

Package ‘dawai’

October 13, 2022

Title Discriminant Analysis with Additional Information

Version 1.2.5

Date 2021-05-03

Author David Conde, Miguel A. Fernandez, Bonifacio Salvador

Maintainer David Conde <dconde@eio.uva.es>

Description In applications it is usual that some additional information is available. This package dawai (an acronym for Discriminant Analysis With Additional Information) performs linear and quadratic discriminant analysis with additional information expressed as inequality restrictions among the populations means. It also computes several estimations of the true error rate.

Depends mvtnorm, ibdreg, boot

Suggests survival

License GPL-2 | GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2021-05-03 06:10:02 UTC

R topics documented:

dawai-package	2
err.est	3
err.est.rlda	3
err.est.rqda	5
predict.rlda	7
predict.rqda	9
rlda	11
rqda	15
Vehicle2	17

Index	19
--------------	-----------

dawai-package

Discriminant analysis with additional information

Description

This package performs linear and quadratic discriminant analysis with additional information expressed as inequality constraints among the populations means and computes several estimations of the true error rate

Details

Package: dawai

Type: Package

Version: 1.2.5

Date: 2021-05-03

License: GPL-2 | GPL-3

For a complete list of functions with individual help pages, use `library(help = "dawai")`.

Author(s)

David Conde, Miguel A. Fernandez, Bonifacio Salvador

Maintainer: David Conde <dconde@eio.uva.es>

References

Conde, D., Fernandez, M. A., Rueda, C., and Salvador, B. (2012). Classification of samples into two or more ordered populations with application to a cancer trial. *Statistics in Medicine*, **31**, 3773-3786.

Conde, D., Fernandez, M. A., Salvador, B., and Rueda, C. (2015). dawai: An R Package for Discriminant Analysis with Additional Information. *Journal of Statistical Software*, **66(10)**, 1-19. URL <http://www.jstatsoft.org/v66/i10/>.

Conde, D., Salvador, B., Rueda, C., and Fernandez, M. A. (2013). Performance and estimation of the true error rate of classification rules built with additional information. An application to a cancer trial. *Statistical Applications in Genetics and Molecular Biology*, **12(5)**, 583-602.

Fernandez, M. A., Rueda, C., Salvador, B. (2006). Incorporating additional information to normal linear discriminant rules. *Journal of the American Statistical Association*, **101**, 569-577.

err.est	<i>Restricted Discriminant Analysis. True Error Rate estimation</i>
---------	---

Description

err.est is a generic function for true error rate estimations of classification rules built with additional information. The function invokes particular *methods* which depend on the `class` of the first argument.

Usage

```
err.est(x, ...)
```

Arguments

x	An object for which true error rate estimations are desired.
...	Additional arguments affecting the true error rate estimations produced.

Value

See the documentation of the particular methods for details of what is produced by each method.

Author(s)

David Conde

See Also

[err.est.rlda](#), [err.est.rqda](#)

err.est.rlda	<i>Restricted Linear Discriminant Analysis. True Error Rate estimation</i>
--------------	--

Description

Estimate the true error rate of linear classification rules built with additional information (in conjunction with `rlda`).

Usage

```
## S3 method for class 'rlda'  
err.est(x, nboot = 50, gamma = x$gamma, prior = x$prior, ...)
```

Arguments

x	An object of class 'rlda'.
nboot	Number of bootstrap samples used to estimate the true error rate of the classification rules.
gamma	A vector of values specifying which rules to take among the ones in x. If unspecified, all rules built with x\$gamma will be used. If present, gamma must be contained in x\$gamma.
prior	The prior probabilities of class membership. If unspecified, x\$prior probabilities are used. If present, the probabilities must be specified in the order of the factor levels.
...	Arguments based from or to other methods.

Details

This function is a method for the generic function `err.est()` for class 'rlda'.

Value

A list with components

call	The (matched) function call.
restrictions	Character vector with the restrictions on the means vector detailed.
prior	The prior probabilities of the classes used.
counts	The number of observations of the classes used.
N	The total number of observations used.
estimationError	Matrix with BT2, BT3, BT2CV and BT3CV true error rate estimates of the rules.

Note

To overcome singularity of the covariance matrices after bootstrapping, the number of observations in each class must be greater than the number of explanatory variables divided by 0.632.

