

1. Variance and Bias (With Diagram, Overfitting & Underfitting Explanation)

Introduction to Bias and Variance

In Machine Learning, when we train a model, our main goal is to make it predict accurately on **new unseen data**. However, models can make two types of errors:

- **Bias Error**
- **Variance Error**

Understanding bias and variance helps us choose the **best fit model**.

What is Bias?

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

- High bias → Model is too simple
- It ignores important patterns
- Leads to **underfitting**

Example:

If we use a straight line to fit curved data, the model will not capture the pattern correctly.

Characteristics of High Bias:

- High training error
- High testing error
- Model is too simple

What is Variance?

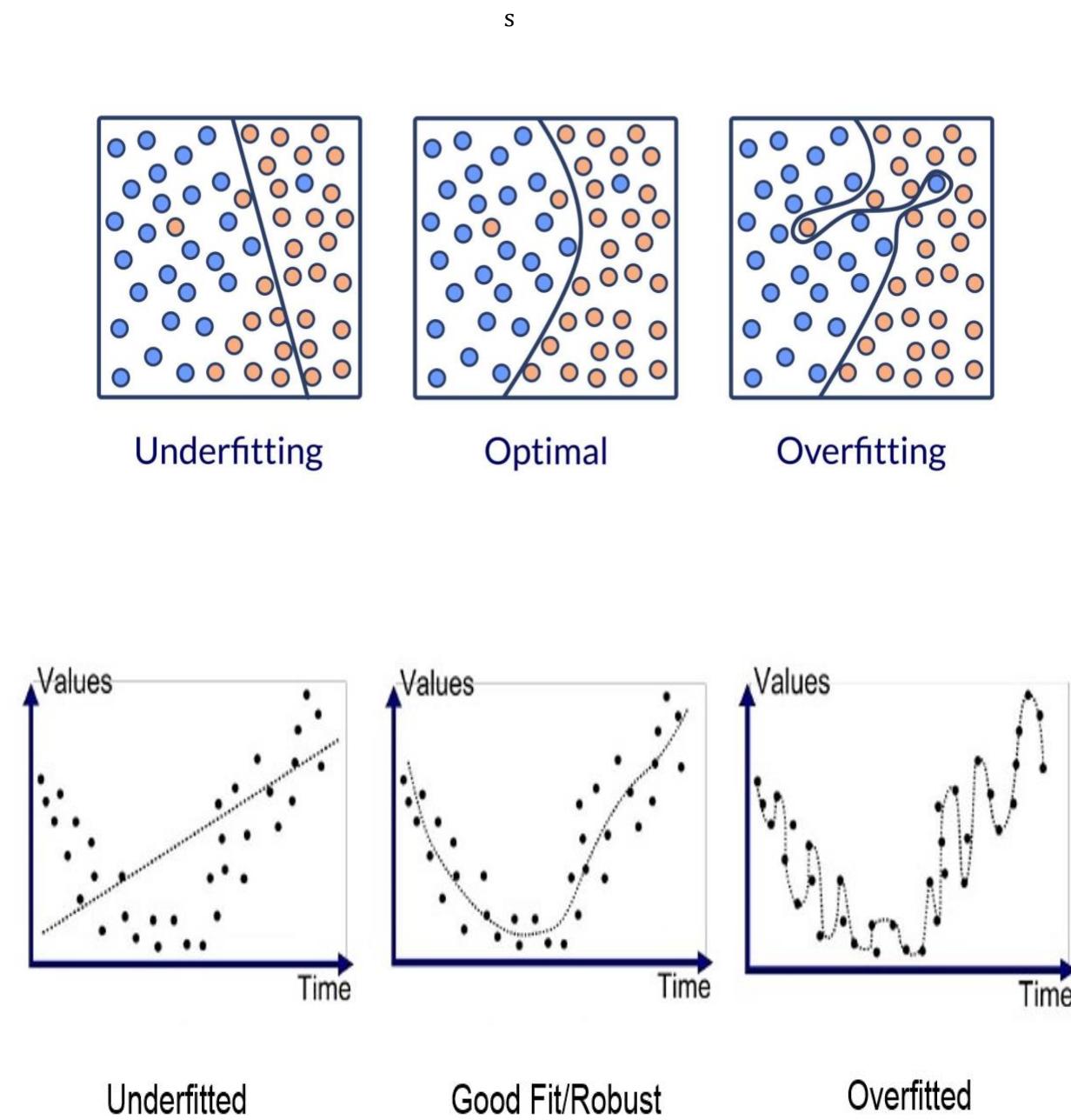
Variance is the error due to the model being too sensitive to small fluctuations in the training data.

