

For the california housing dataset, do:

Dataset description:

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_california_housing.html#sklearn.datasets.fetch_california_housing

https://scikit-learn.org/stable/datasets/real_world.html#california-housing-dataset

(a) Load the dataset.

(b) Make 8 graphs showing the dependence of y on individual x variables i (columns from dataset X). Sign axes of charts.

(c) Split the data into training and testing sets using function `train_test_split` in a proportion of 0.7 data

learning and test 0.3:

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html.

`train_test_split.html`.

(d) Train a LinearRegression model on the training data.

An example of using LinearRegression on another dataset

https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html.

(e) Calculate the MAE (mean absolute error) and MSE (mean squared error) errors on the test data, and

training data, evaluate the quality of this model.

(f) For each of the 8 variables do:

i. Select the selected variable from the X_{train} and X_{test} matrices so that they are single-column matrices

new.

1

ii. Train the LinearRegression model on this selected variable and on the y_{train} vector.

iii. Draw a graph as in point (b) and place regression lines on it (see example of use Linear Regression given in point (d)).

iv. Also calculate and list the MAE and MSE values for this model.

(g) Analyze the results obtained in (e) and (f). Is training the model on single variables reduces or increases the error