

LDA.R

Amisha Sancheti

2020-04-14

```
emp_churn <- read.csv("C:/Users/Amisha Sancheti/Desktop/MITA
sem2/Multivariate Analysis/Project/attritiondata.csv")
library(data.table)
setDT(emp_churn)

#Splitting into test and train data
library(caTools)

partn = floor(0.80*nrow(emp_churn))

set.seed(123)
train_ind = sample(seq_len(nrow(emp_churn)),size = partn)

empchurn_train = emp_churn[train_ind,]
empchurn_test= emp_churn[-train_ind,]

empchurn_train[is.na(empchurn_train)] <- 0

library(MASS)

empchurn_train = subset(empchurn_train, select = -
c(BusinessTravel,MaritalStatus) )

#Executing LDA Model
lda1 <- lda(formula=empchurn_train$Attrition ~., data = empchurn_train)

#Making prediction on test data
pred1 <- predict(lda1, newdata=empchurn_test, type ="response")

lda.pred<- pred1$x

#confusion matrix
cm<-table(pred1$class, empchurn_test$Attrition)
cm

##
##      No Yes
## No  234  28
## Yes  12  20
```

```

#overall accuracy
accuracy <- (sum(diag(cm)) / sum(cm))
accuracy

## [1] 0.8639456

library(pROC)

## Warning: package 'pROC' was built under R version 3.6.3

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var

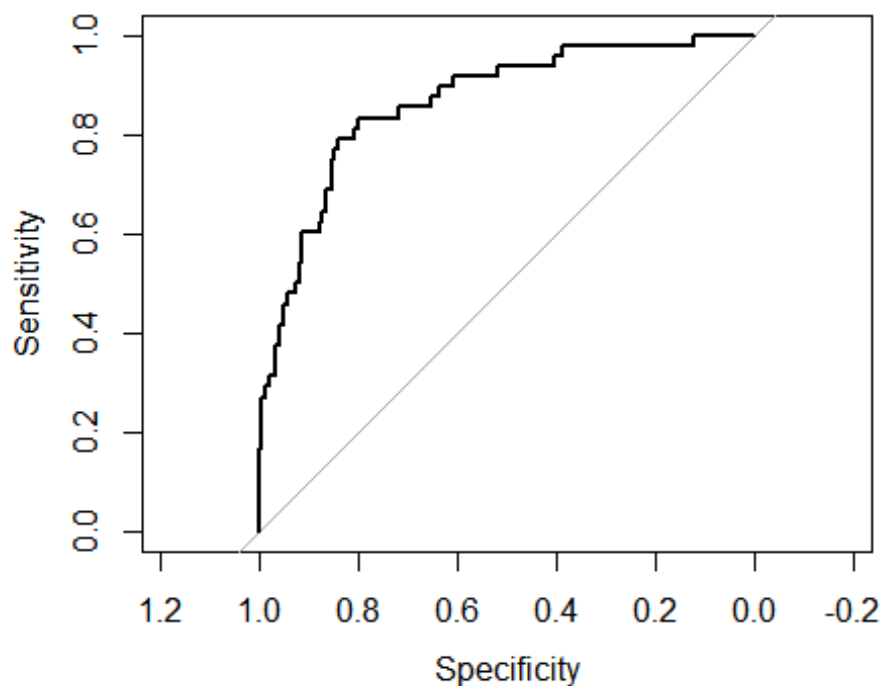
#area under the curve
r <- roc(empchurn_test$Attrition, pred1$posterior[,2])

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

plot.roc(r)

```



```

auc(r)

```

Area under the curve: 0.8652

#The model accuracy for Linear Discriminant Analysis is about 87% and the area under the curve for the same is 0.85.