

LAB_1**Bias–variance trade-off.**

A data scientist is studying the relationship between the number of hours a student studies (x) and the marks obtained in an internal assessment (y).

The observed dataset is given below:

| Study Hours (x) | Marks (y) |
|------------------------|------------------|
| 1 | 52 |
| 2 | 55 |
| 3 | 61 |
| 4 | 70 |
| 5 | 82 |

The true underlying relationship between study hours and marks is **non-linear**, influenced by learning saturation and fatigue effects. To analyze model behavior, train the following two models:

- **Model A** Simple Linear Regression: $\hat{y} = \beta_0 + \beta_1 x$
- **Model B** Polynomial Regression of degree 4: $\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$

Tasks**1. Model Fitting**

- (a) Fit Model A using the least squares method and obtain the regression equation.
- (b) Fit Model B using polynomial regression and determine the polynomial coefficients.

2. Prediction Comparison

Using both models, predict the marks obtained when:

$$x = 6 \text{ hours}$$

3. Training Error Analysis

- (a) Compute the Mean Squared Error (MSE) for Model A on the training data.
- (b) Compute the Mean Squared Error (MSE) for Model B on the training data.

4. Bias–Variance Reasoning

Answer the following with justification:

- Which model has **higher bias**?
- Which model has **higher variance**?