boston

June 6, 2024

[8]:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error,mean_absolute_error,r2_score
```

[24]:

```
dataset = pd.read_csv("HousingData.csv")
dataset = dataset.dropna()
dataset.head(3)
```

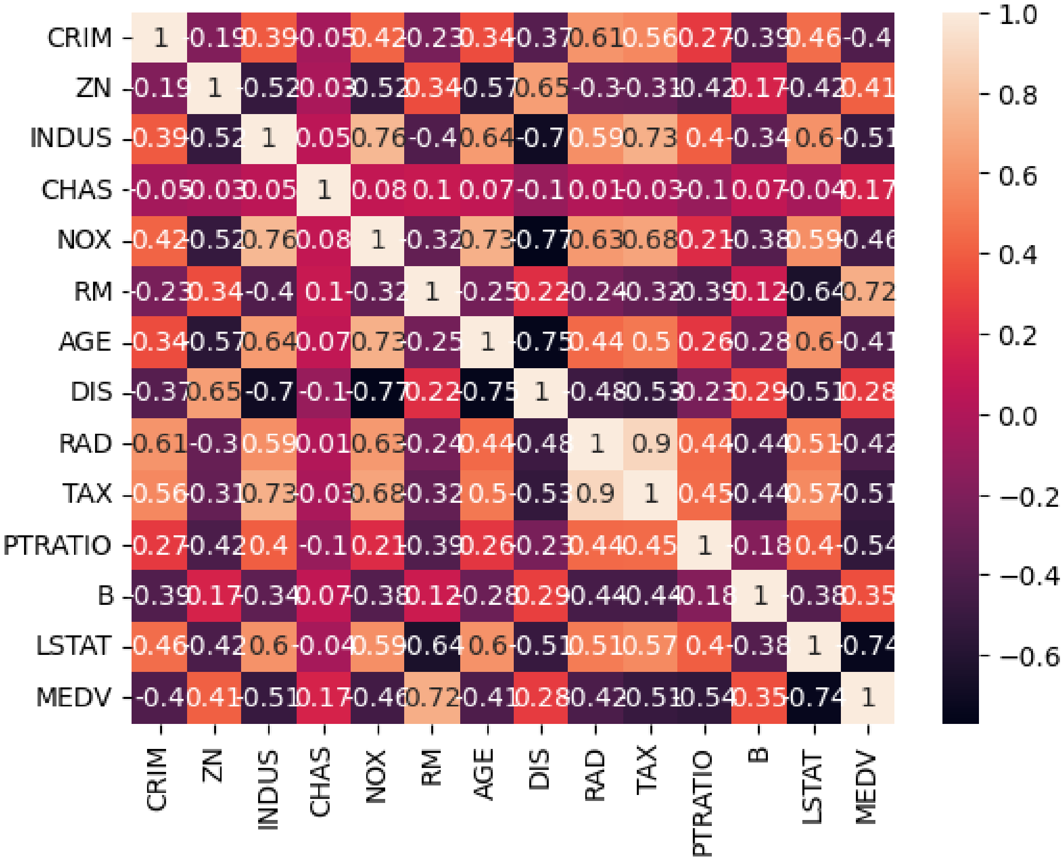
[24]:

```
      CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD  TAX  PTRATIO  \
0  0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900    1  296     15.3
1  0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671    2  242     17.8
2  0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671    2  242     17.8

        B  LSTAT  MEDV
0  396.90   4.98  24.0
1  396.90   9.14  21.6
2  392.83   4.03  34.7
```

[25]:

```
correlation_matrix = dataset.corr().round(2)
import seaborn as sns
sns.heatmap(correlation_matrix,annot=True)
plt.show()
```



[26]:
```
X = pd.DataFrame(np.c_[dataset['LSTAT'],dataset['RM']],columns=['LSTAT','RM'])
Y = dataset['MEDV']
```

[27]:
```
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.
 ↪2,random_state=42)
```

[28]:
```
linear_model = LinearRegression()
linear_model.fit(X_train,Y_train)
Y_pred = linear_model.predict(X_test)
```

[30]:
```
mse = mean_squared_error(Y_test,Y_pred)
mse
rmse = np.sqrt(mse)
mae = mean_absolute_error(Y_test,Y_pred)
re = r2_score(Y_test,Y_pred)
```

[32]:
```
print(mse)
print(rmse)
```

```python
print(mae)
print(re)
```

```
37.739803570357694
6.14327303400701
3.8614905763730554
0.5525619081372957
```

[ ]:

```

```