ASSIGNMENT-1

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

1.Bias

Definition

Bias is the error caused by wrong or overly simple assumptions in the learning algorithm.

- What it means

Model is too simple

Fails to capture true pattern

Underfits the data

- ❖ Characteristics
- ➢ High training error
- ➢ High testing error
- ➢ Model is stable across datasets
- ▪ Example

Using linear regression for highly nonlinear data.

- Underfitting = High Bias + Low Variance

2. Variance

- Definition

Variance is the error caused by the model being too sensitive to small changes in training data.

- What it means,

Model is too complex

Learns noise in training data

Overfits

- ❖ Characteristics
- ➢ Very low training error
- ➢ High testing error
- ➢ Large fluctuations if trained on different samples
- ▪ Example

Very deep decision tree or high-degree polynomial.

Overfitting = Low Bias + High Variance

3. Irreducible Error

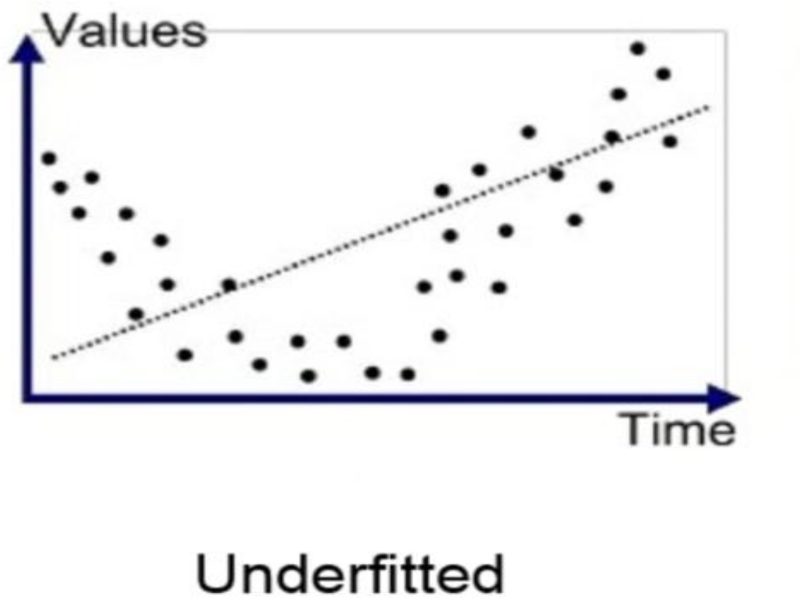
This is noise in the data:

- Measurement errors
- Randomness
- Missing features
- It cannot be reduced even with a perfect model.

Visual Intuition (Dartboard Example)

Imagine predicting a target value as throwing darts at a board:

- High Bias, Low Variance (Underfitting)



Darts tightly grouped

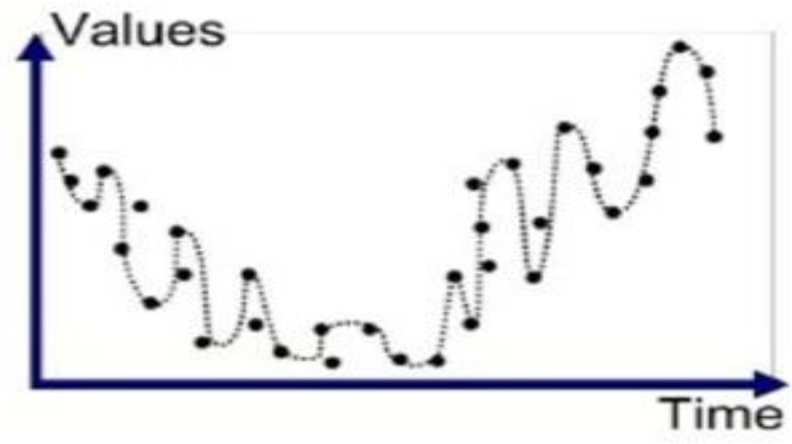
- ✓ But far from center

→ Consistently wrong

- Low Bias, High Variance (Overfitting)

- ✓ Darts scattered
- ✓ Average near center

→ Inconsistent predictions

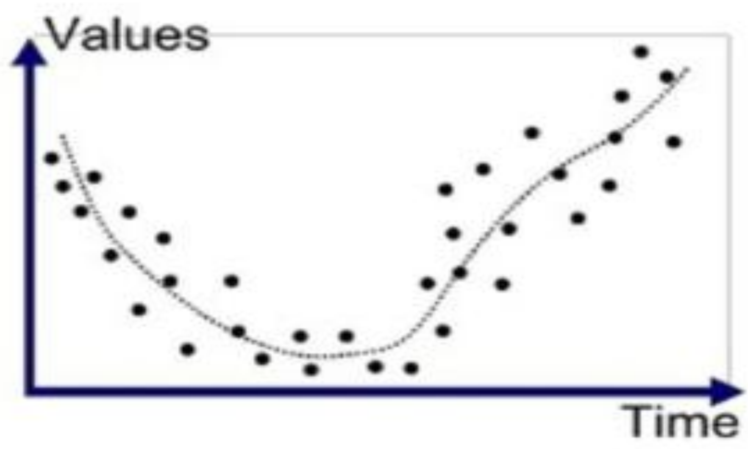


Overfitted

- ✓ Darts scattered
- ✓ Average near center

→ Inconsistent predictions

- ❖ Low Bias, Low Variance (Best Model)



Good Fit/Robust

Darts tightly grouped

Near center

→ Accurate and consistent

Graphical Explanation

As model complexity increases:

- ✓ Bias ↓
- ✓ Variance ↑

Validation error forms a U-shaped curve:

Left side → Underfitting

Middle → Optimal balance

Right side → Overfitting

Mathematical Intuition

If we denote:

True function = $f(x)$

Model prediction = $\hat{y}$

Then:

$$\mathbb{E}[(y - \hat{y})^2] =$$

$$(\text{Bias})^2 + \text{Variance} + \sigma^2$$

Where:

Bias^2 measures systematic error

Variance measures sensitivity

σ^2 is irreducible noise