Package 'OptimalCutpoints'

October 12, 2022

Type Package

Title Computing Optimal Cutpoints in Diagnostic Tests

Version 1.1-5

Date 2021-10-06

Author Monica Lopez-Raton, Maria Xose Rodriguez-Alvarez

Maintainer Monica Lopez Raton <monica.lopez.raton@gmail.com>

Description

Computes optimal cutpoints for diagnostic tests or continuous markers. Various approaches for se-
lecting optimal cutoffs have been implemented, including methods based on cost-benefit analy-
sis and diagnostic test accuracy measures (Sensitivity/Specificity, Predictive Values and Diag-
nostic Likelihood Ratios). Numerical and graphical output for all methods is easily obtained.

Imports stats, graphics

License GPL

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2021-10-07 10:30:05 UTC

R topics documented:

```
OptimalCutpoints-package
```
                    Computing Optimal Cutpoints in Diagnostic Tests

---

Description

Continuous biomarkers or diagnostic tests are often used to discriminate between diseased and healthy populations. In clinical practice, it is necessary to select a cutpoint or discrimination value c which defines the positive and negative test results. Several methods for selecting optimal cutpoints in diagnostic tests have been proposed in the literature depending on the underlying reason for this choice. This package allows the user to compute the optimal cutpoint for a diagnostic test or continuous marker. Various approaches for selecting optimal cutoffs have been implemented, including methods based on cost-benefit analysis and diagnostic test accuracy measures (Sensitivity/Specificity, Predictive Values and Diagnostic Likelihood Ratios) or prevalence. Numerical and graphical output for all methods is easily obtained.

Details

Package:	OptimalCutpoints
Type:	Package
Version:	1.1-5
Date:	2021-10-06
License:	GPL

In the OptimalCutpoints package all these methods have been incorporated in a way designed to be clear and user-friendly for the end-user. For all methods, the optimal cutoff value obtained is always one of the values of the diagnostic marker, and the Receiver Operating Characteristic (ROC) and Predictive ROC (PROC) curves and accuracy measures are empirically estimated. The program only requires a data frame, which can be built from a data-entry file or from something else (a database, direct entry, predictions from another function,...), which must, at minimum, contain the following variables: diagnostic marker; disease status (diseased/healthy); and whether adjustment is to be made for any (categorical) covariate of interest, a variable that indicates the levels of this covariate. A standard-type data input structure is used, with each row of the database indicating a patient/case and each column referring to a variable.

The most important functions in the package are the `optimal.cutpoints()`, `control.cutpoints()`, `summary.optimal.cutpoints()` and `plot.optimal.cutpoints()` functions. The `optimal.cutpoints()` function computes the optimal cutpoint(s) with its accuracy measures, according to the criterion selected. More than one criterion can be chosen for selecting the optimal cutpoint. The `control.cutpoints()` function is used to set several parameters that are specific of each method, such as the cost values or the minimum values for diagnostic accuracy measures. The `summary.optimal.cutpoints()` and `plot.optimal.cutpoints()` functions produce numerical and graphical output, respectively. Numerical output includes information relating to: the optimal cutpoint; the method used for selecting the optimal value, together with the number of optimal cutpoints (in some cases there may be more than one value); and the optimal cutoff(s) and its/their accuracy-measure estimates. Graphical

output includes the plots of the ROC and PROC curves, indicating the optimal cutpoint on these plots.

Author(s)

Monica Lopez-Raton and Maria Xose Rodriguez-Alvarez

Maintainer: Monica Lopez-Raton <monica.lopez.raton@gmail.com>

References

Lopez-Raton, M., Rodriguez-Alvarez, M.X, Cadarso-Suarez, C. and Gude-Sampedro, F. (2014). OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *Journal of Statistical Software* **61**(8), 1–36. doi: 10.18637/jss.v061.i08.

---

control.cutpoints   *Controlling the optimal-cutpoint selection process*

Description

Used to set various parameters controlling the optimal-cutpoint selection process

Usage

```
control.cutpoints(costs.ratio = 1, CFP = 1, CFN = 1,
  valueSp = 0.85, valueSe = 0.85,
  maxSp = TRUE,
  generalized.Youden = FALSE,
  costs.benefits.Youden = FALSE,
  costs.benefits.Efficiency = FALSE,
  weighted.Kappa = FALSE,
  standard.deviation.accuracy = FALSE,
  valueNPV = 0.85, valuePPV = 0.85,
  maxNPV = TRUE,
  valueDLR.Positive = 2,
  valueDLR.Negative = 0.5,
  adjusted.pvalue = c("PADJMS","PALT5","PALT10"),
  ci.SeSp = c("Exact","Quadratic","Wald","AgrestiCoull","RubinSchenker"),
  ci.PV = c("Exact","Quadratic","Wald","AgrestiCoull","RubinSchenker",
  "Transformed","NotTransformed","GartNam"),
  ci.DLR = c("Transformed","NotTransformed","GartNam"))
```

Arguments

costs.ratio   a numerical value meaningful only in the "CB" method. It specifies the costs ratio:
$$CR = \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}$$
where C_{FP}, C_{TN}, C_{FN} and C_{TP} are the costs of False Positive, True Negative, False Negative and True Positive decisions, respectively. The default value is 1.

CFP   a numerical value meaningful only in the "MCT", "Youden" and "MaxKappa"
                           methods. It specifies the cost of a False Positive decision. The default value is
                           1.

CFN   a numerical value meaningful only in the "MCT", "Youden" and "MaxKappa"
                           methods. It specifies the cost of a False Negative decision. The default value is
                           1.

valueSp   a numerical value meaningful only in the "MinValueSp", "ValueSp" and "Min-
                           ValueSpSe" methods. It specifies the (minimum or specific) value set for Speci-
                           ficity. The default value is 0.85.

valueSe   a numerical value meaningful only in the "MinValueSe", "ValueSe" and "Min-
                           ValueSpSe" methods. It specifies the (minimum or specific) value set for Sensi-
                           tivity. The default value is 0.85.

maxSp   a logical value meaningful only in the "MinValueSpSe" method, in a case where
                           there is more than one cutpoint fulfilling the conditions. If TRUE, those of the
                           cutpoints which yield maximum Specificity are computed. Otherwise the cutoff
                           that yields maximum Sensitivity is computed. The default is TRUE.

generalized.Youden
                           a logical value meaningful only in the "Youden" method. If TRUE, the Gener-
                           alized Youden Index is computed. The default is FALSE.

costs.benefits.Youden
                           a logical value meaningful only in the "Youden" method. If TRUE, the optimal
                           cutpoint based on cost-benefit methodology is computed. The default is FALSE.

costs.benefits.Efficiency
                           a logical value meaningful only in the "MaxEfficiency" method. If TRUE, the
                           optimal cutpoint based on cost-benefit methodology is computed. The default is
                           FALSE.

