Package 'abcrlda'

October 17, 2019

Type Package

Title Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis

Version 0.1.1

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

URL https://ieeexplore.ieee.org/document/8720003/

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, y, gamma = 1, cost = c(0.5, 0.5))
```

Arguments

x	Input matrix or data.frame , of dimension nobs x nvars; each row is an observation vector.
y	Class labels. Should be a factor with two levels, or a vector with two distinct values. If y is presented as a vector, it will be coerced into a factor. Length of y has to correspond to number of samples in x.
gamma	Regularization parameter in the following equation $$W_{ABC}^{RLDA} = \gamma(x - \frac{\bar{x}_0 + \bar{x}_1}{2})^T H(\bar{x}_0 - \bar{x}_1) - log(\frac{C_{01}}{C_{10}}) + \omega_{opt}^{\check{}}$$
cost	parameter that controls prioretization of classes. This is a vector of length 1 or 2 where first value is C_{10} (represents prioretization of class 0) and second value if provided is C_{01} (represents prioretization of class 1). Default value is c(0.5, 0.5), so both classes have equal priority and risk essentially becomes equivalent to error rate. If single value is provided it should be normalized to be between 0 and 1 (but not including 0 or 1). This value will be assigned to C_{10} and C_{01} will be equal to $(1 - C_{10})$ In a vector of length 1, values bigger than 0.5 prioretizes correct classification of 0 class while values less than 0.5 prioretizes 1 class.

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 C_{10} + C_{01} == 1.
gamma	Regularization parameter value provided during fitting.
lev	Levels. Corresponds to the labels in y.

Reference

A. Zollanvari, M. Abdirash, A. Dadlani and B. Abibullaev, "Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis for Cost-Sensitive Binary Classification," in IEEE Signal Processing Letters, vol. 26, no. 9, pp. 1300-1304, Sept. 2019. doi: 10.1109/LSP.2019.2918485 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8720003&isnumber=8770167

See Also

Other Other functions in the package: cross_validation, da_risk_estimator, grid_search, 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 = 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	Input matrix or data.frame , of dimension nobs x nvars; each row is an observation vector.
gamma	regularization parameter
cost	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
kfolds	Number of for cross validation algorithm

Value

Returns average error of cross validation

See Also

Other Other functions in the package: abcrlda, da_risk_estimator, grid_search, 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 = 0.75)
risk_estimate_20(model)
```

da_risk_estimator *Double Asymptotic Risk Estimator*

Description

Calculates weighted error based on normalized cost values

Usage

```
da_risk_estimator(object)
```

Arguments

object	An object of class "abcrlda".

Value

Weighted error based on "abcrlda" object

Reference

A. Zollanvari, M. Abdirash, A. Dadlani and B. Abibullaev, "Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis for Cost-Sensitive Binary Classification," in IEEE Signal Processing Letters, vol. 26, no. 9, pp. 1300-1304, Sept. 2019. doi: 10.1109/LSP.2019.2918485 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8720003&isnumber=8770167

See Also

Other Other functions in the package: abcrlda, cross_validation, grid_search, 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 = 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, y, range_gamma, range_cost, method = "estimator",
  nfold = 10)
```

Arguments

x	Input matrix or data.frame , of dimension nobs x nvars; each row is an observation vector.
y	Class labels. Should be a factor with two levels, or a vector with two distinct values. If y is presented as a vector, it will be coerced into a factor. Length of y has to correspond to number of samples in 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"
nfold	number of fold to use with cross-validation

Value

List of best founded parameters

See Also

Other Other functions in the package: abcrlda, cross_validation, da_risk_estimator, 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])
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_cost = cost_range,
                  method = "estimator")
model <- abcrlda(train_data, train_label,
                 gamma = gs$gamma[1], cost = gs$cost[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, newx, out_type = "class", ...)
```

Arguments

object	An object of class "abcrlda".
newx	Matrix of new values for x at which predictions are to be made.
...	Argument used by generic function predict(object, x, ...).
type	Determines a type of output. Two type of input could be provided. If "class" value is provided this will return factor with levels corresponding to lev stored in object. If "raw" value is provided this will return numeric vector with values obtained from discriminant function.

Value

class Class prediction for each observation. raw Raw values.

Reference

A. Zollanvari, M. Abdirash, A. Dadlani and B. Abibullaev, "Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis for Cost-Sensitive Binary Classification," in IEEE Signal Processing Letters, vol. 26, no. 9, pp. 1300-1304, Sept. 2019. doi: 10.1109/LSP.2019.2918485 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8720003&isnumber=8770167

See Also

Other Other functions in the package: abcrlda, cross_validation, da_risk_estimator, grid_search

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 = 0.75)
predict(model, train_data)
```

Index