ASSIGNMENT-1

Task: Variance and bias (diagram, overfit, underfit)- For best fit model should we have low bias or high variance, low bias or low variance, high bias or high variance, low bias or high variance

1. Introduction

Machine learning models are designed to learn patterns from data and make predictions on unseen inputs. However, achieving high accuracy on both training and testing data is challenging. Two fundamental concepts that explain prediction errors in machine learning models are **bias** and **variance**.

Bias represents errors due to overly simplistic assumptions, while variance represents errors due to excessive sensitivity to training data. Understanding the relationship between bias and variance helps in identifying problems such as **underfitting** and **overfitting**, and in building an optimal or best-fit model.

This project explains:

- Bias and variance
- Underfitting and overfitting
- Conditions for a best-fit model

2. Bias and Variance in Machine Learning

Bias

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

Key characteristics of high bias:

- Model is too simple
- Fails to capture complex patterns
- Leads to underfitting

Variance

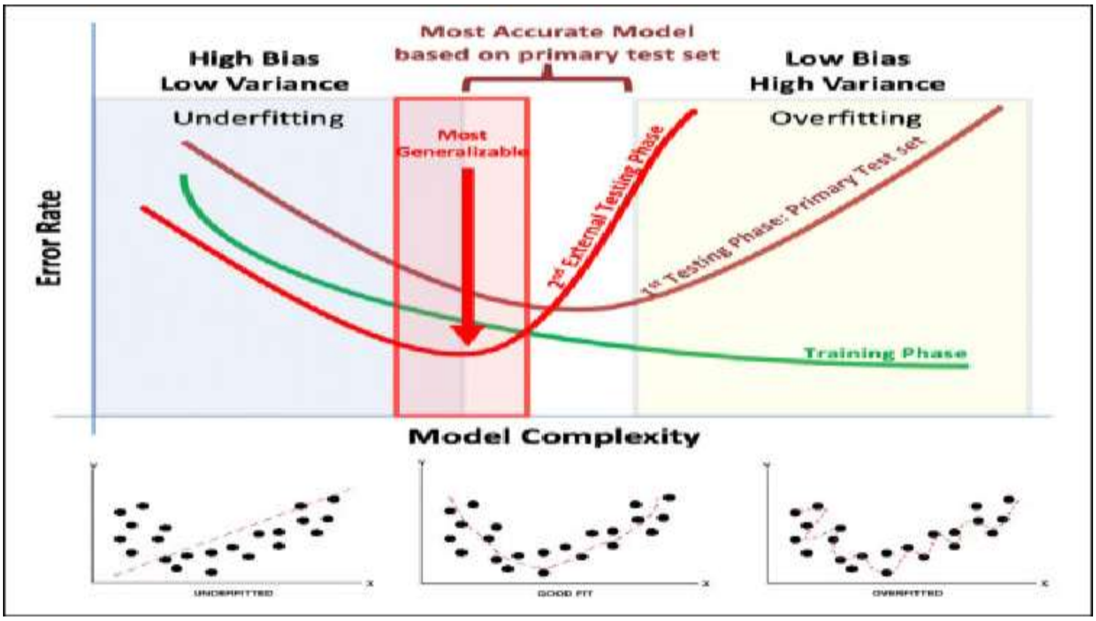
Variance refers to the error caused by a model's sensitivity to small changes in the training data.

Key characteristics of high variance:

- Model is too complex
- Learns noise instead of pattern
- Leads to overfitting

Diagram 1: Illustration of bias and variance in machine learning models.

This diagram visually explains how bias and variance affect model predictions relative to the true target.



3. Underfitting

Definition

Underfitting occurs when a model is **too simple** to learn the underlying structure of the data.

Causes

- High bias
- Insufficient model complexity
- Inadequate feature representation

Characteristics

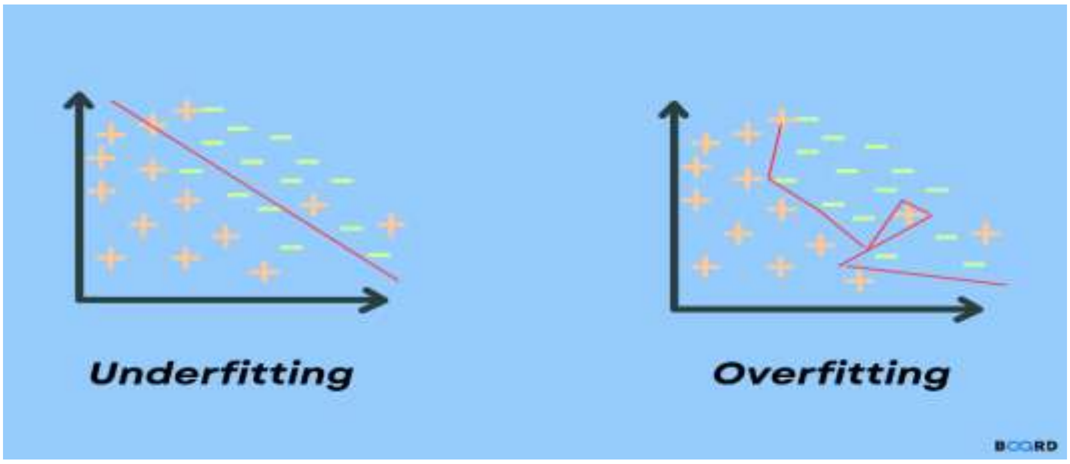
- High training error

- High testing error
- Poor predictive performance

Bias–Variance Condition

High Bias and Low Variance

Diagram 2: Underfitting Representation



4. Overfitting

Definition

Overfitting occurs when a model learns the training data **too well**, including noise and random fluctuations.

