

Name: Vikram Surya R

Reg No: 19BCE1490

Essentials of Data Analytics

Tasks for Week-5: Logistic Regression.

Aim:

To understand the following operations/functions on to perform logistic Regression and perform similar operations on 'Social_Network_Ads' dataset based on given instructions.

Algorithm:

1. Load the dataset and caTools library.
2. Make sure your data meet the assumptions and has non null values
3. Convert the non-numeric values to numeric using the as.factor command for further calculations
4. Build the logistic regression model for the variable Purchased, by using the rest values as independent variables and keep the family to binomial as we are predicting in a yes/no manner.
5. Predict the responses of each user, and build a confusion matrix
6. Calculate the accuracy of the proposed model

Inference:

Since the accuracy of the model is 0.825 it is accepted.

```
> summary(mymodel)
```

Call:
glm(formula = Purchased ~ Age + Gender + EstimatedSalary, family = "binomial",
data = train)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.8432	-0.5390	-0.1462	0.3890	2.3655

Coefficients:

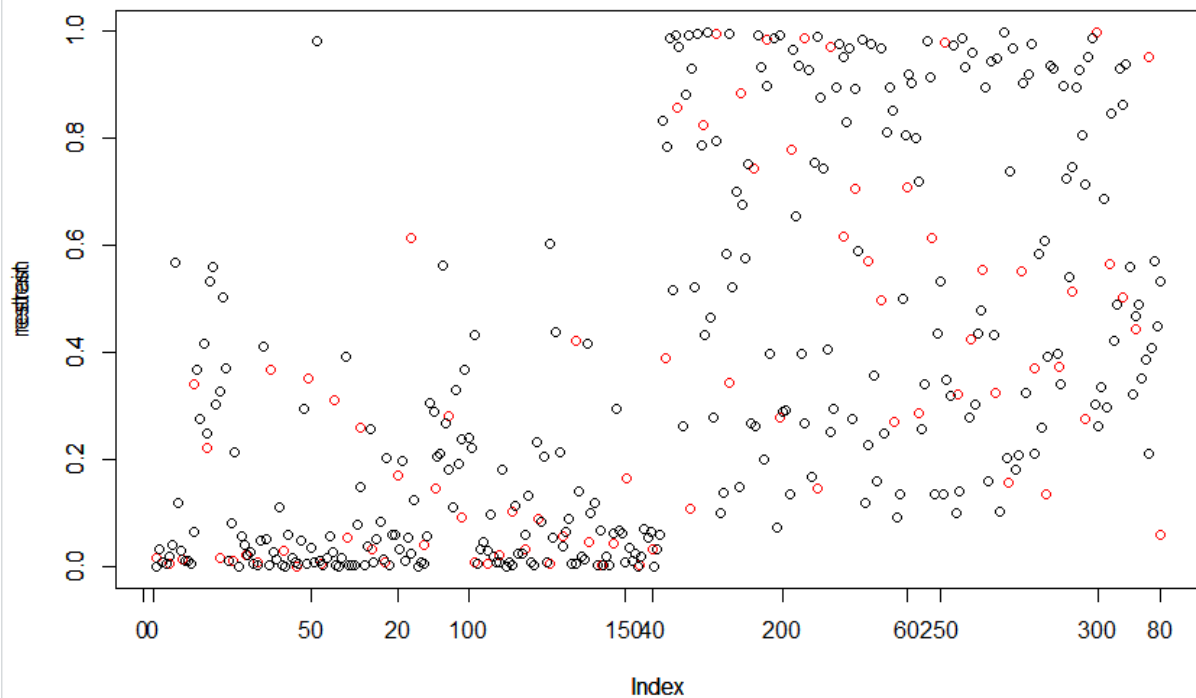
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.217e+01	1.414e+00	-8.604	< 2e-16 ***
Age	2.254e-01	2.744e-02	8.215	< 2e-16 ***
GenderMale	1.599e-01	3.395e-01	0.471	0.638
EstimatedSalary	3.485e-05	6.099e-06	5.714	1.1e-08 ***

