

Assignment #1

Course: *Machine learning*
Due date: *October 16th, 2023*

Assignment

In this assignment you will learn about linear regression. You have to predict the fuel consumption of different cars based on a set of car features.

Write the code to fit the multiple linear regression in Python.

Download the dataset Auto MPG and remove the attribute *cylinders* and *car_name*.

<http://archive.ics.uci.edu/dataset/9/auto+mpg>
Pay attention to potential categorical attributes and missing values.

Split your dataset on train and test data.

Test your linear regression on the Auto dataset and compare it with the linear regression implemented in SciKit-learn.

Diagnose your linear regression with the diagnostic plots.

Is there something you can improve like transform your target variable or remove some attributes/instances? Try different things and refit the linear regression.