binary

Stephen Smitshoek

19/05/2022

```
## Warning: package 'caTools' was built under R version 4.1.3
```

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

Confidence Matrix

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

Accuracy

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

```
library(caTools)

setwd("C:\\Users\\sksmi\\PeytoAccess\\Personal\\Bellevue\\DSC520\\dsc520")

binary_data <- read.csv('data\\binary-classifier-data.csv')
binary_data$label <- as.factor(binary_data$label)

binary_data_split <- sample.split(binary_data, SplitRatio = 0.7)

training_data <- subset(binary_data, binary_data_split==TRUE)
testing_data <- subset(binary_data, binary_data_split==FALSE)

binary_glm <- glm(label ~ x + y, data=training_data, family=binomial())

response <- predict(binary_glm, newdata=testing_data, type="response")

confmatrix <- table(Actual_Value=testing_data$label, Predicted_Value =
response > 0.5)
confmatrix

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