**Task-1**

Generate a dataset for simple linear regression problem (dataset with one input and one output column) using numpy. Add some noise also to avoid perfect or absolute linearity. The dataset must be almost linear (as discussed in the class).

Analyze the data distribution, i.e. check the linearity in data points using 2D plots.

Now implement the simple linear regression from scratch using closed form solution, i.e. the direct formulation for **m** and **c** to compute the weight and offset. Divide the dataset into train and test sets and compute m and c using train set. Now, compute the output for test data using the trained simple linear regression model and compute the Mean Absolute Error, Root Mean Squared Error, R2 score.

Plot the linear regression model on the training set and test set to represent the model fitting on the data points.

Next, use the LinearRegression class of sklearn library to find the **m** and **c** values and compare them with those of the implemented model.

**Task-II**

Download the following dataset (Remove the column **‘extracurricular activities’**) and apply multiple linear regression using the inbuilt method.

[Student Performance (Multiple Linear Regression) (kaggle.com)](https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression?resource=download)

Divide the dataset into train and test sets. After finding out the coefficients for the model, predict the output for the test set and compute the Mean Absolute Error, Root Mean Squared Error, R2 score, and adjusted R2 score upon adding multiple columns.

Find out the column with highest contribution and the R2 score. Now add another column and check the adjusted R2 score. Similarly, check for all inputs whether they are actually important for predicting the output or not by computing the updated adjusted R2 score.