NAME: Penugonda Abhignanendra   **REGNO**:19BCE1677

FACULTY: Dr. Lakshmi Pathi Jakkamputi   **SLOT**: L21, L22

Essential of Data Analytics
Tasks for Week-5: Logistic Regression

Aim: Understand the following operations/functions on to perform logistic Regression and perform similar operations on 'Social_Network_Ads' dataset based on given instructions.

Algorithm:

- the path for data set is set using setwd function.

- Read the dataset Social_Network_Ads.csv using read.csv()

- The char values or the binary data which in integer should be converted into only binary value like levels. To convert them into levels we use factor() command which converts the particular column into different levels according to the categories in the data. So, first we convert gender column.

- Next purchase column is converted using factor().

- We use general linear model to do logistic regression and family should be binomial. The command is as glm(purchase~age+gender+estimatedsalary). Here the predicting column is written on the left side of ~ and all the independent columns on right side.

- To find the summary of the glm we use Summary().

- To find the probability of the each response we use predict() and type="response" to find the probability.

- We use table() for finding the confusion matrix whose prediction is greater than 0.5.

- We find the accuracy of the model by add the few values from confusion matrix those are first row first column and second row second column and dividing it with the total values.

Statistic:

	Estimated	Std. Error	Z value	Pr(>|z|)
(Intercept)	-1.278e+01	1.359e+00	-9.405	< 2e-16 ***
Age	2.370e-01	2.638e-02	8.984	< 2e-16 ***
GenderMale	3.338e-01	3.052e-01	1.094	0.274
EstimatedSalary	3.644e-05	5.473e-06	6.659	2.77e-11 ***

Essential of Data Analytics
Tasks for Week-5: Logistic Regression

Inference:

The accuracy of the model is 85% so we can say that the model is best fit model.

Result:

Confusion matrix:

pred

Act	FALSE	TRUE
0	273	20
1	39	104

Accuracy:

0.8525

Code:

```
setwd("C:/Abhi notes/class3-2/eda/lab")

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

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

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

mymodel <- glm(Purchased ~ Age+Gender+EstimatedSalary, data=mydata,
family='binomial')

summary(mymodel)

res<-predict(mymodel,mydata,type='response')

res

cfmatrix<-table(Act=mydata$Purchased, pred=res>0.5)

cfmatrix

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

Acc
```

NAME: Penugonda Abhignanendra   **REGNO**:19BCE1677

FACULTY: Dr. Lakshmi Pathi Jakkamputi   **SLOT**: L21, L22

Essential of Data Analytics
Tasks for Week-5: Logistic Regression