Variance and Bias (Diagram, over fit, under fit) - 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 and add business problem too.

BIAS AND VARIANCE (Detailed Explanation)

1. What is Bias?

Bias is the error caused by **wrong assumptions** in the learning algorithm.

- Occurs when the model is too simple
- Fails to capture true relationship
- Leads to **Under fitting**

Mathematical Idea:

Bias measures difference between:
$[E[\hat{f}(x)] - f(x)]$

Where:

- ($f(x)$) = True function
- ($\hat{f}(x)$) = Predicted model

High Bias → Model predictions far from actual values.

2. What is Variance?

Variance measures how much model predictions change for different training datasets.

- Occurs when model is too complex
- Captures noise
- Leads to **Over fitting**

Mathematical Idea:

$[Var (\hat{f}(x))]$

High Variance → Model unstable across datasets.

Bias–Variance Tradeoff

Total Error =[Bias^2 + Variance + Irreducible Error]

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

Goal → Find optimal balance.

Diagram Explanation (Conceptual Curve)

Under fitting (High Bias)

- Model too simple (Linear model for nonlinear data)
- Training error: High
- Testing error: High

Graph shape:

- Straight line on curved data

Over fitting (High Variance)

- Model too complex (High-degree polynomial)
- Training error: Very Low
- Testing error: High
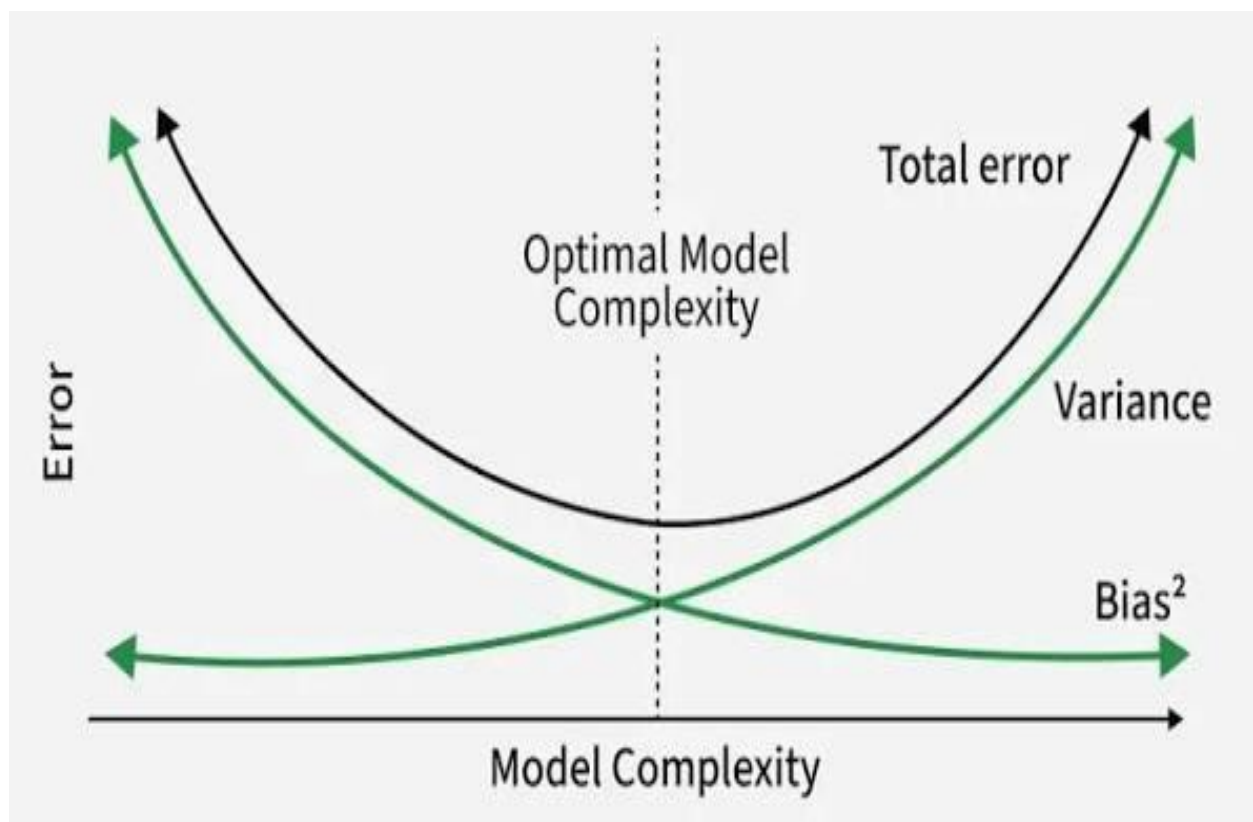
Graph shape:

