Logistic Regression: A Comprehensive Guide

1. Introduction to Logistic Regression

Logistic Regression is a statistical method used for binary classification problems. Unlike linear regression, which predicts continuous values, logistic regression predicts the probability of a class label (e.g., 0 or 1). It is widely used in machine learning, medical diagnosis, fraud detection, and many other domains.

---

2. Why Not Linear Regression for Classification?

Linear regression predicts continuous values, but in classification tasks, we need discrete outputs (e.g., spam or not spam, pass or fail). If we apply linear regression, the predicted values may go beyond 0 and 1, which makes no sense for probability estimation.

Instead, we use **logistic regression**, which applies the **sigmoid function (logistic function)** to restrict outputs between 0 and 1.

---

3. Mathematical Formulation

3.1 The Sigmoid Function

Logistic regression is based on the **sigmoid function**, also known as the **logistic function**, defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

where z is the linear combination of input features:

$$z = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

Here:

- w_0 is the **bias term (intercept)**
- $w_1, w_2, \ldots, w_n$ are the **model weights (coefficients)**
- $x_1, x_2, \ldots, x_n$ are the **feature variables**
- e is the base of the natural logarithm

The sigmoid function transforms the linear combination z into a probability value between 0 and 1.

3.2 Hypothesis Function in Logistic Regression

The logistic regression model predicts the probability that an input belongs to class 1:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n)}}$$

The decision rule for classification:

- If $P(Y = 1|X) \geq 0.5$, predict **class 1**
- If $P(Y = 1|X) < 0.5$, predict **class 0**

---

4. Cost Function (Log Loss)

Unlike linear regression (which uses mean squared error), logistic regression uses the **log loss function**, also known as **binary cross-entropy**, to measure error:

$$J(w) = -\frac{1}{m} \sum_{i=1}^{m} [y_i \log(h_i) + (1 - y_i) \log(1 - h_i)]$$

where:

- m = number of training samples
- h_i = predicted probability for sample i
- y_i = actual label (0 or 1)

Minimizing this cost function ensures that the model correctly classifies data points.

---

5. Gradient Descent for Optimization

To minimize the cost function, we use **gradient descent**, an iterative optimization algorithm:

$$w_j := w_j - \alpha \frac{\partial J(w)}{\partial w_j}$$

where:

- α = learning rate (step size)
- $\frac{\partial J(w)}{\partial w_j}$ = gradient (derivative of cost function)

This updates the weights to minimize classification error.

---

6. Advantages and Disadvantages

6.1 Advantages

✅ Simple and easy to implement
✅ Works well for linearly separable data
✅ Provides probabilistic interpretation of predictions
✅ Efficient for small datasets

6.2 Disadvantages

❌ Assumes a linear decision boundary (not good for complex relationships)
❌ Not ideal for multi-class classification (requires extensions like Softmax regression)
❌ Sensitive to outliers

---

7. Solved Example

Problem Statement

A company wants to predict whether a customer will buy a product ($Y = 1$) or not ($Y = 0$) based on their annual income (X_1) and age (X_2). The logistic regression equation is:

$$P(Y = 1) = \frac{1}{1 + e^{-(w_0 + w_1 X_1 + w_2 X_2)}}$$

Given:

- $w_0 = -5$ (bias)
- $w_1 = 0.03$ (weight for income)
- $w_2 = 0.05$ (weight for age)
- Customer's income = **$70,000$**
- Customer's age = **45 years**

Step 1: Compute Linear Combination z

$$z = (-5) + (0.03 \times 70) + (0.05 \times 45)$$

$$z = -5 + 2.1 + 2.25$$

$$z = -0.65$$

Step 2: Apply Sigmoid Function

$$P(Y = 1) = \frac{1}{1 + e^{0.65}}$$

Approximating $e^{0.65} \approx 1.915$:

$$P(Y = 1) = \frac{1}{1 + 1.915} = \frac{1}{2.915} \approx 0.343$$

Step 3: Interpretation

Since $P(Y = 1) = 0.343 < 0.5$, the model predicts **the customer will not buy the product (Y=0).**

---

8. Exercise

Problem Statement

A bank wants to predict whether a loan applicant will default ($Y = 1$) or not ($Y = 0$) based on their credit score (X_1) and loan amount (X_2). The logistic regression equation is:

$$P(Y = 1) = \frac{1}{1 + e^{-(w_0 + w_1 X_1 + w_2 X_2)}}$$

Given:

- $w_0 = -4$
- $w_1 = 0.02$
- $w_2 = -0.03$
- Applicant's credit score = **680**
- Loan amount = **$25,000$**

Task: Compute the probability $P(Y = 1)$ and determine if the applicant will default.

Exercise: Loan Default Prediction

A bank wants to predict whether a loan applicant will default ($Y = 1$) or not ($Y = 0$) based on their **credit score** (X_1) and **loan amount** (X_2). The logistic regression model is:

$$P(Y = 1) = \frac{1}{1 + e^{-(w_0 + w_1 X_1 + w_2 X_2)}}$$

Given:

- $w_0 = -4$ (bias term)
- $w_1 = 0.02$ (coefficient for credit score)
- $w_2 = -0.03$ (coefficient for loan amount)
- Applicant's **credit score** $X_1 = 680$
- Applicant's **loan amount** $X_2 = 25,000$

Step 1: Compute Linear Combination z

$$z = w_0 + w_1 X_1 + w_2 X_2$$

Substituting the values:

$$z = (-4) + (0.02 \times 680) + (-0.03 \times 25,000)$$

First, compute the individual terms:

$$0.02 \times 680 = 13.6$$

$$-0.03 \times 25,000 = -750$$

Now sum them up:

$$z = -4 + 13.6 - 750$$

$$z = -740.4$$

Step 2: Compute Probability Using the Sigmoid Function

The **sigmoid function** is:

$$P(Y = 1) = \frac{1}{1 + e^{-z}}$$

Substituting $z = -740.4$:

$$P(Y = 1) = \frac{1}{1 + e^{740.4}}$$

Since $e^{740.4}$ **is an extremely large number**, the denominator becomes very large, making the fraction **approach zero**:

$$P(Y = 1) \approx 0$$

Step 3: Interpretation

Since $P(Y = 1) \approx 0$, the probability that the applicant will default on the loan is nearly **zero**.

Final Prediction:

- Since $P(Y = 1) < 0.5$, the model predicts **the applicant will NOT default (Y=0).**

Conclusion

- The model predicts that the applicant is unlikely to default based on their credit score and loan amount.
- The logistic regression model effectively assigns a very low probability of default due to the high negative impact of the large loan amount ($-0.03 \times 25,000$).
- This example illustrates how logistic regression helps in binary classification by estimating probabilities and applying a decision rule.