Package ‘naivebayes’

January 7, 2019

Type Package

Title High Performance Implementation of the Naive Bayes Algorithm

Version 0.9.3

Author Michal Majka

Maintainer Michal Majka <michalmajka@hotmail.com>

Description High performance implementation of the Naive Bayes algorithm.

URL http://github.com/majkamichal/naivebayes

BugReports http://github.com/majkamichal/naivebayes/issues

License GPL-2

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2019-01-07 21:50:51 UTC

R topics documented:

naive_bayes *Naive Bayes Classifier*

Description

naive_bayes is used to fit Naive Bayes model in which predictors are assumed to be independent within each class label.

Usage

```
## Default S3 method:
naive_bayes(x, y, prior = NULL, laplace = 0,
  usekernel = FALSE, ...)

## S3 method for class 'formula'
naive_bayes(formula, data, prior = NULL, laplace = 0,
  usekernel = FALSE, subset, na.action = stats::na.pass, ...)
```

Arguments

x	matrix or dataframe with categorical (character/factor/logical) or metric (numeric) predictors.
y	class vector (character/factor/logical).
formula	an object of class "formula" (or one that can be coerced to "formula") of the form class ~ predictors (class has to be a factor/character/logical).
data	matrix or dataframe with categorical (character/factor/logical) or metric (numeric) predictors.
laplace	value used for Laplace smoothing. Defaults to 0 (no Laplace smoothing).
usekernel	logical; if TRUE, density is used to estimate the densities of metric predictors.
prior	vector with prior probabilities of the classes. If unspecified, the class proportions for the training set are used. If present, the probabilities should be specified in the order of the factor levels.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. By default (na.pass), missing values are not removed from the data and are then omited while constructing tables. Alternatively, na.omit can be used to exclude rows with at least one missing value before constructing tables.
...	other graphical parameters to density (na.rm defaults to TRUE).

Details

Metric predictors are handled by assuming that they follow Gaussian distribution, given the class label. Alternatively, kernel density estimation can be used to estimate their class-conditional distributions. Missing values are not included into constructing tables. Logical variables are treated as categorical (binary) variables.

Value

naive_bayes returns an object of class "naive_bayes" which is a list with following components:

data	list with two components: x (dataframe with predictors) and y (class variable).
tables	list of tables. For each categorical predictor a table with class-conditional probabilities and for each metric predictor a table with means and standard deviations or density objects for each class.
prior	numeric vector with prior probabilities.
laplace	value of Laplace smoothing.
levels	character vector with values of the class variable.
usekernel	logical; TRUE, if the density was used for estimating densities of numeric variables.
call	the call that produced this object.

Author(s)

Michal Majka, <michalmajka@hotmail.com>

See Also

predict.naive_bayes, plot.naive_bayes, tables

Examples

```
data(iris)
nb <- naive_bayes(Species ~ ., data = iris)
plot(nb)

nb_kernel <- naive_bayes(x = iris[-5], y = iris[ ,5], usekernel = TRUE)
plot(nb_kernel)


## Not run:
vars <- 10
rows <- 500000
y <- sample(c("a", "b"), rows, TRUE)

# Only categorical variables
X1 <- as.data.frame(matrix(sample(letters[5:9], vars * rows, TRUE),
                           ncol = vars))
nb_cat <- naive_bayes(x = X1, y = y)
nb_cat
system.time(pred2 <- predict(nb_cat, X1))

# Only numeric variables
X2 <- as.data.frame(matrix(rnorm(vars * rows), ncol = vars))
nb_num <- naive_bayes(x = X2, y = y)
nb_num
system.time(pred2 <- predict(nb_num, X2))
```

```
## End(Not run)
```

---

plot.naive_bayes *Plot Method for naive_bayes Object*

Description

Plot method for objects of class "naive_bayes" designed for a quick look at the marginal probabilities of predictor variables given the class.

