Srushti Ramesh Bhamre BC56

dtype: int64

Linear regression by using Deep Neural network: Implement Boston housing price prediction problem by Linear regression using Deep Neural network. Use Boston House price prediction dataset.

```
In [1]:
             import io
          1
           2 import pandas as pd
           3 import numpy as np
             import matplotlib.pyplot as plt
             import seaborn as sns
          6
             %matplotlib inline
             from sklearn.model selection import train test split
          8 from sklearn.preprocessing import StandardScaler
          9 import tensorflow as tf
         10 from tensorflow import keras
         11 from tensorflow.keras import layers
         12 from sklearn.metrics import mean_absolute_error, r2_score
         13 import warnings
         14 warnings.filterwarnings('ignore')
In [2]:
          1 # Importing DataSet and take a Look at Data
           2 data = pd.read_csv('housing_data - housing_data.csv')
Out[2]:
                CRIM
                       ZN INDUS CHAS
                                        NOX
                                                RM AGE
                                                            DIS RAD TAX PTRATIO
                                                                                        B LSTAT
                                                                                                 MEDV
            0.00632
                      18.0
                             2.31
                                    0.0
                                        0.538 6.575
                                                    65.2 4.0900
                                                                      296
                                                                               15.3 396.90
                                                                                             4.98
                                                                                                   24.0
           1 0.02731
                       0.0
                             7.07
                                    0.0 0.469 6.421
                                                    78.9 4.9671
                                                                   2
                                                                      242
                                                                               17.8 396.90
                                                                                            9.14
                                                                                                   21.6
            2 0.02729
                       0.0
                             7.07
                                    0.0
                                        0.469
                                             7.185
                                                    61.1 4.9671
                                                                   2
                                                                      242
                                                                               17.8 392.83
                                                                                             4.03
                                                                                                   34.7
            3 0.03237
                       0.0
                             2 18
                                                    45.8 6.0622
                                                                   3
                                                                      222
                                                                               18.7 394.63
                                                                                                   33 4
                                    0.0 0.458 6.998
                                                                                            2 94
            4 0.06905
                       0.0
                             2.18
                                        0.458 7.147
                                                    54.2 6.0622
                                                                      222
                                                                               18.7 396.90
                                                                                                   36.2
                                                                                             NaN
          501 0.06263
                       0.0
                            11.93
                                    0.0 0.573 6.593
                                                    69.1 2.4786
                                                                      273
                                                                               21.0 391.99
                                                                                                   22.4
                                                                                            NaN
          502 0.04527
                            11.93
                                                    76.7 2.2875
                                                                               21.0 396.90
                                                                                                   20.6
                       0.0
                                    0.0 0.573 6.120
                                                                      273
                                                                                            9.08
          503 0.06076
                       0.0
                            11.93
                                    0.0 0.573 6.976
                                                    91.0 2.1675
                                                                      273
                                                                               21.0 396.90
                                                                                            5.64
                                                                                                   23.9
                                                                   1 273
          504 0.10959
                       0.0
                            11.93
                                    0.0
                                        0.573 6.794
                                                    89.3 2.3889
                                                                               21.0 393.45
                                                                                             6.48
                                                                                                   22.0
          505 0.04741
                       0.0
                            11.93
                                    0.0 0.573 6.030 NaN 2.5050
                                                                   1 273
                                                                               21.0 396.90
                                                                                            7.88
                                                                                                   11.9
         506 rows × 14 columns
In [3]:
          1 # Handle null values by filling them with the mean of the respective columns
           2 data.fillna(data.mean(), inplace=True)
In [4]:
          1 data.isnull().sum()
Out[4]: CRIM
                     0
         7N
                     0
         INDUS
                     0
         CHAS
                     0
         NOX
                     0
         RM
                     0
         AGE
                     0
                     0
         DIS
         RAD
                     0
         TAX
                     0
         PTRATIO
                     0
         LSTAT
                     0
         MEDV
                     0
```

In [5]: 1 data.describe()

