Machine Learning project
Deadline: Friday 20th september 23:59

Method : regression
Dataset: Life expectancy of the world
https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

In Stage 1:

- Choose and formulate a machine learning problem.
    - predicting life expectancy from BMI, GPD or total government expenditure on health?
- Choose one ML method that is suitable for the problem.
    - Linear regression or polynomial regression
        - regularized or not?

In Stage 2:

- Apply at least 2 ML Methods to the problem you formulated in Stage 1.
- Compare and analyze the results, and choose the best method.

Stage 1 report:

Introduction:

- Explain the background (real-life scenario) of your ML application.
- Briefly outline the structure of the report.

Problem Formulation:

- Formalise the application an ML problem.
- Clearly explain the data points, features and labels of this ML problem.
- Explain the source of the dataset.
  Methods:
- State the number of datapoints, briefly describe the dataset and/or any data preprocessing needed.
- Explain your feature selection process (no theoretical justification needed).
- Describe and explain (why?) your choice of ML model(s)/hypothesis space(s)*, e.g., linear predictors, etc.
- Describe and explain (why?) your choice of loss function(s)*, e.g., logistic loss.
- Explain the process of model validation - how did you split the data into training, validation and test sets. What are the sizes of each set and why did you make such design choice.