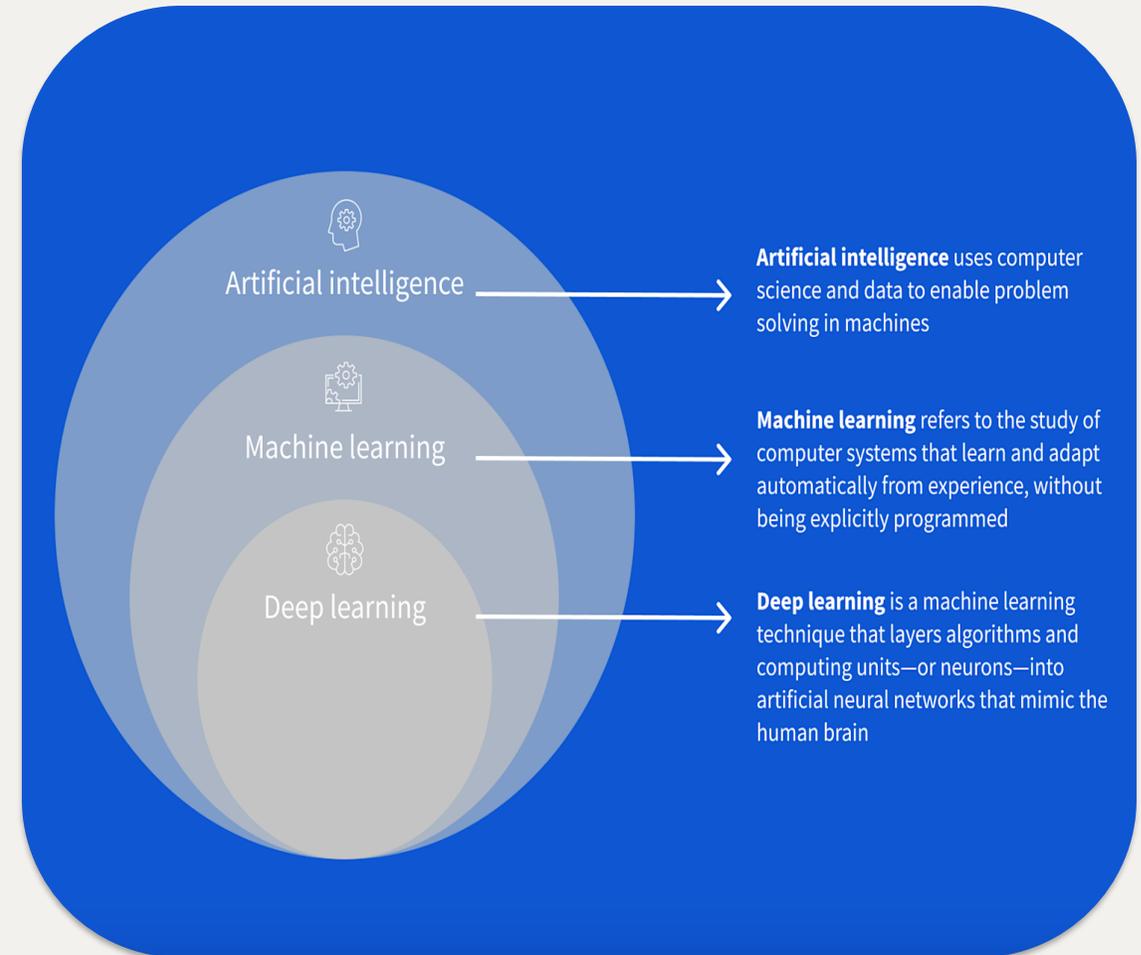**Data Analytics Series**

Data Analytics:

Logistic Regression

23MCT305 - Data Analytics in Automation System

N. KARTHI, AP/MCT

Introduction to Logistic Regression

- Classification algorithm for binary outcomes

- Predicts probability between 0 and 1

- Widely used in data analytics and machine learning

- Foundation for understanding complex classification tasks

Artificial intelligence uses computer science and data to enable problem solving in machines

