <- as.character(x$Hair)  
  
## Create a onehot object  
encoder <- onehot(x)  
  
## Return a list with summary information  
summary(encoder)
```

Index

`make_names`, 2

`onehot`, 2, 2, 3

`predict.onehot`, 3

`print.onehot`, 4

`sas`, 4

`summary.onehot`, 5