Week10_binary_classifier

Akila Selvaraj

2/19/2022

Logistic Regression of Binary Classifier

Splitting training and test dataset

I am splitting the data into training and test dataset to calculate the accuracy of the model.

```
trainIndex <- createDataPartition(bin_class_df$label, p = .67,list = FALSE,times = 1)

bin_class_train <- bin_class_df[trainIndex,]
bin_class_test <- bin_class_df[-trainIndex,]

bin_class_glm <- glm(label ~ x + y , data = bin_class_train, family = binomial(link='logit'))
bin_class_test$model_prob <- predict(bin_class_glm, bin_class_test, type = "response")
```

Summary of model

```
summary(bin_class_glm)
```

```
##
## Call:
## glm(formula = label ~ x + y, family = binomial(link = "logit"),
##     data = bin_class_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3665  -1.1867  -0.9835   1.1536   1.3074
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.413060   0.144422   2.860  0.00424 **
## x           -0.003889   0.002210  -1.760  0.07844 .
## y           -0.005503   0.002242  -2.454  0.01413 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1391.8  on 1003  degrees of freedom
## Residual deviance: 1380.6  on 1001  degrees of freedom
## AIC: 1386.6
##
```

```
## Number of Fisher Scoring iterations: 4
```

Accuracy of the logistic regression classifier

```
fitted.results <- predict(bin_class_glm,newdata=bin_class_test,type='response')
fitted.results <- ifelse(fitted.results > 0.5,1,0)
misClasificError <- mean(fitted.results != bin_class_test$label)
print(paste('Accuracy',1-misClasificError))
```

```
## [1] "Accuracy 0.51417004048583"
```

Accuracy of the logistic regression classifier is 0.53.