Module 6 - Bonus Challenges

§M6: Linear Least Squares Regression

- Below are open-ended bonus challenges; solving them is not required but can help you better understand ML/AI in the context of engineering, and how to use them in practical cases.
- Bonus points earned in all homework assignments will be averaged (6 bonus points for each assignment) and then directly added to your final score to calculate your final letter grade.

Challenge 1.1. For this bonus question, you will apply the feature engineering techniques you've learned to a real-world engineering problem: predicting the compressive strength of concrete. Concrete is widely used in civil engineering, and its properties play a crucial role in building safety. Compressive strength is one of the key properties, depending on factors like age and ingredients. Given the highly nonlinear relationship between compressive strength and these factors, it would be valuable to develop a machine learning model to predict this property and guide the manufacturing process.

`'m06_bonus.xls'` is a dataset adapted from [1]. It includes data points of 1,030 concrete materials. The first 10 columns represent the material features of various concretes, such as cement, blast furnace slag, and fly ash. These features together determine the compressive strength, which is listed in the last column of the sheet.

The goal of this task is to predict compressive strength based on the material features. Please complete the task according to the following requirements: (*6pts*)

1. Fit a linear regression model without any feature engineering. Randomly split the dataset into 80% for training and 20% for testing, using `train_test_split` in `sklearn`. Train your model on the training dataset and evaluate it on the testing set.You may experiment with different polynomial orders or regularization techniques.
2. Analyze the statistical characteristics of the data distribution, then apply any feature engineering techniques you think will be beneficial.
3. Integrate your feature engineering and fit the linear regression model again, using the same model and training setup as before.
4. Compare and discuss the results with and without feature engineering.
5. Submit your Jupyter notebook file with necessary comments.

References

[1] I-Cheng Yeh. *Concrete Compressive Strength*. 2007. URL: https://doi.org/10.24432/C5PK67.