Machine Learning Experiment 7

Guneet Singh Sethi

02913202717

CSE-1

EXPERIMENT -7

Problem Statement-

Develop a machine learning method to predict how people will rate movies, books etc.

Algorithm

2Rs. Algorithm

Linear regression is perhaps one of the most well known and well understood algorithms in statistics and machine learning.

In this post you will discover the linear regression algorithm, how it works and how you can best use it in on your machine learning projects. In this post you will learn:

Why linear regression belongs to both statistics and machine learning.

The many names by which linear regression is known.

The representation and learning algorithms used to create a linear regression model.

How to best prepare your data when modeling using linear regression.

You do not need to know any statistics or linear algebra to understand linear regression. This is a gentle high-level introduction to the technique to give you enough background to be able to use it effectively on your own problems.

Program Screenshots

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
df=pd.read_csv('movies.csv')
```

In [3]:
```
df.corr()
```

Out[3]:

	Audience score %	Profitability	Rotten Tomatoes %	Year
Audience score %	1.000000	0.066278	0.608333	-0.229926
Profitability	0.066278	1.000000	0.042428	-0.182130
Rotten Tomatoes %	0.608333	0.042428	1.000000	-0.093111
Year	-0.229926	-0.182130	-0.093111	1.000000

In [4]:
```
dt= pd.DataFrame(index=range(0,len(df)), columns=['Audience score %','Rotten Tomatoes %'])
for i in range(0,len(dt)):
  dt['Audience score %'][i]= df['Audience score %'][i]
  dt['Rotten Tomatoes %'][i]= df['Rotten Tomatoes %'] [i]
```

In [5]:
```
dt
```

Out[5]:

	Audience score %	Rotten Tomatoes %
0	70	64
1	52	68
2	35	43
3	44	15
4	72	28
...	...	...
72	84	54
73	64	89
74	89	79
75	71	40
76	81	87

77 rows × 2 columns

In [6]:
```
x = dt['Audience score %']
y= dt['Rotten Tomatoes %']
x=np.array(x)
x=x.reshape(-1,1)
y=np.array(y)
y=y.reshape(-1,1)
```

In [7]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
x_train,x_test,y_train,y_test= train_test_split(x,y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(x_train,y_train)
```

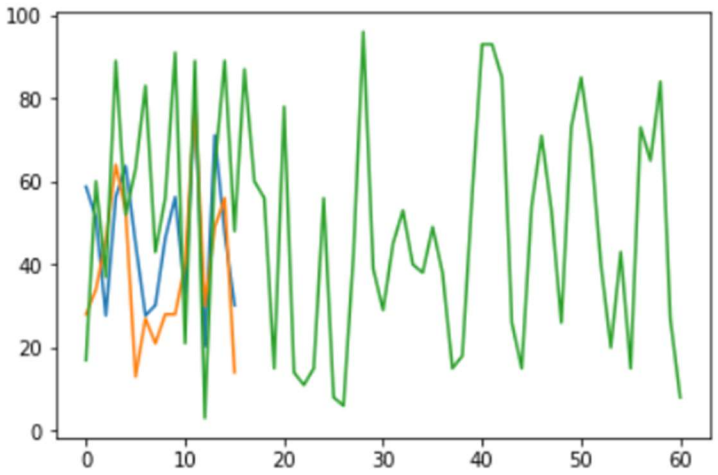
Out[7]:
```
LinearRegression()
```

In [8]:
```
pred = model.predict(x_test)
pred
```

Out[8]:
```
array([[58.70663406],
       [51.26782559],
       [27.71159877],
       [56.22703123],
       [63.6658397 ],
       [45.06881853],
       [27.71159877],
       [30.1912016 ],
       [46.30861994],
       [56.22703123],
       [32.67080442],
       [79.78325805],
       [20.27279031],
       [71.10464817],
       [47.54842135],
       [30.1912016 ]])
```

```
In [9]: plt.plot(pred)
        plt.plot(y_test)
        plt.plot(y_train)
```

```
Out[9]: [<matplotlib.lines.Line2D at 0x1c7ff7020d0>]
```



Github Link-

https://github.com/SethiGuneet/ML-Lab-Work/blob/main/Machine%20Learning%20Experiment%207.ipynb