# Homework 3: Linear Regression

| Contact TA | William Rebelsky (wrebelsky@umass.edu) |
| --- | --- |
| Released date | Thursday, 9/28/2023 |
| Submission deadline | Midnight 11:59 PM, Oct, 10/5/2023 |
| Total marks | 100 |

## Task Descriptions

You should fill in the TODO of [hw3.py](https://drive.google.com/file/d/1l5kfTfmiuTB45nPxa4KGtA3-NUyZLtTZ/view?usp=sharing) to complete the following two tasks. Feel free to define necessary python functions and call these functions in run\_task1() and run\_task2()

### Task 1: Fitting a regression line to the student debt data

Fit a linear model to the US student loan debt dataset (download the dataset [here](https://drive.google.com/file/d/1cXKOVIDaD6wONXD41zZ1M61Mk7VaTERV/view?usp=sharing)) by minimizing the associated linear regression Least Squares problem using gradient descent.

In the CSV file, the first column is the input feature (x) and the second column is the target debt (y). Each row corresponds to a data sample. Without changing any code, you should see output like this:

| (40,) (40,) |
| --- |

Plot the fitted line.

Give the equation of the fitted line.

If this linear trend continues what will the total student debt be in 2030?

### Task 2: Compare the Least Squares and Least Absolute Deviation costs

Fit two linear models to a dataset with outliers (download the dataset [here](https://drive.google.com/file/d/1XfCYuWlQIDR09irNY_RSRQcRbzyZfnHe/view?usp=sharing)). The first linear model should use the least squares loss while the second linear model should use the least deviation loss.

The dataset contains a total of 10 samples.

Without changing any code of hw3.py. You should see an output like this:

| (1, 10) (1, 10) |
| --- |

Plot the fitted lines in the same figure as the original data to compare the models.

Give the equations of the fitted lines.

## 

# Deliverables

* **The completed source code hw3.py**
* **A pdf report**

The source code should be able to run by executing the command: python hw3.py

The PDF report should include:

* Task 1:
  + Single Figure containing the fitted line and the original data
  + The equation of the fitted line
  + Total predicted student debt in 2030
* Task 2:
  + Single figure containing the two fitted lines and the original data
  + The equation of the least squares regression line
  + The equation of the least deviation regression line

# Environment Configuration:

Use the environment from HW2. There are no additional packages needed.