- Very wiggly curve

Best Fit Model

- Balanced complexity
- Training error: Low
- Testing error: Low

Graph shape:

- Smooth curve following data trend

Comparison Table

Condition	Bias	Variance	Training Error	Testing Error	Problem
Under fitting	High	Low	High	High	Too Simple
Over fitting	Low	High	Very Low	High	Too Complex
Best Model	Low	Low	Low	Low	Balanced

Business Problem Examples

1. Healthcare – Disease Prediction

High Bias Model

- Uses only age to predict disease
- Ignores lifestyle, genetics
- Misses many cases
- Leads to incorrect treatment

High Variance Model

- Memorizes patient history
- Works only for past data
- Fails on new patients

Balanced Model

- Uses relevant features
- Predicts accurately
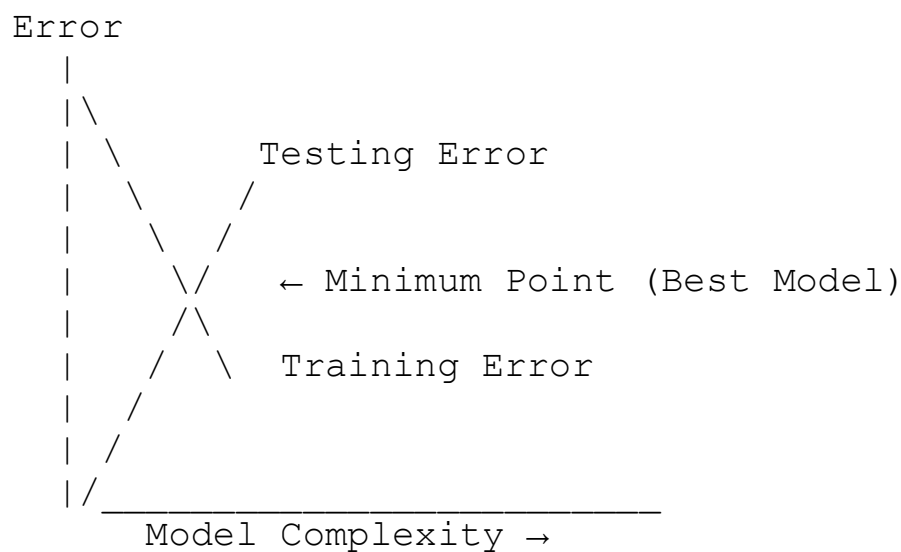- Improves patient care

2. Banking – Credit Risk

- High Bias → Approves risky customers
- High Variance → Rejects good customers randomly
- Balanced → Stable profit & low default rate

3.E-commerce – Recommendation System

- High Bias → Same products recommended to everyone
- High Variance → Recommendations unstable daily
- Balanced → Personalized & consistent suggestions

Error vs. Model Complexity Graph (Important for Exams)

```
Error
  |
  |\
  | \       Testing Error
  |  \    /
  |   \  /
  |    \/   ← Minimum Point (Best Model)
  |    /\
  |   /  \  Training Error
  |  /
  | /
  |/_____
        Model Complexity →
```

- Left side → Underfitting

- Right side → Overfitting
- Middle → Optimal

Final Answer (Direct Exam Answer)

For best fit model, we should have:

Low Bias and Low Variance

Because:

- High Bias causes underfitting
- High Variance causes overfitting
- Balanced model generalizes well