SRINIVAS UNIVERSITY INSTITUTE OF ENGINEERING AND TECHNOLOGY

MODULE 2

DATA ESSENTIAL,TYPES ,BIGDATA ,PROCEESING, AND ETHICS

ASSIGNMENT 2

CHAITRA MARUTI DEVADIG

USN : 01SU24CS031

CSE 'A'

4TH SEM

ASSIGNMENT 2

Variance and Bias in Machine Learning

Explain bias, variance, overfitting, and underfitting. Show these on a bias-variance diagram and identify the best-fit model.

1. Introduction

In machine learning, **bias** and **variance** are two important factors that affect the performance of a model.
They determine whether a model **learns properly**, **overfits**, or **underfits** the data.

> **Bias** → Error due to overly simple assumptions in the learning algorithm.

> **Variance** → Error due to too much sensitivity to training data.

A good model must maintain a **balance between bias and variance**, which is called the **best fit**.

2. Bias in Machine Learning

Bias measures how far the model's predictions are from the actual values.

Types:

High Bias

Model is too simple.

Cannot capture patterns in data

Leads to **underfitting**.

Low Bias

Model fits training data well.

Can capture complex relationships.

3. Variance in Machine Learning

Variance measures how much the model's predictions change when the training data changes.

Types:

High Variance

Model is very sensitive to training data.

Fits noise instead of pattern.

Leads to **overfitting**.

Low Variance

Model predictions are stable.

Generalizes well to new data.

4. Overfitting

Overfitting occurs when a model learns the training data too well, including noise and unnecessary details.

Characteristics:

Very good performance on **training data**.

Poor performance on **test data**.

Model becomes too complex.

Bias–Variance condition:

✔ **Low Bias**
✔ **High Variance**

So correct relation:
Overfitting → Low Bias, High Variance

5. Underfitting

Underfitting occurs when a model is too simple to learn the patterns in the data.

Characteristics:

Poor performance on **training data** and **test data**.

Model cannot capture relationships.

Bias–Variance condition:

✔ **High Bias**
✔ **Low Variance**

So correct relation:
Underfitting → High Bias, Low Variance

6. Best Fit Model (Ideal Model)

A **best fit model** balances both bias and variance.

Properties:

Low Bias → Learns real patterns.

Low Variance → Generalizes well to new data.

Performs well on **both training and test data**.

So:
Best fit → Low Bias, Low Variance (balanced).