

Team Name: Runtime error: Brain not Found

Tanishque Soni

September 1, 2024

1 Introduction

This report covers the implementation of logistic regression from scratch in Python. Logistic regression is a statistical method for binary classification that uses a logistic function to model the probability of a binary outcome.

2 Logistic Regression Overview

Logistic regression models the probability of a binary response based on one or more predictor variables. The model is given by:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(w^T X + b)}}$$

where w represents the weights, b is the bias, and X is the input feature vector.

3 Implementation Details

3.1 Custom Logistic Regression Class

A custom logistic regression class was implemented from scratch to understand the underlying mechanics. The class includes:

- **Initialization:** Sets the learning rate and number of iterations.
- **Sigmoid Function:** Computes the probability using the sigmoid function.
- **Fit Method:** Trains the model using gradient descent by updating weights and bias.
- **Predict Method:** Makes predictions by applying the learned weights and bias.

3.2 Data Preprocessing

The following preprocessing steps were applied to the dataset:

- **Handling Missing Values:** SimpleImputer was used to replace missing values with the mean of each feature.
- **Feature Scaling:** StandardScaler was used to standardize features by removing the mean and scaling to unit variance.

3.3 Evaluation

The model's performance was evaluated using accuracy on both training and test sets. Accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

4 Dataset

The dataset used for training and evaluation contains asteroid data with the following features:

- **Name:** Name of the asteroid.
- **Various Attributes:** Including orbital parameters and physical characteristics.

A sample of the dataset is provided below:

Table 1: Sample Dataset

Name	a	e	i	om	w	q	ad	per_y	data_arc	condition_code
Asteroid 680001	2.690	0.153	15.278	137.354	72.385	2.277	3.104	4.413	37	5
Asteroid 661512	2.795	0.212	0.862	2.162	204.270	2.202	3.389	4.674	58	5

5 Results

The custom logistic regression model was trained and evaluated, demonstrating an understanding of the logistic regression algorithm and the preprocessing steps required to prepare data for machine learning models.

The performance of the logistic regression model on training data is summarized below:

Table 2: Model Accuracy

Metric	Accuracy
Training Accuracy	0.9976
Test Accuracy	0.9975