Author(s)

David Conde

References

- Conde, D., Fernandez, M. A., Rueda, C., and Salvador, B. (2012). Classification of samples into two or more ordered populations with application to a cancer trial. *Statistics in Medicine*, **31**, 3773-3786.
- Conde, D., Fernandez, M. A., Salvador, B., and Rueda, C. (2015). dawai: An R Package for Discriminant Analysis with Additional Information. *Journal of Statistical Software*, **66(10)**, 1-19. URL <http://www.jstatsoft.org/v66/i10/>.

Conde, D., Salvador, B., Rueda, C. , and Fernandez, M. A. (2013). Performance and estimation of the true error rate of classification rules built with additional information. An application to a cancer trial. *Statistical Applications in Genetics and Molecular Biology*, **12(5)**, 583-602.

See Also

[err.est.rlda](#), [predict.rlda](#), [rqda](#), [predict.rqda](#), [err.est.rqda](#)

Examples

```
data(Vehicle2)
levels(Vehicle2$Class)
## "bus" "opel" "saab" "van"

data = Vehicle2[, c("Holl.Ra", "Sc.Var.maxis")]
grouping = Vehicle2$Class
levels(grouping) <- c(3, 1, 1, 2)
## now we can consider the following restrictions:
## mu11 >= mu21 >= mu31
##
## we can specify these restrictions by resexp = "s>1"

set.seed(-1007)
values <- runif(length(rownames(data)))
trainsubset <- values < 0.05
testsubset <- values >= 0.05
obj <- rlda(data, grouping, subset = trainsubset, resexp = "s>1")
pred <- predict(obj, data[testsubset,], grouping = grouping[testsubset],
               prior = c(1/3, 1/3,1/3))
pred$error.rate
err.est(obj, 25, prior = c(1/3, 1/3, 1/3))
```

err.est.rqda	<i>Restricted Quadratic Discriminant Analysis. True Error Rate Estimation</i>
--------------	---

Description

Estimate the true error rate of quadratic classification rules built with additional information (in conjunction with rqda).

Usage

```
## S3 method for class 'rqda'
err.est(x, nboot = 50, gamma = x$gamma, prior = x$prior, ...)
```

Arguments

x	An object of class 'rqda'.
nboot	Number of bootstrap samples used to estimate the true error rate of the classification rules.
gamma	A vector of values specifying which rules to take among the ones in x. If unspecified, all rules built with x\$gamma will be used. If present, gamma must be contained in x\$gamma.
prior	The prior probabilities of class membership. If unspecified, x\$prior probabilities are used. If present, the probabilities must be specified in the order of the factor levels.
...	Arguments based from or to other methods.

Details

This function is a method for the generic function `err.est()` for class 'rqda'.

Value

A list with components

call	The (matched) function call.
restrictions	Character vector with the restrictions on the means vector detailed.
prior	The prior probabilities of the classes used.
counts	The number of observations of the classes used.
N	The total number of observations used.
estimationError	Matrix with BT2, BT3, BT2CV and BT3CV true error rate estimates of the rules.

Note

To overcome singularity of the covariance matrices after bootstrapping, the number of observations in each class must be greater than the number of explanatory variables divided by 0.632.

Author(s)

David Conde

References

- Conde, D., Fernandez, M. A., Rueda, C., and Salvador, B. (2012). Classification of samples into two or more ordered populations with application to a cancer trial. *Statistics in Medicine*, **31**, 3773-3786.
- Conde, D., Fernandez, M. A., Salvador, B., and Rueda, C. (2015). dawai: An R Package for Discriminant Analysis with Additional Information. *Journal of Statistical Software*, **66(10)**, 1-19. URL <http://www.jstatsoft.org/v66/i10/>.

Conde, D., Salvador, B., Rueda, C. , and Fernandez, M. A. (2013). Performance and estimation of the true error rate of classification rules built with additional information. An application to a cancer trial. *Statistical Applications in Genetics and Molecular Biology*, **12(5)**, 583-602.