Usage

```
## S3 method for class 'naive_bayes'
plot(x, which = NULL, ask = FALSE, legend = TRUE,
  legend.box = FALSE, arg.num = list(), arg.cat = list(), ...)
```

Arguments

x	object of class inheriting from "naive_bayes".
which	variables to be plotted (all by default). This can be any valid indexing vector or vector containing names of variables.
ask	logical; if TRUE, the user is asked before each plot, see par(ask=.).
legend	logical; if TRUE a legend will be be plotted.
legend.box	logical; if TRUE a box will be drawn around the legend.
arg.num	other parameters to be passed as a named list to matplot.
arg.cat	other parameters to be passed as a named list to mosaicplot.
...	not used.

Details

Conditional probabilities are visualised by matplot (for metric predictors) and mosaicplot (for categorical predictors). In case of non parametric estimation of class-conditional densities, the bandwidths are reported. Nothing is returned.

Author(s)

Michal Majka, <michalmajka@hotmail.com>

See Also

naive_bayes, predict.naive_bayes, tables

Examples

```
data(iris)
iris2 <- cbind(iris, New = sample(letters[1:3], 150, TRUE))
nb <- naive_bayes(Species ~ ., data = iris2)
plot(nb, ask = TRUE)
plot(nb, which = c(1, 2), ask = TRUE,
     arg.num = list(col = 1:3, lty = 1, main = "Naive Bayes Plot"))
plot(nb, which = "New", arg.cat = list(color = 4:7))
```

---

predict.naive_bayes *Predict Method for naive_bayes Objects*

Description

Classification based on Naive Bayes models.

Usage

```
## S3 method for class 'naive_bayes'
predict(object, newdata = NULL, type = c("class","prob"),
  threshold = 0.001, ...)
```

Arguments

object	object of class inheriting from "naive_bayes".
newdata	matrix or dataframe with categorical (character/factor/logical) or metric (numeric) predictors.
type	if "class", new data points are classified according to the highest posterior probabilities. If "prob", the posterior probabilities for each class are returned.
threshold	value by which zero probabilities corresponding to metric variables are replaced (zero probabilities corresponding to categorical variables can be handled with Laplace smoothing).
...	not used.

Details

Computes conditional posterior probabilities for each class label using the Bayes' rule under the assumption of independence of predictors. If no new data is provided, the data from the object is used. Logical variables are treated as categorical (binary) variables. Predictors with missing values are not included into the computation of posterior probabilities.

Value

predict.naive_bayes returns either a factor with class labels corresponding to the maximal conditional posterior probabilities or a matrix with class label specific conditional posterior probabilities.

Author(s)

Michal Majka, <michalmajka@hotmail.com>

See Also

naive_bayes, plot.naive_bayes, tables

Examples

```
ind_iris <- sample(1:nrow(iris), size = round(0.3 * nrow(iris)))
iris_train <- iris[-ind_iris, ]
iris_test <- iris[ind_iris, ]

nb_iris <- naive_bayes(Species ~ ., iris_train)
predict(nb_iris, iris_test)
head(predict(nb_iris, iris_test, type = "prob"))


## Not run:
vars <- 10
rows <- 500000
y <- sample(c("a", "b"), rows, TRUE)

# Only categorical variables
X1 <- as.data.frame(matrix(sample(letters[5:9], vars * rows, TRUE),
                           ncol = vars))
nb_cat <- naive_bayes(x = X1, y = y)
nb_cat
system.time(pred2 <- predict(nb_cat, X1))

# Only numeric variables
X2 <- as.data.frame(matrix(rnorm(vars * rows), ncol = vars))
nb_num <- naive_bayes(x = X2, y = y)
nb_num
system.time(pred2 <- predict(nb_num, X2))

## End(Not run)
```

---

tables *Browse Tables of Naive Bayes Classifier*

Description

Auxiliary function for "naive_bayes" objects for easy browsing tables.

Usage

```
tables(object, which = NULL)
```

Arguments

object	object of class inheriting from "naive_bayes".
which	tables to be showed (all by default). This can be any valid indexing vector or vector containing names of variables.

Details

Default print method for "naive_bayes" objects shows at most five first tables. The auxiliary function tables returns by default all tables.

Value

list with tables.

Author(s)

Michal Majka, <michalmajka@hotmail.com>

See Also

naive_bayes, predict.naive_bayes, plot.naive_bayes

Examples

```
data(iris)
nb <- naive_bayes(Species ~ ., data = iris)
tables(nb, "Sepal.Length")
tables(nb, 1:2)
```

Index