weighted.Kappa   a logical value meaningful only in the "MaxKappa" method. If TRUE, the
                           Weighted Kappa Index is computed. The default is FALSE.

standard.deviation.accuracy
                           a logical value meaningful only in the "MaxEfficiency" method. If TRUE, stan-
                           dard deviation associated with accuracy (or efficiency) at the optimal cutpoint is
                           computed. The default is FALSE.

valueNPV   a numerical value meaningful only in the "MinValueNPV", "ValueNPV" and
                           "MinValueNPVPPV" methods. It specifies the minimum value set for Negative
                           Predictive Value. The default value is 0.85.

valuePPV   a numerical value meaningful only in the "MinValuePPV", "ValuePPV" and
                           "MinValueNPVPPV" methods. It specifies the minimum value set for Positive
                           Predictive Value. The default value is 0.85.

maxNPV   a logical value meaningful only in the "MinValueNPVPPV" method, in a case
                           where there is more than one cutpoint fulfilling the conditions. If TRUE, those
                           of the cutpoints which yield the maximum Negative Predictive Value are com-
                           puted. Otherwise the cutoff that yields the maximum Positive Predictive Value
                           is computed. The default is TRUE.

valueDLR.Positive
                           a numerical value meaningful only in the "ValueDLR.Positive" method. It speci-
                           fies the value set for the Positive Diagnostic Likelihood Ratio. The default value
                           is 2.

valueDLR.Negative
> a numerical value meaningful only in the "ValueDLR.Negative" method. It specifies the value set for the Negative Diagnostic Likelihood Ratio. The default value is 0.5.

adjusted.pvalue
> a character string meaningful only in the "MinPvalue" method. It specifies the method for adjusting the p-value, i.e., "PADJMS" for the Miller and Siegmund method, and "PALT5", "PALT10" for the Altman method (see details). The default is "PADJMS".

ci.SeSp
> a character string meaningful only when the argument ci.fit of the optimal.cutpoints function is TRUE. It indicates how the confidence interval for Sensitivity and Specificity measures is estimated. Options are "Exact" (Clopper and Pearson 1934), "Quadratic" (Fleiss 1981), "Wald" (Wald and Walfowitz 1939), "Agresti-Coull" (Agresti and Coull 1998) and "RubinSchenker" (Rubin and Schenker 1987) (see details). The default is "Exact".

ci.PV
> a character string meaningful only when the argument ci.fit of the optimal.cutpoints function is TRUE. It indicates how the confidence interval for Predictive Values is estimated. Options are "Exact" (Clopper and Pearson 1934), "Quadratic" (Fleiss 1981), "Wald" (Wald and Walfowitz 1939), "AgrestiCoull" (Agresti and Coull 1998), "RubinSchenker" (Rubin and Schenker 1987), "Transformed" (Simel et al. 1991), "NotTransformed" (Koopman 1984) and "GartNam" (Gart and Nam 1988) (see details). The default is "Exact".

ci.DLR
> a character string meaningful only when the argument ci.fit of the function optimal.cutpoints is TRUE. It indicates how the confidence interval for Diagnostic Likelihood Ratios is estimated. Options are "Transformed" (Simel et al. 1991), "NotTransformed" (Koopman 1984) and "GartNam" (Gart and Nam 1988)(see details). The default is "Transformed".

Details

The value yielded by this function is used as the control argument of the optimal.cutpoints() function.

Several methods for correcting the increase in type-I error associated with the "MinPvalue" criterion have been proposed. In this package, two methods for adjusting the p-value have been implemented, i.e., the Miller and Siegmund (1982) and Altman (1994) methods. The first of these ("PADJMS" option) uses the minimum observed p-value ($pmin$) and the proportion (ϵ) of sample data which is below the lowest (ϵ_{low}) (or above the highest, ϵ_{high}) cutpoint considered:

$$p_{acor} = \phi(z)(z - \frac{1}{z})log\left(\frac{\epsilon_{high}(1 - \epsilon_{low})}{(1 - \epsilon_{high})\epsilon_{low}}\right) + 4\frac{\phi(z)}{z}$$

where z is the $(1 - pmin/2)$ quantile of the standard normal distribution and ϕ its corresponding density function. The second method is a simplification of the above formula, which considers specific values for ϵ: with $\epsilon = \epsilon_{low} = \epsilon_{high} = 5\%$ ("PALT5" option): $p_{alt5} = -3.13 p_{min}(1 + 1.65 ln(p_{min}))$ with $\epsilon = \epsilon_{low} = \epsilon_{high} = 10\%$ ("PALT10" option): $p_{alt10} = -1.63 p_{min}(1 + 2.35 ln(p_{min}))$. These approaches work well for low $pmin$ values (0.0001<$pmin$<0.1) and are easy to apply.

For inference performed on Sensitivity and Specificity measures (which are proportions), some of the most common confidence intervals have been considered. If $pr = x/n$ is the proportion to be estimated and 1-α is the confidence level, the options are as follows:

"Exact": The exact confidence interval of Clopper and Pearson (1934) based on the exact distribution of a proportion:

$$\left[\frac{x}{(n-x+1)F_{\alpha/2,2(n-x+1),2x}+x}, \frac{(x+1)F_{\alpha/2,2(x+1),2(n-x)}}{(n-x)+(x+1)F_{\alpha/2,2(x+1),2(n-x)}} \right]$$

where $F_{\alpha/2,a,b}$ is the (1-α/2) quantile of a Fisher-Snedecor distribution with a and b degrees of freedom. Note that the "exact" method cannot be applied when x or n-x is equal to zero, since the quantile of the Fisher-Snedecor distribution is not defined for zero degrees of freedom. In that cases, the program returns a NaN for the limit of the confidence interval that could not be computed.

"Quadratic": Fleiss' quadratic confidence interval (Fleiss 1981). It is based on the asymptotic normality of the estimator of a proportion but adding a continuity correction. This approach is valid in a situation where x and $n-x$ are greater than 5:

$$\frac{1}{n+z_{1-\alpha/2}^2} \left[(x \mp 0.5) + \frac{z_{1-\alpha/2}^2}{2} \mp z_{1-\alpha/2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{(x \mp 0.5)(n-x \mp 0.5)}{n}} \right]$$

where $z_{1-\alpha/2}$ is the (1-α/2) quantile of the standard normal distribution.

