

Assignment #1

Course: *Machine learning*

Date: *October 13th, 2025*

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. Name your file Your-Surname_assignment1.ipynb.

Write the code to fit the multiple linear regression in Python. Use the closed form solution.

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 (see Lecture 2).

Comment on your results.

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 (see correlations)? Maybe some x have a different relationship to y than linear. Try different things and refit the linear regression. Comment the results.