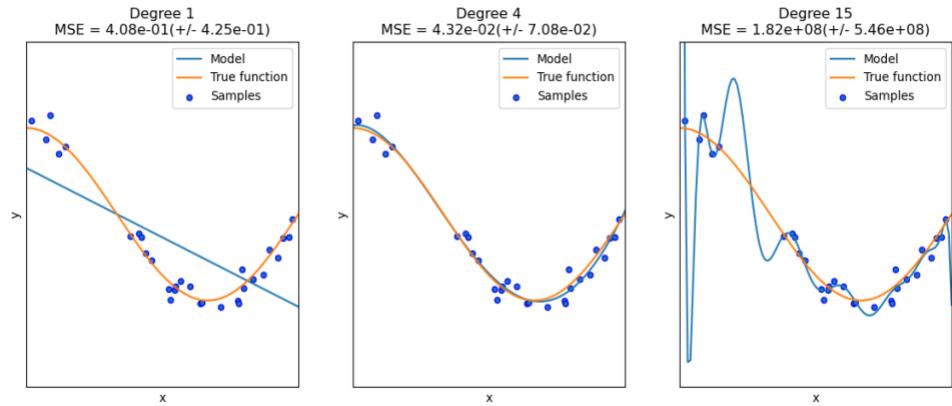
- High variance → Model is too complex
- It memorizes training data
- Leads to **overfitting**

Characteristics of High Variance:

- Very low training error
- Very high testing error
- Model is too complex

Underfitting (High Bias)





What is Underfitting?

Underfitting happens when:

- The model is too simple
- It cannot capture the underlying trend of data

Cause:

- High Bias
- Low Variance

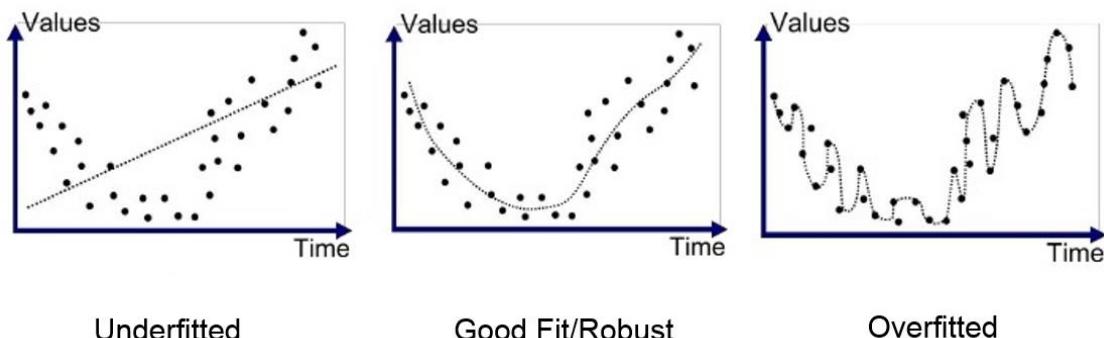
Example:

Using a linear model for non-linear data.

Result:

- Poor performance on both training and test data.

Overfitting (High Variance)



What is Overfitting?

