Creado por:

Isabel Maniega

Decission Tree Regression

In [1]:
```
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```
# pip install scikit-learn
```

In [3]:
```
import numpy as np
from sklearn import datasets, linear_model
import matplotlib.pyplot as plt
import pandas as pd
```

In [4]:
```
boston = datasets.load_boston()
```

In [5]:
```
boston.data
```

Out[5]:
```
array([[6.3200e-03, 1.8000e+01, 2.3100e+00, ..., 1.5300e+01, 3.9690e+02,
        4.9800e+00],
       [2.7310e-02, 0.0000e+00, 7.0700e+00, ..., 1.7800e+01, 3.9690e+02,
        9.1400e+00],
       [2.7290e-02, 0.0000e+00, 7.0700e+00, ..., 1.7800e+01, 3.9283e+02,
        4.0300e+00],
       ...,
       [6.0760e-02, 0.0000e+00, 1.1930e+01, ..., 2.1000e+01, 3.9690e+02,
        5.6400e+00],
       [1.0959e-01, 0.0000e+00, 1.1930e+01, ..., 2.1000e+01, 3.9345e+02,
        6.4800e+00],
       [4.7410e-02, 0.0000e+00, 1.1930e+01, ..., 2.1000e+01, 3.9690e+02,
        7.8800e+00]])
```

In [6]:
```
print('Nombre de columnas:')
print(boston.feature_names)
```

```
Nombre de columnas:
['CRIM' 'ZN' 'INDUS' 'CHAS' 'NOX' 'RM' 'AGE' 'DIS' 'RAD' 'TAX' 'PTRATIO'
 'B' 'LSTAT']
```

In [7]:
```
df = pd.DataFrame(boston.data, columns=boston.feature_names)
df
```

Out[7]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33
...	...	...	...	...	...	...	...	...	...	...	...	...	...
501	0.06263	0.0	11.93	0.0	0.573	6.593	69.1	2.4786	1.0	273.0	21.0	391.99	9.67
502	0.04527	0.0	11.93	0.0	0.573	6.120	76.7	2.2875	1.0	273.0	21.0	396.90	9.08
503	0.06076	0.0	11.93	0.0	0.573	6.976	91.0	2.1675	1.0	273.0	21.0	396.90	5.64
504	0.10959	0.0	11.93	0.0	0.573	6.794	89.3	2.3889	1.0	273.0	21.0	393.45	6.48
505	0.04741	0.0	11.93	0.0	0.573	6.030	80.8	2.5050	1.0	273.0	21.0	396.90	7.88

506 rows × 13 columns

In [8]:
```
print("Informacion en el dataset:")
print(boston.keys())
```

```
Informacion en el dataset:
dict_keys(['data', 'target', 'feature_names', 'DESCR', 'filename', 'data_module'])
```

In [9]:
```
print("Características del dataset:")
print(boston.DESCR)
```

```
Características del dataset:
.. _boston_dataset:

Boston house prices dataset
---------------------------

**Data Set Characteristics:**

    :Number of Instances: 506

    :Number of Attributes: 13 numeric/categorical predictive. Median Value (attribute 14) is usually the targe
t.

    :Attribute Information (in order):
        - CRIM     per capita crime rate by town
        - ZN       proportion of residential land zoned for lots over 25,000 sq.ft.
        - INDUS    proportion of non-retail business acres per town
        - CHAS     Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
        - NOX      nitric oxides concentration (parts per 10 million)
        - RM       average number of rooms per dwelling
        - AGE      proportion of owner-occupied units built prior to 1940
        - DIS      weighted distances to five Boston employment centres
        - RAD      index of accessibility to radial highways
        - TAX      full-value property-tax rate per $10,000
        - PTRATIO  pupil-teacher ratio by town
        - B        1000(Bk - 0.63)^2 where Bk is the proportion of black people by town
        - LSTAT    % lower status of the population
        - MEDV     Median value of owner-occupied homes in $1000's

    :Missing Attribute Values: None

    :Creator: Harrison, D. and Rubinfeld, D.L.

This is a copy of UCI ML housing dataset.
https://archive.ics.uci.edu/ml/machine-learning-databases/housing/


This dataset was taken from the StatLib library which is maintained at Carnegie Mellon University.

The Boston house-price data of Harrison, D. and Rubinfeld, D.L. 'Hedonic
prices and the demand for clean air', J. Environ. Economics & Management,
vol.5, 81-102, 1978.   Used in Belsley, Kuh & Welsch, 'Regression diagnostics
...', Wiley, 1980.   N.B. Various transformations are used in the table on
pages 244-261 of the latter.

The Boston house-price data has been used in many machine learning papers that address regression
problems.

.. topic:: References

   - Belsley, Kuh & Welsch, 'Regression diagnostics: Identifying Influential Data and Sources of Collinearity',
Wiley, 1980. 244-261.
   - Quinlan,R. (1993). Combining Instance-Based and Model-Based Learning. In Proceedings on the Tenth Internat
ional Conference of Machine Learning, 236-243, University of Massachusetts, Amherst. Morgan Kaufmann.
```

