Package 'abcrlda'

October 4, 2019

Type Package

Title Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis

Version 0.1.1

Author Dmitriy Fedorov <dmitriy.fedorov@nu.edu.kz>

Maintainer Dmitriy Fedorov <dmitriy.fedorov@nu.edu.kz>

Description This package offers methods to perform
asymptotically bias-corrected regularized linear discriminant analysis
for cost-sensitive binary classification.

Imports stats

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

R topics documented:

---

abcrlda | *Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis for Cost-Sensitive Binary Classification*

Description

Performs Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis

Usage

```
abcrlda(x, grouping, gamma = 1, cost = c(0.5, 0.5))
```

1

Arguments

x	Matrix or data.frame of observations.
grouping	Grouping variable. A vector of numeric values 0 and 1 is recommended. Length has to correspond to nrow(x).
gamma	Regularization parameter.
cost	Parameter that controls priorety of class 0.

Value

An object of class "rrlda" is returned which can be used for class prediction (see predict())

a	Slope of a discriminant hyperplane. W(x) = a'x + m.
m	Bias term. W(x) = a'x + m.
cost	Normilized cost such that cost_10 + cost_01 == 1.
gamma	Regularization parameter.
lev	Levels. Corresponds to the groups.

See Also

Other abcrlda binary classifier: cross_validation, predict.abcrlda

Examples

```
data(iris)
train_data <- iris[which(iris[, ncol(iris)] == ”virginica” |
                         iris[, ncol(iris)] == ”versicolor”), 1:4]
train_label <- factor(iris[which(iris[, ncol(iris)] == ”virginica” |
                                 iris[, ncol(iris)] == ”versicolor”), 5])
model <- abcrlda(train_data, train_label, gamma = 0.5, cost_10 = 0.75)
predict(model, train_data)
```

---

cross_validation	*Cross Validation*

Description

Cross Validation

Usage

```
cross_validation(x, grouping, gamma = 1, cost = c(0.5, 0.5),
  kfolds = 10)
```

Arguments

x	Matrix or data.frame of observations.
grouping	Grouping variable. A vector of numeric values 0 and 1 is recommended. Length has to correspond to nrow(x).
gamma	regularization parameter
kfolds	Number of for cross validation algorithm
C_10	parameter that controls prioretization of classes. It's value should be between 0 and 1 (0 < cost_10 < 1) Values bigger than 0.5 prioretizes correct classification of 0 class while values less than 0.5 prioretizes 1 class

Value

Returns average error of cross validation

See Also

Other abcrlda binary classifier: `abcrlda`, `predict.abcrlda`

Examples

```
data(iris)
train_data <- iris[which(iris[, ncol(iris)] == "virginica" |
                         iris[, ncol(iris)] == "versicolor"), 1:4]
train_label <- factor(iris[which(iris[, ncol(iris)] == "virginica" |
                                 iris[, ncol(iris)] == "versicolor"), 5])
model <- abcrlda(train_data, train_label, gamma = 0.5, cost_10 = 0.75)
risk_estimate_20(model)
```

---

grid_search	*Grid Search*

Description

Performs grid search based on cross validation or error estimation formula.

Usage

```
grid_search(x, grouping, range_gamma, range_cost, method = "estimator",
  k_fold = 10)
```

Arguments

x	Matrix or data.frame of observations.
grouping	Grouping variable. A vector of numeric values 0 and 1 is recommended. Length has to correspond to nrow(x).
range_gamma	vector of gamma values to check
range_cost	[1 x n] vector or [2 x n] matrix of cost values to check
method	selects method to evaluete error. "estimator" and "cross"
k_fold	number of fold to use with cross-validation

Value

List of best founded parameters

Examples

```
data(iris)
train_data <- iris[which(iris[, ncol(iris)] == "virginica" |
                        iris[, ncol(iris)] == "versicolor"), 1:4]
train_label <- factor(iris[which(iris[, ncol(iris)] == "virginica" |
                                iris[, ncol(iris)] == "versicolor"), 5])
cost_range <- seq(0.1, 0.9, by = 0.2)
gamma_range <- c(0.1, 1, 10, 100, 1000)

gs <- grid_search(train_data, train_label,
                  range_gamma = gamma_range,
                  range_C_10 = cost_range,
                  method = "estimator")
model <- abcrlda(train_data, train_label,
                 gamma = gs$gamma[1], cost_10 = gs$C_10[1])
predict(model, train_data)
```

---

predict.abcrlda   *Class Prediction for abcrlda objects*

Description

Computes class predictions for new data based on a given abcrlda object

Usage

```
## S3 method for class 'abcrlda'
predict(object, x, type = "class", ...)
```

Arguments

object	An object of class "abcrlda".
x	New data for which the classes are to predict
...	Argument used by generic function predict(object, x, ...).

Value

class Class prediction for each observation. raw Raw values.

See Also

Other abcrlda binary classifier: abcrlda, cross_validation

Examples

```
data(iris)
train_data <- iris[which(iris[, ncol(iris)] == "virginica" |
                         iris[, ncol(iris)] == "versicolor"), 1:4]
train_label <- factor(iris[which(iris[, ncol(iris)] == "virginica" |
                                 iris[, ncol(iris)] == "versicolor"), 5])
model <- abcrlda(train_data, train_label, gamma = 0.5, cost_10 = 0.75)
predict(model, train_data)
```

---

risk_estimate_20   *Risk Estimator*

Description

Calculates weighted error based on normalized cost values

Usage

```
risk_estimate_20(object)
```

Arguments

object   An object of class "abcrlda".

Value

Weighted error based on "abcrlda" object

Examples

```
data(iris)
train_data <- iris[which(iris[, ncol(iris)] == "virginica" |
                         iris[, ncol(iris)] == "versicolor"), 1:4]
train_label <- factor(iris[which(iris[, ncol(iris)] == "virginica" |
                                 iris[, ncol(iris)] == "versicolor"), 5])
model <- abcrlda(train_data, train_label, gamma = 0.5, cost_10 = 0.75)
risk_estimate_20(model)
```

Index