Overfitting happens when:

- The model learns noise from training data
- It performs well only on training data

Cause:

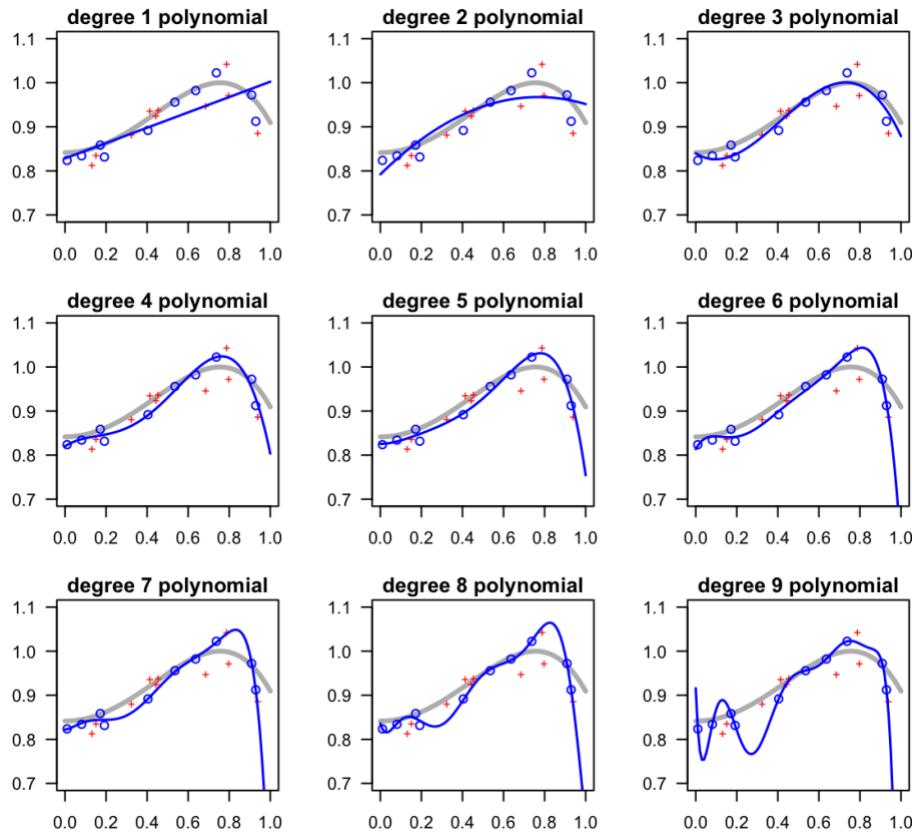
- Low Bias
- High Variance

Example:

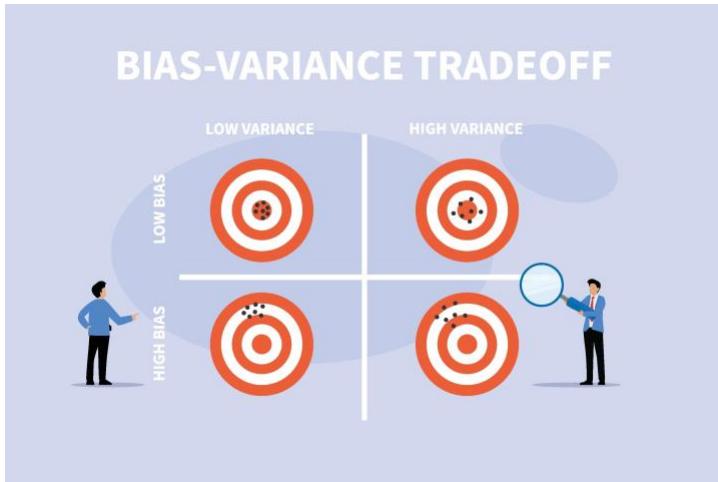
Using a very high-degree polynomial curve for simple data.

Result:

- Very low training error
- Very high testing error



Bias–Variance Tradeoff



There is a tradeoff between bias and variance.

- As model complexity increases:
 - Bias decreases
 - Variance increases

The goal is to find a balance.

Best Fit Model – What Should It Have?

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

Correct Answer:

Best Fit Model → Low Bias and Low Variance

Because:

- Low bias → Model captures true pattern
- Low variance → Model generalizes well to new data

This point lies in the middle of the bias-variance tradeoff curve.

Summary Table

Model Type	Bias	Variance	Problem
Underfitting	High	Low	Too simple
Overfitting	Low	High	Too complex
Best Fit Model	Low	Low	Balanced

Conclusion

Bias and variance are two important sources of error in machine learning models.

- High bias leads to underfitting.
- High variance leads to overfitting.
- The best model maintains a balance between both.

Therefore, **the best fit model should have Low Bias and Low Variance.**