Logistic Regression: Built from Scratch with Gradient Descent

Tejinder Singh Hunjan, Mohit Kamkhalia, Sashank Mishra

Abstract

Logistic Regression is a widely-used algorithm for binary classification problems. In this report, we explore the mathematical foundations of Logistic Regression and implement it from scratch using Gradient Descent, without relying on external libraries.

1. Introduction

Classification tasks are central to many machine learning applications. Logistic Regression provides a simple yet effective approach to binary classification by modeling the probability that a given input belongs to a particular class using a sigmoid function.

2. Sigmoid Function

The core of Logistic Regression is the sigmoid function, which maps any real-valued number into the range (0, 1):

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad \text{where } z = w \cdot x + b$$

Here, w is the weight vector, x is the input vector, and b is the bias term.

3. Hypothesis Function

The predicted output $\hat{y}$ is the probability that the input belongs to the positive class:

$$\hat{y} = \sigma(w \cdot x + b)$$

4. Loss Function (Binary Cross Entropy)

To evaluate the performance of the model, we use the Binary Cross Entropy loss:

$$J(w, b) = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

5. Gradient Derivation

To minimize the loss function, we compute the gradients with respect to w and b:

$$\frac{\partial J}{\partial w} = \frac{1}{n} \sum_{i=1}^{n} x_i (\hat{y}_i - y_i)$$

$$\frac{\partial J}{\partial b} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)$$

6. Parameter Updates

Using the gradients and a learning rate α, the parameters are updated as follows:

$$w := w - \alpha \cdot \frac{\partial J}{\partial w}$$

$$b := b - \alpha \cdot \frac{\partial J}{\partial b}$$

7. Training Process

1. Initialize weights and bias to zero or small random values.
2. Compute the prediction $\hat{y}$ using the sigmoid function.
3. Calculate the loss using Binary Cross Entropy.
4. Compute gradients and update parameters.
5. Repeat for a fixed number of iterations or until convergence.

8. Prediction

After training, prediction is done using the sigmoid output:

$$\text{If } \hat{y} \geq 0.5, \text{ predict 1}; \quad \text{otherwise, predict 0}$$

9. Conclusion

Logistic Regression, though simple, is foundational for classification problems. Implementing it from scratch deepens our understanding of model training, optimization, and decision boundaries in classification tasks.