MIW (eng) - mini-project 4 (22.04.20)

Linear Regression

1. Read in the data from a file *daneXX.txt* that's been assigned to you via the link, where XX is the dataset number:

   https://docs.google.com/spreadsheets/d/1eyqkibVdV5LXjAoKUhVNvWHixhPk2iw3zyYUi_2LPUM/edit#gid=1925742827

2. Suggest the split of that data into *training* and *test* datasets.

3. Suggest a parametric linear model **Model 1**. Find the model's parameters by applying the method of least squares on the training dataset.

4. Verify the correctness of **Model 1**.

5. Suggest a more complex parametric model **Model 2**. Find the model's parameters by applying the method of least squares on the training dataset.

6. Verify the correctness of **Model 2**.

7. Compare both models.

It has to be implemented **from scratch**. You can only use libraries: **NumPy**, **math**, **pandas.read_csv** and **matplotlib**. If you want to use any additional libraries, send me an email to ihalych@pja.edu.pl or send me a message in Microsoft Teams.

Deadline rules:

1. By 22.04.20 - max amount of points is 10
2. By 29.04.20 - max amount of points is 10
3. By 06.05.20 - max amount of points is 5
4. Later - 0 points

Correctness of a model:

You can try one of the metrics to evaluate and compare the models:

https://towardsdatascience.com/regression-an-explanation-of-regression-metrics-and-what-can-go-wrong-a39a9793d914

- Mean Squared Error(MSE)
- Root-Mean-Squared-Error(RMSE).
- Mean-Absolute-Error(MAE).
- R^2 or Coefficient of Determination.
- Adjusted R^2