See Also

[err.est](#), [rqda](#), [predict.rqda](#), [rlda](#), [predict.rlda](#), [err.est.rlda](#)

Examples

```
data(Vehicle2)
levels(Vehicle2$Class)
## "bus" "opel" "saab" "van"

data = Vehicle2[, c("Kurt.Maxis", "Holl.Ra", "Sc.Var.maxis")]
grouping = Vehicle2$Class
levels(grouping) <- c(3, 1, 1, 2)
## now we can consider the following restrictions:
## mu11 >= mu21 >= mu31
## mu12 >= mu22 >= mu32
##
## we can specify these restrictions by resextext = "s>1,2"

set.seed(5561)
values <- runif(length(rownames(data)))
trainsubset <- values < 0.05
testsubset <- values >= 0.05
obj <- rqda(data, grouping, subset = trainsubset, resextext = "s>1,2")
pred <- predict(obj, data[testsubset,], grouping = grouping[testsubset],
               prior = c(1/3, 1/3,1/3))
pred$error.rate
err.est(obj, 25, prior = c(1/3, 1/3, 1/3))
```

predict.rlda

*Restricted Linear Discriminant Analysis. Multivariate Observations
Classification*

Description

Classify multivariate observations with linear classification rules built with additional information in conjunction with rlda.

Usage

```
## S3 method for class 'rlda'
predict(object, newdata, prior = object$prior,
        gamma = object$gamma, grouping = NULL, ...)
```

Arguments

object	An object of class 'rlda'.
newdata	A data frame of cases to be classified, containing the variables used on creating object. A vector will be interpreted as a row vector.
prior	The prior probabilities of class membership. If unspecified, object\$prior probabilities are used. If present, the probabilities must be specified in the order of the factor levels.
gamma	A vector of values specifying which rules to take among the ones in object. If unspecified, all rules built with object\$gamma will be used. If present, gamma must be contained in object\$gamma.
grouping	A numeric vector or factor with numeric levels specifying the class for each observation. If present, true error rate will be estimated from newdata.
...	Arguments based from or to other methods.

Details

This function is a method for the generic function predict() for class 'rlda'.

Value

A list with components

call	The (matched) function call.
class	Matrix with the classification for each rule (in columns).
prior	The prior probabilities of the classes used.
posterior	Array with the posterior probabilities of the classes for each rule.
error.rate	True error rate estimation (when grouping specified) for each rule, based on newdata.

Note

If there are missing values in newdata, corresponding observations will not be classified.

If there are missing values in grouping, corresponding observations will not be considered on calculating the true error rate.

Author(s)

David Conde

References

- Conde, D., Fernandez, M. A., Rueda, C., and Salvador, B. (2012). Classification of samples into two or more ordered populations with application to a cancer trial. *Statistics in Medicine*, **31**, 3773-3786.
- Conde, D., Fernandez, M. A., Salvador, B., and Rueda, C. (2015). dawai: An R Package for Discriminant Analysis with Additional Information. *Journal of Statistical Software*, **66(10)**, 1-19. URL <http://www.jstatsoft.org/v66/i10/>.

See Also

[rlda](#), [err.est.rlda](#), [rqda](#), [predict.rqda](#), [err.est.rqda](#)

Examples

```

data(Vehicle2)
levels(Vehicle2$Class)
## "bus" "opel" "saab" "van"

data <- Vehicle2
levels(data$Class) <- c(4, 2, 1, 3)
## classes ordered by increasing size
##
## according to variable definitions, we can
## consider the following restrictions on the means vectors:
## mu11, mu21 >= mu31 >= mu41
## mu12, mu22 >= mu32 >= mu42
##
## we have 6 restrictions, 3 predictors and 4 classes, so
## resmatrix must be a 6 x 12 matrix:

A <- matrix(0, ncol = 12, nrow = 6)
A[t(matrix(c(1, 1, 2, 2, 3, 4, 4, 5, 5, 7, 6, 8), nrow = 2))] <- -1
A[t(matrix(c(1, 7, 2, 8, 3, 7, 4, 8, 5, 10, 6, 11), nrow = 2))] <- 1

set.seed(983)
values <- runif(dim(data)[1])
trainsubset <- values < 0.2
testsubset <- values >= 0.2
obj <- rlda(Class ~ Kurt.Maxis + Holl.Ra + Sc.Var.maxis,
            data, subset = trainsubset, gamma = c(0, 0.5, 1),
            resmatrix = A)
pred <- predict(obj, newdata = data[testsubset,],
               grouping = data[testsubset, "Class"],
               prior = rep(1/4, 4))

pred$error.rate
## we can see that the test error rate of the restricted
## rules decrease with gamma:
##
##               gamma=0  gamma=0.5  gamma=1
## True error rate (%): 40.86957  39.71014  39.71014

```

predict.rqda

Restricted Quadratic Discriminant Analysis. Multivariate Observations Classification

Description

Classify multivariate observations with quadratic classification rules built with additional information in conjunction with rqda.