Out[5]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTR#
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000
mean	3.611874	11.211934	11.083992	0.069959	0.554695	6.284634	68.518519	3.795043	9.549407	408.237154	18.455
std	8.545770	22.921051	6.699165	0.250233	0.115878	0.702617	27.439466	2.105710	8.707259	168.537116	2.164
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	1.129600	1.000000	187.000000	12.600
25%	0.083235	0.000000	5.190000	0.000000	0.449000	5.885500	45.925000	2.100175	4.000000	279.000000	17.400
50%	0.290250	0.000000	9.900000	0.000000	0.538000	6.208500	74.450000	3.207450	5.000000	330.000000	19.050
75%	3.611874	11.211934	18.100000	0.000000	0.624000	6.623500	93.575000	5.188425	24.000000	666.000000	20.200
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	12.126500	24.000000	711.000000	22.000

In [6]:

1 data.info() 2 data.shape

<class 'pandas.core.frame.DataFrame'> RangeIndex: 506 entries, 0 to 505 Data columns (total 14 columns):

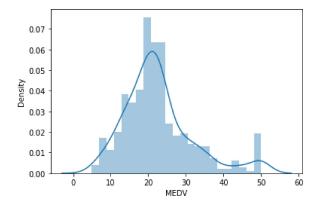
Data	COTUMITS	(LOCAL 14 COLUMNIS	٥).					
#	Column	Non-Null Count	Dtype					
0	CRIM	506 non-null	float64					
1	ZN	506 non-null	float64					
2	INDUS	506 non-null	float64					
3	CHAS	506 non-null	float64					
4	NOX	506 non-null	float64					
5	RM	506 non-null	float64					
6	AGE	506 non-null	float64					
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	В	506 non-null	float64					
12	LSTAT	506 non-null	float64					
13	MEDV	506 non-null	float64					
dtypes: float64(12), int64(2)								

memory usage: 55.5 KB

Out[6]: (506, 14)

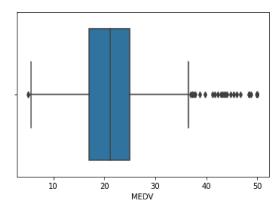
In [7]: 1 import seaborn as sns 2 sns.distplot(data.MEDV)

Out[7]: <AxesSubplot:xlabel='MEDV', ylabel='Density'>



```
In [8]:
         1 sns.boxplot(data.MEDV)
```

Out[8]: <AxesSubplot:xlabel='MEDV'>



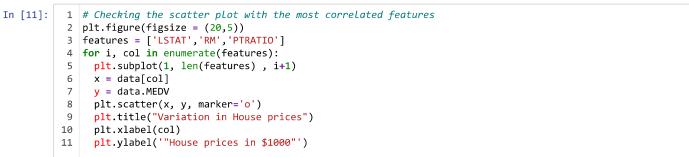
```
In [9]:
         1 correlation = data.corr()
         2 correlation.loc['MEDV']
```

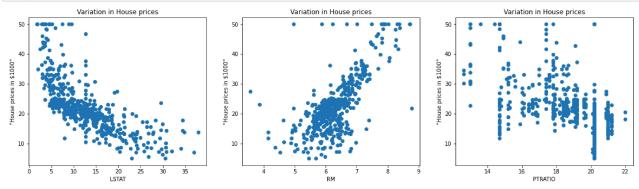
```
Out[9]: CRIM
                  -0.379695
        ZN
                   0.365943
        INDUS
                  -0.478657
        CHAS
                   0.179882
        NOX
                  -0.427321
        RM
                   0.695360
                  -0.380223
        AGE
                   0.249929
        DIS
        RAD
                  -0.381626
        TAX
                  -0.468536
                  -0.507787
        PTRATIO
        В
                   0.333461
        LSTAT
                  -0.721975
        MEDV
                   1.000000
```

