Fitting a Logistic Regression Model to the Thoracic Surgery Binary Dataset

Dan Siegel

# Fitting the training data to a logistic regression model using the glm() function and to calculate the following classifier metrics.

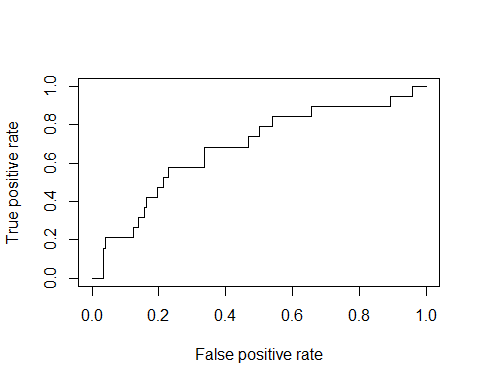
set.seed(123)  
n<- nrow(surg)  
shuffled <- surg[sample(n),]  
train <- shuffled[1:round(0.7 \* n),]  
test <- shuffled[(round(0.7 \* n) + 1):n,]  
testing\_survived <- glm(Risk1Yr ~ DGN + PRE4 + PRE5 + PRE6 + PRE7 + PRE8 +   
 PRE9 + PRE10 + PRE11 + PRE14 + PRE17 + PRE19 +   
 PRE25 + PRE30 + PRE32 + AGE, data=train)  
predicted\_data <- predict(testing\_survived, newdata = test)

## Precision, recall, and F1 score for the model

pred <- prediction(predicted\_data, test$Risk1Yr)  
RP.perf <- performance(pred, "prec", "rec")  
f1\_score <- performance(pred,"f")  
ROC.perf <- performance(pred, "tpr", "fpr")

## Plotting the receiver operating characteristic (ROC) curve using the test dataset and the area under the curve (AUC) value.

plot (ROC.perf)



auc.tmp <- performance(pred,"auc");  
auc <- as.numeric(auc.tmp@y.values)  
auc

## [1] 0.6816221