Usage

```
## S3 method for class 'rqda'
predict(object, newdata, prior = object$prior,
        gamma = object$gamma, grouping = NULL, ...)
```

Arguments

object	An object of class 'rqda'.
newdata	A data frame of cases to be classified, containing the variables used on creating object. A vector will be interpreted as a row vector.
prior	The prior probabilities of class membership. If unspecified, object\$prior probabilities are used. If present, the probabilities must be specified in the order of the factor levels.
gamma	A vector of values specifying which rules to take among the ones in object. If unspecified, all rules built with object\$gamma will be used. If present, gamma must be contained in object\$gamma.
grouping	A numeric vector or factor with numeric levels specifying the class for each observation. If present, true error rate will be estimated from newdata.
...	Arguments based from or to other methods.

Details

This function is a method for the generic function predict() for class 'rqda'.

Value

A list with components

call	The (matched) function call.
class	Matriarchx with the classification for each rule (in columns).
prior	The prior probabilities of the classes used.
posterior	Array with the posterior probabilities of the classes for each rule.
error.rate	True error rate estimation (when grouping specified) for each rule, based on newdata.

Note

If there are missing values in newdata, corresponding observations will not be classified.

If there are missing values in grouping, corresponding observations will not be considered on calculating the true error rate.

Author(s)

David Conde

References

Conde, D., Fernandez, M. A., Rueda, C., and Salvador, B. (2012). Classification of samples into two or more ordered populations with application to a cancer trial. *Statistics in Medicine*, **31**, 3773-3786.

Conde, D., Fernandez, M. A., Salvador, B., and Rueda, C. (2015). dawai: An R Package for Discriminant Analysis with Additional Information. *Journal of Statistical Software*, **66(10)**, 1-19. URL <http://www.jstatsoft.org/v66/i10/>.

See Also

[rqda](#), [err.est.rqda](#), [rlda](#), [predict.rlda](#), [err.est.rlda](#)

Examples

```
data(Vehicle2)
levels(Vehicle2$Class)
## "bus" "opel" "saab" "van"

data <- Vehicle2[, 1:4]
grouping = Vehicle2$Class
levels(grouping) <- c(4, 2, 1, 3)
## classes ordered by increasing size
##
## according to variable definitions, we can consider
## the following restrictions on the means vectors:
## mu11 >= mu21 >= mu31 >= mu41
## mu12 >= mu22 >= mu32 >= mu42
## mu13 >= mu23 >= mu33 >= mu43
##
## we can specify these restrictions by restruct = "s>1,2,3"

set.seed(7964)
values <- runif(dim(data)[1])
trainsubset <- values < 0.2
testsubset <- values >= 0.2
obj <- rqda(data, grouping, subset = trainsubset,
            gamma = (1:5)/5, restruct = "s>1,2,3")
pred <- predict(obj, newdata = data[testsubset,],
               grouping = grouping[testsubset])
pred$error.rate
## we can see that the test error rate of the restricted
## rules decrease with gamma:
##
##           gamma=0.2 gamma=0.4 gamma=0.6 gamma=0.8 gamma=1
## True error rate (%): 40.14815 39.85185 39.85185 39.11111 39.11111
```

Description

Build linear classification rules with additional information expressed as inequality restrictions among the populations means.

Usage

```
rlda(x, ...)

## S3 method for class 'matrix'
rlda(x, ...)

## S3 method for class 'data.frame'
rlda(x, grouping, ...)

## S3 method for class 'formula'
rlda(formula, data, ...)

## Default S3 method:
rlda(x, grouping, subset = NULL, resmatrix = NULL, restruct = NULL,
      gamma = c(0, 1), prior = NULL, ...)
```