Name: MEDV, dtype: float64

Out[10]: <AxesSubplot:>







```
In [12]:
          1 # Splitting the dependent feature and independent feature
          2 #X = data[['LSTAT','RM','PTRATIO']]
          3 X = data.iloc[:,:-1]
          4 y= data.MEDV
In [13]:
         1 import numpy as np
          2 from sklearn.model_selection import train_test_split
          4 # Assuming you have data stored in some variables X and y
          5 # Splitting data into training and testing sets
          6 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
          8 # Now you can proceed with the code you provided
          9 # Importing necessary Libraries
         10 from sklearn.linear_model import LinearRegression
         11 from sklearn.preprocessing import StandardScaler
         12
         13 # Scaling the features
         14 scaler = StandardScaler()
         15 X_train_scaled = scaler.fit_transform(X_train)
         16 X_test_scaled = scaler.transform(X_test)
         17
         18 mean = X_train.mean(axis=0)
         19 std = X_train.std(axis=0)
         20 X_train = (X_train - mean) / std
         21 X_test = (X_test - mean) / std
         22 #Linear Regression
         23
         24 | from sklearn.linear_model import LinearRegression
         25 regressor = LinearRegression()
         26 #Fitting the model
         27 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
         28 regressor.fit(X_train,y_train)
Out[13]: LinearRegression
         LinearRegression()
In [14]: | 1 #Prediction on the test dataset
          2 y_pred = regressor.predict(X_test)
          3 # Predicting RMSE the Test set results
          4 from sklearn.metrics import mean_squared_error
          5 rmse = (np.sqrt(mean_squared_error(y_test, y_pred)))
          6 print(rmse)
         5.001766890194158
In [15]:
          1 from sklearn.metrics import r2_score
          2 r2 = r2_score(y_test, y_pred)
          3 print(r2)
```

0.6588520195508162

```
In [16]:
          1 import keras
          2 from keras.layers import Dense
          3 | from keras.models import Sequential
          4 from sklearn.preprocessing import StandardScaler
          5 import matplotlib.pyplot as plt
          7 # Assuming X train and X test are defined and initialized previously
          8 # Assuming y_train is also defined and initialized
          10 # Scaling the dataset
         11 sc = StandardScaler()
         12 X train = sc.fit transform(X train)
         13 X_test = sc.transform(X_test)
         14
         15 # Creating the neural network model
         16 model = Sequential()
         17 model.add(Dense(128, activation='relu', input_dim=13))
          18 model.add(Dense(64, activation='relu'))
          19 model.add(Dense(32, activation='relu'))
          20 model.add(Dense(16, activation='relu'))
          21 model.add(Dense(1))
         22
         23 # Compiling the model
          24 |model.compile(optimizer='adam', loss='mean_squared_error', metrics=['mae'])
         25
          26 # Visualizing the model architecture
         27 keras.utils.plot_model(model, to_file='model.png', show_shapes=True, show_layer_names=True)
          28
          29 # Assuming you have defined your training data X_train and y_train
          30 history = model.fit(X_train, y_train, epochs=100, validation_split=0.05)
          31
          32 # Plotting the training and validation loss
         plt.plot(history.history['loss'], label='Training Loss')
          34 plt.plot(history.history['val_loss'], label='Validation Loss')
          35 plt.xlabel('Epochs')
          36 plt.ylabel('Loss')
          37 plt.legend()
         38 plt.show()
         39
```

```
You must install pydot (`pip install pydot`) and install graphviz (see instructions at https://graphviz.gitl
ab.io/download/) (https://graphviz.gitlab.io/download/)) for plot_model to work.
Epoch 1/100
val_mae: 20.7200
Epoch 2/100
12/12 [==============] - 0s 9ms/step - loss: 525.0057 - mae: 20.9747 - val_loss: 399.8764 -
val mae: 18.5050
Epoch 3/100
val mae: 13.2538
Epoch 4/100
val_mae: 4.9532
Epoch 5/100
mae: 4.4646
Epoch 6/100
mae: 4.6085
Epoch 7/100
12/12 [=================== ] - 0s 9ms/step - loss: 27.4871 - mae: 3.8358 - val loss: 53.4959 - val
_mae: 4.6053
Epoch 8/100
_mae: 4.6640
Epoch 9/100
mae: 4.5293
Epoch 10/100
12/12 [==============] - 0s 10ms/step - loss: 18.5831 - mae: 3.1421 - val_loss: 49.0089 - va
1 mae: 4.3896
Epoch 11/100
12/12 [========================= ] - 0s 12ms/step - loss: 16.9307 - mae: 2.9703 - val_loss: 47.4259 - va
1_mae: 4.2535
Epoch 12/100
mae: 4.1555
Epoch 13/100
12/12 [==============] - 0s 10ms/step - loss: 14.9110 - mae: 2.7769 - val_loss: 43.3636 - va
l mae: 4.0691
Epoch 14/100
12/12 [=============] - 0s 10ms/step - loss: 14.2704 - mae: 2.7415 - val_loss: 42.3949 - va
1_mae: 4.0346
Epoch 15/100
mae: 3.9858
Epoch 16/100
_mae: 3.8408
Epoch 17/100
12/12 [============== ] - 0s 11ms/step - loss: 12.9651 - mae: 2.6520 - val_loss: 40.7240 - va
1_mae: 3.9047
Epoch 18/100
1 mae: 3.9262
Epoch 19/100
1_mae: 3.7524
Epoch 20/100
mae: 3.7707
Epoch 21/100
l_mae: 3.8018
Epoch 22/100
mae: 3.6912
Epoch 23/100
12/12 [==============] - 0s 10ms/step - loss: 11.2924 - mae: 2.5149 - val_loss: 37.3225 - va
1_mae: 3.6865
Epoch 24/100
12/12 [==============] - 0s 9ms/step - loss: 11.1942 - mae: 2.4082 - val_loss: 36.1898 - val
mae: 3.6670
Epoch 25/100
```