"Wald": Wald's confidence interval (Wald and Wolfowitz 1939) with a continuity correction. It is based on maximum-likelihood estimation of a proportion, and adds a continuity correction. This approach is valid where x and $n-x$ are greater than 20:

$$\hat{pr} \mp z_{1-\alpha/2} \sqrt{\frac{\hat{pr}(1-\hat{pr})}{n}} + \frac{1}{2n}$$

"AgrestiCoull": The confidence interval proposed by Agresti and Coull (1998). It is a score confidence interval that does not use the standard calculation for the binomial proportion:

$$\frac{\hat{pr} + \frac{z_{1-\alpha/2}^2}{2n} \mp z_{1-\alpha/2} \sqrt{\frac{\hat{pr}(1-\hat{pr}) + \frac{z_{1-\alpha/2}^2}{4n}}{n}}}{1 + \frac{z_{1-\alpha/2}^2}{n}}$$

"RubinSchenker": Rubin and Schenker's logit confidence interval (1987). It uses logit transformation and Bayesian arguments with an a priori Jeffreys distribution.

$$logit \left[logit\left(\frac{x+0.5}{n+1}\right) \mp \frac{z_{1-\alpha/2}}{\sqrt{(n+1)\left(\frac{x+0.5}{n+1}\right)\left(1-\frac{x+0.5}{n+1}\right)}} \right]$$

where the $logit$ function is $logit(q) = log\left(\frac{q}{1-q}\right)$.

Since Diagnostic Likelihood Ratios represent a ratio between two probabilities, obtaining a confidence interval for them is less direct than it is for Sensitivity and Specificity. Let $pr_1 = x_1/n_1$ be the proportion in the numerator and $pr_2 = x_2/n_2$, the proportion in the denominator. Based on the logarithmic transformation of the Likelihood Ratio ("Transformed" option), the 100(1-α)% confidence interval is (Simel et al., 1991):

$$exp \left[ln\left(\frac{\widehat{pr}_1}{\widehat{pr}_2}\right) \mp z_{1-\alpha/2} \sqrt{\frac{1-\widehat{pr}_1}{n_1\widehat{pr}_1} + \frac{1-\widehat{pr}_2}{n_2\widehat{pr}_2}} \right]$$

These confidence intervals tend to perform better than do untransformed confidence intervals (Koopman 1984) (″NotTransformed″ option) because the distribution of the Likelihood Ratios is asymmetric (Simel et al., 1991; Roldan Nofuentes and Luna del Castillo, 2007):

$$\frac{\widehat{pr}_1}{\widehat{pr}_2} \mp \sqrt{\frac{\widehat{pr}_1(1 - \widehat{pr}_1)}{n_1 \widehat{pr}_2^2} + \frac{\widehat{pr}_1^2 \widehat{pr}_2(1 - \widehat{pr}_2)}{n_2 \widehat{pr}_2^4}}$$

Another confidence interval (″GartNam″ option) is based on the calculation of the interval for the ratio between two independent proportions (Gart and Nam, 1988). The following quadratic equation must be solved:

$$\frac{\left(\widehat{pr}_1 - \frac{pr_1}{pr_2}\widehat{pr}_2\right)^2}{\frac{\widehat{pr}_1(1 - \widehat{pr}_1)}{n_1} + \frac{\left(\frac{pr_1}{pr_2}\right)^2 \widehat{pr}_2(1 - \widehat{pr}_2)}{n_2}} = z_{1-\alpha/2}^2$$

Inference of the Predictive Values depends on the type of study, i.e., whether cross-sectional(prevalence can be estimated on the basis of the sample) or case-control. In the former case, the approaches for computing the confidence intervals of the Predictive Values are exactly the same as for the Sensitivity and Specificity measures. However, in a case control study, where prevalence is not estimated from the sample, the confidence intervals are based on the intervals of the Likelihood Ratios. Hence, once a prevalence estimator $\hat{p}$ is computed and substituting each limit of these intervals into the expressions

$$\left(1 + \frac{1 - \hat{p}}{\hat{p}\widehat{DLR}^+}\right)^{-1}$$

and

$$\left(1 + \frac{\hat{p}}{1 - \hat{p}}\widehat{DLR}^-\right)^{-1}$$

confidence intervals for the Positive and Negative Predictive Values are obtained, where $DLR+$ and $DLR-$ are the Positive and Negative Diagnostic Likelihood Ratios, respectively.

Value

A list with components for each of the possible arguments.

Author(s)

Monica Lopez-Raton and Maria Xose Rodriguez-Alvarez

References

Agresti, A. and Coull, B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician* **52**, 119–126.

Altman, D.G., Lausen, B., Sauerbrei, W. and Schumacher, M. (1994). Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute* **86**(11), 829–835.

Clopper, C. and Pearson, E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404–413.

Fleiss, J.L. (1981). Statistical methods for rates and proportions. John Wiley & Sons, New York.

Gart, J.J. and Nam, J. (1998). Aproximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness. *Biometrics* **44**, 323–338.

Koopman PAR (1984). Confidence limits for the ratio of two binomial proportions. *Biometrics* **40**, 513–517.

Miller, R. and Siegmund, D. (1982). Maximally selected chi square statistics. *Biometrics* **38**, 1011–1016.

Roldan Nofuentes, J.A. and Luna del Castillo, J.D. (2007). Comparing of the likelihood ratios of two binary diagnostic tests in paired designs. *Statistics in Medicine* **26**, 4179–4201.

Rubin, D.B. and Schenker, N. (1987). Logit-based interval estimation for binomial data using the Jeffreys prior. *Sociological Methodology* **17**, 131–144.

Simel, D.L., Samsa, G.P. and Matchar, D.B. (1991). Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *Journal of Clinical Epidemiology* **44**(8), 763–770.

Wald A, Wolfowitz J (1939). Confidence limits for continuous distribution functions. *The Annals of Mathematical Statistics* **10** 105–118.

See Also

optimal.cutpoints

Examples

```
library(OptimalCutpoints)
data(elas)

###########################################################
# Youden Index Method ("Youden"): Covariate gender
###########################################################
optimal.cutpoint.Youden<-optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov =
"gender", control = control.cutpoints(), ci.fit = TRUE, conf.level = 0.95, trace = FALSE)

summary(optimal.cutpoint.Youden)

# Change the method for computing the confidence interval
# of Sensitivity and Specificity measures
optimal.cutpoint.Youden<-optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(ci.SeSp = "AgrestiCoull"), ci.fit = TRUE, conf.level = 0.95,
trace = FALSE)

summary(optimal.cutpoint.Youden)

# Compute the Generalized Youden Index
optimal.cutpoint.Youden<-optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(generalized.Youden = TRUE), ci.fit = TRUE, conf.level = 0.95,
trace = FALSE)
```

```
summary(optimal.cutpoint.Youden)
```

---

elas	*Leukocyte Elastase Data*

Description

The `elas` data set was obtained from the Cardiology Department at the Galicia General Hospital (Santiago de Compostela, Spain). This study was conducted to assess the clinical usefulness of leukocyte elastase determination in the diagnosis of coronary artery disease (CAD).

Usage

```
data(elas)
```

Format

A data frame with 141 observations on the following 3 variables.

`elas` leukocyte elastase. Numerical vector

`status` true disease status (presence/absence of coronary artery disease). Numerical vector (0=absence, 1=presence)

`gender` patient's gender. Factor with `Male` and `Female` levels

Source

Amaro, A., Gude, F., Gonzalez-Juanatey, R., Iglesias, C., Fernandez-Vazquez, F., Garcia-Acuna, J. and Gil, M. (1995). Plasma leukocyte elastase concentration in angiographically diagnosed coronary artery disease. *European Heart Journal* **16**, 615–622.