Machine learning refers to the study of computer systems that learn and adapt automatically from experience, without being explicitly programmed

Deep learning is a machine learning technique that layers algorithms and computing units—or neurons—into artificial neural networks that mimic the human brain

Linear vs Logistic Regression
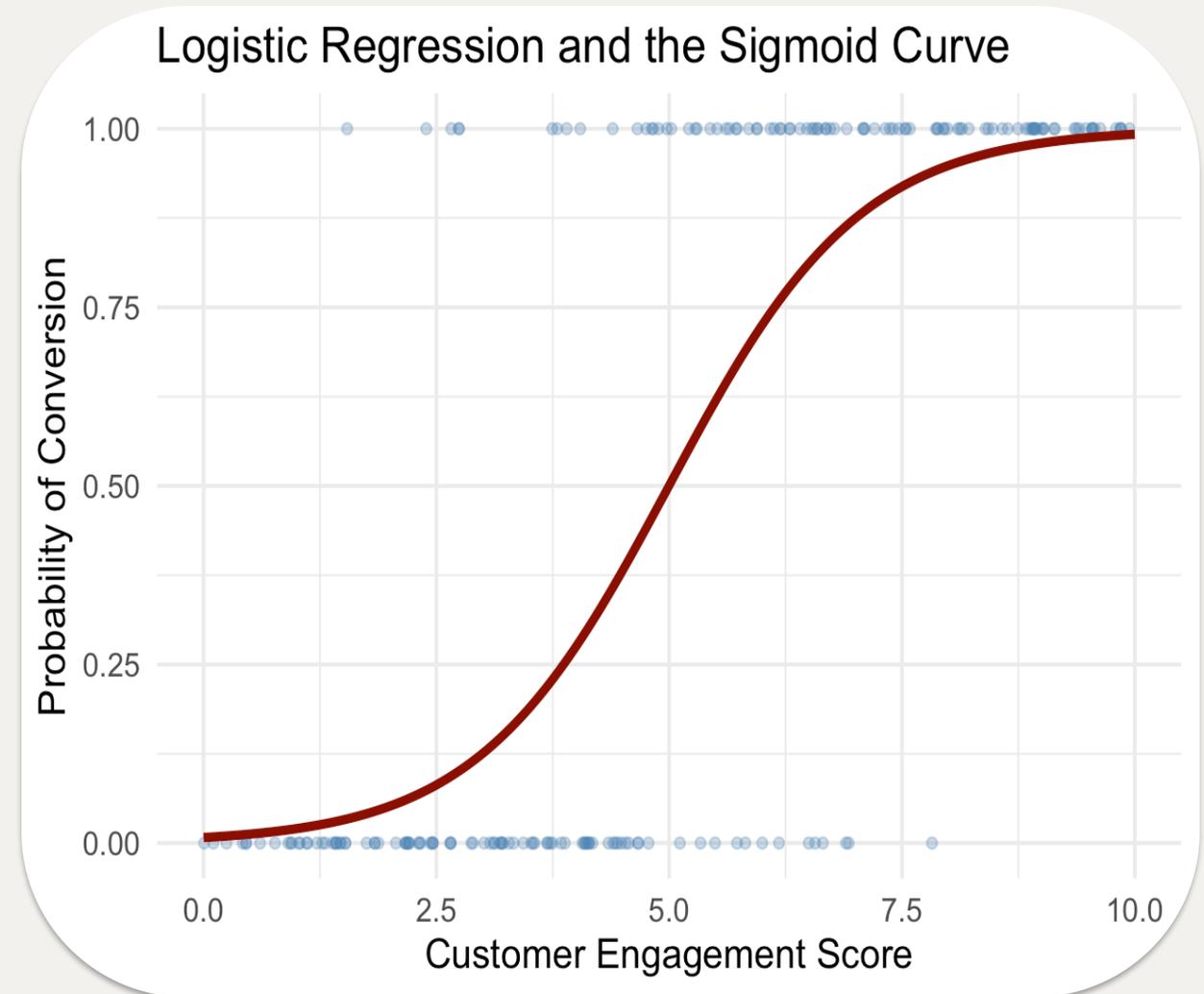
Linear Regression