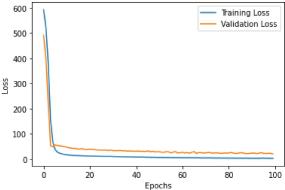
```
1_mae: 3.6580
Epoch 26/100
_mae: 3.6747
Epoch 27/100
mae: 3.5572
Epoch 28/100
mae: 3.6871
Epoch 29/100
_mae: 3.6116
Epoch 30/100
12/12 [============== ] - 0s 7ms/step - loss: 10.5037 - mae: 2.4386 - val loss: 36.0907 - val
mae: 3.6182
Epoch 31/100
mae: 3.6011
Epoch 32/100
_mae: 3.5193
Epoch 33/100
12/12 [=============] - 0s 9ms/step - loss: 9.3183 - mae: 2.2621 - val_loss: 33.1790 - val_
mae: 3.5449
Epoch 34/100
mae: 3.6106
Epoch 35/100
mae: 3.5442
Epoch 36/100
mae: 3.5653
Epoch 37/100
mae: 3.4381
Epoch 38/100
mae: 3.5358
Epoch 39/100
12/12 [============== ] - 0s 10ms/step - loss: 8.4396 - mae: 2.1538 - val loss: 32.1248 - val
_mae: 3.5076
Epoch 40/100
mae: 3.4163
Epoch 41/100
_mae: 3.4657
Epoch 42/100
_mae: 3.5412
Epoch 43/100
mae: 3.3762
Epoch 44/100
mae: 3.4593
Epoch 45/100
mae: 3.4678
Epoch 46/100
mae: 3.5630
Epoch 47/100
mae: 3.3674
Epoch 48/100
_mae: 3.4752
Epoch 49/100
mae: 3.3669
Epoch 50/100
mae: 3.3653
```

```
mae: 3.3936
Epoch 52/100
_mae: 3.3383
Epoch 53/100
12/12 [==============] - 0s 9ms/step - loss: 6.0513 - mae: 1.8575 - val_loss: 27.0754 - val_
mae: 3.3404
Epoch 54/100
mae: 3.4202
Epoch 55/100
mae: 3.4191
Epoch 56/100
12/12 [==============] - 0s 8ms/step - loss: 5.9944 - mae: 1.8457 - val_loss: 24.6683 - val_
mae: 3.2299
Epoch 57/100
mae: 3.4076
Epoch 58/100
_mae: 3.4728
Epoch 59/100
12/12 [=============] - 0s 9ms/step - loss: 5.4385 - mae: 1.7720 - val_loss: 23.6976 - val_
mae: 3.1610
Epoch 60/100
mae: 3.3188
Epoch 61/100
```

