

Assignment #1

Course: *Machine learning*
Due date: *October 13th, 2024*

Assignment

In this assignment you will learn about linear regression.

Download the dataset from Učilnica and perform data preprocessing (e.g. deal with missing values).

Use Jupyter notebook to run your code and write comments.

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

Split your dataset on train and test data. As the data is temporal do not perform random shuffle. Remove the Datum from the dataset. Your target variable is PM10.

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

Diagnose your linear regression with the diagnostic plots.

Can you make some improvements to your linear regression fit?

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