- Predicts **continuous values**

- Output can be **any real number**

- Linear relationship

- Best fit line minimizes error
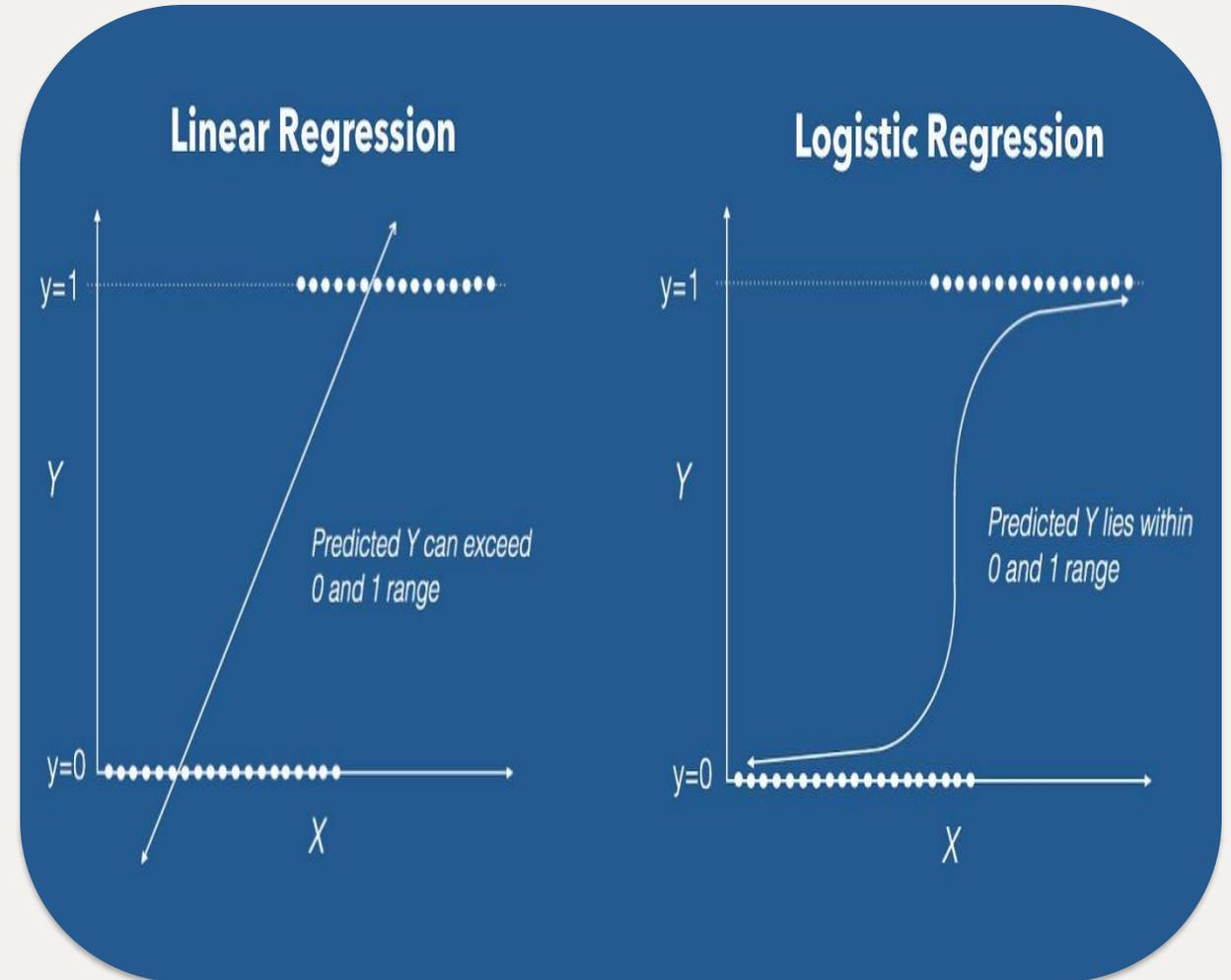
Understanding the Sigmoid Function

$$S(z) = 1 / (1 + e_{-z})$$

- Maps any input to range **[0, 1]**

- S-shaped (sigmoidal) curve

- Threshold at 0.5 for classification

- Asymptotes at **0 and 1**

Logistic Regression and the Sigmoid Curve

Types of Logistic Regression

- **Binary Logistic Regression**
  - Two possible outcomes (**Yes/No**)
  - Example: Spam detection