Causes

- Excessive model complexity
- Too many parameters
- Small training dataset

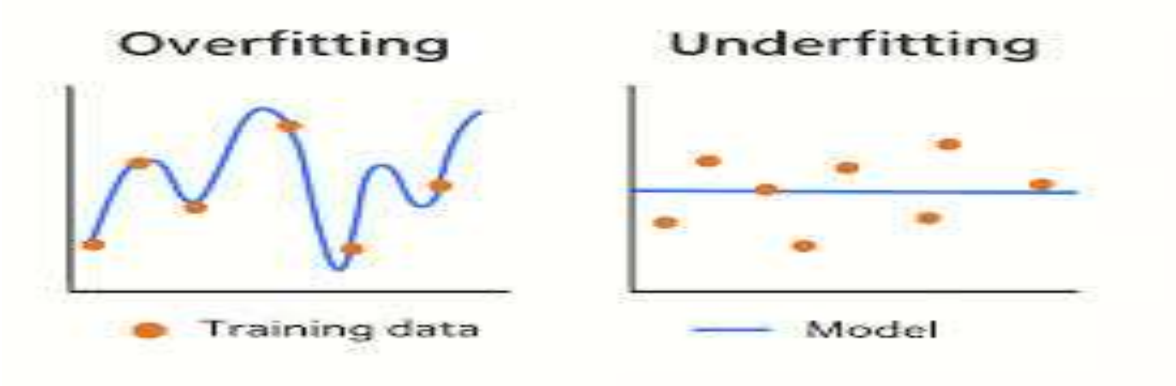
Characteristics

- Very low training error
- High testing error
- Poor generalization

Bias–Variance Condition

Low Bias and High Variance

Diagram 3: Overfitting Representation



5. Best Fit Model (Optimal Model)

Definition

A best-fit model successfully learns the underlying data pattern and generalizes well to unseen data.

Characteristics

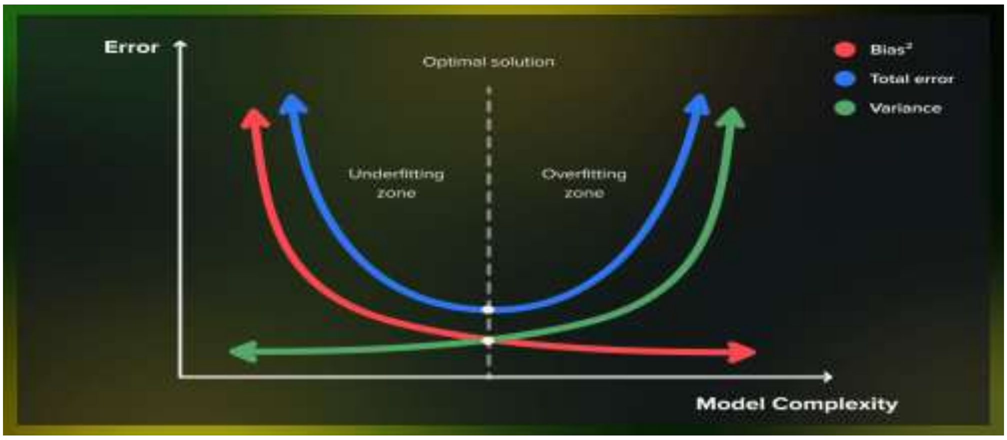
- Low training error
- Low testing error
- Stable and reliable predictions

Bias–Variance Condition

✅ **Low Bias and Low Variance**

This balance ensures that the model is neither too simple nor too complex.

Diagram 4: Bias–Variance Trade off Curve



6. Comparison of Bias–Variance Scenarios

Bias	**Variance**	**Model Behaviour**	**Result**
High	Low	Too simple	Underfitting
Low	High	Too complex	Overfitting
High	High	Unstable	Poor model
Low	**Low**	Balanced	**Best fit**

8. Conclusion

This project demonstrates that the performance of a machine learning model is fundamentally governed by the balance between **bias** and **variance**. Models with **high bias** fail to capture underlying data patterns and result in **underfitting**, while models with **high variance** become overly sensitive to training data and lead to **overfitting**. Both conditions produce poor generalization and are unsuitable for real-world applications.

The analysis confirms that an **optimal or best-fit model** is achieved only when **both bias and variance are low**. This balanced condition minimizes total prediction error and enables the model to learn meaningful patterns while remaining stable when exposed to unseen data. The bias–variance tradeoff curve further validates that neither extremely simple nor extremely complex models provide optimal performance.

Final Result:

The most effective machine learning models are those that maintain **low bias and low variance**, ensuring high accuracy, robustness, and reliable generalization. Understanding and managing the bias–variance tradeoff is therefore essential for designing efficient and practical machine learning systems.