Ex.No.4 Implement Linear and Logistic Regression

Aim
To implement linear and logistic regression for Sales dataset.

Pre-lab Discussion Theory
Linear Regression:

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.
It consists of 3 stages –
(1) analyzing the correlation and directionality of the data
(2) estimating the model, i.e., fitting the line
(3) evaluating the validity and usefulness of the model.
The linear regression model provides a sloped straight line representing the relationship between the variables. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Types of Linear Regression

Linear regression can be further divided into two types of the algorithm:
SimpleLinearRegression
If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

Multiple Linear regression

If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

The main function to calculate values of coefficients Initialize the parameters.
Initialize the parameters.
Predict the value of a dependent variable by given an independent variable.
Calculate the error in prediction for all data points.
Calculate partial derivative w.r.t a0 and a1.
Calculate the cost for each number and add them.
Update the values of a0 and a1.

Logistic Regression:

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.

Type of Logistic Regression:

On the basis of the categories, Logistic Regression can be classified into three types:

Binomial: In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.

Multinomial: In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"

Ordinal: In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

Applications of Logistic Regression

1. Predicting a probability of a person having a heart attack
2. Predicting a customer’s propensity to purchase a product or halt a subscription.
3. Predicting the probability of failure of a given process or product.

Steps in Logistic Regression:
1. Data Pre-processing step
2. Fitting Logistic Regression to the Training set
3. Predicting the test result
4. Test accuracy of the result(Creation of Confusion matrix)
5. Visualizing the test set result.

PROGRAM:

******SIMPLE LINEAR REGRESSION******

```
dataset = read.csv("data-marketing-budget-12mo.csv", header=T, colClasses = c("numeric",
"numeric", "numeric"))
 head(dataset,5)

//////Simple Regression/////

simple.fit = lm(Sales~Spend,data=dataset)
summary(simple.fit)
```

OUTPUT:
SIMPLE LINEAR REGRESSION:

```
Call:
lm(formula = Sales ~ Spend, data = dataset)

Residuals:
   Min     1Q Median     3Q    Max
 -3385  -2097    258   1726   3034

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1383.4714  1255.2404   1.102    0.296
Spend         10.6222     0.1625  65.378 1.71e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2313 on 10 degrees of freedom
Multiple R-squared:  0.9977, Adjusted R-squared:  0.9974
F-statistic:  4274 on 1 and 10 DF,  p-value: 1.707e-14
```

PROGRAM:

******MULTIPLE LINEAR REGRESSION ******

```
multi.fit = lm(Sales~Spend+Month, data=dataset)
 summary(multi.fit)
```

OUTPUT:

MULTIPLE LINEAR REGRESSION

```
Call:
lm(formula = Sales ~ Spend + Month, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1793.73 -1558.33    -1.73  1374.19  1911.58

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -567.6098  1041.8836  -0.545  0.59913
Spend         10.3825     0.1328  78.159 4.65e-14 ***
Month        541.3736   158.1660   3.423  0.00759 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1607 on 9 degrees of freedom
Multiple R-squared:  0.999,	Adjusted R-squared:  0.9988
F-statistic:  4433 on 2 and 9 DF,  p-value: 3.368e-14
```

PROGRAM:

****Logistic Regression ****

#selects some column from mtcars

```
 input<- mtcars [,c("am","cyl","hp","wt")]
 print(head(input))
input<- mtcars [,c("am","cyl","hp","wt")]
am.data =glm(formula = am ~ cyl+hp+wt,data = input,family = binomial)
print(summary(am.data))
```

OUTPUT:

LOGISTIC REGRESSION:

```
> print(head(input))
                  am cyl  hp    wt
Mazda RX4          1   6 110 2.620
Mazda RX4 Wag      1   6 110 2.875
Datsun 710         1   4  93 2.320
Hornet 4 Drive     0   6 110 3.215
Hornet Sportabout  0   8 175 3.440
Valiant            0   6 105 3.460
```

```
Call:
glm(formula = am ~ cyl + hp + wt, family = binomial, data = input)

Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-2.17272  -0.14907  -0.01464   0.14116   1.27641  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) 19.70288    8.11637   2.428   0.0152 *
cyl          0.48760    1.07162   0.455   0.6491  
hp           0.03259    0.01886   1.728   0.0840 .
wt          -9.14947    4.15332  -2.203   0.0276 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43.2297  on 31  degrees of freedom
Residual deviance:  9.8415  on 28  degrees of freedom
AIC: 17.841

Number of Fisher Scoring iterations: 8
```

Result:

Thus, the implementation of linear and logistic regression for sales data set was implemented successfully.