

Bias and Variance in Machine Learning

Problem Statement: 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.

Introduction

In Machine Learning, model performance depends on two important concepts called Bias and Variance. Understanding these concepts helps us build models that generalize well on unseen data.

The goal of every machine learning model is to achieve a balance between bias and variance, known as the Bias–Variance Tradeoff.

Bias:

Bias refers to the error introduced when a model makes overly simple assumptions about the data.

Characteristics of High Bias

- Model is too simple
- Cannot capture complex patterns
- Leads to underfitting
- High training error
- High testing error

Example

Using a straight-line model for data that actually follows a curved pattern.

Variance:

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

Characteristics of High Variance

- Model is too complex
- Learns noise from training data
- Performs poorly on new data
- Very low training error
- High testing error

Example

A model that tries to pass through every data point.

Underfitting

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

Properties

High Bias

- Low Variance
- Poor performance on training data
- Poor performance on testing data

Reason

Model complexity is insufficient.

Overfitting

Overfitting occurs when the model learns both patterns and noise in the training data.

Properties

Low Bias

High Variance

Excellent training accuracy

Poor testing accuracy

Reason

Model complexity is too high.

Best Fit Model

A Best Fit Model balances bias and variance effectively.

Properties

- Low Bias
- Low Variance
- Good generalization ability
- -Performs well on unseen data

This is the desired condition in Machine Learning.

Bias–Variance Tradeoff

- Increasing model complexity:

Decreases Bias

Increases Variance

- Decreasing model complexity:

Increases Bias

Decreases Variance

The goal is to find an optimal balance between them.

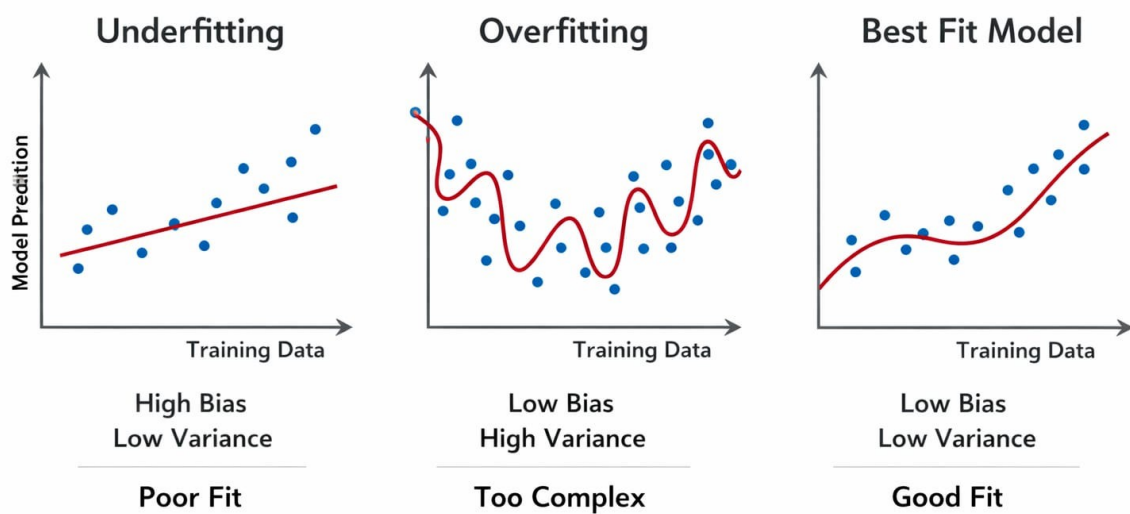


Diagram Explanation

The diagram shows three situations:

1. Underfitting – Model is too simple and fails to capture patterns.
2. Overfitting – Model is too complex and learns noise.
3. Best Fit Model – Model captures the true relationship without over-learning.

Conclusion

**“For an ideal machine learning model, we must achieve:
Low Bias and Low Variance”**

This balance ensures better prediction accuracy and model generalization.

