

# Student Score Predictor - Simple Linear Regression

## 1. Introduction

In this project, we predict a student's marks based only on the number of hours they studied. We use Simple Linear Regression, which finds the best straight line to represent the relationship between 'Hours Studied' and 'Marks Scored'.

The equation of a straight line is:

$$y = m \cdot x + c$$

where:

y = predicted marks

x = hours studied

m = slope (change in marks per extra hour)

c = intercept (marks when hours = 0)

## 2. How Linear Regression Works

1. The model tries different values of m and c.
2. For each pair, it predicts marks for all data points.
3. It compares predicted marks to actual marks and calculates the error for each point:  
$$\text{error} = \text{actual} - \text{predicted}$$
4. These errors are squared to avoid negatives cancelling positives and to penalize large errors more.
5. The Mean Squared Error (MSE) is calculated:  
$$\text{MSE} = (\text{sum of all squared errors}) / (\text{number of points})$$
6. The combination of m and c with the smallest MSE is chosen as the best line.

## 3. Why We Use Mean Squared Error (MSE)

MSE ensures that we are minimizing the average squared distance between our prediction line and the actual points. This leads to a line that is as close as possible to all points in the dataset.

## 4. Steps to Build the Project

# Student Score Predictor - Simple Linear Regression

Steps to build the Student Score Predictor:

1. Import libraries (pandas, matplotlib, scikit-learn).
2. Load and explore the dataset.
3. Visualize the relationship with a scatter plot.
4. Split data into training and test sets.
5. Train the Linear Regression model to find  $m$  and  $c$ .
6. Plot the regression line over the training data.
7. Test the model on unseen data and calculate the  $R^2$  score to measure performance.
8. Predict marks for new input hours.