References

Amaro, A., Gude, F., Gonzalez-Juanatey, R., Iglesias, C., Fernandez-Vazquez, F., Garcia-Acuna, J. and Gil, M. (1995). Plasma leukocyte elastase concentration in angiographically diagnosed coronary artery disease. *European Heart Journal* **16**, 615–622.

Examples

```
data(elas)
summary(elas)
```

| optimal.cutpoints | *Computing Optimal Cutpoints in diagnostic tests* |

Description

optimal.cutpoints calculates optimal cutpoints in diagnostic tests. Several methods or criteria for selecting optimal cutoffs have been implemented, including methods based on cost-benefit analysis and diagnostic test accuracy measures (Sensitivity/Specificity, Predictive Values and Diagnostic Likelihood Ratios) or prevalence (Lopez-Raton et al. 2014).

Usage

```
optimal.cutpoints(X, ...)
## Default S3 method:
optimal.cutpoints(X, status, tag.healthy, methods, data, direction = c("<", ">"),
categorical.cov = NULL, pop.prev = NULL, control = control.cutpoints(),
ci.fit = FALSE, conf.level = 0.95, trace = FALSE, ...)
## S3 method for class 'formula'
optimal.cutpoints(X, tag.healthy, methods, data, direction = c("<", ">"),
categorical.cov = NULL, pop.prev = NULL, control = control.cutpoints(),
ci.fit = FALSE, conf.level = 0.95, trace = FALSE, ...)
```

Arguments

X	either a character string with the name of the diagnostic test variable (then method 'optimal.cutpoints.default' is called), or a formula (then method 'optimal.cutpoints.formula' is called). When 'X' is a formula, it must be an object of class "formula". Right side of ~ must contain the name of the variable that distinguishes healthy from diseased individuals, and left side of ~ must contain the name of the diagnostic test variable.
status	a character string with the name of the variable that distinguishes healthy from diseased individuals. Only applies for the method 'optimal.cutpoints.default').
tag.healthy	the value codifying healthy individuals in the status variable .
methods	a character vector selecting the method(s) to be used: "CB" (cost-benefit method); "MCT" (minimizes Misclassification Cost Term); "MinValueSp" (a minimum value set for Specificity); "MinValueSe" (a minimum value set for Sensitivity); "ValueSe" (a value set for Sensitivity); "MinValueSpSe" (a minimum value set for Specificity and Sensitivity); "MaxSp" (maximizes Specificity); "MaxSe" (maximizes Sensitivity);"MaxSpSe" (maximizes Sensitivity and Specificity simultaneously); "MaxProdSpSe" (maximizes the product of Sensitivity and Specificity or Accuracy Area); "ROC01" (minimizes distance between ROC plot and point (0,1)); "SpEqualSe" (Sensitivity = Specificity); "Youden" (Youden Index); "MaxEfficiency" (maximizes Efficiency or Accuracy, similar to minimize Error Rate); "Minimax" (minimizes the most frequent error); "MaxDOR" (maximizes Diagnostic Odds Ratio); "MaxKappa" (maximizes Kappa Index);

"MinValueNPV" (a minimum value set for Negative Predictive Value); "Min-ValuePPV" (a minimum value set for Positive Predictive Value); "ValueNPV" (a value set for Negative Predictive Value);"ValuePPV" (a value set for Positive Predictive Value);"MinValueNPVPPV" (a minimum value set for Predictive Values); "PROC01" (minimizes distance between PROC plot and point (0,1)); "NPVEqualPPV" (Negative Predictive Value = Positive Predictive Value); "MaxN-PVPPV" (maximizes Positive Predictive Value and Negative Predictive Value simultaneously); "MaxSumNPVPPV" (maximizes the sum of the Predictive Values); "MaxProdNPVPPV" (maximizes the product of Predictive Values); "Val-ueDLR.Negative" (a value set for Negative Diagnostic Likelihood Ratio); "Val-ueDLR.Positive" (a value set for Positive Diagnostic Likelihood Ratio); "Min-Pvalue" (minimizes p-value associated with the statistical Chi-squared test which measures the association between the marker and the binary result obtained on using the cutpoint); "ObservedPrev" (The closest value to observed prevalence); "MeanPrev" (The closest value to the mean of the diagnostic test values); or "PrevalenceMatching" (The value for which predicted prevalence is practically equal to observed prevalence).

data	a data frame containing all needed variables.
direction	character string specifying the direction to compute the ROC curve. By default individuals with a test value lower than the cutoff are classified as healthy (negative test), whereas patients with a test value greater than (or equal to) the cutoff are classified as diseased (positive test). If this is not the case, however, and the high values are related to health, this argument should be established at ">".
categorical.cov	a character string with the name of the categorical covariate according to which optimal cutpoints are to be calculated. The default is NULL (no categorical covariate).
pop.prev	the value of the disease's prevalence. The default is NULL (prevalence is estimated on the basis of sample prevalence). It can be a vector indicating the prevalence values for each categorical covariate level.
control	output of the control.cutpoints function.
ci.fit	a logical value. If TRUE, inference is performed on the accuracy measures at the optimal cutpoint. The default is FALSE.
conf.level	a numerical value with the confidence level for the construction of the confidence intervals. The default value is 0.95.
trace	a logical value. If TRUE, information on progress is shown. The default is FALSE.
...	further arguments passed to or from other methods. None are used in this method.

Details

Continuous biomarkers or diagnostic tests are often used to discriminate between diseased and healthy populations. In clinical practice, it is necessary to select a cutpoint or discrimination value c which defines positive and negative test results. Several methods for selecting optimal cutpoints in diagnostic tests have been proposed in the literature depending on the underlying reason for this

choice. In this package, thirty-two criteria are available. Before describing the methods in detail, mention should be made of the following notation: C_{FP}, C_{TN}, C_{FN} and C_{TP} are the costs of False Positive, True Negative, False Negative and True Positive decisions, respectively; p is disease prevalence; Se is Sensitivity; and Sp is Specificity.

"CB": Criterion based on cost-benefit methodology by means of calculating the slope of the ROC curve at the optimal cutoff as

$$ S = \frac{1-p}{p} CR = \frac{1-p}{p} \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}} $$

(McNeill et al. 1975; Metz et al. 1975; Metz 1978). This method thus weighs the relative costs of the different predictions in the diagnosis. By default, the costs ratio is 1, and this is the `costs.ratio` argument in the `control.cutpoints` function.

"MCT": Criterion based on the minimization of the Misclassification Cost Term (MCT) defined as

$$ MCT(c) = \frac{C_{FN}}{C_{FP}} p(1 - Se(c)) + (1 - p)(1 - Sp(c)) $$

(Smith 1991; Greiner 1995,1996). By default, $C_{FN} = C_{FP} = 1$, and these are the `CFN` and `CFP` arguments in the `control.cutpoints` function.

