Course Project 1: Developing Regression Model

A series of experiments were carried out at SPCE by a PG student to study rate of crack growth in a standard compact tension specimen of carbon steel, subjected to cyclic loading. Each experiment was conducted with a different combination of specimen thickness, maximum amplitude of load and frequency of loading. The objective of experiment was to study the influence of loading frequency and range of stress intensity factor (ΔK is a parameter which represents crack tip stress field) on the crack growth rate.

You have been provided with a complete dataset of around 10,400 observations (crack-growth.csv) obtained from few of the experiments. In the dataset, following three features are relevant for the course project.

1. 'da/dN' : crack growth rate in m/cycle
2. 'Delta K' : range of stress intensity factor in $MPa\sqrt{m}$
3. 'Frequency' : Frequency of loading in Hz (1 Hz, 25 Hz and 50 Hz)

Your job is to develop a regression model which will predict the crack growth rate ('da/dN') for given values of 'Delta K' and 'Frequency'.

Each of you will submit code for your model separately. Once submitted, it will be evaluated by me against a dataset (obtained from remaining experiments) reserved for this purpose.

In this course we will have two course projects, this being the first one. Students with the best scores, considering their performance for the two projects taken together, will be awarded with the best performance certificate.

The criteria for evaluating model's performance are as follows. In case of close match in first criterion, second criterion will be used as a tiebreaker.

- First criterion: R^2 score (higher is better)
- Second criterion: Time to train the model (lower is better)

NOTES:

1 You can use only one of the following models (or a combination of the following models) provided at https://scikit-learn.org/stable/supervised_learning.html#supervised-learning
   1.a 1.1.1 Ordinary least squares
   1.b 1.1.2 Ridge regression
   1.c 1.1.17 Polynomial regression
2 Your model performance (on the test data used by me for final evaluation) will vary based on values of parameters "test_size" and "random_state" of 'train_test_split' considered by you for generating the training set. You may submit your code with fixed values of these parameters; if not, the best performance obtained against suitable number of trials will be considered for evaluation.
3 Before you submit your final model, you will get a SINGLE chance to obtain R^2 score of your model against the reserved test data. You may use this information to fine tune your model, if necessary. This score of your model will be made available for all students.