Automatically Find Optimal Threshold Point in ROC Curve using ROCit package in R

Yash Patil [Follow]

Aug 11 · 3 min read

Interpreting Binary Classifier with R using ROCit Package

Sensitivity (or recall or true positive rate), false-positive rate, specificity, precision (or positive predictive value), negative predictive value, misclassification rate, accuracy, F-score- these are popular metrics for assessing the performance of the binary classifier for a certain threshold. These metrics are calculated at certain threshold values. The receiver operating characteristic (ROC) curve is a common tool for assessing the overall diagnostic ability of the binary classifier. Unlike depending on a certain threshold, the area under ROC curve (also known as AUC), is a summary statistic about how well a binary classifier performs overall for the classification task. ROCit package provides flexibility to easily evaluate threshold-bound metrics. Also, the ROC curve, along with AUC can be obtained using different methods, such as empirical, binormal, and non-parametric. ROCit encompasses a wide variety of methods for constructing a confidence interval of the ROC curve and AUC. ROCit also features the option of constructing an empirical gains table, which is a handy tool for direct marketing. The package offers options for commonly used visualization, such as ROC curve, KS plot, lift plot. Along with an in-built default graphics setting, there are rooms for manual tweak by providing the necessary values as function arguments. ROCit is a powerful tool offering a range of things, yet it is very easy to use.

Performance metrics of a binary classifier

Various performance metrics for binary classifiers are available that are cutoff specific. Following metrics can be called for via measure argument:

- ACC: Overall accuracy of classification.

- MIS: Misclassification rate.

- SENS: Sensitivity.

- SPEC: Specificity.

- PREC: Precision.

- REC: Recall. Same as sensitivity.

- PPV: Positive predictive value.

- NPV: Positive predictive value.

- TPR: True positive rate.

- FPR: False positive rate.

- TNR: True negative rate.

- FNR: False-negative rate.

- pDLR: Positive diagnostic likelihood ratio.

- nDLR: Negative diagnostic likelihood ratio.

- FSCR: F-score.

ROC curve estimation

rocit is the main function of the ROCit package. With the diagnostic score and the class of each observation, it calculates true positive rate (sensitivity) and false positive rate (1-Specificity) at convenient cutoff values to construct the ROC curve. The function returns "rocit" object, which can be passed as arguments for other S3 methods.

Diabetes data contains information on 403 subjects from 1046 subjects who were interviewed in a study to understand the prevalence of obesity, diabetes, and other cardiovascular risk factors in central Virginia for African Americans. According to Dr John Hong, Diabetes Mellitus Type II (adult onset diabetes) is associated most strongly with obesity. The waist/hip ratio may be a predictor in diabetes and heart disease. DM II is also associated with hypertension — they may both be part of "Syndrome X". The 403 subjects were the ones who were screened for diabetes. Glycosylated hemoglobin > 7.0 is usually taken as a positive diagnosis of diabetes.

In the data, the dtest variable indicates whether glyhb is greater than 7 or not.

```r
data("Diabetes")
## first, fit a logistic model
logistic.model <- glm(as.factor(dtest)~
chol+age+bmi,
data = Diabetes,
family = "binomial")
## make the score and class
class <- logistic.model$y
# score = log odds
score <- logit(logistic.model$fitted.values)
## rocit object
rocit_emp <- rocit(score = score,
class = class,
method = "emp")
rocit_bin <- rocit(score = score,
class = class,
method = "bin")
rocit_non <- rocit(score = score,
class = class,
method = "non")
summary(rocit_emp)
summary(rocit_bin)
summary(rocit_non)
## Plot ROC curve
plot(rocit_emp, col = c(1,"gray50"),
```
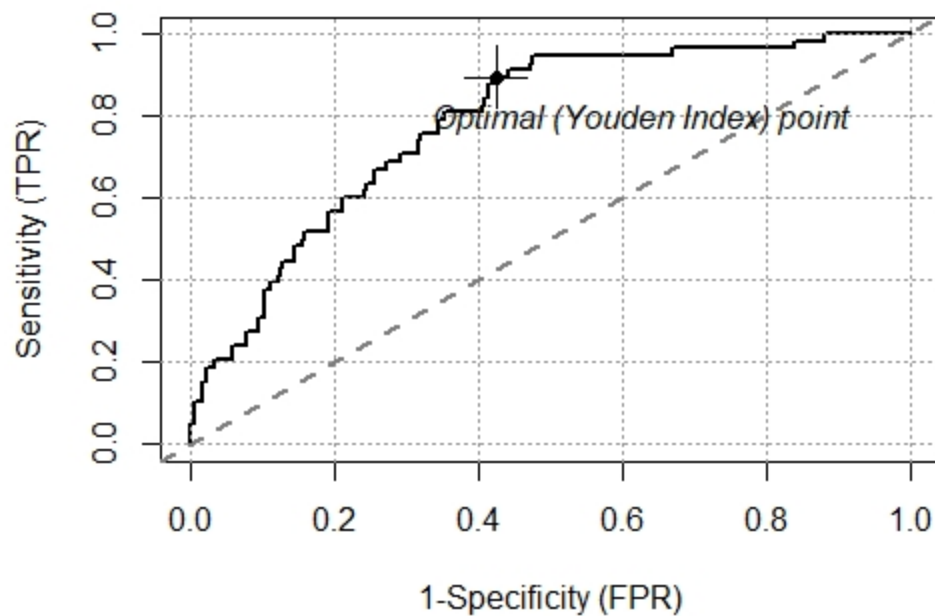
```
legend = FALSE, YIndex = FALSE)
```

The above program will create the logistic regression model for "Diabetes" dataset and the rocit object which is created gives us the graph between FPR and TPR with optimum cutoff value using the Youden Index as follows

References:

1. https://MedicalBiostatistics.com

2. https://rdrr.io/cran/ROCit/f/vignettes/my-vignette.Rmd

Roc   Auc   Roc Curve

Medium

About   Help   Legal

Get the Medium app

Download on the
App Store

GET IT ON
Google Play