"MinValueSp": Criterion based on setting a minimum value for Specificity and maximizing Sensitivity, subject to this condition (Shaefer 1989; Vermont et al. 1991; Gallop et al. 2003). Hence, in a case where there is more than one cutpoint fulfilling this condition, those which yield maximum Sensitivity are chosen. If several cutpoints still remain, those yielding the greatest Specificity are chosen. By default, the minimum value for Specificity is 0.85, and this is the `valueSp` argument in the `control.cutpoints` function.

"MinValueSe": Criterion based on setting a minimum value for Sensitivity and maximizing Specificity, subject to this condition (Shaefer 1989; Vermont et al. 1991; Gallop et al. 2003). Hence, in a case where there is more than one cutpoint fulfilling this condition, those which yield maximum Specificity are chosen. If several cutpoints still remain, those yielding the greatest Sensitivity are chosen. By default, the minimum value for Sensitivity is 0.85, and this is the `valueSe` argument in the `control.cutpoints` function.

"ValueSp": Criterion based on setting a particular value for Specificity (Rutter and Miglioretti 2003). In a case where there is more than one cutpoint fulfilling this condition, those which yield maximum Sensitivity are chosen.

"ValueSe": Criterion based on setting a particular value for Sensitivity (Rutter and Miglioretti 2003). In a case where there is more than one cutpoint fulfilling this condition, those which yield maximum Specificity are chosen.

"MinValueSpSe": Criterion based on setting minimum values for Sensitivity and Specificity measures (Shaefer 1989). In a case where there is more than one cutpoint fulfilling these conditions, those which yield maximum Sensitivity or maximum Specificity are chosen. The user can select one of these two options by means of the `maxSp` argument in the `control.cutpoints` function. If TRUE (the default value), the cutpoint/s yielding maximum Specificity is/are computed. If there are still several cutpoints which maximize the chosen measure, those which also maximize the other measure are chosen.

"MaxSp": Criterion based on maximization of Specificity (Bortheiry et al. 1994; Hoffman et al. 2000; Alvarez-Garcia et al. 2003). If there is more than one cutpoint fulfilling this condition, those which yield maximum Sensitivity are chosen.

"MaxSe": Criterion based on maximization of Sensitivity (Filella et al. 1995; Hoffman et al. 2000; Alvarez-Garcia et al. 2003). If there is more than one cutpoint fulfilling this condition, those which yield maximum Specificity are chosen.

"MaxSpSe": Criterion based on simultaneously maximizing Sensitivity and Specificity (Riddle and Stratford 1999; Gallop et al. 2003).

"MaxProdSpSe": Criterion based on maximizing the product of Specificity and Sensitivity (Lewis et al. 2008). This criterion is the same as the method based on maximization of the Accuracy Area (Greiner 1995, 1996) defined as

$$AA(c) = fracTP(c)TN(c)(TP(c) + FN(c))(FP(c) + TN(c))$$

where TP, TN, FN and FP are the number of True Positives, True Negatives, False Negatives and False Positives classifications, respectively.

"ROC01": Criterion of the point on the ROC curve closest to the point (0,1), i.e, upper left corner of the unit square (Metz 1978; Vermont et al. 1991).

"SpEqualSe": Criterion based on the equality of Sensitivity and Specificity (Greiner et al. 1995; Hosmer and Lemeshow 2000; Peng and So 2002). Since Specificity may not be exactly equal to Sensitivity, the absolute value of the difference between them is minimized.

"Youden": Criterion based on Youden's Index (Youden 1950; Aoki et al. 1997; Shapiro 1999; Greiner et al. 2000) defined as $YI(c) = max_c(Se(c) + Sp(c) - 1)$. This is identical (from an optimization point of view) to the method that maximizes the sum of Sensitivity and Specificity (Albert and Harris 1987; Zweig and Campbell 1993) and to the criterion that maximizes concordance, wich is a monotone function of the AUC, defined as

$$AUC(c) = \frac{Se(c) - (1 - Sp(c)) + 1}{2}$$

(Begg et al. 2000; Gonen and Sima 2008). Costs of misclassifications can be considered in this criterion and for using the Generalized Youden Index: $GYI(c) = max_c(Se(c) + rSp(c) - 1$ (Geisser 1998; Greiner et al. 2000; Schisterman et al. 2005), where

$$r = \frac{1-p}{p}\frac{C_{FN}}{C_{FP}}$$

. If the generalized.Youden argument in the control.cutpoints function is TRUE, Generalized Youden Index is computed. The default is FALSE. The CFN and CFP arguments in the control.cutpoints function indicate the cost values, and by default, $C_{FN} = C_{FP} = 1$. Moreover, the optimal cutpoint based on Youden's Index can be computed by means of cost-benefit methodology (see "CB" method), with the slope of the ROC curve at the optimal cutoff being $S = 1$ for the Youden Index and $S = \frac{1-p}{p}\frac{C_{FN}}{C_{FP}}$ for the Generalized Youden Index. If the costs.benefits.Youden argument in the control.cutpoints function is TRUE, the optimal cutpoint based on cost-benefit methodology is computed. By default, it is FALSE.

"MaxEfficiency": Criterion based on maximization of the Efficiency, Accuracy, Validity Index or percentage of cases correctly classified defined as $Ef(c) = pSe(c) + (1-p)Sp(c)$ (Feinstein 1975; Galen 1986; Greiner 1995, 1996). This criterion is similar to the criterion based on minimization of the Misclassification Rate which measures the error in cases where diseased and disease-free patients are misdiagnosed (Metz 1978). It is defined as $ER(c) = p(1-Se(c)) + (1-p)(1-Sp(c))$. Moreover, the optimal cutpoint based on this method can be computed by means of cost-benefit methodology (see "CB" method), with the slope of the ROC curve at the optimal cutoff being

$S = \frac{1-p}{p}$. If the `costs.benefits.Efficiency` argument in the `control.cutpoints` function is TRUE, the optimal cut-point based on cost-benefit methodology is computed. By default, it is FALSE.

"Minimax": Criterion based on minimization of the most frequent error (Hand 1987): $min_c(max(p(1 - Se(c)), (1 - p)(1 - Sp(c))))$. In a case where there is more than one cutpoint fulfilling this condition, those which yield maximum Sensitivity or maximum Specificity are chosen. The user can select one of these two options by means of the `maxSp` argument in the `control.cutpoints` function. If TRUE (the default value), the cutpoint/s yielding maximum Specificity is/are computed. If there are still several cutpoints which maximize the chosen measure, those which also maximize the other measure are chosen.

"MaxDOR": Criterion based on maximizating the Diagnostic Odds Ratio (DOR), defined as

$$DOR(c) = \frac{Se(c)}{(1 - Se(c))} \frac{Sp(c)}{(1 - Sp(c))}$$

(Kraemer 1992; Greiner et al. 2000; Boehning et al. 2011).

