

Experiment – 3

Student Name: Vivek Kumar

UID: 21BCS8129

Branch: BE-CSE(LEET)

Section/Group: WM-20BCS-616/A

Semester: 5th

Date of Performance: 31/08/2022

Subject Name: Machine Learning Lab

Subject Code: 20CSP-317

1. Aim/Overview of the practical:

Implement Linear Regression on any data set.

2. Task to be done/ Which logistics used:

Implement Linear Regression on any data set using sklearn.

3. Algorithm/Flowchart (For programming-based labs):

4. Steps for experiment/practical/Code:

```
from google.colab import drive
drive.mount('/content/drive')

import pandas as pd
import numpy as np

data=pd.read_csv('/content/drive/MyDrive/Data/Students_data.csv')

data.head()

data.isnull()

data.isnull().sum()

import matplotlib.pyplot as plt
import seaborn as sn
import sklearn

data.plot.scatter('race', 'Algebra')

x=data[['GPA', 'Algebra', 'Calculus1', 'Calculus2', 'Statistics', 'Probability',
'Measure', 'Functional_analysis', 'from4', 'y']]
y=data['race']

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=0.4)

lm=LinearRegression()
lm.fit(x_train,y_train)

predictions=lm.predict(x_test)

plt.scatter(y_test,predictions)
plt.xlabel('y test')
plt.ylabel('predicted y')

from sklearn import metrics

print('MAE:',metrics.mean_absolute_error(y_test,predictions))
print('MSE:',metrics.mean_squared_error(y_test,predictions))
print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,predictions)))

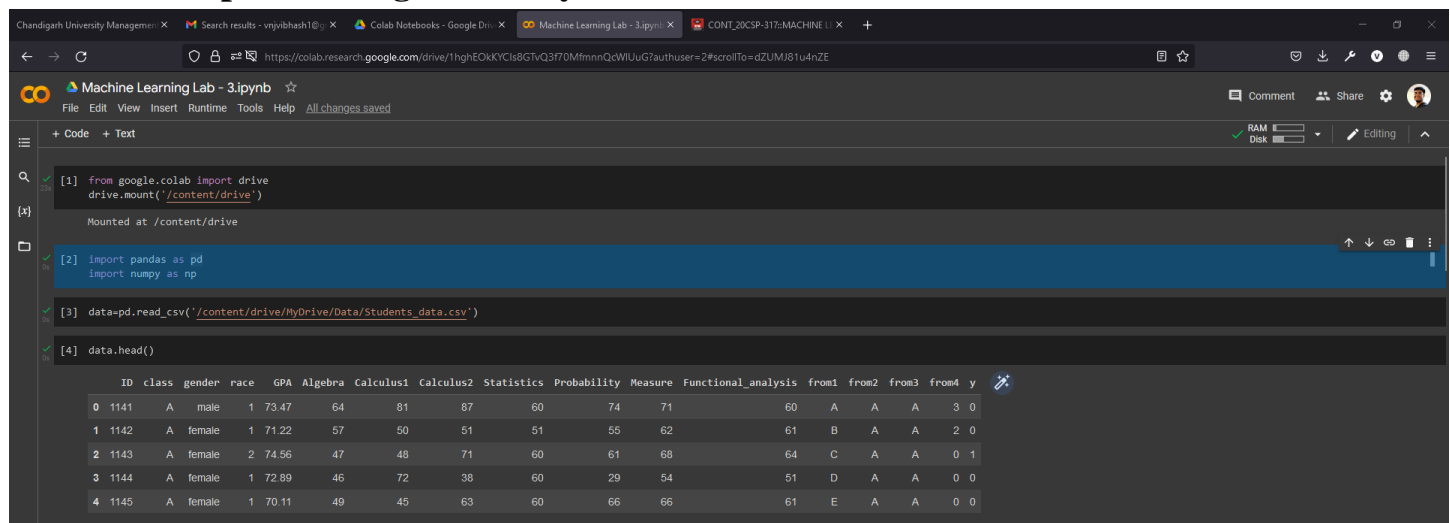
sn.distplot((y_test-predictions),bins=50)

coefficient=pd.DataFrame(lm.coef_,x.columns)
coefficient.columns=['coefficient']
coefficient
```

5. Observations/Discussions/ Complexity Analysis:

In this I have implemented simple Liner regression based on the data by splitting into two parts such as training and testing as well as I have displayed the plotted the graph and displayed coefficient of all the related columns.

