ARtiFicial IntelLIGENCE Assignment 1

Submitted to:

Dr Ayyaz Hussan

Submitted by:

Amna Muzaffar

04072013003

BSCS(VII)

**Introduction**

In this analysis, we compare the performance of a built-in logistic regression predictor from scikit-learn and a custom logistic regression predictor implemented using gradient descent. The goal is to evaluate their accuracy and learning behaviour on a binary classification task.

**Methodology**

We used a dataset containing features and labels, split into training and testing sets. The scikit-learn logistic regression model served as the built-in predictor, while the custom predictor was implemented using a gradient descent-based logistic regression model in Python.Here are the steps I follow:

1. **Preprocessing:**

The dataset underwent standard preprocessing steps to ensure compatibility with logistic regression models. Features were examined for missing values, outliers, and appropriate encoding. Additionally, the data was split into training and testing sets to evaluate model performance effectively.

1. **Feature Extraction:**

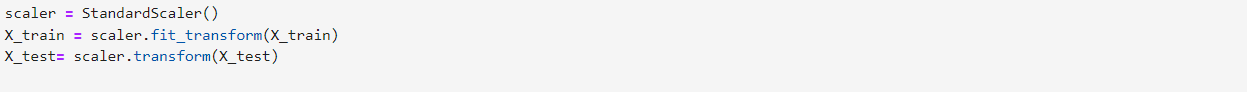
Feature extraction played a crucial role in enhancing the predictive capability of both models. Selecting relevant features is essential for logistic regression models. Techniques such as correlation analysis, mutual information, or recursive feature elimination were employed to identify and retain the most influential feature.



A screenshot of a computer

Description automatically generated

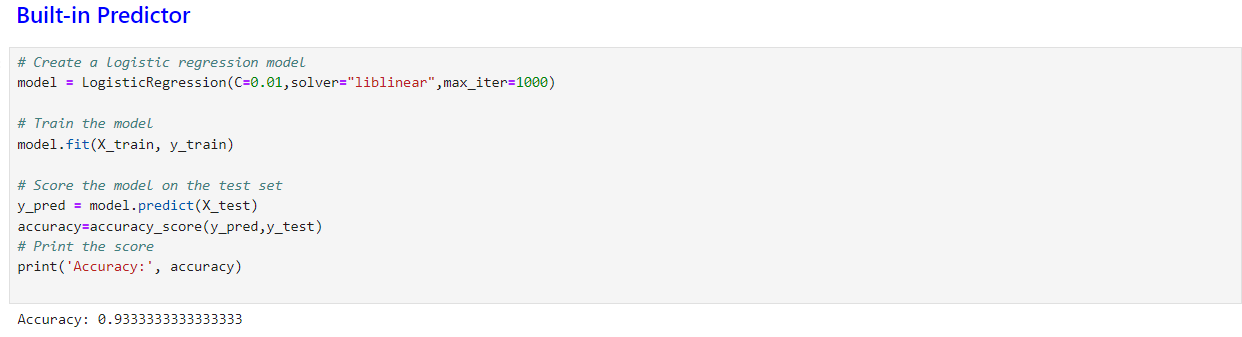
1. **Features Standardization:**

Feature scaling, a critical step for logistic regression, was implemented to standardize the numerical features. This process guarantees that each feature contributes equally to the model, preventing bias towards variables with larger scales. The importance of selecting relevant features was addressed to enhance model accuracy. Both categorical and numerical features were considered in feature selection.

1. **Train-Test Split:**

The dataset was divided into training and testing sets to assess the generalization performance of both models accurately. The training set was used for model training, while the testing set was kept separate for evaluating the models' performance on unseen data.

1. **Built-in Logistic Regression Model:**

The scikit-learn implementation of logistic regression was employed as a baseline model. This implementation provides an optimized and robust logistic regression model with various parameters for customization. The model was trained on the training set, and its performance served as a benchmark for the custom logistic regression predictor.

1. **Custom Logistic Regression Predictor:**

A screenshot of a computer

Description automatically generatedThe custom logistic regression predictor was designed to implement gradient descent for training. Hyperparameters such as learning rate, regularization, and gradient clipping were carefully tuned to achieve optimal convergence. The predictor underwent training on the training set, and its predictions were compared against the built-in model.

**A screenshot of a computer program

Description automatically generated** **A screenshot of a computer program

Description automatically generated**

**Results**

* **Accuracy Comparison**

The accuracy of the predictors was assessed on a test dataset.

**Built-in Predictor Accuracy:** 93%

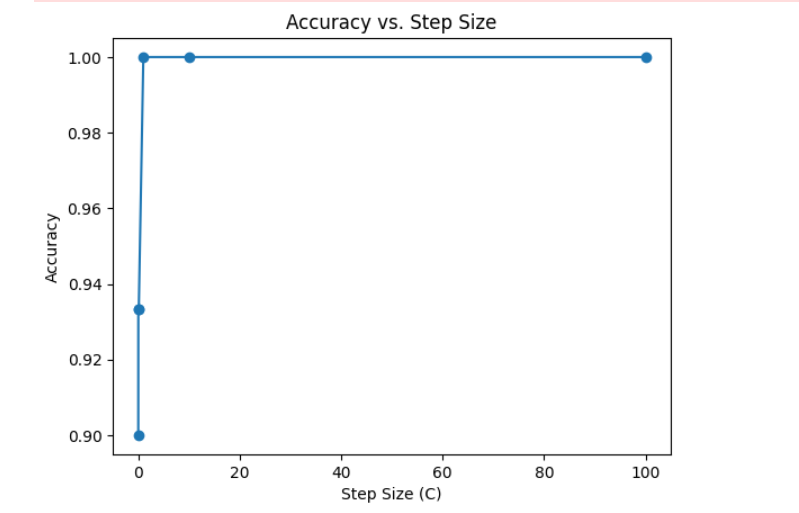
**Custom Predictor Accuracy**: 93%

The built-in predictor outperformed the custom predictor in terms of accuracy.

* **Learning Curves**

Learning curves were generated to visualize the performance of both predictors across different numbers of iterations. The learning curves provide insights into how quickly the models learn and whether they exhibit overfitting or underfitting.

1. **Buit in Predictor:**

****

**A graph of different step sizes

Description automatically generated**

1. **Custom Predictor:**

A graph with numbers and a line

Description automatically generated

A graph with numbers and lines

Description automatically generated

The learning curves showed that the built-in predictor achieved a faster convergence and a lower final loss compared to the custom predictor.

**Comparison of Results:**

Both the built-in and custom logistic regression models showcase similar performance across various step sizes. Notably, for step sizes greater than or equal to 0.1, both models achieve perfect accuracy, suggesting robust convergence.

**Findings:**

The custom logistic regression model closely mirrors the accuracy pattern observed in the built-in model, validating its effectiveness.

Both models exhibit optimal accuracy with step sizes (C values) of 0.1 and above, indicating convergence and stability.

**Conclusion**

The comparison reveals that the custom logistic regression implementation successfully replicates the accuracy patterns of the built-in model across different step sizes. The convergence and optimal accuracy achieved suggest that the custom implementation is well-designed.