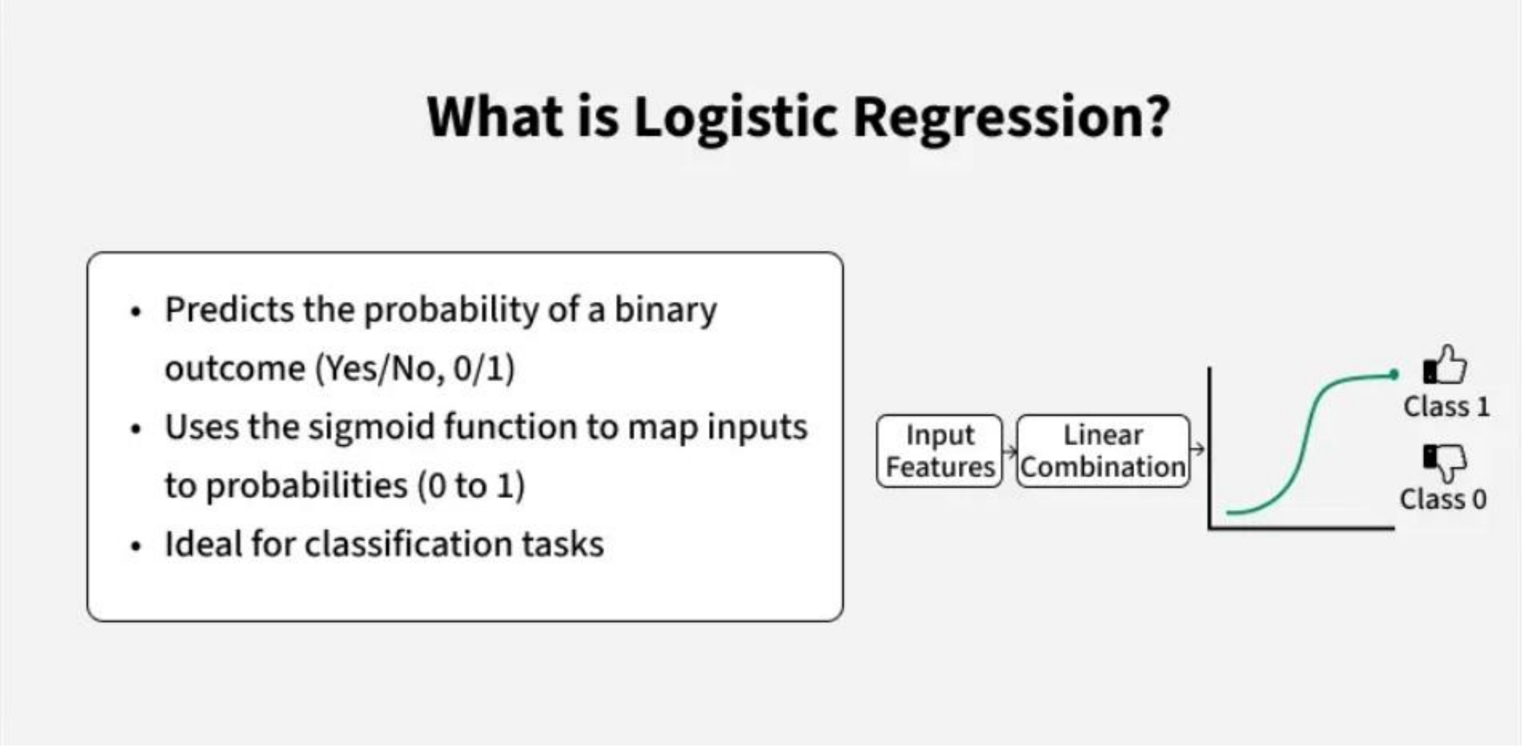
Logistic Regression

Logistic Regression is a supervised machine learning algorithm used for classification problems. Unlike linear regression which predicts continuous values it predicts the probability that an input belongs to a specific class. It is used for binary classification where the output can be one of two possible categories such as Yes/No, True/False or 0/1. It uses sigmoid function to convert inputs into a probability value between 0 and 1.



Logistic Regression Workflow

Step 1:Data Collection & Preparation

- Gather dataset (features + target)
- Handle missing values, encode categorical data
- Split into training and test sets

Step 2: Model Setup

- Define logistic regression model
- Decide features to use
- Initialize model parameters (weights & bias)

Step 3:Training (Fitting)

- Compute $z = w_0 + w_1x_1 + w_2x_2$ ...
- Apply Sigmoid Function → get probability $\hat{y}$
- Calculate Log Loss (Cost Function)
- Optimize parameters using Gradient Descent

Note: Logistic Regression models the probability of $y = 1$as:

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}, z = wX + b$$

Where:
w= weight (slope)
b= bias (intercept)
X= input feature (cell size)

Cost function

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right]$$

Step 4:Prediction

For new input data:

- If $P \geq 0.5$: Predict Class = 1
- Else : Class = 0

Step 5:Model Evaluation

- Compare predictions with actual labels
- Use metrics:
  - Accuracy
  - Confusion Matrix
  - Precision, Recall, F1-score
  - ROC-AUC

Advantages

- Simple and interpretable
- Works well for linearly separable data
- Outputs probabilities

Disadvantages

- Assumes linear decision boundary
- Not suitable for complex nonlinear datasets