Evaluation-Charts-for-Classification.R

patriciamaya

2020-10-06

```r
#EVALUATION CHARTS
#Example using logistic regression mode
library(ISLR)

## Warning: package 'ISLR' was built under R version 4.0.2

data("Carseats")
attach(Carseats)
High <- as.factor(ifelse(Sales >= 8, "YES", "NO"))
Data <- data.frame(Carseats, High)
Data <- Data[,-1]
colnames(Data)[11] <- "Target"

indx <- sample(2, nrow(Data), replace=T, prob= c(0.8,0.2))
train <- Data[indx==1, ]
test <- Data[indx==2, ]

logitModel <- glm(Target ~ ., data= train, family="binomial")
#predicted probabibilities
predictions<- predict(logitModel, newdata= test, type ="response")
predicted_class <- as.factor(ifelse(predictions >= 0.5, "YES", "NO"))
Actual <- test$Target

library(ROCR)

## Warning: package 'ROCR' was built under R version 4.0.2

#2 main functions: prediction & performance
#pred <- prediction(True labels, predicted probabilities for +ve class)
#performance(pred, y-axis, x-axis)
    #gain chart: performance(pred, "tpr", "rpp")
    #response chart: performance(pred, "ppv", "rpp")
    #ROC curve: performance(pred, "tpr", "fpr")
pred <- prediction(predictions, Actual)

#****GAIN CHART****
#shows sensitivity versus rate of positive predictions.
perf_gain<- performance(pred, "tpr", "rpp")
plot(perf_gain)
```
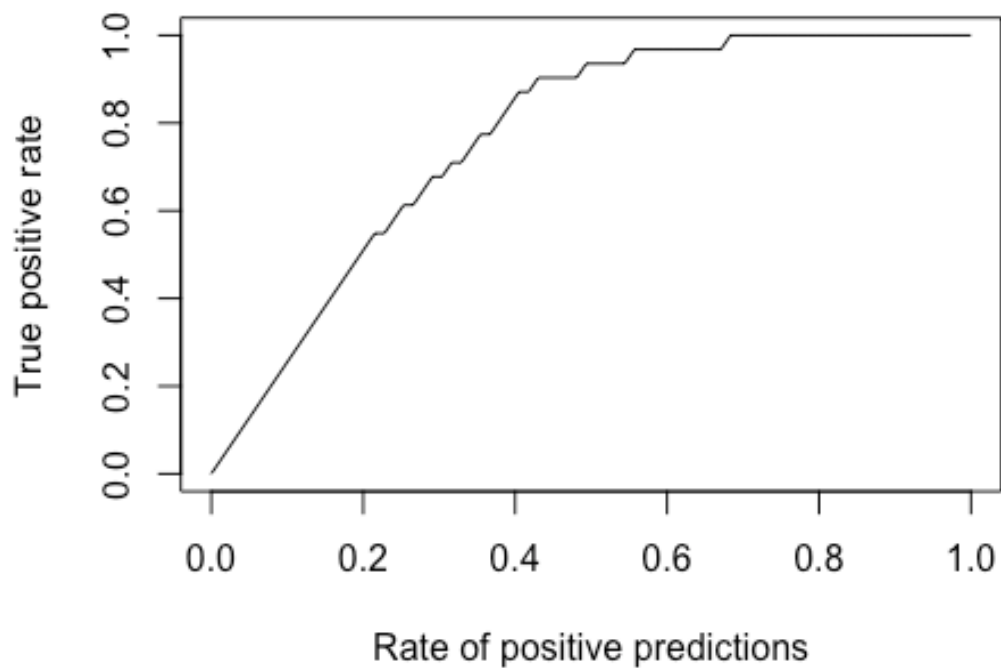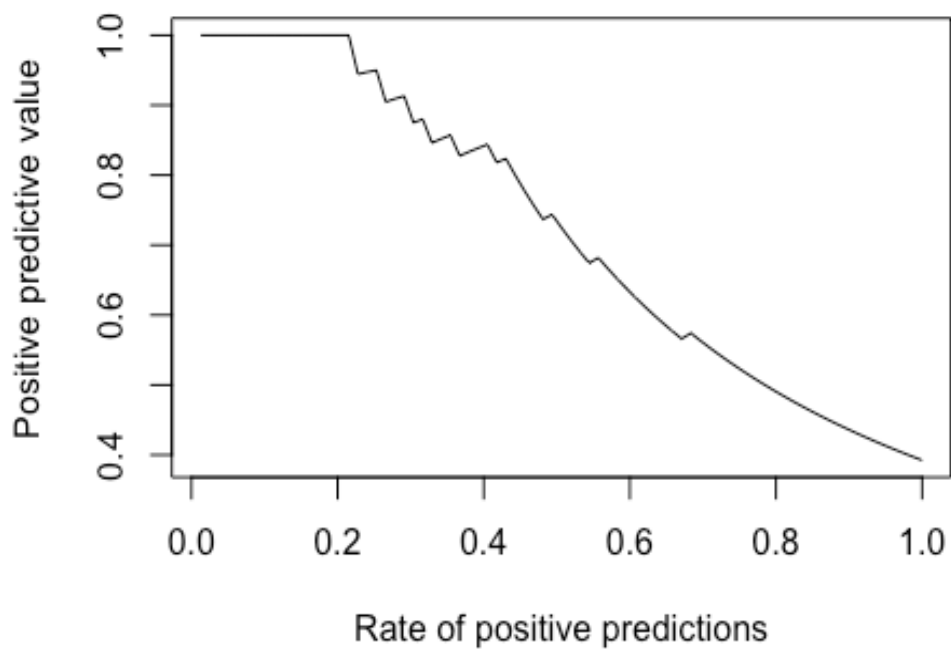
```
#****RESPONSE CHART****
perf_response <- performance(pred, "ppv", "rpp")
plot(perf_response)
```

```
#****LIFT CHART****
#Lift is a measure of the effectiveness of a predictive model calculated as
the
#ratio between the results obtained with and without the predictive model.
