

ASSIGNMENT 10.2.2

Kiran Komati

2021-05-22

```
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_knit$set(root.dir = 'C:/Users/kiran/dsc520')
```

Load the Data

```
bcd<-read.csv('data/binary-classifier-data.csv')
```

a) Fit a logistic regression model to the binary-classifier-data.csv dataset

```
bcd_glm<-glm(formula=label ~ x + y,family = binomial(), data = bcd)
summary(bcd_glm)
```

```
##
## Call:
## glm(formula = label ~ x + y, family = binomial(), data = bcd)
##
## 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
```

b) The dataset (found in binary-classifier-data.csv) contains three variables; label, x, and y. The label variable is either 0 or 1 and is the output we want to predict using the x and y variables.

i) What is the accuracy of the logistic regression classifier?

```
bcd$predicted.proBABILITIES<-fitted(bcd_glm)
res <- predict(bcd_glm,type="response")
confMatrix <- table(Actual_Value=bcd$label,Predicted_Value=res>0.5)
confMatrix
```

```
##               Predicted_Value
## Actual_Value FALSE TRUE
##           0    429   338
##           1    286   445
```

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

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