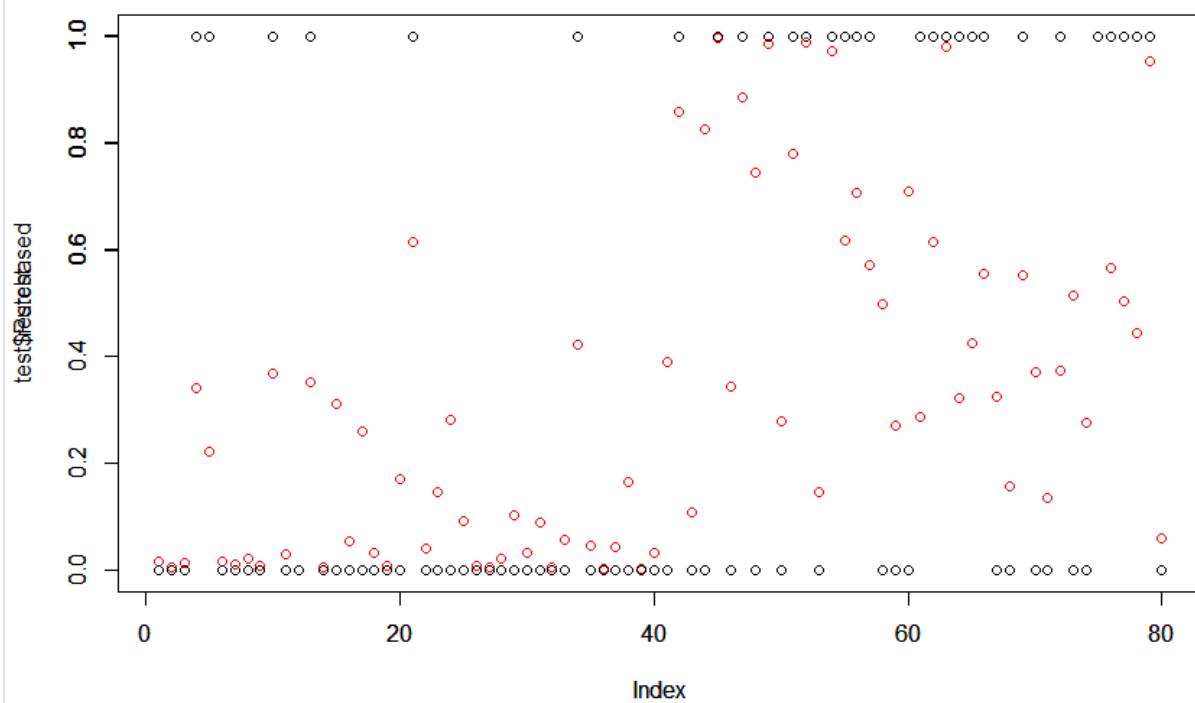
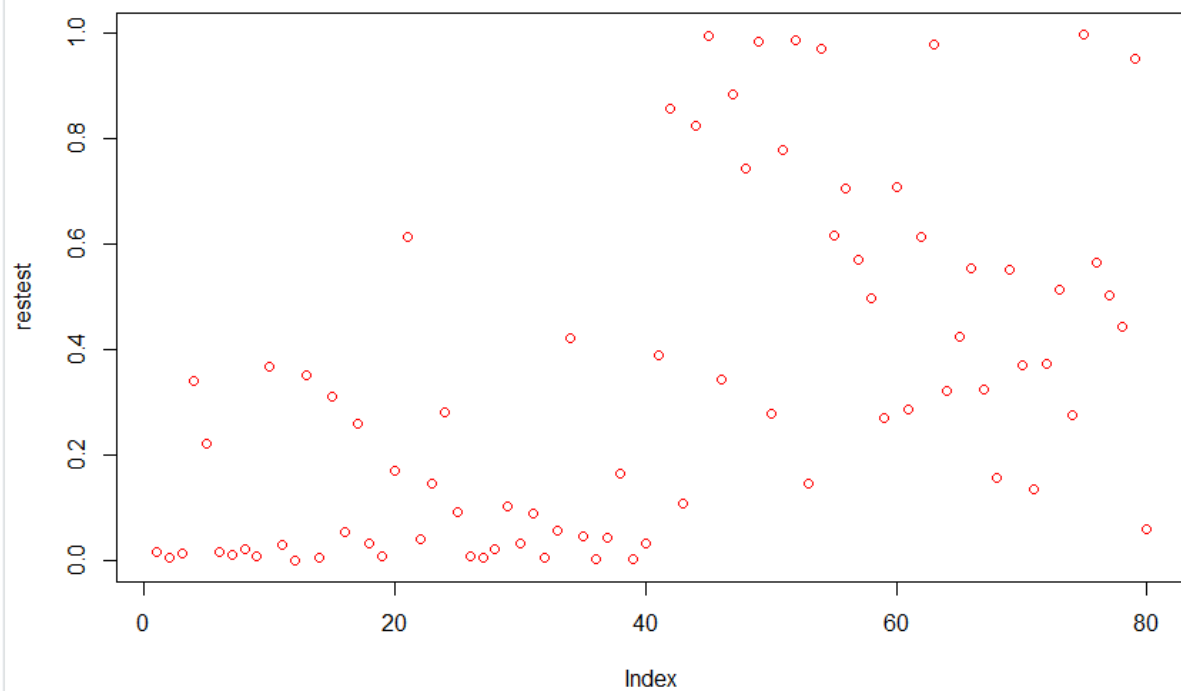
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 416.79 on 319 degrees of freedom
Residual deviance: 221.37 on 316 degrees of freedom
AIC: 229.37