- **Multinomial Logistic Regression**
  - Three or more unordered outcomes
  - Example: Predicting product type

- **Ordinal Logistic Regression**
  - Three or more ordered outcomes
  - Example: Customer rating scale

Types of Logistic Regression Models

	Binomial Logistic Regression	Multinomial Logistic Regression	Ordinal Logistic Regression
Number of Categories for Response Variable	2	3 or more	3 or more
Does Order of Categories Matter?	No	No	Yes

Key Concepts & Terminology

- **Decision Boundary**
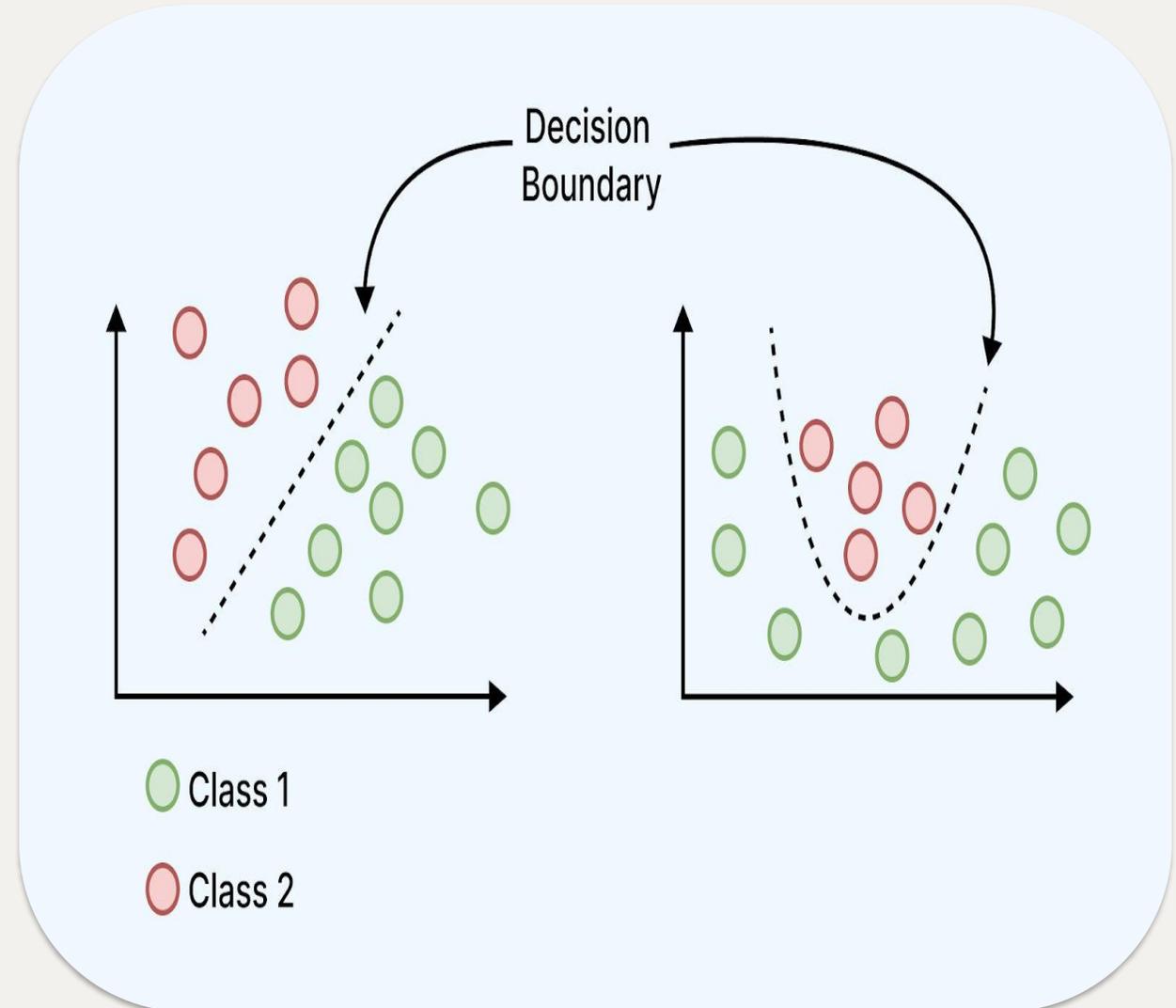
  Separates different classes

