BIAS AND VARIANCE

THE HEART OF MACHINE LEARNING ACCURACY

PRESENTATION BY
KASI RAJAN

BIAS AND VARIANCE IN MACHINE LEARNING

Model Evaluation & The Problem of Error

WE EVALUATE MACHINE LEARNING MODELS USING METRICS LIKE MSE, PRECISION, RECALL, AND ROC.

BIAS AND VARIANCE ARE CRUCIAL CONCEPTS THAT HELP US UNDERSTAND THE SOURCE OF MODEL ERROR, GUIDING PARAMETER TUNING AND MODEL SELECTION.

WHAT IS BIAS?

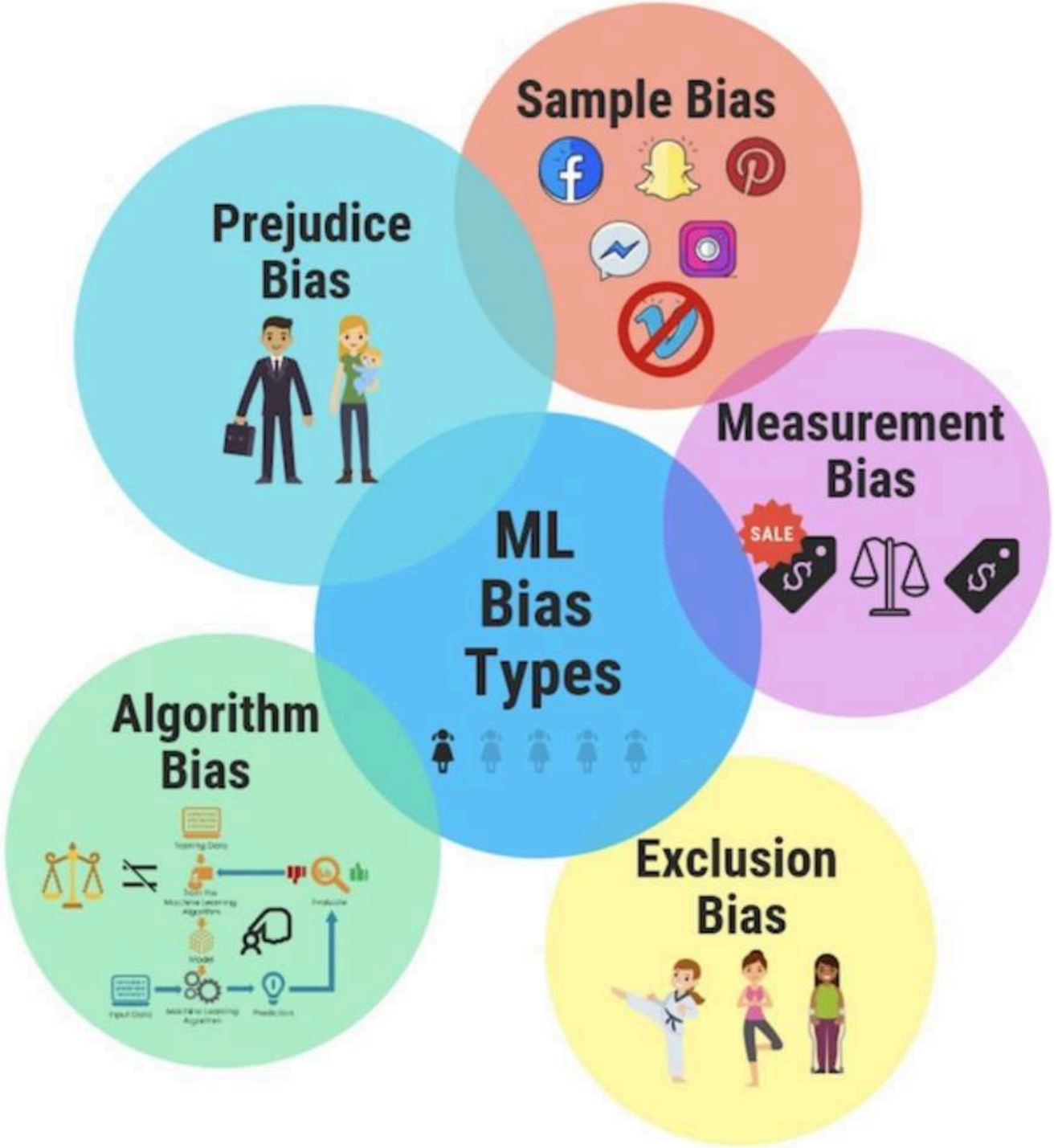
ERROR DUE TO OVERSIMPLIFICATION.

