ASSIGNMENT

Variance and Bias in Machine Learning

Bias and Variance are two important sources of error in Machine Learning models. Understanding them helps us build models that generalize well to new data.

1. What is Bias?

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

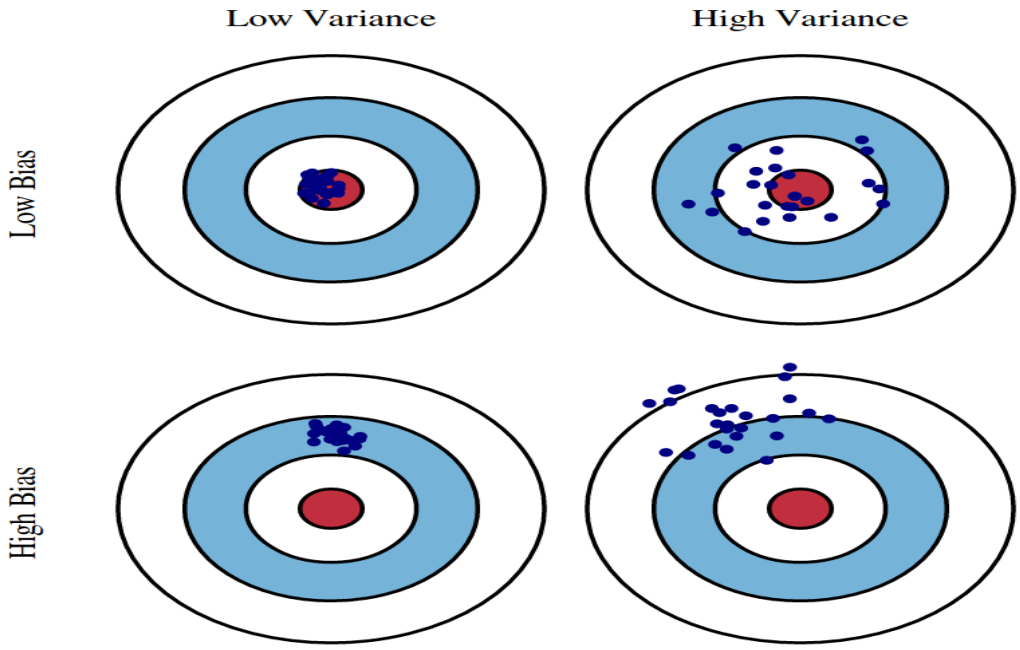
- High Bias → Model is too simple
- It cannot capture the underlying pattern
- Leads to **Underfitting**

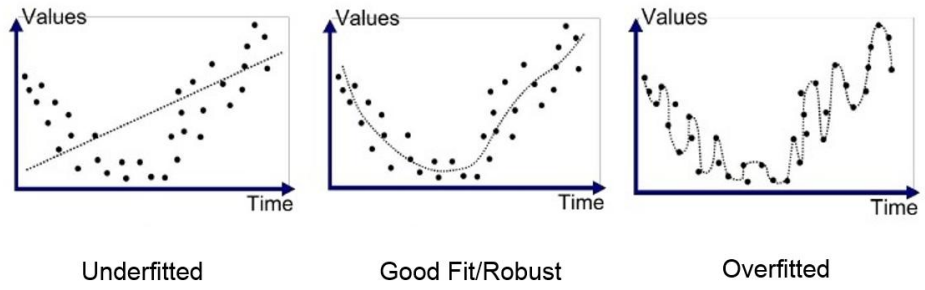
2. What is Variance?

Variance is the error caused due to excessive sensitivity to training data.

- High Variance → Model is too complex
- It fits noise in the training data
- Leads to **Overfitting**

3. Bias vs Variance Diagram





Explanation of Diagram:

1. **High Bias (Underfitting)**
    - Model too simple
    - High training error
    - High testing error
2. **High Variance (Overfitting)**
    - Model too complex
    - Low training error
    - High testing error

4. Underfitting

Underfitting happens when the model is unable to learn the pattern from data.

Characteristics:

- High Bias
- Low Variance
- Poor performance on training and testing data

5. Overfitting

Overfitting happens when the model memorizes training data including noise.

Characteristics:

- Low Bias
- High Variance
- Excellent training accuracy
- Poor testing accuracy

6. Bias–Variance Tradeoff

- Increasing model complexity ↓ Bias but ↑ Variance
- Decreasing model complexity ↑ Bias but ↓ Variance

The goal is to find a balance between them.

For a **best fit model**, we need:

Low Bias and Low Variance

Explanation:

- Low Bias → Model captures true pattern
- Low Variance → Model generalizes well to new data

Other combinations:

Bias	Variance	Result
High Bias	Low Variance	Underfitting
Low Bias	High Variance	Overfitting
High Bias	High Variance	Poor Model
Low Bias	Low Variance	✅ Best Model

7. Conclusion

A good machine learning model should not be too simple and not too complex.

The ideal model maintains a proper balance between bias and variance, minimizing both training and testing errors.

Thus, the best fit model has **Low Bias and Low Variance**.