Assignment 1 : Linear Regression

Linear regression is based on the relationship of dependent and independent variables. Using linear regression we can predict the dependent variable using the independent variable by using Y=MX+C.

Part 1 : Linear Regression using Pseudoinverse

In this we use the formula weights m = (X^T.X)^-1.(X^T).Y also known as pseudo inverse.
In this I have used the data set which is provided by sklearn which is of **Boston house price prediction.** This can be directly downloaded from sklearn library also for calculating the inverse I have used the PINV function provided by numpy library.
To run the Program, first upload the .ipynb file.
Description of Dataset:

1. Experimental data used for **Prediction** (house price in boston) from different independent features.
2. This dataset contains 13 features and a target variable: 'CRIM' 'ZN' 'INDUS' 'CHAS' 'NOX' 'RM' 'AGE' 'DIS' 'RAD' 'TAX' 'PTRATIO' 'B' 'LSTAT'
3. Target Variable: price or target
4. This Dataset contains 506 rows and 14 columns with 13 features and 1 target.
5. No missing values.

I have used the LSTAT feature as an independent feature for prediction of target; any other feature can also be used. All the required steps are already written in the .ipynb file, Different Splits are also Provided; minimum mean squared error is found as 13%.
This shows pseudo inverse is not a good fit for the predictions.

Part 1 : Linear Regression using gradient descent

In this we use the formula weights Y=MX+C and optimize the weight w or m using the gradients of M and C.
In this I have used the data set which is provided by sklearn which is of **Boston house price prediction.** This can be directly downloaded/loaded from sklearn library also for calculating the inverse I have used the PINV function provided by numpy library.
To run the Program, first upload the .ipynb file.
Description of Dataset:

6. Experimental data used for **Prediction** (house price in boston) from different independent features.
7. This dataset contains 13 features and a target variable: 'CRIM' 'ZN' 'INDUS' 'CHAS' 'NOX' 'RM' 'AGE' 'DIS' 'RAD' 'TAX' 'PTRATIO' 'B' 'LSTAT'
8. Target Variable: price or target
9. This Dataset contains 506 rows and 14 columns with 13 features and 1 target.
10. No missing values.

I have used the LSTAT feature as an independent feature for prediction of target; any other feature can also be used. All the required steps are already written in the .ipynb file, Different Splits are also Provided; minimum mean squared error is found as 1%.

This shows that gradient descent is best fit for predictions.