Logistic Regression

Breast Cancer Classification

Educational Project, by Saba, 28 April 2025.

Project Goal

- Implement logistic regression from scratch
- Practice fundamental Machine Learning concepts
- Predict whether a tumor is benign or malignant based on the extracted medical measurements

Table of contents

Introduction

- Logistic Regression is a supervised machine learning algorithm
- It is used for binary classification problems
- Predicts the probability that a given input belongs to a certain class
- Based on the sigmoid function to map outputs between 0 and 1

Data

- Breast Cancer Wisconsin
- 569 patients
- 30 features: radius, texture, perimeter
- 1 target: diagnosis (benign = 0, malignant = 1)

Weighted Sum

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

Sigmoid Function

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

Cost Function

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

Gradient Computation

$$\frac{\partial J(w,b)}{\partial w} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)x^{(i,j)}$$

$$\frac{\partial J(w,b)}{\partial b} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)$$

Gradient Descent

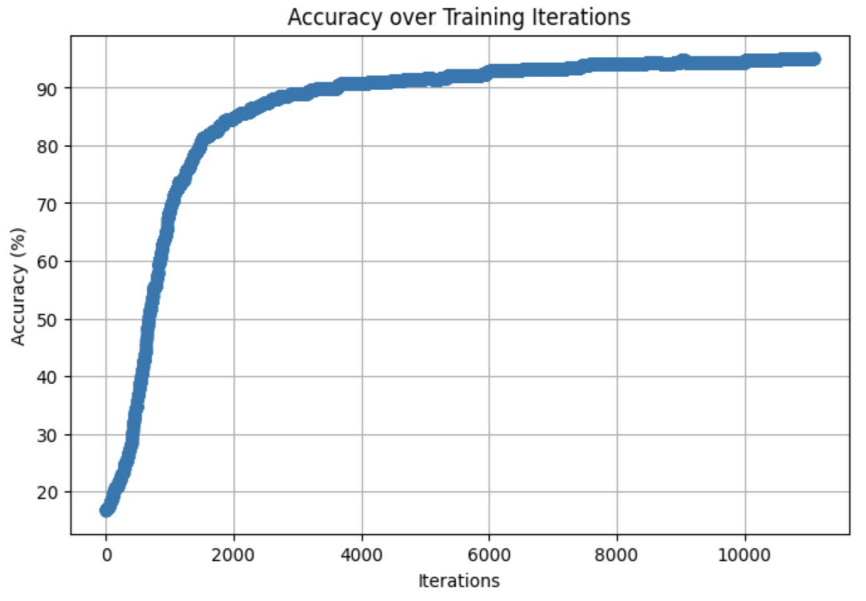
$$w_j := w_j - \alpha \frac{\partial J(w, b)}{\partial w_j}$$

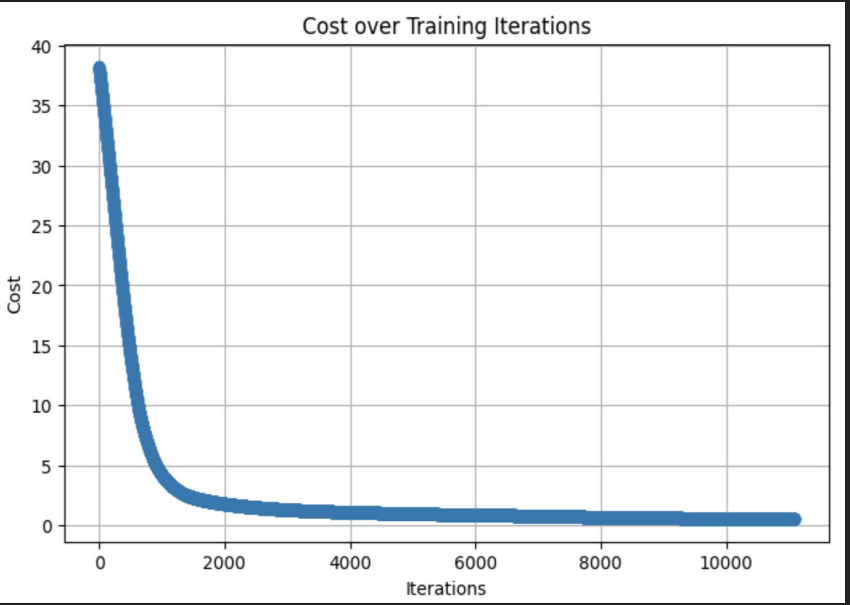
$$b := b - \alpha \frac{\partial J(w, b)}{\partial b}$$

Model Training

- Randomly initialize weights and bias
- Set the iteration counter to zero
- Calculate the initial cost and initial accuracy
- While the accuracy is less than 95%:
    - Perform one step of gradient descent to update weights and bias
    - Recompute the cost and the accuracy
    - Record the number of iterations, cost values, and accuracy values
- Stop training once the desired accuracy is reached

Evaluation


Accuracy over Training Iterations
Accuracy (%)
Iterations
20
30
40
50
60
70
80
90
0
2000
4000
6000
8000
10000


Cost over Training Iterations
Cost
Iterations
0
5
10
15
20
25
30
35
40
0
2000
4000
6000
8000
10000

Conclusion

- Logistic regression implemented manually
- Key ML concepts understood
- 95% test accuracy achieved
- Strengthened machine learning fundamentals
- Open to feedback and improvements
- Successfully developed an algorithm that achieves the goal of predicting whether a tumor is benign or malignant based on the extracted medical measurements