- THE ERROR FROM WRONG ASSUMPTIONS.

- CAUSED BY AN OVERLY SIMPLISTIC MODEL THAT FAILS TO CAPTURE THE UNDERLYING PATTERNS IN THE DATA.

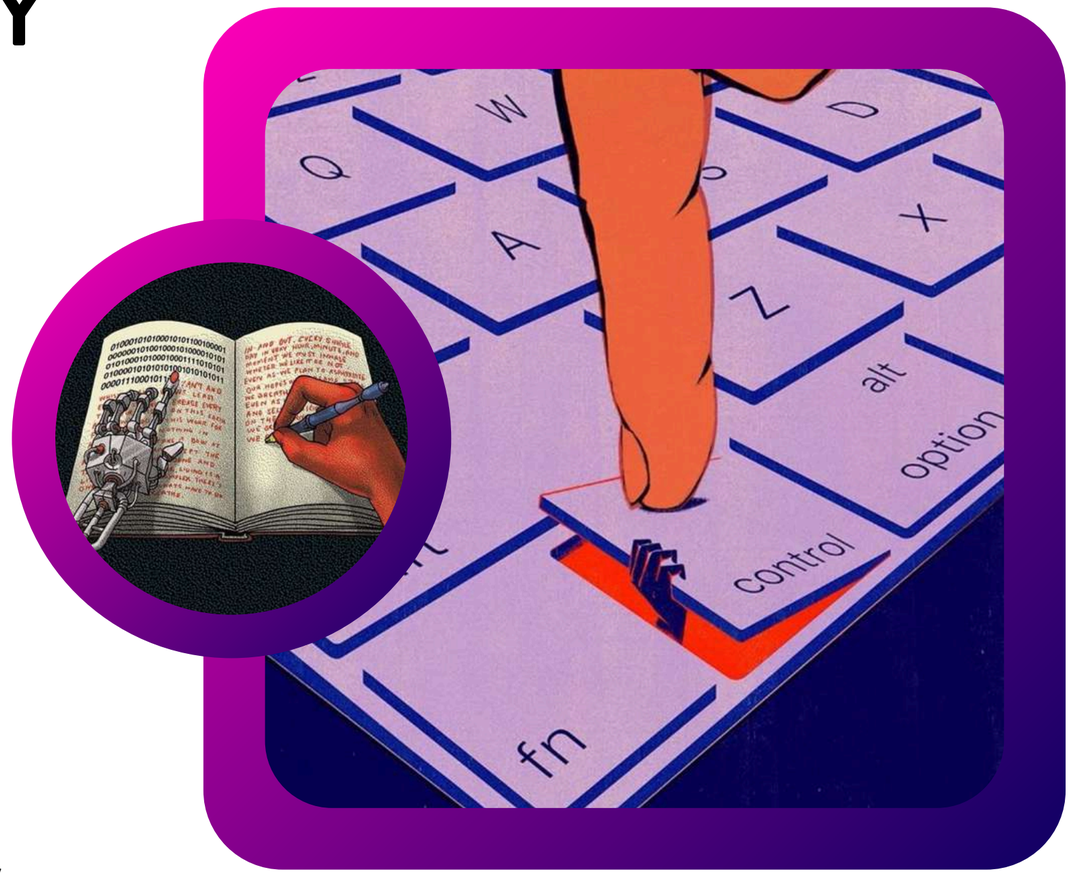
- EXAMPLE: ASSUMING DATA IS LINEAR WHEN IT ACTUALLY FOLLOWS A COMPLEX FUNCTION.

