Understanding Logistic Regression

What Is Classification in ML?

Classification is a type of supervised learning where a model assigns input data to predefined categories (classes).

The model is trained using labeled data, where:

- **Features (X) represent input variables**
- **Labels (y) represent the correct class**

The model learns a decision boundary that separates different classes.

Types of Classification

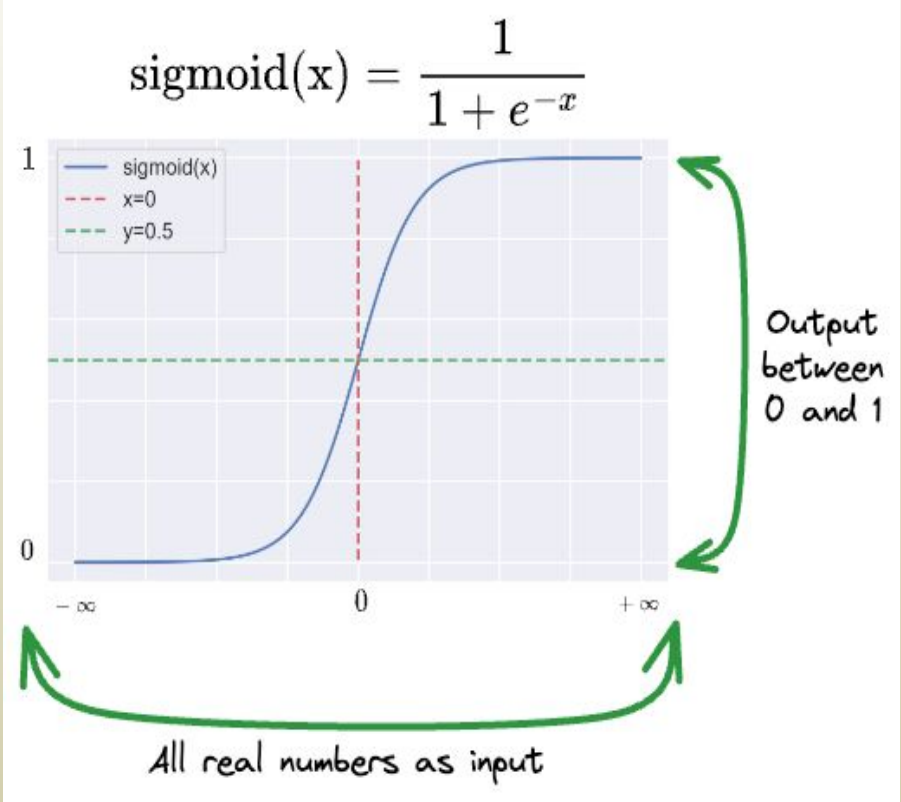
- **Binary Classification: Two classes (e.g., Spam / Not Spam)**
- **Multiclass Classification: More than two classes (e.g., digits 0–9)**
- **Multilabel Classification: Multiple labels per instance (e.g., image tags)**

What is Logistic Regression?

- Logistic Regression is a supervised machine learning algorithm used primarily for classification tasks, especially binary classification.It models the relationship between input features and the probability of a class using a logistic (sigmoid) function, producing outputs between 0 and 1.

- It works by computing a weighted sum of the input features and passing it through the sigmoid function, then applying a threshold (commonly 0.5) to decide the class label.

- Because Logistic Regression learns a linear decision boundary, it is efficient, interpretable, and works well when the relationship between features and the target is approximately linear.



When to Employ Logistic Regression

- Binary classification problems
- When model interpretability is important
- Datasets with linearly separable classes
- When you need probability outputs
- Small to medium-sized datasets

Limitations of the Logistic Regression

- Assumes a linear relationship between features and the log-odds
- Performs poorly with nonlinear decision boundaries
- Sensitive to outliers
- Can struggle with highly correlated features
- Requires proper feature scaling for best performance

Evaluation Metrics for Logistic Regression

Accuracy – overall correctness

Precision – correctness of positive predictions

Recall – ability to find all positives

F1 Score – balance between precision & recall

Confusion Matrix – detailed error analysis