GEO5017 A1
Linear Regression

Question 1. Basic model (20%)

Suppose we have observed N data points $\{x_i, t_i\}_{i=1\ldots N}$, answer the following questions:

1) What do we need before we can formulate a linear regression problem to describe the distribution of the data? Use no more than two sentences to answer this question.(10%)

2) What would be the optimal solution? Use no more than three sentences to answer this question. (10%)

Question 2. Programming: drone (60%)

We are testing a tracking program and we evaluate it with the help of a drone. The drone sends estimates of its velocity and the tracking program estimates its global position concerning the drone's initial position (before flying).

1) The tracker yields these tracked positions at a frequency of $1Hz$:

$$\mathcal{T} = \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1.08 \\ 1.68 \\ 2.38 \end{pmatrix} \begin{pmatrix} -0.83 \\ 1.82 \\ 2.49 \end{pmatrix} \begin{pmatrix} -1.97 \\ 0.28 \\ 2.15 \end{pmatrix} \begin{pmatrix} -1.31 \\ -1.51 \\ 2.59 \end{pmatrix} \begin{pmatrix} 0.57 \\ -1.91 \\ 4.32 \end{pmatrix} \right\}$$

   Plot the trajectory through these data points with your tool of choice. (10%)

2) Use the Polynomial Regression method introduced in the lecture to answer the following questions and provide details in your report. You can choose your favorite programming language to formulate and solve the regression problems (10%), and you **must** use your own implementation of the Gradient Descent solver. (10%)

   (a) Assuming the drone flies with constant speed, what speed does it have? What is the residual error (i.e., sum-of-squares error) of the estimation? (10%)

   (b) Now assume that the drone flies with constant acceleration. What is the residual error (i.e., the sum of squared error) now? Is the error higher or lower? Why, and can you still improve your model? (10%)

   (c) According to our last model, what is the drone's most likely position in the next second? Plot this position together with previous positions. (10%)

Submission (Due: Wednesday, March 6th)

Please compress all the following into a single archive titled **GEO5017_A1_Group_XX.zip** (where 'XX' is your group ID and can be found here) and submit it to BrightSpace:

- **A report (max 2 pages) (20%)**
  - The report must contain the answers and necessary explanations.
  - The report must contain a short description of who did what.
    - Please focus on explaining the mathematical formulations of your models and solutions, but try to avoid unnecessary details like how the data is stored in your code and how a curve is plotted.
- **Source code** (for Question 2)
  - The source code, archived in a 'code' subfolder. The code should build, run, and reproduce your results without changes.
    - For Python code, **only** '*.py' files are accepted. If you use Jupyter notebook for development, make sure you submit '*.py' files ('*.ipynb' files will not be accepted). For C++ code, **only** '*.cpp' files are accepted.
    - If you have multiple source code files that should run in a specific order, **do** name them in a way such that they are well ordered (e.g., '1_data_fitting.py', '2_solver.py', '3_visualization.py'). If no such information is provided, we will **expect a 'main.py' file** as the only entry point.
    - Provide a 'ReadMe.txt' file to briefly explain how to run the code and reproduce the results, e.g., dependence of external libraries/packages, path to data, and where to find the results in case you save results or figures into files.
  - [*optional*] Provide a link to the GitHub repository (only if you use GitHub) in the 'Experiment' section of your report. You are encouraged to collaborate with your teammates on GitHub.

Note: The report should be as concise as possible (preferably within 2 pages excluding figures, tables, and references), but it should provide sufficient information to re-implement your methods and to reproduce your results. Irrelevant descriptions and discussion may lead to the deduction of points.