- LEADS TO: UNDERFITTING.



WHAT IS VARIANCE?

ERROR DUE TO SENSITIVITY TO TRAINING DATA.



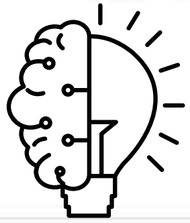
- THE ERROR FROM SENSITIVITY TO NOISE.

- CAUSED BY AN OVERLY COMPLEX MODEL THAT LEARNS THE TRAINING DATA TOO WELL, INCLUDING ITS NOISE AND RANDOM FLUCTUATIONS.

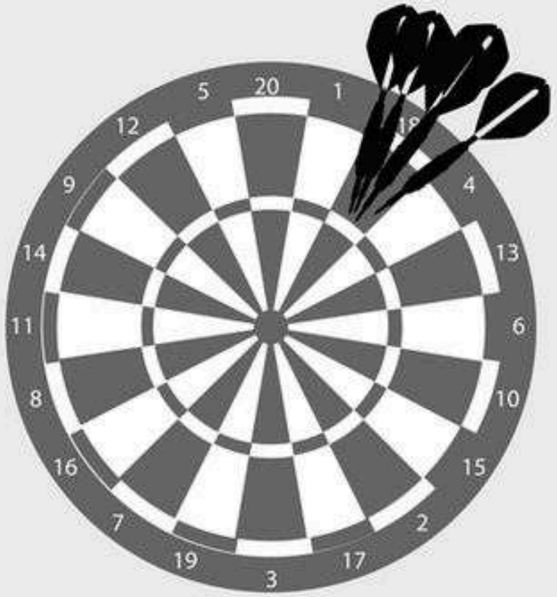
- LEADS TO: OVERFITTING.

Bias vs. Variance: Key Differences

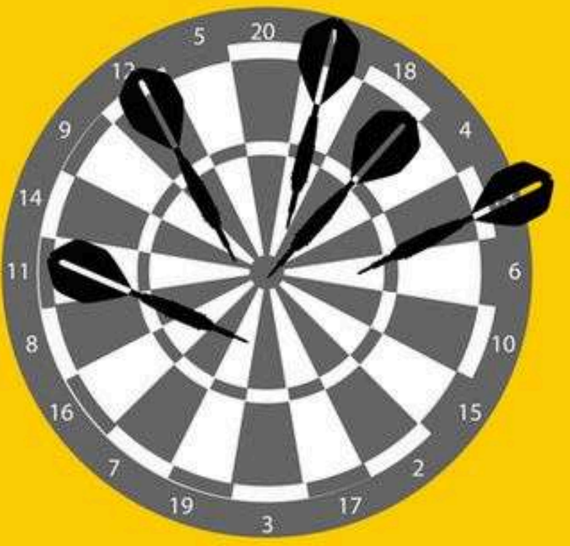
Feature	Bias	Variance
Core Concept	Error from **oversimplifying** the problem	Error from **overcomplicating** the problem
Model Behavior	Makes strong **incorrect assumptions** about the data	**Overreacts** to minor fluctuations in the training data
Resulting Problem	✅ **Underfitting** • Model is too simple • High error on training & test data	✅ **Overfitting** • Model is too complex • Low error on training, high error on test data
Impact on Accuracy	**Consistently inaccurate** (Misses the mark)	**Inconsistently inaccurate** (Wildly different results)
How to Reduce	• Use a more complex model • Add more features • Reduce regularization	• Use a simpler model • Get more training data • Increase regularization



High Bias
Low Variance



High Variance
Low Bias



The Goal: Find the perfect balance to minimize total error!

Bias-Variance Tradeoff

In machine learning, the bias-variance tradeoff describes the relationship between the complexity of a model and its ability to generalize.

High Bias

A high bias model is too simple and cannot capture the underlying patterns in the data.

