Package 'ROCt632'

October 18, 2011

Type Package

Title Estimation of prognostic capacity of microarray data.

Version 0.1

Date 2011-10-17

Author Y. Foucher <Yohann.Foucher@univ-nantes.fr>

Maintainer Y. Foucher <Yohann.Foucher@univ-nantes.fr>

Description This package computes different time-dependent ROC curves using the coss-validation, the 0.632 and the 0.632+ estimators.

License GPL (>=2)

LazyLoad yes

Depends splines, survival, penalized, survivalROC

Imports splines, survival, penalized, survivalROC

URL www.r-project.org, www.divat.fr

R topics documented:

`ROCt632-package` *Estimation of prognostic capacity of microarray data.*

Description

This package can be used for different bootstrap corrections of overfitting in order to estimate the time-dependent ROC curves.

Details

Package:	ROCt632
Type:	Package
Version:	0.1
Date:	2011-10-17
License:	GPL (>=2)
LazyLoad:	yes

Compute different time-dependent ROC curve using the coss-validation, the 0.632 and the 0.632+ estimators.

ROCt	This function performs the different estimations of time-dependent ROC curves.
auc	This function computes the area under ROC curve using the trapezoidal rule.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

References

R. Danger and Y. Foucher. Time dependent ROC curves for the estimation of true prognostic capacity of microarray data. 2011. Submited.

See Also

URL: http://www.divat.fr

`auc` *Area under ROC curve from sensitivities and specificities*

Description

This function computes the area under ROC curve using the trapezoidal rule.

Usage

```
auc(sens, spec)
```

Arguments

`sens`	A numeric vector with the sensitivities
`spec`	A numeric vector with the sensitivities

Details

This function computes the area under ROC curve using the trapezoidal rule. The value of the area is directly returned.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

Examples

```
se.temp <- c(0, 0.5, 0.5, 1)
sp.temp <- c(1, 0.5, 0.5, 0)
auc(se.temp, sp.temp)
```

DLBCLgenes *The data concerning the gene expressions of the DLBCL patients*

Description

A matrix with the 7399 gene expressions of the 240 DLBCL patients.

Usage

```
data(DLBCLgenes)
```

Format

A matrix with 240 observations (raws) with the 7399 genes (columns).

Details

Rosenwald et al. (2002) have evaluated tumor samples from 240 DLBCL patients treated with anthracycline based therapy. Te missing data were replaced according to the mean expression level of the nearest 8 genes.

Source

the data is published at http://llmpp.nih.gov/lymphoma/data.shtml.

References

Rosenwald et al. The use of molecular proffiling to predict survival after chemotherapy for diffuse large-b-cell lymphoma. New England Journal of Medicine, 346(25):1937-47, 2002.

Alizadeh et al. Distinct types of diffuse large b-cell lymphoma identiffied by gene expression proffiling. Nature, 403(6769):503-11, 2000.

Examples

```
data(DLBCLpatients)
data(DLBCLgenes)

### Patients survival according to the two subgroups defined by using
### the median of the first gene
plot(survfit(Surv(t, f) ~ I(DLBCLgenes[,1] > median(DLBCLgenes[,1])),
 data = DLBCLpatients), xlab="Survival time (in years)",
 ylab="Patient survival", mark.time=FALSE)
```

DLBCLpatients *The data concerning the clinical information of the DLBCL patients*

Description

A data frame with 240 DLBCL patients. The time-to-event is the time to patient death. This time can be right-censored.

Usage

```
data(DLBCLpatients)
```

Format

A data frame with 240 observations (raws) with the 8 following variables (columns).

`ident` This numeric vector represents the key for patient identication

`t` This numeric vector represents the follow up times (until death or censoring)

`f` This numeric vector represents the failure indicators at the follow-up end (1=death, 0=alive)

Details

Rosenwald et al. (2002) evaluated tumor samples from 240 DLBCL patients treated with anthracycline based therapy. They confirmed the existence of the two DLBCL subgroups described previously, GCB-like and ABC-like. The overall survival was significantly different among the subgroups, with 5-year survivals of 60% for the GCB-like and 35% for ABC-like subgroups. An additional third subtype was described with a 5-year survival of 39%.

Source

The data is published at http://llmpp.nih.gov/lymphoma/data.shtml.

References

Rosenwald et al. The use of molecular proffiling to predict survival after chemotherapy for diffuse large-b-cell lymphoma. New England Journal of Medicine, 346(25):1937-47, 2002.

Alizadeh et al. Distinct types of diffuse large b-cell lymphoma identiffied by gene expression proffiling. Nature, 403(6769):503-11, 2000.

Examples

```
data(DLBCLpatients)

### Kaplan and Meier estimation of the patients survival
plot(survfit(Surv(t, f) ~ 1, data = DLBCLpatients),
 xlab="Survival time (in years)", ylab="Patient survival",
 mark.time=FALSE)
```

ROCt632 *Estimation of true prognostic capacity of microarray data using the coss-validation, the 0.632 and the 0.632+ estimators of time-dependent ROC curves*

Description

This function performs estimations of time-dependent ROC curves according to different bootstrap-based approaches.

Usage

```
ROCt632(times, failures, features, N.boot,
 precision, prop, pro.time)
```

Arguments

times	A numeric vector with the follow up times.
failures	A numeric vector with the event indicators (0=right censored, 1=event).
features	A matrix with the observed features. The number of raw is the number of individuals (length of the arguments times and failures).
N.boot	Number of bootstrap iterations.
precision	The quantiles of the predictor used for computing each point of the time dependent ROC curve.
prop	This is the proportion of the nearest neighbors used in the Akritas estimator. The estimation will be based on 2*lambda (1 lambda on the left and 1 lambda on the right) of the total sample size.
pro.time	The value(s) of prognostic time represent(s) the maximum delay(s) for which the capacity of the variable is evaluated. The same unit than the one used in the argument times.

Details

This function computes time-dependent ROC curve with right-censoring data based on the 0.632+ estimator. The Akritas approach (nearest neighbor's estimation) is used for ensuring monotone and increasing ROC curve. The theory was defined by Heagerty, Lumley and Pepe (Biometrics, 2000).

Value

The function returns a data frame. The raw(s) represent(s) the value(s) of the prognostic time. train is the mean of the areas obtained by using the individuals included in the bootstrap samples (training). valid is the mean of the areas obtained by using the individuals not included in the bootstrap samples (cross-validation). s632 is the mean of the areas obtained by using the simple 0.632 estimator. p632 is the mean of the areas obtained by using the 0.632+ estimator.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

References

R. Danger and Y. Foucher. Time dependent ROC curves for the estimation of true prognostic capacity of microarray data. 2011. Submited.

Examples

```
# import and attach the data example

data(DLBCLpatients)
data(DLBCLgenes)

ROCt632(times=DLBCLpatients$t, failures=DLBCLpatients$f,
 features=DLBCLgenes, N.boot=2,
 precision=seq(0.01, 0.99, by=0.02),
 prop=0.02, pro.time=5)
```

Index