"MaxKappa": Criterion based on maximization of the Kappa Index (Cohen 1960; Greiner et al. 2000). Kappa makes full use of the information in the confusion matrix to assess the improvement over chance prediction. Costs of misclassifications can be considered in this criterion and for using the Weighted Kappa Index (Kraemer 1992; Kraemer et al. 2002) defined as

$$PK(c) = \frac{p(1 - p)(Se(c) + Sp(c) - 1)}{p(p(1 - Se(c)) + (1 - p)Sp(c))r + (1 - p)(pSe(c) + (1 - p)(1 - Sp(c)))(1 - r)}$$

where

$$r = \frac{C_{FP}}{C_{FP} + C_{FN}}$$

. If the `weighted.Kappa` argument in the `control.cutpoints` function is TRUE, the Weighted Kappa Index is computed. The default value is FALSE. The `CFN` and `CFP` arguments in the `control.cutpoints` function indicate the cost values, and by default, $C_{FP} = C_{FN} = 1$.

"MinValueNPV": Criterion based on setting a minimum value for Negative Predictive Value (Vermont et al. 1991). In a case where there is more than one cutpoint fulfilling this condition, those which yield the maximum Positive Predictive Value are chosen. If several cutpoints still remain, those yielding the highest Negative Predictive Value are chosen. By default, the minimum value for Negative Predictive Value is 0.85 and this is the `valueNPV` argument in the `control.cutpoints()` function.

"MinValuePPV": Criterion based on setting a minimum value for Positive Predictive Value (Vermont et al. 1991). In a case where there is more than one cutpoint fulfilling this condition, those which yield the maximum Negative Predictive Value are chosen. If several cutpoints still remain, those yielding the highest Positive Predictive Value are chosen. By default, the minimum value for Positive Predictive Value is 0.85, and this is specified by the `valuePPV` argument in the `control.cutpoints()` function.

"ValueNPV": Criterion based on setting a particular value for Negative Predictive Value. In a case where there is more than one cutpoint fulfilling this condition, those which yield maximum Positive Predictive Value are chosen.

"ValuePPV": Criterion based on setting a particular value for Positive Predictive Value. In a case where there is more than one cutpoint fulfilling this condition, those which yield maximum Negative Predictive Value are chosen.

"MinValueNPVPPV": Criterion based on setting minimum values for Predictive Values (Vermont et al. 1991). In a case where there is more than one cutpoint fulfilling these conditions, those which yield the maximum Negative or maximum Positive Predictive Value are chosen. The user can select one of these two options by means of the maxNPV argument in the control.cutpoints function. If TRUE (the default value), the cutpoint/s yielding maximum Negative Predictive Value is/are computed. If there are still several cutpoints which maximize the chosen measure, those which also maximize the other measure are chosen.

"PROC01": Criterion of the point on the PROC curve closest to the point (0,1), i.e., upper left corner of the unit square (Vermont et al. 1991; Gallop et al. 2003).

"NPVEqualPPV": Criterion based on the equality of Predictive Values (Vermont et al. 1991). Since the Positive Predictive Value may not be exactly equal to the Negative Predictive Value, the absolute value of the difference between them is minimized.

"MaxNPVPPV": Criterion based on simultaneously maximizing Positive Predictive Value and Negative Predictive Value.

"MaxSumNPVPPV": Criterion based on maximizing the sum of Positive Predictive Value and Negative Predictive Value.

"MaxProdNPVPPV": Criterion based on maximizing the product of Positive Predictive Value and Negative Predictive Value.

"ValueDLR.Negative": Criterion based on setting a particular value for the Negative Diagnostic Likelihood Ratio (Boyko 1994; Rutter and Miglioretti 2003). The default value is 0.5, and it is specified by the valueDLR.Negative argument in the control.cutpoints function.

"ValueDLR.Positive": Criterion based on setting a particular value for the Positive Diagnostic Likelihood Ratio (Boyko 1994; Rutter and Miglioretti 2003). The default value is 2, and it is specified by the valueDLR.Positive argument in the control.cutpoints function.

"MinPvalue": Criterion based on the minimum p-value associated with the statistical Chi-squared test which measures the association between the marker and the binary result obtained on using the cutpoint (Miller and Siegmund 1982; Lausen and Schumacher 1992; Altman et al. 1994; Mazumdar and Glasman 2000).

"ObservedPrev": Criterion based on setting the closest value to observed prevalence, i.e., $c/max_c|c - p|$, with p being prevalence estimated from the sample. This criterion is thus indicated/valid in cases where the diagnostic test takes values in the interval (0,1), and it is a useful method in cases where preserving prevalence is of prime importance (Manel et al. 2001).

"MeanPrev": Criterion based on setting the closest value to the mean of the diagnostic test values. This criterion is usually used in cases where the diagnostic test takes values in the interval (0,1), i.e., the mean probability of ocurrence, e.g., based on the results of a statistical model(Manel et al. 2001; Kelly et al. 2008).

"PrevalenceMatching": Criterion based on the equality of sample and predicted prevalence: $pSe(c) + (1 - p)(1 - Sp(c))$ where p is the prevalence estimated from the sample (Manel et al. 2001; Kelly et al. 2008). This criterion is usually used in cases where the diagnostic test takes values in the interval (0,1), i.e., the predicted probability, e.g., based on a statistical model.

Value

Returns an object of class "optimal.cutpoints" with the following components:

methods   a character vector with the value of the methods argument used in the call.

levels.cat	a character vector indicating the levels of the categorical covariate if the `categorical.cov` argument in the `optimal.cutpoints` function is not NULL.
call	the matched call.
data	the data frame with the variables used in the call.

For each of the methods used in the call, a list with the following components is obtained:

`"measures.acc"`	a list with all possible cutoffs, their associated accuracy measures (Sensitivity, Specificity, Predictive Values, Diagnostic Likelihood Ratios and Area under ROC Curve, AUC), the prevalence and the sample size for both healthy and diseased populations.
`"optimal.cutoff"`	a list with the optimal cutoff(s) and its/their associated accuracy measures (Sensitivity, Specificity, Predictive Values, Diagnostic Likelihood Ratios and the number of False Positive and False Negative decisions).

The following components only appear in some methods:

`"criterion"`	the value of the method considered for selecting the optimal cutpoint for each cutoff.
`"optimal.criterion"`	the optimal value of the method considered for selecting the optimal cutpoint, i.e., the value of the criterion at the optimal cutpoint.

Author(s)

Monica Lopez-Raton and Maria Xose Rodriguez-Alvarez

References

Albert, A. and Harris, E.K. (1987). Multivariate Interpretation of Clinical Laboratory Data. Marcel Dekker, New York, NY.

Altman, D.G., Lausen, B., Sauerbrei, W. and Schumacher, M. (1994). Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute* **86** (11), 829–835.

