# importing dependencies

import numpy as np

import pandas as pd

from matplotlib import pyplot as plt

from sklearn.datasets import load\_boston

# understanding the dataset

boston = load\_boston()

print(boston.DESCR)

.. \_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 target.

: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 blacks 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 International Conference of Machine Learning, 236-243, University of Massachusetts, Amherst. Morgan Kaufmann.

# access data attributes

dataset = boston.data

for name, index in enumerate(boston.feature\_names):

print(index, name)

CRIM 0

ZN 1

INDUS 2

CHAS 3

NOX 4

RM 5

AGE 6

DIS 7

RAD 8

TAX 9

PTRATIO 10

B 11

LSTAT 12

# reshaping data

data = dataset[:,12].reshape(-1,1)

# shape of the data

np.shape(dataset)

(506, 13)

# ensuing that matplotlib is working inside the notebook

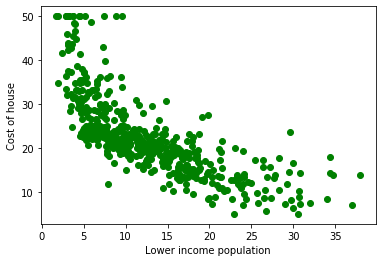
%matplotlib inline

plt.scatter(data, target, color='green')

plt.xlabel('Lower income population')

plt.ylabel('Cost of house')

plt.show()



# regression

from sklearn.linear\_model import LinearRegression

# creating a regression scale

reg = LinearRegression()

# fit in model

reg.fit(data, target)

LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=None, normalize=False)

# prediction

pred = reg.predict(data)

# ensuing that matplotlib is working inside the notebook

%matplotlib inline

plt.scatter(data, target, color='red')

plt.plot(data, pred, color='green')

plt.xlabel('Lower income population')

plt.ylabel('Cost of house')

plt.show()

