#Logistic regression

#load the data

quality <- read.csv("https://raw.githubusercontent.com/TarekDib03/Analytics/master/Week3%20-%20Logistic%20Regression/Data/quality.csv")

View(quality)

#1 - low quality care

#0 - high quality care

# Number of patients received good and bad care

table(quality$PoorCare)

#baseline accuracy = ?

98/(98+33)

#split the data

library(caTools)

set.seed(88)

split <- sample.split(quality$PoorCare, SplitRatio = 0.75)

split

#creating training and testing dataset

quality\_train <- subset(quality, split == TRUE)

quality\_test <- subset(quality, split == FALSE)

names(quality\_train)

#Run logistic regression model

qualitylog <- glm (PoorCare ~ ERVisits + OfficeVisits, data = quality\_train, family = binomial)

qualitylog

summary(qualitylog)

Predicttrain <- predict(qualitylog, type = "response")

Predicttrain

View(quality\_train)

#confusion matrix

table(actual\_values = quality\_train$PoorCare, predicted\_values = Predicttrain > 0.5)

# Predicted 0 Predicted 1

# Actual 0 TN FP

# Actual 1 FN TP

#right classification can be obtained by

(71+6)/(3+19+71+6)

View(quality\_train)

#confusion matrix

table(actual\_values = quality\_train$PoorCare, predicted\_values = Predicttrain > 0.5)

# Predicted 0 Predicted 1

# Actual 0 TN FP

# Actual 1 FN TP

#right classification can be obtained by

(69+10)/(71+3+19+6)

table(actual\_values = quality\_train$PoorCare, predicted\_values = Predicttrain > 0.4)

#Roc Curve

#False positive rate and True positive rate

#install.packages("ROCR")

library(ROCR)

#predict function

ROCRred <- prediction(Predicttrain, quality\_train$PoorCare)

ROCRPref <- performance(ROCRred, "tpr", "fpr")

#plot ROC curve

plot(ROCRPref)

#Add colors

plot(ROCRPref, colorize = TRUE)

#To get the better cutoff value

plot(ROCRPref, colorize = TRUE, print.cutoffs.at = seq(0.2, by = 0.2), text.adj = c(-0.2, 1.7))

# Apply LM for Test Data

View(quality\_test)

predicttest <- predict(qualitylog, newdata = quality\_test, type = "response")

table(Actual\_values = quality\_test$PoorCare, predicted\_values = predicttest >0.4)

(19+3)/(16+8+5+3)

quality\_test$poorcareP <- predict(qualitylog, newdata = quality\_test, type = "response")

View(quality\_test)