Alvarez-Garcia, G. et al. (2003). Influence of age and purpose for testing on the cut-off selection of serological methods in bovine neosporosis. *Veterinary Research* **34**, 341–352.

Aoki, K., Misumi, J., Kimura, T., Zhao, W. and Xie, T. (1997). Evaluation of cutoff levels for screening of gastric cancer using serum pepsinogens and distributions of levels of serum pepsinogens I, II and Of PG I/PG II ratios in a gastric cancer case-control study. *Journal of Epidemiology* **7**, 143–151.

Begg, C.B., Cramer, L.D., Venkatraman, E.S. and Rosai, J. (2000). Comparing tumour staging and grading systems: a case study and a review of the issues, using thymoma as a model. *Statistics in Medicine* **19**, 1997–2014.

Boehning, D., Holling, H. and Patilea, V. (2011). A limitation of the diagnostic-odds ratio in determining an optimal cut-off value for a continuous diagnostic test. *Statistical Methods in Medical Research*, **20**(5), 541–550.

Bortheiry, A.L., Malerbi, D.A. and Franco, L.J. (1994). The ROC curve in the evaluation of fasting capillary blood glucose as a screening test for diabetes y IGT. *Diabetes Care* **17**, 1269–1272.

Boyko, E.J. (1994). Ruling out or ruling in disease with the most sensitive or specific diagnostic test: short cut or wrong turn?. *Medical Decision Making* **14**, 175–179.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ Psychol Meas* **20**, 37–46.

Feinstein, S.H. (1975). The accuracy of diver sound localization by pointing. *Undersea. Biomed.Res* **2**(3), 173–184.

Filella, X., Alcover, J., Molina, R. et al. (1995). Clinical usefulness of free PSA fraction as an indicator of prostate cancer. *Int. J. Cancer* **63**, 780–784.

Galen, R.S. (1986). Use of predictive value theory in clinical immunology. In: N.R.Rose, H. Friedmann and J.L. Fahey (Eds.), Manual of Clinical Laboratory Immunology. American Society of Microbiology. Washington, DC, pp 966-970.

Gallop, R.J., Crits-Christoph, P., Muenz, L.R. and Tu, X.M. (2003). Determination and Interpretation of the Optimal Operating Point for ROC Curves Derived Through Generalized Linear Models. *Understanding Statistics* **2**(4), 219–242.

Geisser, S. (1998). Comparing two tests used for diagnostic or screening processes. *Statistics Probability Letters* **40**, 113–119.

Gonen, M. and Sima, C. (2008). Optimal cutpoint estimation with censored data. Memorial Sloan-Kettering Cancer Center Department of Epidemiology and Biostatistics Working Paper Series.

Greiner, M. (1995). Two-graph receiver operating characteristic (TG-ROC): a Microsoft-EXCEL template for the selection of cut-off values in diagnostic tests. *Journal of Immunological Methods* **185**(1),145–146.

Greiner, M. (1996). Two-graph receiver operating characteristic (TG-ROC): update version supports optimisation of cut-off values that minimise overall misclassification costs. *J. Immunol. Methods* **191**, 93–94.

Greiner, M., Pfeiffer, D. and Smith, R.D. (2000). Principals and practical application of the receiver operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine* **45**, 23–41.

Hand, D. (1987). Screening vs Prevalence Estimation. *Applied Statistics* **36**, 1–7.

Hoffman, R.M., Clanon, D.L., Littenberg, B., Frank, J.J. and Peirce, J.C. (2000). Using the Free-to-total Prostate-specific Antigen Ratio to Detect Prostate Cancer in Men with Nonspecific Elevations of Prostate-specific Antigen Levels. *J. Gen. Intern Med* **15**, 739–748.

Hosmer, D.W. and Lemeshow, S. (2000). Applied Logistic Regression. Wiley-Interscience, New York, USA.

Kelly, M.J., Dunstan, F.D., Lloyd, K. and Fone, D.L. (2008). Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: a comparison of five different methods. *BMC Psychiatry* **8**, 10.

Kraemer, H.C. (1992). Risk ratios, odds ratio, and the test QROC. In: Evaluating medical tests. Newbury Park, CA: SAGE Publications, Inc.; pp 103–113.

Kraemer, H.C., Periyakoil, V.S. and Noda, A. (2002). Kappa coefficients in medical research. *Statistics in Medicine* **21**, 2109–2129.

Lausen, B. and Schumacher, M. (1992). Maximally selected rank statistics. *Biometrics* **48**, 73–85.

Lewis, J.D., Chuai, S., Nessel, L., Lichtenstein, G.R., Aberra, F.N. and Ellenberg, J.H. (2008). Use of the Noninvasive Components of the Mayo Score to Assess Clinical Response in Ulcerative Colitis. *Inflamm Bowel Dis* **14**(12), 1660–1666.

Lopez-Raton, M., Rodriguez-Alvarez, M.X, Cadarso-Suarez, C. and Gude-Sampedro, F. (2014). OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *Journal of Statistical Software* **61**(8), 1–36. doi: 10.18637/jss.v061.i08.

Manel, S., Williams, H. and Ormerod, S. (2001). Evaluating Presence-Absence Models in Ecology: the Need to Account for Prevalence. *Journal of Applied Ecology* **38**, 921–931.

Mazumdar, M. and Glassman, J.R. (2000). Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Statistics in Medicine* **19**, 113–132.

McNeill, B.J., Keeler, E. and Adelstein, S.J. (1975). Primer on certain elements of medical decision making, with comments on analysis ROC. *N. Engl. J Med* **293**, 211–215.

Metz, C.E., Starr, S.J., Lusted, L.B. and Rossmann, K. (1975). Progress in evaluation of human observer visual detection performance using the ROC curve approach. In: Raynaud C, Todd-Pokropek AE eds. Information processing in scintigraphy. Orsay, France: CEA, 420–436.

Metz, CE. (1978). Basic principles of ROC analysis. *Seminars Nucl. Med.* **8**, 283–298.

Miller, R. and Siegmund, D. (1982). Maximally selected chi square statistics. *Biometrics* **38**, 1011–1016.

Navarro, J.B., Domenech, J.M., de la Osa, N. and Ezpeleta, L. (1998). El analisis de curvas ROC en estudios epidemiologicos de psicopatologia infantil: aplicacion al cuestionario CBCL. *Anuario de Psicologia* **29** (1), 3–15.

Peng, C.Y.J. and So, T.S.H. (2002). Logistic Regression Analysis and Reporting: A Primer. *Understanding Statistics* **1**(1), 31–70.

Riddle, D.L. and Stratford, P.W. (1999). Interpreting validity indexes for diagnostic tests: an illustration using the Berg Balance Test. *Physical Therapy* **79**, 939–950.

Rutter, C.M. and Miglioretti, D.L. (2003). Estimating the accuracy of psychological scales using longitudinal data. *Biostatistics* **4**(1), 97–107.