Arguments

formula	A formula of the form $\text{groups} \sim x_1 + x_2 + \dots$. That is, the response is the grouping factor and the right hand side specifies the (non-factor) discriminators.
data	Data frame from which variables specified in formula are to be taken.
x	(Required if no formula is given as the principal argument.) A data frame or matrix containing the explanatory variables.
grouping	(Required if no formula is given as the principal argument.) A numeric vector or factor with numeric levels specifying the class for each observation.
subset	An index vector specifying the cases to be used in the training sample.
resmatrix	A matrix specifying the linear restrictions on the mean vectors: $\text{resmatrix} \%*\% \mu \leq 0$, where $\mu = c(\mu_1, \mu_2, \dots)$ and μ_i is the mean vector of class i . If unspecified, <code>restruct</code> will be required (and <code>resmatrix</code> established accordingly).
restruct	(Required if no <code>resmatrix</code> argument is given.) A character string from which <code>resmatrix</code> will be calculated. The first element must be either "s" (simple order) or "t" (tree order: $\mu_1 \geq \mu_2, \mu_1 \geq \mu_3 \dots$). The second element must be either "<" (increasing componentwise order) or ">" (decreasing componentwise order). The rest of the elements must be numbers from 1 to the number of explanatory variables, separated by commas, specifying among which variables the restrictions hold. For example, "s<1,3" will stand for $\mu_{11} \leq \mu_{21} \leq \mu_{31} \leq \dots, \mu_{13} \leq \mu_{23} \leq \mu_{33} \leq \dots$.
gamma	A vector of values in the unit interval that determine the classification rules with additional information (see references).
prior	The prior probabilities of class membership. If unspecified, the class proportions for the training set are used. If present, the probabilities must be specified in the order of the factor levels.

... Arguments passed to or from other methods.

Details

Specifying the prior will affect the classification and error unless over-ridden in `predict.rlda` and `err.est.rlda`, respectively.

Value

An object of class 'rlda' containing the following components:

<code>call</code>	The (matched) function call.
<code>trainset</code>	Matrix with the training set used (first columns) and the class for each observation (last column).
<code>restrictions</code>	Edited character string with the linear restrictions on the mean vectors detailed.
<code>resmatrix</code>	The matrix with the restrictions on the mean vectors used.
<code>prior</code>	Prior probabilities of class membership used.
<code>counts</code>	The number of observations of the classes used.
<code>N</code>	The total number of observations used.
<code>samplemeans</code>	Matrix with the sample means in rows.
<code>samplevariances</code>	Array with the sample covariance matrices of the classes.
<code>gamma</code>	Gamma values used.
<code>spooled</code>	Pooled covariance matrix.
<code>estimatedmeans</code>	Array with the estimated means for each classification rule.
<code>apparent</code>	Apparent error rate for each classification rule.

Note

This function may be called giving either a formula and data frame, or a data frame and grouping factor, or a matrix and grouping factor as the first two arguments. All other arguments are optional.

Classes must be identified, either in a column of data or in the grouping vector, by natural numbers varying from 1 to the number of classes. The number of classes must be greater than 1.

If there are missing values in either data, x or grouping, corresponding observations will be deleted.

To overcome singularity of the covariance matrices, the number of observations in each class must be greater or equal than the number of explanatory variables.

Author(s)

David Conde

References

Conde, D., Fernandez, M. A., Rueda, C., and Salvador, B. (2012). Classification of samples into two or more ordered populations with application to a cancer trial. *Statistics in Medicine*, **31**, 3773-3786.

Conde, D., Fernandez, M. A., Salvador, B., and Rueda, C. (2015). dawai: An R Package for Discriminant Analysis with Additional Information. *Journal of Statistical Software*, **66(10)**, 1-19. URL <http://www.jstatsoft.org/v66/i10/>.

Fernandez, M. A., Rueda, C., Salvador, B. (2006). Incorporating additional information to normal linear discriminant rules. *Journal of the American Statistical Association*, **101**, 569-577.

See Also

[predict.rlda](#), [err.est.rlda](#), [rqda](#), [predict.rqda](#), [err.est.rqda](#)

Examples

```
data(Vehicle2)
levels(Vehicle2$Class)
## "bus" "opel" "saab" "van"

data <- Vehicle2
levels(data$Class) <- c(4, 2, 1, 3)
## classes ordered by increasing size
##
## according to variable definitions, we can
## consider the following restrictions on the means vectors:
## mu11, mu21 >= mu31 >= mu41
## mu12, mu22 >= mu32 >= mu42
##
## we have 6 restrictions, 3 predictors and 4 classes, so
## resmatrix must be a 6 x 12 matrix:

A <- matrix(0, ncol = 12, nrow = 6)
A[t(matrix(c(1, 1, 2, 2, 3, 4, 4, 5, 5, 7, 6, 8), nrow = 2))] <- -1
A[t(matrix(c(1, 7, 2, 8, 3, 7, 4, 8, 5, 10, 6, 11), nrow = 2))] <- 1

set.seed(983)
values <- runif(dim(data)[1])
trainsubset <- values < 0.2
obj <- rlda(Class ~ Kurt.Maxis + Holl.Ra + Sc.Var.maxis,
            data, subset = trainsubset, gamma = c(0, 0.5, 1),
            resmatrix = A)

obj
## we can see that the apparent error rate of the restricted
## rules decrease with gamma:
## gamma=0 gamma=0.5 gamma=1
## 42.30769 41.66667 41.02564
```

Description

Build quadratic classification rules with additional information expressed as inequality restrictions among the populations means.

