SRINIVAS UNIVERSITY
INSTITUTE OF ENGINEERING & TECHNOLOGY
Srinivas Campus,Mukka,Surathkal,Mangaluru-574 146

"ASSIGNMENT"

Submitted by:

NAME	TOTADA CHANDRASHEKAR
USN	01SU23CS214

Subject:
Fundamentales of AI and ML

Subject code:

24SBT110

Submitted to:

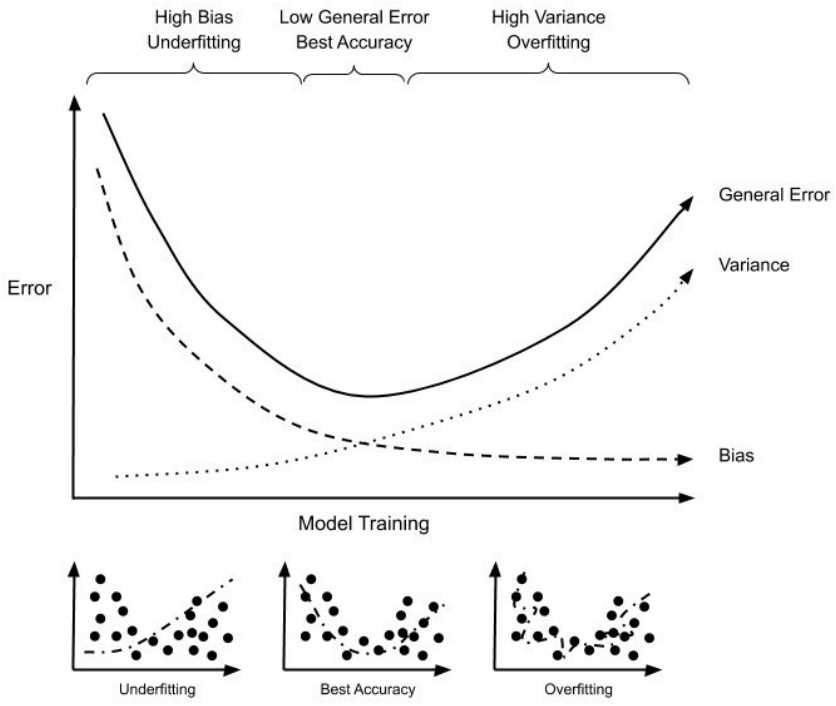
Prof,Mahesh Kumar

Academic Year 2026

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

Bias, Variance, Overfitting & Underfitting



Introduction

- Machine learning models aim to generalize well to unseen data.
- Two main sources of error: Bias and Variance.
- Understanding these helps to choose or design the best-fit model.

. Bias

- Bias is the error from oversimplifying the model.

- High bias → underfitting → model cannot capture patterns.
- Low bias → good pattern learning.
- Example: A straight line trying to fit curved data will perform badly.
- Effect: Both training and testing errors are high when bias is high.

Variance

- Variance is the error from model being too sensitive to training data.
- High variance → overfitting → model learns noise along with patterns.
- Low variance → stable predictions across different datasets.
- Example: A very complex curve fitting every data point in a small dataset.
- Effect: Low training error but high testing error

Underfitting vs Overfitting vs Best Fit

Model Type	Bias	Variance	What Happens	Result
Underfit	High	Low	Too simple	Poor performance
Overfit	Low	High	Too complex	Good training, bad test
Best Fit	Low	Low	Just right	Good generalization

- Underfitting: Model does not learn enough.
- Overfitting: Model learns noise and patterns.
- Best Fit: Balanced bias and variance → lowest error.

Bias-Variance Trade-off

- Total error = Bias^2 + Variance + Irreducible Error.

- Increasing model complexity: o Bias decreases o Variance increases
- Decreasing complexity: o Variance decreases o Bias increases
- Goal: Minimize total error → find the best-fit model.

For the best-fit model, should we have:

- Low bias AND low variance

Explanation :

- Low bias → model captures true patterns.
- Low variance → model generalizes well and avoids overfitting.
- Achieves the best performance on unseen data.

2. Take one Domain and draw the graph (Normal distribution) (Empirical rule)

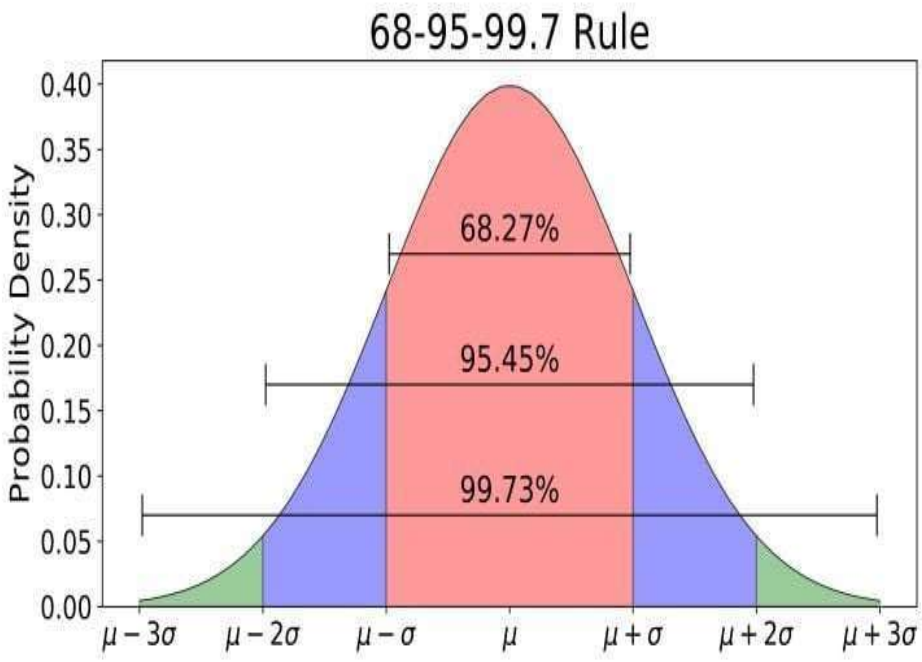
Domain – Student Exam Scores

Domain: Marks scored by students in a class (0–100 marks)

