ASSIGNMENT 3

Name: Aakanksha Bhondve

Grno: 21810939

Rollno: 322004

Comp B1

Aim –

Build a Data model in Python for the given dataset and apply Linear Regression/Logistic Regression. Infer the result using accuracy score.

Theory –

Linear regression is used for regression or to predict continuous values whereas logistic regression can be used both in classification and regression problems but it is widely used as a classification algorithm. Regression models aim to project value based on independent features. The main difference that makes both different from each other is when the dependent variables are binary logistic regression is considered and when dependent variables are continuous then linear regression is used.

Linear Regression –

Every person must have come across linear models when they were at school. Mathematics taught us about linear models. It is the same model that is used widely in predictive analysis now. It majorly tells about the relationship between a target that is a dependent variable and predictors using a straight line. Linear regression is basically of two types that are Simple and Multiple Regression.

Logistic Regression –

It is an algorithm that can be used for regression as well as classification tasks but it is widely used for classification tasks. The response variable that is binary belongs either to one of the classes. It is used to predict categorical variables with the help of dependent variables. Consider there are two classes and a new data point is to be checked which class it would belong to. Then algorithms compute probability values that range from 0 and 1.

Some differences between linear and logistic regression –

Linear regression is used for predicting the continuous dependent variable using a given set of independent features whereas Logistic Regression is used to predict the categorical.

Linear regression is used to solve regression problems whereas logistic regression is used to solve classification problems.

In Linear regression, the approach is to find the best fit line to predict the output whereas in the Logistic regression approach is to try for S curved graphs that classify between the two classes that are 0 and 1.

The method for accuracy in linear regression is the least square estimation whereas for logistic regression it is maximum likelihood estimation.

Code and Output –

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model
```

In [3]:
```
df = pd.read_csv('homeprices.csv')
df.head()
```

Out[3]:

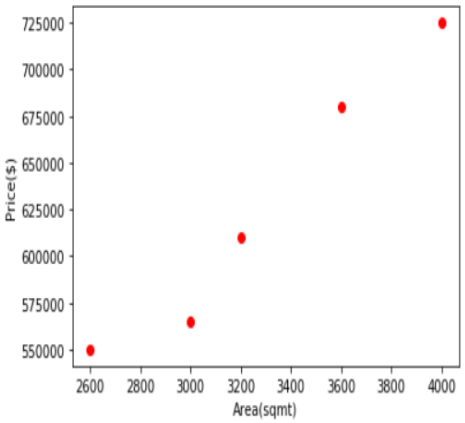
	area	price
0	2600	550000
1	3000	565000
2	3200	610000
3	3600	680000
4	4000	725000

In [4]:
```
x = df['area']
y = df['price']
plt.scatter(x, y, color='red')
plt.xlabel('Area(sqmt)')
plt.ylabel('Price($)')
```

Out[4]: Text(0, 0.5, 'Price($)')

In [4]:
```
x = df['area']
y = df['price']
plt.scatter(x, y, color='red')
plt.xlabel('Area(sqmt)')
plt.ylabel('Price($)')
```

Out[4]: Text(0, 0.5, 'Price($)')



Training

In [12]:
```
reg = linear_model.LinearRegression()  #Creating a linear regression object called 'reg'
reg.fit(df[['area']], df['price'])  # training the model using fit() and passing our data(homeprices) to it.
```

Out[12]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

Predicting

In [7]:
```
reg.predict([[5000]])  #You pass the value of the independent variable in the form of a 2D array.
```

Out[7]: array([859554.79452055])

Accuracy

In [14]:
```
accuracy = reg.score(df[['area']],df['price'])
print(accuracy*100,'%')
```

95.84301138199486 %

In [8]:
```
#  y = mx + c
reg.coef_  #This gives 'm'(coefficient).
```

Out[8]: array([135.78767123])

In [9]:
```
reg.intercept_  #This gives 'c'(y-intercept).
```

Out[9]: 180616.43835616432

In [10]:

```
x = df['area']
y = df['price']
plt.scatter(x, y, color='red')
plt.plot(x, reg.predict(df[['area']]), color='k') #You plot area(x) and the predicted prices(y) for the values of area given
plt.xlabel('Area(sqmt)')
plt.ylabel('Price($)')
```

Out[10]: Text(0, 0.5, 'Price($)')