**Submitted by: Adithya Acharya**
**USN : 01SU24CS011**
**Branch: CSE**
**Section : A section**
**Subject: Fundamentals of AI and ML**
**Topic : Assignment 1**
**Submitted to : Prof. Mahesh Kumar**

# Variance and Bias (Diagram, Overfitting, Underfitting)

## 1. Introduction to Bias and Variance

In Machine Learning, one of the most important concepts for building a good predictive model is the Bias–Variance Tradeoff. When we train a model using data, we want it to perform well not only on training data but also on unseen data (test data). The model's error mainly comes from two sources:

1. Bias
2. Variance

**Understanding these two helps us avoid two major problems:**

1. Underfitting
2. Overfitting

The goal of a machine learning model is to find a balance between bias and variance so that it gives accurate predictions.

### 1. What is Bias?

Definition:

Bias is the error that occurs due to overly simple assumptions in the learning algorithm.

In simple words:

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

A high bias model is too simple and does not capture the patterns in data properly.

**Characteristics of High Bias:**

1. Oversimplified model
2. Misses important patterns
3. High training error

Example:

If we try to fit a straight line to highly curved data, the model will not capture the curve properly. This is high bias.

2. **What is Variance?**

Definition:

Variance is the error that occurs due to too much sensitivity to small changes in training data.

In simple words:

Variance measures how much the model's predictions change when trained on different data samples.

A high variance model learns too much from training data, including noise.

**Characteristics of High Variance:**

1. Very complex model
2. Fits training data perfectly
3. Low training error
4. High testing error
5. Leads to Overfitting

Example:

If we use a very complex polynomial curve that passes through every training point exactly, it may perform poorly on new data. This is high variance.

## 4. Underfitting (High Bias, Low Variance)

Definition:

Underfitting occurs when the model is too simple to capture the underlying structure of the data.

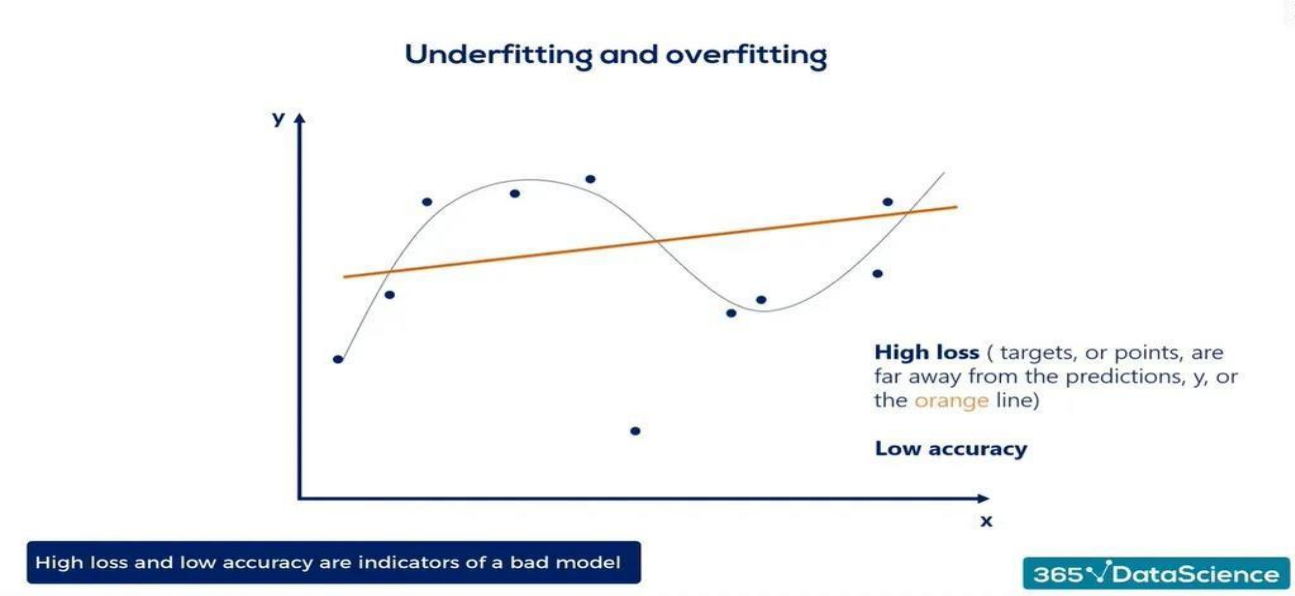
**Causes:**

1. Model is too simple
2. Not enough training
3. Poor feature selection

**Effects:**

1. High training error
2. High testing error
3. Poor performance

**Graphical Representation**




The straight line does not follow the data pattern. Key Point:
Underfitting = High Bias + Low Variance

5.Overfitting (Low Bias, High Variance)

Definition:

Overfitting occurs when the model learns training data too well, including noise and outliers.

**Causes:**

1.Model too complex

2.many parameters

3.Small dataset

**Effects:**

1.Very low training error
2,High testing error

3.Poor generalization

**Best Fit Model (Ideal Case)**

The best model is one that balances bias and variance.

It:

1,Captures real patterns

2.Ignores noise

3.Performs well on new data

**Ideal Condition:**

Low Bias + Low Variance

**This gives:**

1.Low training error

2.Low testing error

3.Good generalization

## 6. Bias-Variance Tradeoff Diagram

The Bias-Variance tradeoff can be explained using a U-shaped curve.

Explanation:

1. Left side: Model too simple → High Bias → Underfitting
2. Right side: Model too complex → High Variance → Overfitting
3. Middle point: Optimal Model → Low Bias + Low Variance

As model complexity increases:

1. Bias decreases
2. Variance increases

Total error = $Bias^2$ + Variance + Irreducible error

### 1. Comparison Table

| Condition | Bias | Variance | Training Error | Testing Error | Problem Type |
|---|---|---|---|---|---|
| underfitting | High | Low | High | High | Too simple model |
| overfitting | Low | High | Low | High | Too complex model |
| Best fit | Low | Low | Low | Low | Balanced model |

## 2. How to Reduce Bias and Variance

To Reduce High Bias (Underfitting):

1. Use more complex model
2. Add more features
3. Train longer
4. Reduce regularization

## 2. Real-World Example

Consider predicting house prices.

If we use a very simple linear model ignoring many factors → Underfitting (High Bias).

If we use a very complex 20-degree polynomial → Overfitting (High Variance).

If we use a balanced regression model considering important features → Best Fit.

## Conclusion

Bias and Variance are two fundamental sources of error in machine learning models. A model with high bias underfits the data, while a model with high variance overfits the data. The key goal in model building is to achieve a balance between these two, known as the Bias-Variance Tradeoff.

The best model is achieved when we maintain:

Low Bias

Low Variance

This ensures:

Good learning of data patterns

Strong generalization ability

Accurate predictions on unseen data

Understanding bias and variance is essential for designing efficient machine