Usage

```
rqda(x, ...)

## S3 method for class 'matrix'
rqda(x, ...)

## S3 method for class 'data.frame'
rqda(x, grouping, ...)

## S3 method for class 'formula'
rqda(formula, data, ...)

## Default S3 method:
rqda(x, grouping, subset = NULL, resmatrix = NULL, restruct = NULL,
      gamma = c(0, 1), prior = NULL, ...)
```

Arguments

formula	A formula of the form $\text{groups} \sim x_1 + x_2 + \dots$. That is, the response is the grouping factor and the right hand side specifies the (non-factor) discriminators.
data	Data frame from which variables specified in formula are to be taken.
x	(Required if no formula is given as the principal argument.) A data frame or matrix containing the explanatory variables.
grouping	(Required if no formula is given as the principal argument.) A numeric vector or factor with numeric levels specifying the class for each observation.
subset	An index vector specifying the cases to be used in the training sample.
resmatrix	A matrix specifying the linear restrictions on the mean vectors: $\text{resmatrix} \%*\% \mu \leq 0$, where $\mu = c(\mu_1, \mu_2, \dots)$ and μ_i is the mean vector of class i . If unspecified, restruct will be required (and resmatrix established accordingly).
restruct	(Required if no resmatrix argument is given.) A character string from which resmatrix will be calculated. The first element must be either "s" (simple order) or "t" (tree order: $\mu_1 \geq \mu_2, \mu_1 \geq \mu_3 \dots$). The second element must be either "<" (increasing componentwise order) or ">" (decreasing componentwise order). The rest of the elements must be numbers from 1 to the number of explanatory variables, separated by commas, specifying among which variables the restrictions hold. For example, "s<1,3" will stand for $\mu_{11} \leq \mu_{21} \leq \mu_{31} \leq \dots, \mu_{13} \leq \mu_{23} \leq \mu_{33} \leq \dots$

gamma	A vector of values in the unit interval that determine the classification rules with additional information (see references).
prior	The prior probabilities of class membership. If unspecified, the class proportions for the training set are used. If present, the probabilities must be specified in the order of the factor levels.
...	Arguments passed to or from other methods.

Details

Specifying the prior will affect the classification and error unless over-ridden in `predict.rlda` and `err.est.rlda`, respectively.

Value

An object of class 'rqda' containing the following components:

call	The (matched) function call.
trainset	Matrix with the training set used (first columns) and the class for each observation (last column).
restrictions	Edited character string with the linear restrictions on the mean vectors detailed.
resmatrix	The matrix with the restrictions on the mean vectors used.
prior	Prior probabilities of class membership used.
counts	The number of observations of the classes used.
N	The total number of observations used.
samplemeans	Matrix with the sample means in rows.
samplevariances	Array with the sample covariance matrices of the classes.
gamma	Gamma values used.
estimatedmeans	Array with the estimated means for each classification rule.
apparent	Apparent error rate for each classification rule.

Note

This function may be called using either a formula and data frame, or a data frame and grouping factor, or a matrix and grouping factor as the first two arguments. All other arguments are optional.

Classes must be identified, either in a column of data or in the grouping vector, by natural numbers varying from 1 to the number of classes. The number of classes must be greater than 1.

If there are missing values in either data, x or grouping, corresponding observations will be deleted.

To overcome singularity of the covariance matrices, the number of observations in each class must be greater or equal than the number of explanatory variables.

Author(s)

David Conde

References

Conde, D., Fernandez, M. A., Rueda, C., and Salvador, B. (2012). Classification of samples into two or more ordered populations with application to a cancer trial. *Statistics in Medicine*, **31**, 3773-3786.

Conde, D., Fernandez, M. A., Salvador, B., and Rueda, C. (2015). dawai: An R Package for Discriminant Analysis with Additional Information. *Journal of Statistical Software*, **66(10)**, 1-19. URL <http://www.jstatsoft.org/v66/i10/>.