6. Result/Output/Writing Summary:



The screenshot shows a Google Colab notebook titled "Machine Learning Lab - 3.ipynb". The code executed includes mounting the drive, importing pandas and numpy, reading a CSV file, and displaying the head of the data. The output shows the data is mounted at /content/drive and the head of the data is displayed as follows:

	ID	class	gender	race	GPA	Algebra	Calculus1	Calculus2	Statistics	Probability	Measure	Functional_analysis	from1	from2	from3	from4	y
0	1141	A	male	1	73.47	64	81	87	60	74	71	60	A	A	A	3	0
1	1142	A	female	1	71.22	57	50	51	51	55	62	61	B	A	A	2	0
2	1143	A	female	2	74.56	47	48	71	60	61	68	64	C	A	A	0	1
3	1144	A	female	1	72.89	46	72	38	60	29	54	51	D	A	A	0	0
4	1145	A	female	1	70.11	49	45	63	60	66	66	61	E	A	A	0	0

Machine Learning Lab - 3.ipynb

data.isnull()

	ID	class	gender	race	GPA	Algebra	Calculus1	Calculus2	Statistics	Probability	Measure	Functional_analysis	from1	from2	from3	from4	y
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...
100	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
101	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
102	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
103	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
104	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

105 rows x 17 columns

[6] data.isnull().sum()

```

ID      0
class   0
gender   0
race     0
GPA      0
Algebra  0
Calculus1 0
Calculus2 0
Statistics 0
Probability 0
Measure  0
Functional_analysis 0
from1    0
from2    0
from3    0
from4    0
y        0
  
```

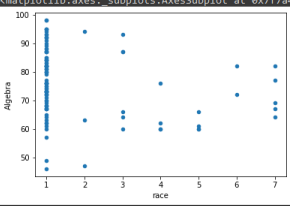
completed at 10:52 AM

Machine Learning Lab - 3.ipynb

dtype: int64

[7] import matplotlib.pyplot as plt
import seaborn as sns
import sklearn

[8] data.plot.scatter('race', 'Algebra')



[9] x=data[['GPA', 'Algebra', 'Calculus1', 'Calculus2', 'Statistics', 'Probability', 'Measure', 'Functional_analysis', 'from1', 'y']]
y=data['race']

[10] from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

[11] x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.4)

[12] lm=LinearRegression()
lm.fit(x_train,y_train)

LinearRegression()

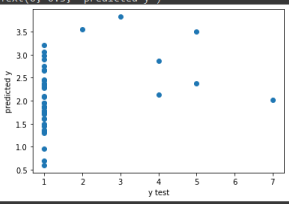
completed at 10:52 AM

Machine Learning Lab - 3.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

[14] `plt.scatter(y_test,predictions)`
`plt.xlabel('y_test')`
`plt.ylabel('predicted y')`



Text(0, 0.5, 'predicted y')

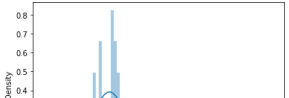
from sklearn import metrics

```
print('MAE:', metrics.mean_absolute_error(y_test, predictions))
print('MSE:', metrics.mean_squared_error(y_test, predictions))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, predictions)))
```

MAE: 1.1978101930448362
MSE: 2.129168217337046
RMSE: 1.4591669608621604

[16] `sn.distplot((y_test-predictions),bins=50)`

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: 'distplot' is a deprecated function and will be removed in a future version. Please adapt your code to use either 'displot' (a figure-level function) or 'distplot' (an axes-level function).



[17] `coefficient=pd.DataFrame(lm.coef_.x.columns)`
`coefficient.columns=['coefficient']`
`coefficient`

	coefficient
GPA	-0.107618
Algebra	-0.035548
Calculus1	0.021286
Calculus2	0.011049
Statistics	-0.021190
Probability	0.018117
Measure	-0.002306
Functional_analysis	0.019356
from4	-0.426621
y	0.148745

Learning outcomes (What I have learnt):

1. Data Splitting
2. Linear regression concept
3. Plotting the prediction.

Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.			
2.			
3.			