practical-exam-06

May 23, 2023

[ ]:

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

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
```

1 Problem Statement 6

Create a Linear Regression Model using Python to predict home prices using the Boston Housing Dataset. The Boston Housing dataset contains information about various houses in Boston through different parameters. There are 506 samples and 14 feature variables in this dataset.

[ ]:

```
import pandas as pd
```

[160]:

```
data = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/exam_datasets/6.
  ↪housing_data.csv')
```

```
<IPython.core.display.HTML object>
```

[165]:

```
data.dropna()
```

```
<IPython.core.display.HTML object>
```

[165]:

```
         CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD  TAX  \
0     0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900    1  296
1     0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671    2  242
2     0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671    2  242
3     0.03237   0.0   2.18   0.0  0.458  6.998  45.8  6.0622    3  222
5     0.02985   0.0   2.18   0.0  0.458  6.430  58.7  6.0622    3  222
..        …    …     …     …      …      …     …       …  …   …
499   0.17783   0.0   9.69   0.0  0.585  5.569  73.5  2.3999    6  391
500   0.22438   0.0   9.69   0.0  0.585  6.027  79.7  2.4982    6  391
502   0.04527   0.0  11.93   0.0  0.573  6.120  76.7  2.2875    1  273
503   0.06076   0.0  11.93   0.0  0.573  6.976  91.0  2.1675    1  273
504   0.10959   0.0  11.93   0.0  0.573  6.794  89.3  2.3889    1  273

     PTRATIO       B  LSTAT  MEDV
0       15.3  396.90   4.98  24.0
```

```
1       17.8  396.90   9.14  21.6
2       17.8  392.83   4.03  34.7
3       18.7  394.63   2.94  33.4
5       18.7  394.12   5.21  28.7
..       …      …     …    …
499     19.2  395.77  15.10  17.5
500     19.2  396.90  14.33  16.8
502     21.0  396.90   9.08  20.6
503     21.0  396.90   5.64  23.9
504     21.0  393.45   6.48  22.0

[394 rows x 14 columns]
```

[166]:

```python
data.info()
```

```
<IPython.core.display.HTML object>

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 506 entries, 0 to 505
Data columns (total 14 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   CRIM     486 non-null    float64
 1   ZN       486 non-null    float64
 2   INDUS    486 non-null    float64
 3   CHAS     486 non-null    float64
 4   NOX      506 non-null    float64
 5   RM       506 non-null    float64
 6   AGE      486 non-null    float64
 7   DIS      506 non-null    float64
 8   RAD      506 non-null    int64
 9   TAX      506 non-null    int64
 10  PTRATIO  506 non-null    float64
 11  B        506 non-null    float64
 12  LSTAT    486 non-null    float64
 13  MEDV     506 non-null    float64
dtypes: float64(12), int64(2)
memory usage: 55.5 KB
```

[173]:

```python
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score,␣
 ↪recall_score
```

```
<IPython.core.display.HTML object>
```

[180]:

```python
data = data.dropna()
X = data.drop('MEDV', axis=1)
```

```python
y = data['MEDV']


X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
lr = LinearRegression()
lr.fit(X_train, y_train)

y_pred = lr.predict(X_test)
```

```
<IPython.core.display.HTML object>
```

[181]:

```python
r2 = lr.score(X_test, y_test)
print(f'R^2 score: {r2:.2f}')
```

```
<IPython.core.display.HTML object>
```

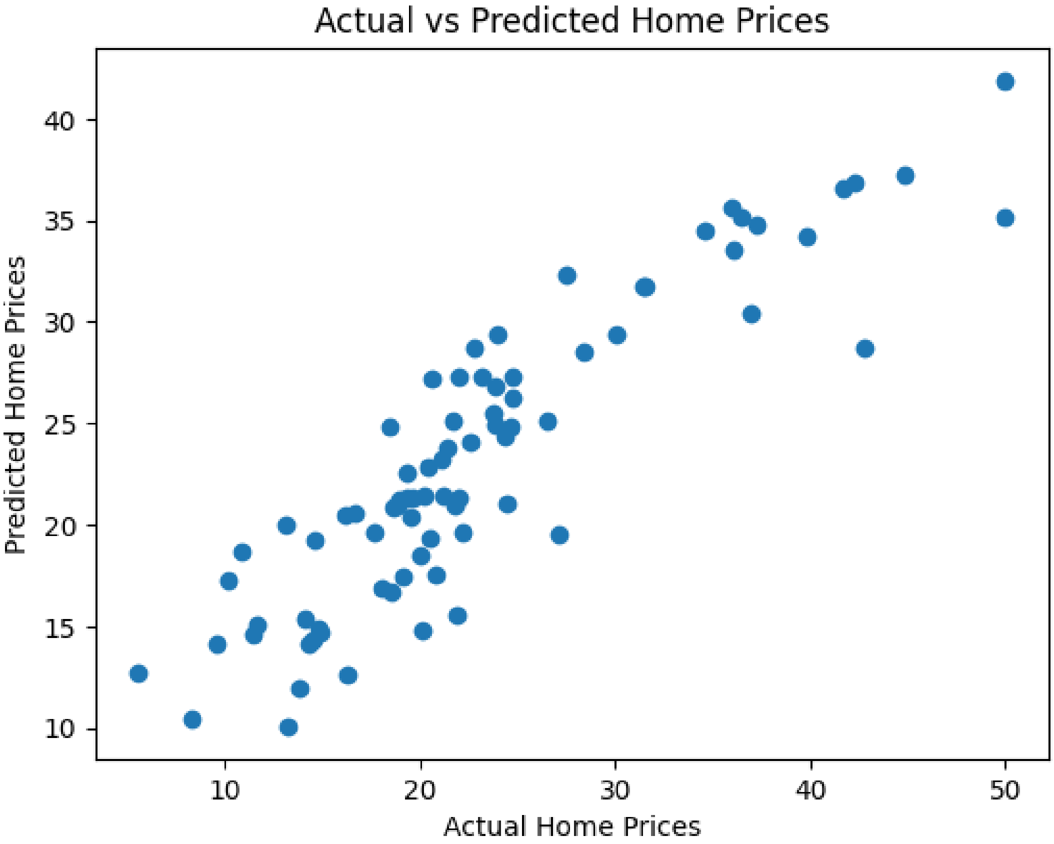
```
R^2 score: 0.79
```

[182]:

```python
import matplotlib.pyplot as plt

# Plot the predicted and actual home prices
plt.scatter(y_test, y_pred)
plt.xlabel('Actual Home Prices')
plt.ylabel('Predicted Home Prices')
plt.title('Actual vs Predicted Home Prices')
plt.show()
```

```
<IPython.core.display.HTML object>
```


Actual vs Predicted Home Prices
Predicted Home Prices
40
35
30
25
20
15
10
10
20
30
40
50
Actual Home Prices