



## Programming Assignment 2: Polynomial Regression

### Description

In this assignment you are required to find online a dataset with multiple features which is suitable for polynomial regression (you will need to create polynomial features from the original features and use them in the training or otherwise your model will not perform well). Upload this dataset to your Google Drive account and create a link to it that allows anyone to download it and use this link in your Colab Notebook. This way, we can test your work without having to worry about the dataset directory. Your assignment will have the following parts:

#### Part 1:

Apply linear regression to your dataset from the sklearn library (it is your choice whether to scale the feature for this part or not). The purpose here is to prove that the features are not enough to model the problem and there is a need to create polynomial features. This what will make finding the dataset is not straight forward. Of course you will need to report the performance metrics.

#### Part 2:

Create polynomial features and append them to your dataset. Then apply an appropriate feature scaling, train a polynomial regression, and report the appropriate performance metrics. In this scenario, we want you to showcase to models, one that is just good and one that has high variance. Then compare in a table these models and the one in part 1 (high bias).

#### Part 3:

Redo part 2 but without scaling your features and compare the results with part 2. Discuss the results and include your thoughts.

Note that you may need split your dataset into train, validation, and test sets to be able to solve the assignment.

The above parts (1, 2, and 3) and the performance metrics, discussion, description of the dataset, the source of the dataset and link to it, and all details should be contained in a single CoLab Notebook.