

*Department of Computer Engineering***Machine Learning Lab BE Computer (Semester-VII)****Experiment No.2 : Multiple Linear Regression**

Aim- To study, understand and implement a multiple linear regression algorithm.

Theory-

Multiple linear regression attempts to model the relationship between **two or more features** and a response by fitting a linear equation to the observed data. Clearly, it is nothing but an extension of simple linear regression.

It has a dataset with **p** features(or independent variables) and one response(or dependent variable). Also, the dataset contains **n** rows/observations.

It is to be observed on any dataset having more than one input parameter to predict one output variable which is with continuous value.

Code –

```
▶ from google.colab import files
  a = files.upload()

Choose Files cars.csv
• cars.csv(text/csv) - 922 bytes, last modified: 7/30/2022 - 100% done
Saving cars.csv to cars.csv
```

```
[2] import pandas
    a = pandas.read_csv("cars.csv")
    print(a)
```

```

x =a[['Weight' , 'Volume']]
y =a['CO2']
from sklearn import linear_model,metrics
r = linear_model.LinearRegression()
r.fit(x,y)
print('coef:',r.coef_)
k = r.predict([[2000,174611]])
print(k)

```

Results-

✓ 42s

```

from google.colab import files
a = files.upload()

```

📁 Choose Files cars.csv

- **cars.csv**(text/csv) - 922 bytes, last modified: 7/30/2022 - 100% done

Saving cars.csv to cars.csv

✓ 0s

```

[2] import pandas
a = pandas.read_csv("cars.csv")
print(a)

```

	Car	Model	Volume	Weight	CO2
0	Toyoty	Aygo	1000	790	99
1	Mitsubishi	Space Star	1200	1160	95
2	Skoda	Citigo	1000	929	95
3	Fiat	500	900	865	90
4	Mini	Cooper	1500	1140	105
5	VW	Up!	1000	929	105
6	Skoda	Fabia	1400	1109	90
7	Mercedes	A-Class	1500	1365	92
8	Ford	Fiesta	1500	1112	98
9	Audi	A1	1600	1150	99
10	Hyundai	I20	1100	980	99
11	Suzuki	Swift	1300	990	101

8	Ford	Fiesta	1500	1112	98
9	Audi	A1	1600	1150	99
10	Hyundai	I20	1100	980	99
11	Suzuki	Swift	1300	990	101
12	Ford	Fiesta	1000	1112	99
13	Honda	Civic	1600	1252	94
14	Hundai	I30	1600	1326	97
15	Opel	Astra	1600	1330	97
16	BMW	1	1600	1365	99
17	Mazda	3	2200	1280	104
18	Skoda	Rapid	1600	1119	104
19	Ford	Focus	2000	1328	105
20	Ford	Mondeo	1600	1584	94
21	Opel	Insignia	2000	1428	99
22	Mercedes	C-Class	2100	1365	99
23	Skoda	Octavia	1600	1415	99
24	Volvo	S60	2000	1415	99
25	Mercedes	CLA	1500	1465	102
26	Audi	A4	2000	1490	104
27	Audi	A6	2000	1725	114
28	Volvo	V70	1600	1523	109
29	BMW	5	2000	1705	114
30	Mercedes	E-Class	2100	1605	115
31	Volvo	XC70	2000	1746	117
32	Ford	B-Max	1600	1235	104
33	BMW	216	1600	1390	108
34	Opel	Zafira	1600	1405	109

```

25 Mercedes CLA 1500 1465 102
26 Audi A4 2000 1490 104
27 Audi A6 2000 1725 114
28 Volvo V70 1600 1523 109
29 BMW 5 2000 1705 114
30 Mercedes E-Class 2100 1605 115
31 Volvo XC70 2000 1746 117
32 Ford B-Max 1600 1235 104
33 BMW 216 1600 1390 108
34 Opel Zafira 1600 1405 109
35 Mercedes SLK 2500 1395 120

```

```

[ ] x =a[['Weight' , 'Volume']]
y =a['CO2']
from sklearn import linear_model,metrics
r = linear_model.LinearRegression()
r.fit(x,y)
print('coef:',r.coef_)
k = r.predict([[2000,174611]])
print(k)

```

```

coef: [0.00755095 0.00780526]
[1457.68043601]

```

```

/usr/local/lib/python3.7/dist-packages/sklearn/base.py:451: UserWarning: X does not have valid feature names, but LinearRegression was fitted
  "X does not have valid feature names, but"

```

Discussion-

Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable and to predict the outcome of a response variable.

Applications-

- Trend lines: A trend line represents the variation in quantitative data with the passage of time (like GDP, oil prices, etc.). These trends usually follow a linear relationship. Hence, linear regression can be applied to predict future values. However, this method suffers from a lack of scientific validity in cases where other potential changes can affect the data.
- Economics: Linear regression is the predominant empirical tool in economics. For example, it is used to predict consumer spending, fixed investment spending, inventory investment, purchases of a country's exports, spending on imports, the demand to hold liquid assets, labor demand, and labor supply.
- Finance: The capital price asset model uses linear regression to analyze and quantify the systematic risks of an investment.
- 4. Biology: Linear regression is used to model causal relationships between parameters in biological systems.

Conclusion-

The concept of a multiple linear regression is studied and implemented using python built-in functions with standard dataset.

References-

1. http://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html
2. <http://www.statisticssolutions.com/assumptions-of-linear-regression/>

