Classification in ML

What is Classification in Machine Learning?

Classification is a type of **Supervised Learning** where the goal is to assign input data to one or more predefined categories or labels. It is used when the target variable is **categorical** (e.g., Yes/No, Spam/Not Spam, Dog/Cat).

How Does Classification Work?

A classification model learns from labeled training data and **maps input features (X) to a target category (Y)**:

$$Y = f(X)$$

Where:

- Y = Predicted category (class label)

- X = Input features (e.g., text, images, numerical data)

- f = Function (learned by the model)

📌 **Example:**

- **Email Classification** → Predict **Spam (1) or Not Spam (0)**

- **Medical Diagnosis** → Predict **Disease A, B, or C** based on symptoms

- **Sentiment Analysis** → Predict **Positive/Negative/Neutral**

Types of Classification Problems

1️⃣ Binary Classification

- **Only two possible classes** (e.g., 0 or 1, Yes or No).

- **Examples**:

    - Spam Detection (Spam vs. Not Spam)

- Fraud Detection (Fraudulent vs. Legitimate)

2️⃣ Multiclass Classification

- **More than two categories**, but each data point belongs to **only one class**.
- **Examples**:
  - Handwritten Digit Recognition (0-9)
  - Flower Classification (Setosa, Versicolor, Virginica)

Common Classification Algorithms

Algorithm	Type	Best For
Logistic Regression	Binary	Simple problems with linear decision boundaries
Decision Tree	Binary/Multiclass	Interpretable, non-linear patterns
Random Forest	Binary/Multiclass	Complex problems, reduces overfitting
Support Vector Machine (SVM)	Binary/Multiclass	Works well with small datasets
Naïve Bayes	Binary/Multiclass	Text classification (Spam Detection)
K-Nearest Neighbors (KNN)	Binary/Multiclass	Small datasets, non-linear problems
Neural Networks (Deep Learning)	Any	Complex patterns, image/text classification

When to Use Classification?

🔷 When the target variable has **distinct categories**.

🔷 When you need a **decision-based outcome** (e.g., Yes/No, Good/Bad).

🔷 When automating tasks like **fraud detection, disease diagnosis, sentiment analysis**.