Week 10_2

Janine Par

2022-05-20

```r
#Fit a Logistic Regression Model

## Set the working directory to the root of your DSC 520 directory
setwd("C:/Users/janin/OneDrive/Documents/R_repo/dsc520/")

## Load the `data/binaryclassifierdata`
binaryclass_df <- read.csv("data/binary-classifier-data.csv", header=TRUE, comment.char = "@")

str(binaryclass_df)
```

```
## 'data.frame':    1498 obs. of  3 variables:
##  $ label: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ x    : num  70.9 75 73.8 66.4 69.1 ...
##  $ y    : num  83.2 87.9 92.2 81.1 84.5 ...
```

```r
head (binaryclass_df)
```

```
##   label        x        y
## 1     0 70.88469 83.17702
## 2     0 74.97176 87.92922
## 3     0 73.78333 92.20325
## 4     0 66.40747 81.10617
## 5     0 69.07399 84.53739
## 6     0 72.23616 86.38403
```

```r
#Fit a logistic regression model to the binary-classifier-data.csv dataset
# The dataset (found in binary-classifier-data.csv) contains three variables; label, x, and y. The labe

binaryclass.model <- glm(label~x+y, data=binaryclass_df, family=binomial())

summary (binaryclass.model)
```

```
##
## Call:
## glm(formula = label ~ x + y, family = binomial(), data = binaryclass_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3728  -1.1697  -0.9575   1.1646   1.3989
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.424809   0.117224   3.624  0.00029 ***
## x           -0.002571   0.001823  -1.411  0.15836
## y           -0.007956   0.001869  -4.257 2.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2075.8  on 1497  degrees of freedom
## Residual deviance: 2052.1  on 1495  degrees of freedom
## AIC: 2058.1
##
## Number of Fisher Scoring iterations: 4
```

```r
#What is the accuracy of the logistic regression classifier?

#Split data
bcsplit <- sample.split(binaryclass_df, SplitRatio = 0.8)

bcsplit_train <- subset(binaryclass_df,tssplit='True')
bcsplit_train

bcsplit_test <- subset(binaryclass_df,tssplit='Fales')
bcsplit_test

#Predict
res.train <- predict(binaryclass.model,bcsplit_train,type ="response")
res.train

res.test <- predict(binaryclass.model,bcsplit_test,type ="response")
res.test
```

```r
confmatrix <- table(Actual_value=bcsplit_train$label, Predicted_Value= res.train > 0.5)

(confmatrix [[1,1]] + confmatrix [[2,2]])/sum(confmatrix)
```

```
## [1] 0.5834446
```

Accuracy of the model is 58.3%