****

**ABHINAV VIJAYAKUMAR**

**19BCE1311**

**CSE3506 – ESSENTIALS OF DATA ANALYTICS LAB-5**

**DR. LAKSHMI PATHI JAKKAMPUTI (L21 + L22)**

**--------------------------------------------------------------------------------------------**

**Tasks for Week-5: Logistic Regression**

Understand the following operations/functions on to perform logistic Regression and perform similar operations on ‘Social\_Network\_Ads’ dataset based on given instructions.

**Aim**: To perform logistic Regression and perform similar operations on the ‘Social\_Network\_Ads’ dataset.

**Algorithm:**

**1.** Import the dataset and load the caTools library.

**2.** Split the data using split function into test and train data in a ratio=0.8.

**3.** Convert the purchased and Gender variable to categorical variable using as.factor.

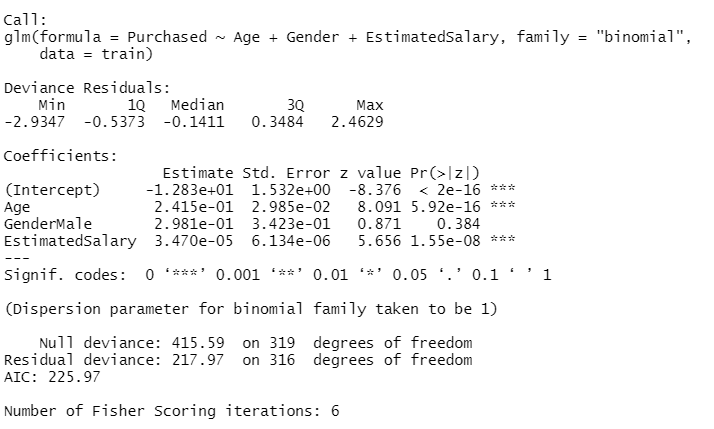
**4.** Apply the generalized linear model using glm command for the dependent and independent variables and print the summary.

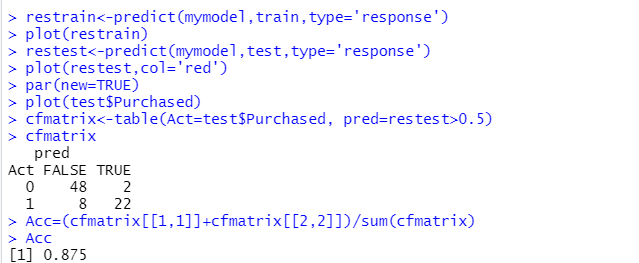
**5.** Using the trained model, predict the output for the test data and observe the accuracy and plot the graphs.

**6.** Generate and Display the confusion matrix.

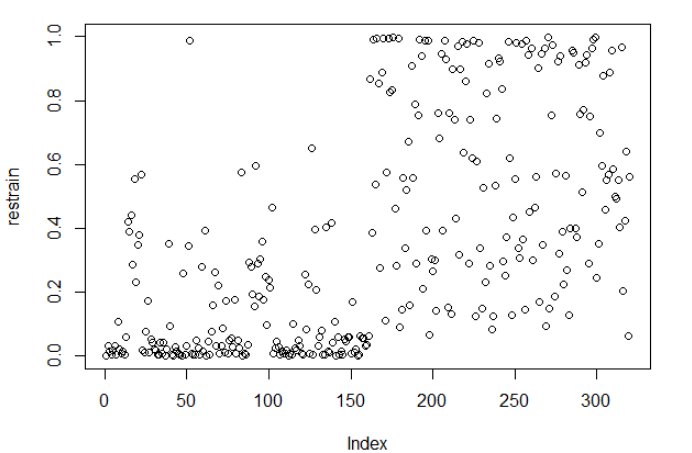
**Statistics:**

**1. Summary of Applied Model:**

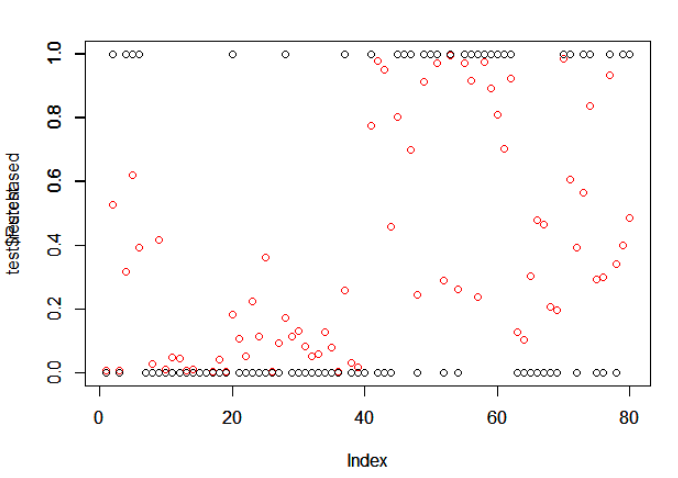
****

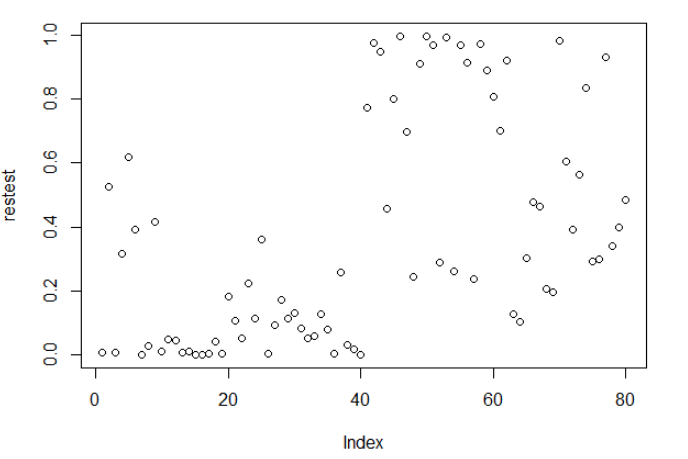
****

**2. Graph of Predicted train data:**

****

**3. Graph of Predicted test**

****

****

**Inference:**

**1.** As seen in the summary of glm model, the profit value (Pr(>|z|)) is less than 0.05 for all variables apart from gender male. Hence, all but one are accepted in the model**.**

**2.** The accuracy of the trained model is observed to be 0.875 i.e., 87.5%.

**3.** There are a total of 80 objectives as seen from the confusion matrix.

**4.** The graph shows the actual and predicted values of the trained model for the train and test data.

**Program:**

rm(list=ls())

setwd("C:/Users/Abhinav Vijayakumar/Desktop/VIT Academics/Sem 6/EDA/LAB/LAB 5")

mydata<-read.csv("Social\_Network\_Ads.csv")

library(caTools)

splitd<-sample.split(mydata,SplitRatio = 0.8)

train=subset(mydata,splitd=="TRUE")

test=subset(mydata,splitd=="FALSE")

train

mydata$Gender<-as.factor(mydata$Gender)

mydata$Purchased<-as.factor(mydata$Purchased)

mymodel <- glm(Purchased ~ Age+Gender+EstimatedSalary, data=train,

family='binomial')

summary(mymodel)

restrain<-predict(mymodel,train,type='response')

plot(restrain)

restest<-predict(mymodel,test,type='response')

plot(restest,col='red')

par(new=TRUE)

plot(test$Purchased)

cfmatrix<-table(Act=test$Purchased, pred=restest>0.5)

cfmatrix

Acc=(cfmatrix[[1,1]]+cfmatrix[[2,2]])/sum(cfmatrix)

Acc

plot(restest)