Shaefer, H. (1989). Constructing a cut-off point for a quantitative diagnostic test. *Statistics in Medicine* **8**, 1381–1391.

Schisterman, E.F., Perkins, N.J., Liu, A. and Bondell, H. (2005). Optimal cutpoint and its corresponding Youden index to discriminate individuals using pooled blood samples. *Epidemiology* **16**, 73–81.

Shapiro, D.E. (1999). The interpretation of diagnostic tests. *Statistical Methods in Medical Research* **8**, 113–134.

Smith, R.D. (1991). Evaluation of diagnostic tests. In: R.D. Smith (Ed.), Veterinary Clinical Epidemiology. Butter-worth-Heinemann. Stoneham, pp 29–43.

Vermont J, Bosson JL, Francois P, Robert C, Rueff A, Demongeot J. (1991). Strategies for graphical threshold determination. *Computer Methods and Programs in Biomedicine* **35**, 141–150.

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer* **3**, 32–35.

Zweig, M.H., Campbell, G. (1993). Receiver-operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* **39**, 561–577.

See Also

control.cutpoints, summary.optimal.cutpoints

Examples

```
library(OptimalCutpoints)
data(elas)
####################
# marker: elas
# status: status
# categorical covariates:
# gender
####################

############################################################
# Youden Index Method ("Youden"): Covariate gender
############################################################
# Defaut method
optimal.cutpoint.Youden <- optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(), ci.fit = FALSE, conf.level = 0.95, trace = FALSE)

summary(optimal.cutpoint.Youden)

plot(optimal.cutpoint.Youden)

# Formula method
optimal.cutpoint.Youden <- optimal.cutpoints(X = elas ~ status, tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(), ci.fit = FALSE, conf.level = 0.95, trace = FALSE)

#  Inference on the test accuracy measures
optimal.cutpoint.Youden <- optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(), ci.fit = TRUE, conf.level = 0.95, trace = FALSE)

summary(optimal.cutpoint.Youden)

################################################################################
# Sensitivity equal to Specificity Method ("SpEqualSe"): Covariate gender
################################################################################
optimal.cutpoint.SpEqualSe <- optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "SpEqualSe", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(), ci.fit = TRUE, conf.level = 0.95, trace = FALSE)

summary(optimal.cutpoint.SpEqualSe)

plot(optimal.cutpoint.SpEqualSe)
```

---

plot.optimal.cutpoints

Default optimal.cutpoints plotting

Description

On the basis of an `optimal.cutpoints` object, three plots are currently available: (1) a plot of the Receiver Operating Characteristic (ROC) curve; (2) a plot of the Predictive ROC (PROC) curve; and, in some methods, (3) a plot of the values of the optimal criterion used as a function of the cutoffs.

Usage

```
## S3 method for class 'optimal.cutpoints'
plot(x, legend = TRUE, which = c(1,2), ...)
```

Arguments

x	an object of class `optimal.cutpoint` as produced by `optimal.cutpoints()`.
legend	a logical value for including the legend of optimal coordinates with specific characteristics. The default is TRUE.
which	a numeric vector with the required plots. By default, both the ROC and the PROC curves are plotted.
...	further arguments passed to method `plot.default`.

Author(s)

Monica Lopez-Raton and Maria Xose Rodriguez-Alvarez

See Also

`optimal.cutpoints`, `control.cutpoints`

Examples

```
library(OptimalCutpoints)
data(elas)
###########################################################
# Youden Index method ("Youden"): Covariate gender
###########################################################
optimal.cutpoint.Youden<-optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(), ci.fit = TRUE, conf.level = 0.95, trace = FALSE)

# Plot by default
plot(optimal.cutpoint.Youden)

#  Not including the optimal coordinates
plot(optimal.cutpoint.Youden, legend = FALSE)
# Change the colour
plot(optimal.cutpoint.Youden, col = "blue")
```

print.optimal.cutpoints

Print method for optimal.cutpoints objects

Description

Default print method for objects fitted with optimal.cutpoints() function. A short summary is printed with: the call to the optimal.cutpoints() function; the optimal cutpoint(s) and the value of the Area Under the ROC Curve (AUC) for each categorical covariate level (if the categorical.cov argument of the optimal.cutpoints function is not NULL).

Usage

```
## S3 method for class 'optimal.cutpoints'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

x	an object of class optimal.cutpoints as produced by optimal.cutpoints().
digits	controls number of digits printed in the output.
...	further arguments passed to or from other methods. None are used in this method.

Author(s)

Monica Lopez-Raton and Maria Xose Rodriguez-Alvarez

See Also

optimal.cutpoints, summary.optimal.cutpoints

Examples

```
library(OptimalCutpoints)
data(elas)
############################################################
# Youden Index Method ("Youden"): Covariate gender
############################################################
optimal.cutpoint.Youden<-optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(), ci.fit = TRUE, conf.level = 0.95, trace = FALSE)

optimal.cutpoint.Youden

print(optimal.cutpoint.Youden)
```

summary.optimal.cutpoints

Summary method for optimal.cutpoints objects

Description

Produces a summary of a `optimal.cutpoints` object. The following are printed: the call to the `optimal.cutpoints()` function; the optimal cutpoint(s) obtained with the method(s) selected; its/their accuracy measures and the area under ROC curve (AUC) estimates at each categorical covariate level (if the `categorical.cov` argument in the `optimal.cutpoints()` function is not NULL). If `optimal.cutpoints()` was called with the `ci.fit = TRUE` argument, confidence intervals for accuracy measures at the optimal cutpoint are also printed.

Usage

```
## S3 method for class 'optimal.cutpoints'
summary(object, ...)
```

Arguments

`object`	an object of class `optimal.cutpoints` as produced by `optimal.cutpoints()`
`...`	further arguments passed to or from other methods. None are used in this method.

Details

The `summary.optimal.cutpoints` function produces a list of summary information for a fitted `optimal.cutpoints` object. The result depends on the three arguments, namely, `methods`, `categorical.cov` and `ci.fit` of the `optimal.cutpoints()` function used in the optimal cutpoints computing process.

Value

Returns an object of class "summary.optimal.cutpoints" with the same components as the `optimal.cutpoints` function (see [optimal.cutpoints](#)) plus:

`p.table`	a list with all the numerical information to be shown on the screen.

Author(s)

Monica Lopez-Raton and Maria Xose Rodriguez-Alvarez

See Also

[optimal.cutpoints](#)

Examples

```
library(OptimalCutpoints)
data(elas)
###########################################################
# Youden Index Method ("Youden"): Covariate gender
###########################################################
optimal.cutpoint.Youden<-optimal.cutpoints(X = "elas", status = "status", tag.healthy = 0,
methods = "Youden", data = elas, pop.prev = NULL, categorical.cov = "gender",
control = control.cutpoints(), ci.fit = TRUE, conf.level = 0.95, trace = FALSE)

summary(optimal.cutpoint.Youden)
```

Index