In [10]:
```
print("Cantidad de datos:")
print(boston.data.shape)
```

```
Cantidad de datos:
(506, 13)
```

In [11]:
```
# Seleccionamos todas las columnas del data
X = boston.data
```

In [12]:
```
y = boston.target
```

In [13]:
```
from sklearn.model_selection import train_test_split
# Separo los datos de "train" entrenamiento y "test" prueba para probar los algoritmos
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

In [14]:
```
from sklearn.tree import DecisionTreeRegressor

adr = DecisionTreeRegressor(max_depth=5)
```

In [15]:
```
adr.fit(X_train, y_train)
```

Out[15]:
```
▼ DecisionTreeRegressor
DecisionTreeRegressor(max_depth=5)
```

In [16]:
```
y_pred = adr.predict(X_test)
y_pred
```

Out[16]:
```
array([ 9.58857143,  9.58857143, 18.46666667,  9.58857143, 20.29295775,
       50.        , 22.84252874, 22.74      , 22.84252874, 18.46666667,
       34.335     , 14.152     , 14.152     , 22.84252874, 34.335     ,
       18.46666667, 22.84252874, 18.46666667, 18.46666667, 31.32      ,
       20.29295775, 20.29295775, 14.152     , 22.84252874, 22.84252874,
       18.46666667, 40.43333333, 22.84252874, 20.29295775, 14.03333333,
       14.152     ,  9.58857143, 20.29295775, 22.84252874, 17.11666667,
       31.32      ,  9.58857143, 31.32      , 34.335     , 17.11666667,
       50.        , 18.46666667,  9.58857143, 22.84252874, 20.29295775,
       17.11666667, 20.29295775, 22.84252874, 20.29295775, 22.74      ,
        9.58857143, 26.03142857, 18.46666667, 30.36153846, 22.84252874,
       20.29295775, 26.03142857, 17.11666667,  9.58857143, 22.84252874,
       22.84252874, 34.335     , 20.29295775, 14.152     , 22.84252874,
       17.11666667, 14.1       , 18.46666667, 30.36153846,  9.58857143,
       50.        , 20.29295775, 20.29295775,  9.58857143, 22.84252874,
       14.152     , 18.46666667, 20.29295775, 22.84252874, 31.32      ,
       46.04375   , 14.152     ,  9.58857143, 46.04375   , 30.36153846,
       34.335     , 20.29295775, 22.84252874, 40.43333333, 26.03142857,
       18.46666667, 22.84252874, 22.84252874, 20.29295775, 14.152     ,
       22.84252874, 22.84252874, 18.46666667, 17.11666667, 50.        ,
       20.29295775, 30.36153846])
```

In [17]:
```
y_test
```

Out[17]:
```
array([ 5. , 13.3, 14.5, 10.2, 24.3, 10.4, 20.6, 21.7, 25.3, 11.7, 35.4,
       14.9, 23.2, 20.1, 37.2, 19.5, 23.2, 27.1, 13.3, 33.8, 20.6, 23.4,
       16.4, 25. , 23.1, 21.7, 43.8, 22.9, 21.1, 17.8, 15.2, 11.8, 20.9,
       16.1, 17.1, 36.5, 11. , 31.6, 36.2, 16.2, 27.9, 18.7,  9.7, 18.2,
       21.8, 17.4, 24.7, 22.2, 17.1, 20.3, 11.3, 25. , 16. , 28.2, 25. ,
       20. , 28.4, 15.6, 10.2, 21.6, 21.4, 32.7, 18.2, 20. , 19.5, 13.1,
       13.2, 19.1, 32. , 12.3, 50. , 21.8, 19. , 10.4, 20.2, 11.7, 20.5,
       16.1, 21.9, 31. , 44. , 16.8, 10.5, 50. , 28.4, 35.4, 19.9, 23.8,
       50. , 29.9, 15. , 24.4, 23.7, 20.3, 18.4, 20.5, 25. , 21.7, 14.3,
       15. , 16.8, 30.8])
```

In [18]:
```
print("Datos del Modelo de árboles de decision Regresión")
print()

print("Precisión del modelo:")
print(adr.score(X_train, y_train))
```

```
Datos del Modelo de árboles de decision Regresión

Precisión del modelo:
0.9231395333262339
```

Creado por:

Isabel Maniega