High Variance

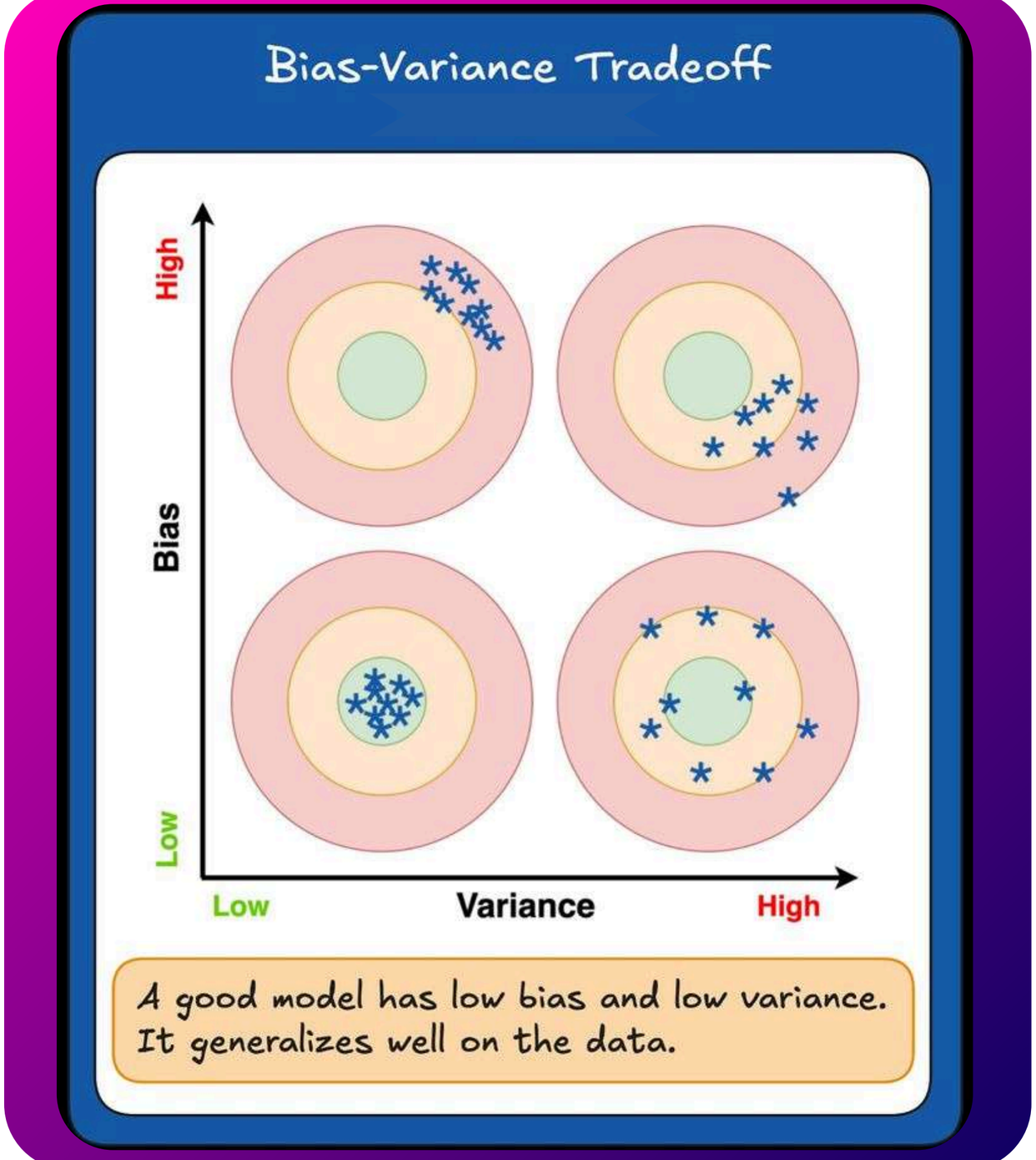
A high variance model is too complex and overfits the training data, leading to poor generalization.