- **Odds Ratio**

  Measure of association

- **Logit Function**

  Natural log of odds

- **Maximum Likelihood Estimation**

  Parameter estimation method

Decision Boundary

○ Class 1

○ Class 2

Mind Map: Logistic Regression Overview

Logistic Regression

- **Types**

  Binary, Multinomial, Ordinal

- **Key Concepts**

  Sigmoid, Decision Boundary, MLE

- **Applications**

  Healthcare, Finance, Marketing

- **Advantages**

  Simple, Interpretable, Efficient

- **Limitations**

  Linear boundary, Outlier sensitivity

Knowledge Representation

Logical Approaches
- Propositional Logic
  - Simple Statements
  - Logical Connectives
  - Truth Tables
- Predicate Logic
  - Predicates and Arguments
  - Quantifiers (∀, ∃)
  - Logical Inference

Reasoning Methods
- Forward Chaining (Data-driven)
  - Rule Application
  - Deductive Reasoning
- Backward Chaining (Goal-driven)
  - Goal Decomposition
  - Subgoal Generation
- Resolution
  - Clausal Form
  - Inference Rules
- Unification

Structured Representations
- Slot-Filler Structures
  - Weak Structures (Simple Attributes)
  - Strong Structures (Inheritance, Relations)
- Semantic Networks
  - Nodes and Links
  - IS-A Relationships
- Frames and Scripts

Challenges and Issues
- Knowledge Acquisition
- Representational Adequacy
- Computational Complexity
- Ambiguity and Uncertainty

Applications of Logistic Regression

- **Healthcare**

  Disease diagnosis and prediction

- **Finance**

  Credit scoring and risk assessment

- **Marketing**

  Customer churn prediction

- **Technology**

  Email spam detection

Applications of Binary Logistic Regression

Customer Churn Analysis

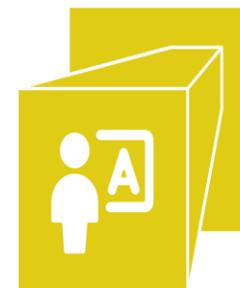Customer churn analysis is simple yet crucial for business strategy.

Disease Prediction

Disease prediction requires complex analysis for high-impact outcomes.

Basic Classification Tasks

Basic classification tasks are straightforward with minimal impact.

Advanced Marketing Strategies

Advanced marketing strategies involve complex methods with limited impact.

Advantages & Limitations

Advantages

✓ Simple to implement and understand

✓ Provides probability scores

✓ Efficient for linearly separable data

✓ No assumptions about data distribution

✓ Highly interpretable coefficients

Limitations

✗ Linear decision boundary only

✗ Sensitive to outliers

✗ Limited for non-linear relationships

✗ Requires large dataset for accuracy

Summary & Recap

- Logistic regression predicts probabilities

- Uses sigmoid function for classification

- Three types: Binary, Multinomial, Ordinal

- Widely applicable across industries

- Balance of simplicity and effectiveness

Logistic regression serves as a fundamental classification algorithm, balancing simplicity with powerful predictive capabilities for binary and multi-class problems.

- 1.Business Understanding
- 2.Data Understanding
- 3.Data Preparation
- 4.Modelling
- 5.Evaluation
- 6.Deployment