perf_lift <- performance(pred,"lift","rpp")
plot(perf_lift)
```
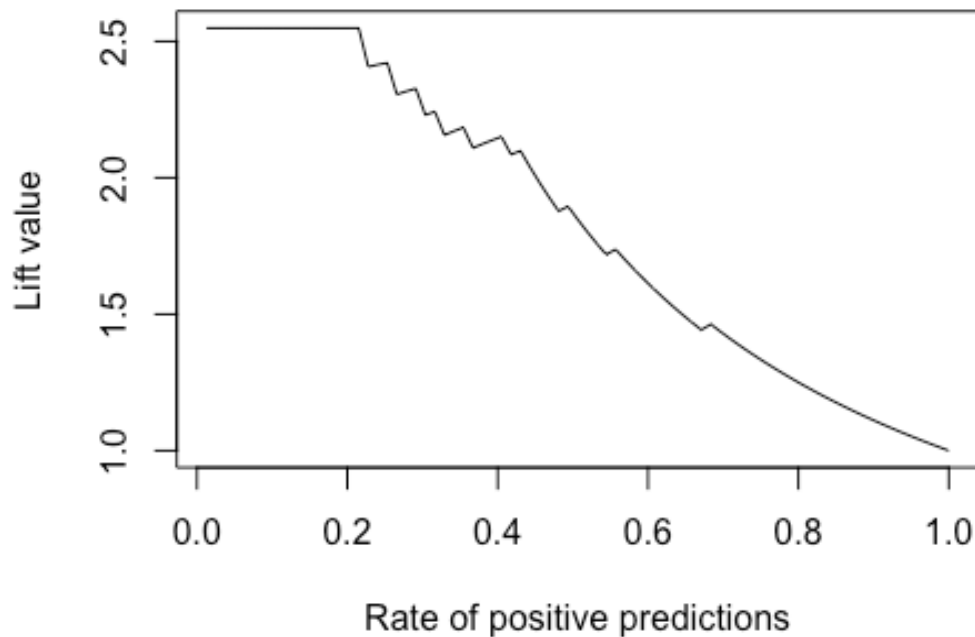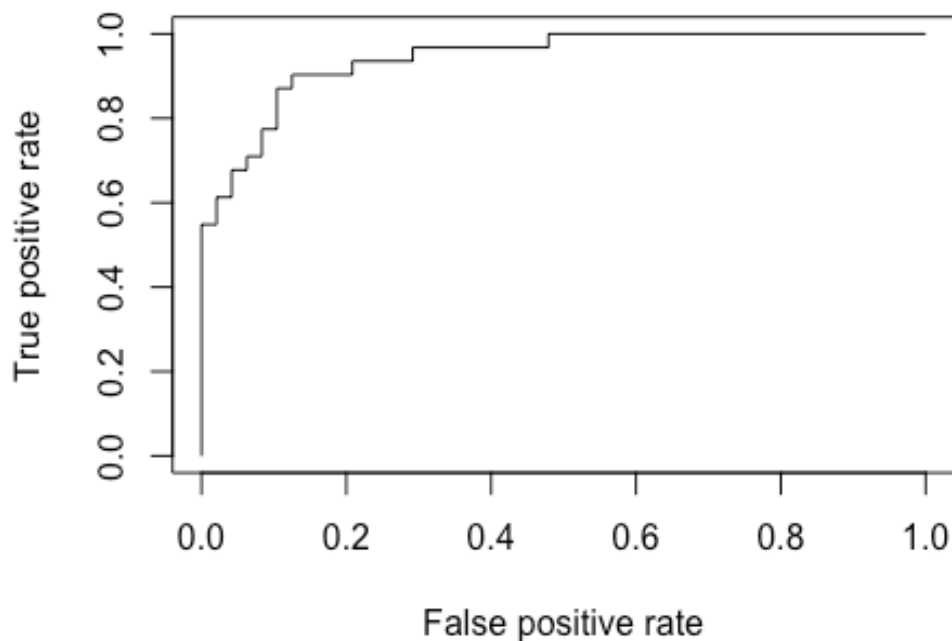


```
#****ROC CURVE****
perf_roc<- performance(pred, "tpr", "fpr")
plot(perf_roc)
```

```r
#****AUC- area under the ROC curve****
auc <- performance(pred, "auc")
auc <- unlist(slot(auc, "y.values"))
    #best auc is 1
    #random model is 0.5


#acc: Accuracy. Estimated as: (TP+TN)/(P+N).
#err: Error rate. Estimated as: (FP+FN)/(P+N).
#fpr: False positive rate. Estimated as: FP/N.
#tpr: True positive rate. Estimated as: TP/P.
#sens: Sensitivity. Same as tpr.
#fnr: False negative rate. Estimated as: FN/P.
#tnr: True negative rate. Estimated as: TN/N.
#spec: Specificity. Same as tnr.
#ppv: Positive predictive value. Estimated as: TP/(TP+FP).
#prec: Precision. Same as ppv.
#npv: Negative predictive value. Estimated as: TN/(TN+FN).
#rpp: Rate of positive predictions. Estimated as: (TP+FP)/(TP+FP+TN+FN).
#rnp: Rate of negative predictions. Estimated as: (TN+FN)/(TP+FP+TN+FN).
#odds: Odds ratio. (TP*TN)/(FN*FP). Note that odds ratio produces Inf or NA
values for all cutoffs corresponding to FN=0 or FP=0. This can substantially
decrease the plotted cutoff region.
```