Epoch 51/100

```
mae: 3.4120
Epoch 62/100
12/12 [==============] - 0s 8ms/step - loss: 5.2426 - mae: 1.7450 - val_loss: 24.2072 - val_
mae: 3.1952
Epoch 63/100
mae: 3.3223
Epoch 64/100
mae: 3.1827
Epoch 65/100
mae: 3.3379
Epoch 66/100
12/12 [==============] - 0s 8ms/step - loss: 4.9866 - mae: 1.7151 - val_loss: 29.6811 - val_
mae: 3.4808
Epoch 67/100
mae: 3.0447
Epoch 68/100
mae: 3.2718
Epoch 69/100
mae: 3.2645
Epoch 70/100
mae: 3.2512
Epoch 71/100
mae: 3.1603
Epoch 72/100
12/12 [========================== ] - 0s 9ms/step - loss: 4.7020 - mae: 1.6812 - val_loss: 26.5144 - val_
mae: 3.3664
Epoch 73/100
mae: 3.2771
Epoch 74/100
mae: 3.1621
Epoch 75/100
mae: 3.2213
Epoch 76/100
mae: 3.2016
Epoch 77/100
mae: 3.1584
Fnoch 78/100
mae: 3.0758
Epoch 79/100
mae: 3.2229
Epoch 80/100
mae: 3.2169
Epoch 81/100
mae: 3.2552
Epoch 82/100
mae: 3.2781
Epoch 83/100
mae: 3.1065
Epoch 84/100
mae: 3.0606
Epoch 85/100
mae: 3.1206
Epoch 86/100
12/12 [========================= ] - 0s 9ms/step - loss: 3.5110 - mae: 1.3960 - val_loss: 25.4949 - val_
```

```
mae: 3.2908
Epoch 87/100
mae: 3.1356
Epoch 88/100
mae: 3.0376
Epoch 89/100
mae: 2.9952
Epoch 90/100
mae: 3.1246
Epoch 91/100
12/12 [============== ] - 0s 7ms/step - loss: 3.2493 - mae: 1.3628 - val loss: 23.5151 - val
mae: 3.1295
Epoch 92/100
12/12 [========================= ] - 0s 7ms/step - loss: 3.1388 - mae: 1.3269 - val_loss: 23.6589 - val_
mae: 3.0611
Epoch 93/100
12/12 [============] - 0s 7ms/step - loss: 3.1233 - mae: 1.3342 - val_loss: 21.1233 - val_
mae: 2.9461
Epoch 94/100
12/12 [=============] - 0s 8ms/step - loss: 3.1661 - mae: 1.3309 - val_loss: 23.6154 - val_
mae: 3.1297
Epoch 95/100
mae: 3.1956
Epoch 96/100
mae: 3.1080
Epoch 97/100
12/12 [=============] - 0s 8ms/step - loss: 3.2998 - mae: 1.3654 - val_loss: 21.8552 - val_
mae: 3.0664
Epoch 98/100
12/12 [=============] - 0s 7ms/step - loss: 3.2627 - mae: 1.3816 - val_loss: 22.6286 - val_
mae: 3.0797
Epoch 99/100
mae: 3.0708
Epoch 100/100
12/12 [============= ] - 0s 9ms/step - loss: 2.8196 - mae: 1.2830 - val loss: 20.1691 - val
mae: 2.8952
```



```
In [18]:
           1 #Comparison with traditional approaches
           2 #First let's try with a simple algorithm, the Linear Regression:
           3 from sklearn.metrics import mean_absolute_error
           4 lr_model = LinearRegression()
           5 lr_model.fit(X_train, y_train)
           6 y_pred_lr = lr_model.predict(X_test)
           7 mse_lr = mean_squared_error(y_test, y_pred_lr)
           8 mae_lr = mean_absolute_error(y_test, y_pred_lr)
          9 print('Mean squared error on test data: ', mse_lr)
10 print('Mean absolute error on test data: ', mae_lr)
          11 from sklearn.metrics import r2_score
          12 r2 = r2_score(y_test, y_pred)
          13 print(r2)
          Mean squared error on test data: 25.01767202384286
          Mean absolute error on test data: 3.1499233573458034
          0.8378310222849991
In [19]: 1 # Predicting RMSE the Test set results
           2 from sklearn.metrics import mean_squared_error
```

4 print(rmse) 3.448545294335442

```
In [20]:
         1 # Make predictions on new data
          2 import sklearn
          3 new_data = scaler.transform([[0.1, 10.0, 5.0, 0, 0.4, 6.0, 50, 6.0, 1, 400, 20, 300, 10]]) # Scaling new
          4 prediction = model.predict(new_data)
          5 print("Predicted house price:", prediction)
```

```
1/1 [======== ] - 0s 246ms/step
Predicted house price: [[14.101335]]
```

3 rmse = (np.sqrt(mean_squared_error(y_test, y_pred)))