Number of Fisher Scoring iterations: 6

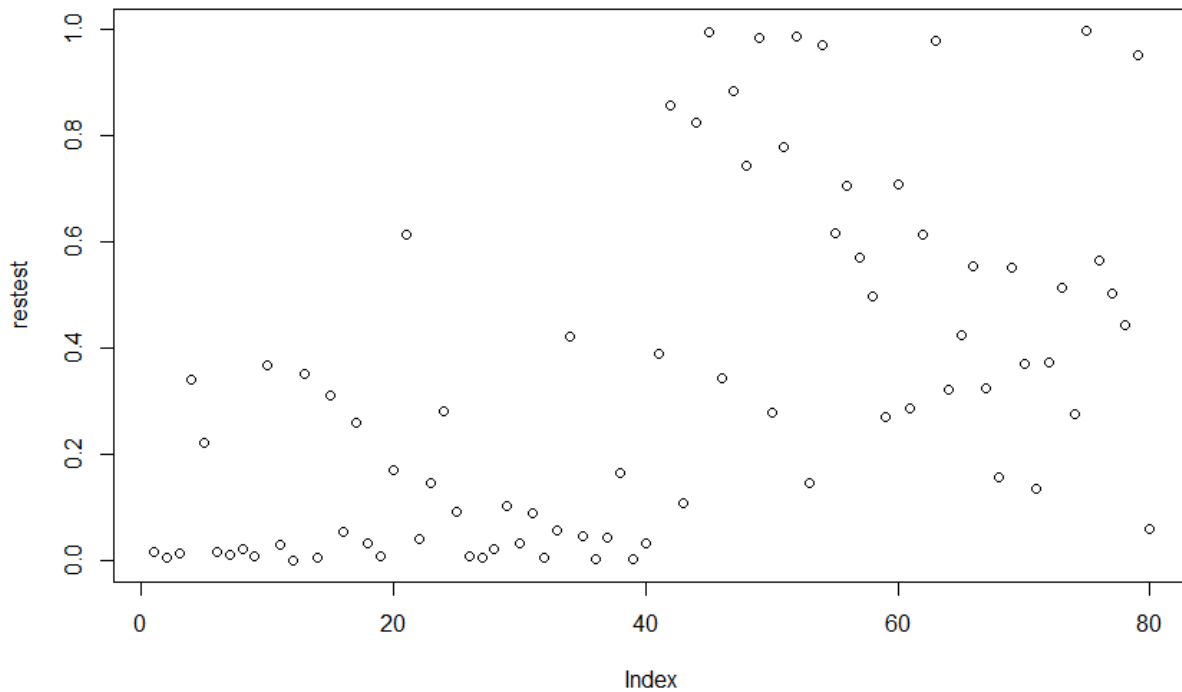




```

> cfmatrix<-table(Act=test$Purchased, pred=retest>0.5)
> cfmatrix
      pred
Act FALSE TRUE
0      47    4
1      10   19
> Acc=(cfmatrix[[1,1]]+cfmatrix[[2,2]])/sum(cfmatrix)
> Acc
[1] 0.825

```



CODE:

```

rm(list=ls())

setwd("C:/Users/VIKRAM SURYA/Desktop/EDA_LAB")

mydata<-read.csv("Social_Network_Ads.csv")

install.packages('caTools')

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)
```

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

```
plot(retest,col='red')
```

```
par(new=TRUE)
```

```
plot(test$Purchased)
```

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

```
cfmatrix
```

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

```
Acc
```

```
plot(retest)
```