 Assume the scores are normally distributed.

- Mean (μ): 70
- Standard Deviation (σ): 10

Normal Distribution Graph



Empirical Rule (68-95-99.7 Rule)

- About 68% of students' scores lie within 1σ of the mean: 60–80 marks
- About 95% lie within 2σ of the mean: 50–90 marks
- About 99.7% lie within 3σ of the mean: 40–100 marks

- Normal distribution is bell-shaped and symmetric around the mean.
- Most data points fall close to the mean, fewer as you move away.
- The Empirical Rule helps in predicting the probability of values in the distribution.

How to Label Your Graph for Assignment

- X-axis: Student marks
- Y-axis: Probability density / frequency
- Mark mean (μ = 70) at the center
- Shade areas for 1σ, 2σ, 3σ ranges with 68%, 95%, 99.7% labels

3. If $\mu = 55$, $\sigma_{4a} = 4$, $\sigma_{4\beta} = 10$, $\sigma_4 = c$ 15, In this which is better

Problem

- Mean: $\mu = 55$
- Standard deviations: o $\sigma_{4a} = 4$ o $\sigma_{4\beta} = 10$ o $\sigma_4 c = 15$

Which is "better"?

Step 1: Understanding σ

- Standard deviation (σ) measures the spread of data around the mean.
- Smaller σ → data is tightly clustered around the mean → more consistent / predictable.
- Larger σ → data is spread out → more variability / less consistent.

Example:

- σ = 4 → most scores are close to 55
- σ = 10 → scores are more spread
- σ = 15 → scores vary a lot

Step 2: Use Empirical Rule (68-95-99.7%)

For μ = 55:

σ	68% range (±1σ)	95% range (±2σ)	99.7% range (±3σ)
4	51–59	47–63	43–67
10	45–65	35–75	25–85
15	40–70	25–85	10–100

Observation:

- Smaller σ → narrower range, scores are more predictable.
- Larger σ → wider range, scores are more spread out → less reliable.

Step 3: Which is "Better"?

Better means more consistent and less variability.

- $\sigma_{4_a} = 4 \rightarrow$ best, scores are tightly clustered around 55
- $\sigma_{4_\beta} = 10 \rightarrow$ moderate
- $\sigma_{4_C} = 15 \rightarrow$ worst, very spread out

Answer: $\sigma_{4_a} = 4$ is the best distribution

Step 4: Graphical Representation (Optional for Assignment)

Draw three bell curves with the same mean $\mu = 55$:

- σ_{4_a} = narrowest curve
- σ_{4_β} = medium curve
- σ_{4_C} = widest curve

- This visually shows that smaller σ = more concentrated / better.

4. Take one domain and build business Understanding

E-Commerce Customer Purchases

Domain: Online retail store – customer purchase data

1. Business Objective

- Understand customer purchasing behavior to:
  - Increase sales
  - Improve customer retention
  - Optimize marketing campaigns
- Example questions:
  - Which products are most popular?
  - Which customers are likely to make repeat purchases?
  - What is the average order value?

Key Metrics / KPIs

- Revenue per customer
- Frequency of purchase
- Average order value
- Customer segmentation (high-value vs low-value customers)

Business Understanding Steps

1. Identify the problem:
    - Increase sales by targeting high-value customers.
2. Determine data requirements:
    - Customer profile: ID, age, gender, location
    - Purchase history: date, product, quantity, price
    - Marketing interactions: emails, promotions
3. Define success criteria:
    - Increase repeat purchase rate by 20%
    - Increase average revenue per customer
4. Set AI/ML goal:
    - Predict which customers are likely to buy again
    - Segment customers based on purchase behavior

How AI Helps in This Domain

- Predictive Analytics:
    - AI models (e.g., logistic regression, decision trees, random forest) can predict which customers are likely to make repeat purchases.
- Customer Segmentation:

    - AI clustering algorithms (e.g., K-Means, DBSCAN) can group customers into high-value, medium-value, and low-value segments.
- Personalized Recommendations:
    - AI-based recommendation systems suggest products to customers based on past purchases and preferences, increasing cross-selling and upselling.
- Churn Prediction:
    - AI detects customers at risk of leaving and helps in creating targeted retention campaigns.
- Marketing Optimization:
    - AI helps in deciding the best offer or promotion for each customer, maximizing ROI on campaigns.

Example Output / Insights Expected

- Segmentation: High-value, medium-value, low-value customers Predicted repeat buyers: Identify who is likely to purchase again
- Recommendations:
    - Personalized promotions to high-value customers o Discounts for medium-value customers to increase engagement o Reduce marketing spend on low-value customers