Page No.
Date. 30/03/2022

Assignment-05 [Regression]

In[1]: from sklearn.datasets import load_boston

Loading in built dataset 'Boston'.

In[2]: boston = load_boston()

The Boston housing prices dataset ----

---- warnings.warn(msg, category = FutureWarning)

In[3]: print(boston.DESCR)

Boston house prices dataset

- LSTAT % lower status of the population.

In[4]: import pandas as pd

In[5]: data = pd.DataFrame(boston.data, columns = boston.feature_names)
print(boston.keys())

dict_keys(['data', 'target', 'feature_names', 'DESCR', 'filename', 'data_module'])

In[6]: data['MEDV'] = pd.DataFrame(boston.target)

In[7]: data.head()

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PIRATIO
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	24.2.0	178
...	---	---	---	---	---	---	---	---	---	---	---
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7

Teacher's Signature .....

Page No.
Date.

In[8]: `pd.DataFrame(data.corr().round(2))`

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
CRM	1.00	-0.20	0.41	-0.06	0.42	-0.22	0.35	-0.38	0.63	0.58	0.29	-0.39	0.46
ZN	-0.20	1.00	-0.53	-0.04	-0.52	0.31	-0.57	0.66	-0.31	-0.31	-039	0.18	-0.41
...													
MEDV	-0.39	0.36	-0.48	0.18	-0.43	0.70	-0.38	0.25	-0.38	-0.47	-0.51	0.33	-0.74

In[9]:
```
x = data['RM']
y = data['MEDV']
```

In[10]: `pd.DataFrame([x,y]).transpose().head()`

	RM	MEDV
0	6.575	24.0
1	6.421	21.6
2	7.185	34.7
3	6.998	33.4
4	7.147	36.2

In[11]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3)
print(type(x_train))
print(type(y_train))
```

```
<class 'pandas.core.series.Series'>
<class 'pandas.core.series.Series'>
```

Teacher's Signature .....

In [12]:
```
x_train = pd.DataFrame(x_train)
y_train = pd.DataFrame(x_test)
```

In [13]:
```
from sklearn.linear_model import LinearRegression.
model = LinearRegression()
model.fit(x_train, y_train)
y_pred =( model.predict(x_test))
```

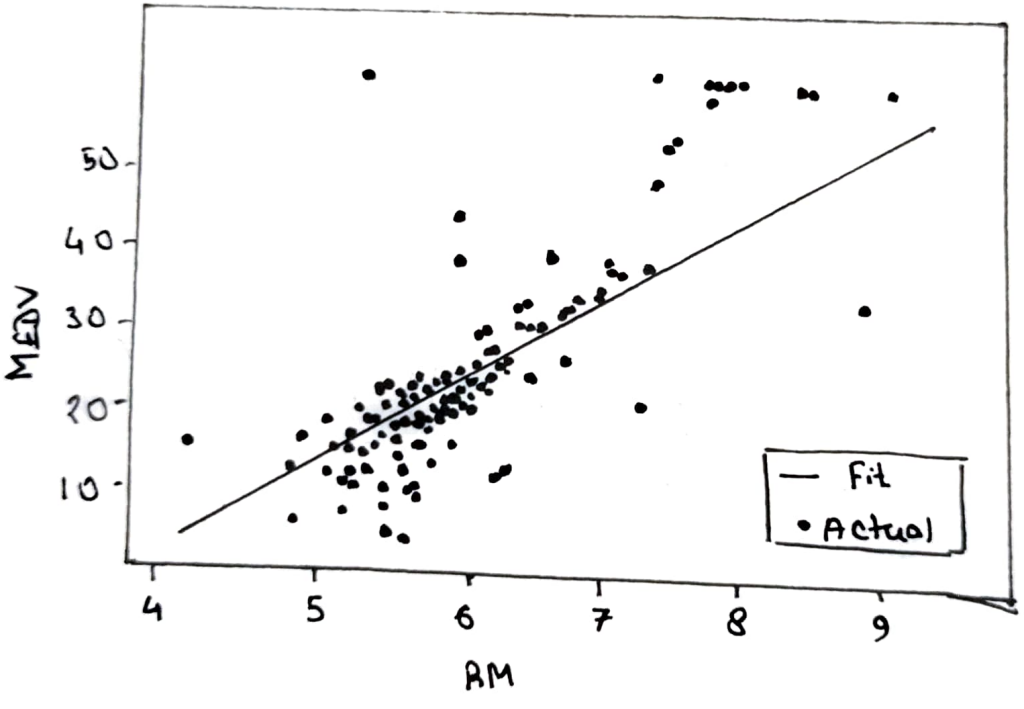
In [14]:
```
from sklearn.metrices import mean_squared_error
from math import sqrt.
print(sqrt(mean_squared_error(y_pred, y_test)))
```

6.078213546201176

In [15]:
```
from matplotlib import pyplot as plt.
% matplotlib inline.
```

In [16]:
```
plt.scatter(x_test, y_test, label = 'Actual')
plt.plot(x_test, y_pred, color='red', label='Fit')
plt.xlabel('RM')
plt.ylabel('MEDV')
plt.legend()
plt.show()
```

Teacher's Signature


MEDV
50
40
30
20
10
4
5
6
7
8
9
RM
Fit
Actual

Date.

In[17]: from sklearn.datasets import load_diabetes
diabetes = load_diabetes()

In[18]: data = pd.DataFrame(diabetes.data, columns=diabetes,feature_na
data['Target'] = pd.DataFrame(diabetes.target)

In[19]: data.head()

	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6	Target
0	0.038076	0.050680	0.061696	0.218	-0.044	0.034	-0.043	-0.002	0.019	-0.0176	151.0
4	0.005383	-0.044642	-0.036	0.021	0.003	0.001	0.008	-0.002	-0.03	-0.04	135.

In[20]: print(data.corr().round(2))

	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6	Target
age	1.00	0.17	0.19	0.34	0.26	0.22	-0.08	0.20	0.27	0.30	0.19
Target	0.19	0.04	0.59	0.44	0.21	0.17	-0.39	0.43	0.57	0.38	1.00