Fernandez, M. A., Rueda, C., Salvador, B. (2006). Incorporating additional information to normal linear discriminant rules. *Journal of the American Statistical Association*, **101**, 569-577.

See Also

[predict.rqda](#), [err.est.rqda](#), [rlda](#), [predict.rlda](#), [err.est.rlda](#)

Examples

```
data(Vehicle2)
levels(Vehicle2$Class)
## "bus" "opel" "saab" "van"

data <- Vehicle2[, 1:4]
grouping = Vehicle2$Class
levels(grouping) <- c(4, 2, 1, 3)
## classes ordered by increasing size
##
## according to variable definitions, we can consider
## the following restrictions on the means vectors:
## mu11 >= mu21 >= mu31 >= mu41
## mu12 >= mu22 >= mu32 >= mu42
## mu13 >= mu23 >= mu33 >= mu43
##
## we can specify these restrictions by restruct = "s>1,2,3"

set.seed(7964)
values <- runif(dim(data)[1])
trainsubset <- values < 0.2
obj <- rqda(data, grouping, subset = trainsubset,
            gamma = (1:5)/5, restruct = "s>1,2,3")
obj
## we can see that the apparent error rate of the restricted
## rules increase with gamma:
## gamma=0.2 gamma=0.4 gamma=0.6 gamma=0.8 gamma=1
## 30.40936 30.99415 30.99415 30.99415 31.57895
```

Description

The purpose is to classify a given silhouette as one of four types of vehicle, using a set of features extracted from the silhouette. The vehicle may be viewed from one of many different angles. The features were extracted from the silhouettes by the HIPS (Hierarchical Image Processing System) extension BINATTS, which extracts a combination of scale independent features utilising both classical moments based measures such as scaled variance, skewness and kurtosis about the major/minor axes and heuristic measures such as hollows, circularity, rectangularity and compactness.

Four "Corgie" model vehicles were used for the experiment: a double decker bus, Cheverolet van, Saab 9000 and an Opel Manta 400. This particular combination of vehicles was chosen with the expectation that the bus, van and either one of the cars would be readily distinguishable, but it would be more difficult to distinguish between the cars.

Usage

```
data(Vehicle2)
```

Format

A data frame with 846 observations on 4 variables, all numerical and one nominal defining the class of the objects.

[,1]	Skew.maxis	Skewness about minor axis
[,2]	Kurt.Maxis	Kurtosis about major axis
[,3]	Holl.Ra	Hollows ratio: (area of hollows)/(area of bounding polygon)
[,4]	Sc.Var.maxis	Scaled variance along minor axis: (2nd order moment about minor axis)/area
[,5]	Class	Type

Source

- Creator: Drs.Pete Mowforth and Barry Shepherd, Turing Institute, Glasgow, Scotland.

These data have been taken from the UCI Repository Of Machine Learning Databases at

- <http://archive.ics.uci.edu/ml/index.php>

and were converted to R format by Evgenia Dimitriadou.

References

Turing Institute Research Memorandum TIRM-87-018 "Vehicle Recognition Using Rule Based Methods" by Siebert, JP (March 1987).

Newman, D.J. & Hettich, S. & Blake, C.L. & Merz, C.J. (1998). UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science.

Examples

```
data(Vehicle2)
summary(Vehicle2)
```

Index

* **classif**

dawai-package, 2
err.est, 3
err.est.rlda, 3
err.est.rqda, 5
predict.rlda, 7
predict.rqda, 9
rlda, 11
rqda, 15

* **datasets**

Vehicle2, 17

* **multivariate**

dawai-package, 2
err.est, 3
err.est.rlda, 3
err.est.rqda, 5
predict.rlda, 7
predict.rqda, 9
rlda, 11
rqda, 15

* **package**

dawai-package, 2

class, 3

dawai (dawai-package), 2

dawai-package, 2

err.est, 3, 5, 7

err.est.rlda, 3, 3, 7, 9, 11, 14, 17

err.est.rqda, 3, 5, 5, 9, 11, 14, 17

predict.rlda, 5, 7, 7, 11, 14, 17

predict.rqda, 5, 7, 9, 9, 14, 17

rlda, 5, 7, 9, 11, 11, 17

rqda, 5, 7, 9, 11, 14, 15

Vehicle2, 17