

# PART-1(Bias-Variance Tradeoff in Machine Learning)

## Impact of Increasing Model Complexity on Bias and Variance

As we increase the complexity of a machine-learning model by adding more features or including higher-order polynomial terms in a regression model, the effects on bias and variance can be described through the **\*\*bias-variance trade-off\*\***.

### 1. Bias

Bias refers to the error introduced by approximating a real-world problem with a simpler model.

- Low-complexity models (e.g., linear models) typically have **high bias**, as they underfit the data and fail to capture underlying patterns. - As model complexity increases, bias **decreases**, since the model can fit more complex patterns.

### 2. Variance

Variance refers to the model's sensitivity to small fluctuations in the training data.

- High-complexity models, especially those with many features or higher-order terms, have **high variance** since they can overfit the training data, learning patterns that do not generalize well to unseen data. - As model complexity increases, variance **increases**, as the model becomes too sensitive to the specific training data.

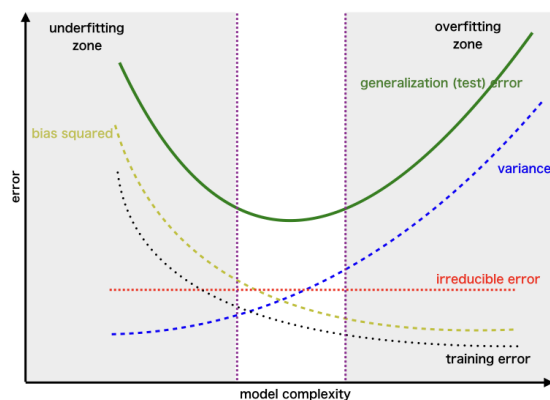
### 3. Bias-Variance Tradeoff

The bias-variance tradeoff represents the balance between bias and variance:

- Initially, as complexity increases, both training and test errors decrease due to a reduction in bias.
- After a certain point, increasing complexity results in overfitting, which causes an increase in test error due to higher variance.

## Graphical Representation

The bias-variance tradeoff can be represented graphically as shown below:



In this graph: - The **x-axis** represents model complexity. - The **y-axis** represents the error (or loss). - The **training error** decreases as complexity increases. - The **test error** initially decreases but then increases due to overfitting. - **Bias** starts high and decreases, while **variance** starts low and increases.

## Mathematical Formulation

The total error can be decomposed into three parts:

$$\text{Total Error} = \underbrace{\text{Bias}^2}_{\text{underfitting}} + \underbrace{\text{Variance}}_{\text{overfitting}} + \underbrace{\text{Irreducible Error}^2}_{\text{Noise}}$$

- **Bias**: Measures how far the model's predictions are from the true values.
- **Variance**: Measures the model's sensitivity to changes in the training data.
- **Irreducible error**: Represents the inherent noise in the data.