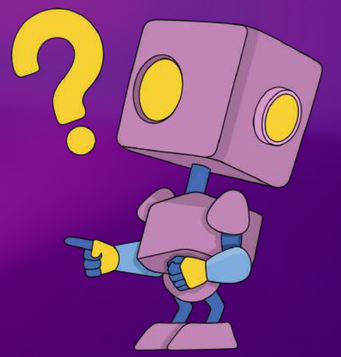
Finding Balance

The goal is to find a model with an optimal balance between bias and variance, resulting in good generalization.

Model Mastery

Understanding the bias-variance tradeoff is key to building effective machine learning models that generalize well to new data.





DIFFERENT COMBINATIONS OF BIAS & VARIANCE

UNDERSTANDING MODEL PERFORMANCE THROUGH THE BIAS-VARIANCE LENS

1. HIGH BIAS, LOW VARIANCE

- ✅ UNDERFITTING
- THE MODEL IS TOO SIMPLE AND MAKES STRONG INCORRECT ASSUMPTIONS.
- IT IS CONSISTENTLY WRONG BUT STABLE ACROSS DIFFERENT DATASETS.
- RESULT: POOR PERFORMANCE ON BOTH TRAINING AND TEST DATA.

2. LOW BIAS, HIGH VARIANCE

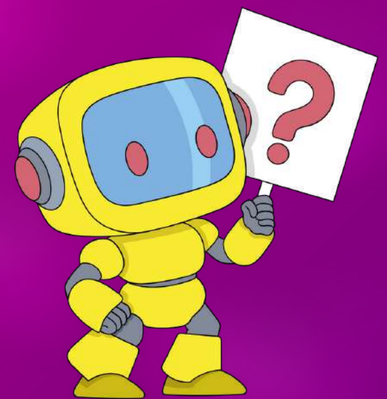
- ✅ OVERFITTING
- THE MODEL IS TOO COMPLEX AND MEMORIZES THE TRAINING DATA NOISE.
- IT IS SENSITIVE TO SMALL FLUCTUATIONS IN THE TRAINING SET.
- RESULT: EXCELLENT ON TRAINING DATA, POOR ON TEST DATA.

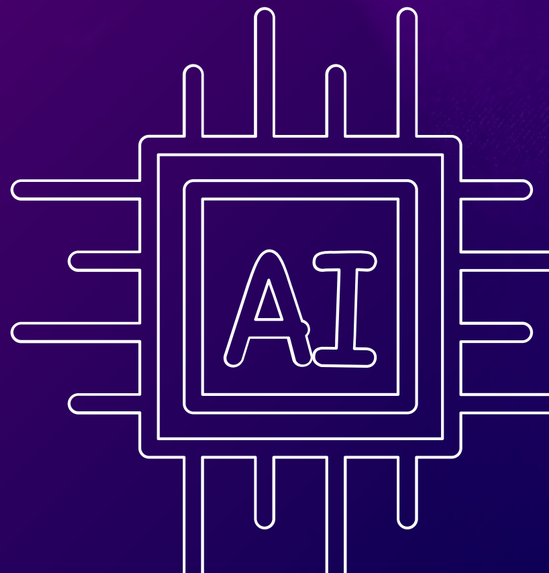
3. HIGH BIAS, HIGH VARIANCE

- ❌ WORST CASE SCENARIO
- THE MODEL FAILS TO LEARN PATTERNS AND IS HIGHLY UNSTABLE.
- PRODUCES INCONSISTENT AND INACCURATE PREDICTIONS.
- RESULT: POOR PERFORMANCE EVERYWHERE WITH HIGH INSTABILITY.

4. LOW BIAS, LOW VARIANCE

- 🎯 IDEAL SCENARIO
- THE MODEL CAPTURES TRUE PATTERNS WITHOUT MEMORIZING NOISE.
- GENERALIZES WELL TO NEW, UNSEEN DATA.
- RESULT: CONSISTENT AND ACCURATE PREDICTIONS (THE ULTIMATE GOAL!).






THANK YOU