**Logistic Regression Log**

The goal of this project is to implement logistic regression from scratch and compare it with Scikit-Learn’s built-in model. This helps in understanding how logistic regression works and evaluating its performance on different datasets.

**Data Understanding**

Generated synthetic classification datasets using Scikit-Learn's `make\_classification()`. The datasets varied in:

* Number of features (2, 5, and 10)
* Number of classes (binary and multi-class)
* Number of informative features

**Data Preparation**

* Split data into training (80%) and testing (20%)
* Standardized features using `StandardScaler` to improve model performance

**Modeling**

* Custom Logistic Regression:

Implemented logistic regression from scratch using gradient descent

Defined sigmoid function, cost function, and parameter updates

Trained the model using different learning rates and epochs

* Scikit-Learn Logistic Regression:

Used `LogisticRegression` from Scikit-Learn as a benchmark

Trained on the same dataset for comparison

**Evaluation**

Used accuracy score to measure model performance

Results from the model execution:

* Custom Logistic Regression Accuracy: 0.4200
* Scikit-Learn Logistic Regression Accuracy: 0.7150

The Scikit-Learn model performed significantly better due to built-in optimizations, while the custom model's accuracy suggests a need for hyperparameter tuning or further improvements.

**Findings and Work Log Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Phase** | **Work Done** | **Duration** | **Difficulty Level (1-10)** |
| Dataset Generation | Created synthetic datasets using make\_classification() | 15 mins | 3 |
| Data Preprocessing | Split data, applied feature scaling using StandardScaler | 20 mins | 4 |
| Custom Model Definition | Implemented logistic regression using NumPy | 50 mins | 7 |
| Model Training (Custom) | Trained custom logistic regression using gradient descent | 60 mins | 8 |
| Model Evaluation (Custom) | Predicted and calculated accuracy for the custom model | 15 mins | 2 |
| Scikit-Learn Model | Implemented and trained LogisticRegression model | 10 mins | 5 |
| Model Evaluation (Sklearn) | Predicted and calculated accuracy using Scikit-Learn | 15 mins | 5 |
| Results Comparison | Compared performance of custom vs. Scikit-Learn models | 10 mins | 4 |

**Conclusion**

This project provided insights into how logistic regression works and how a custom implementation compares to a standard library. The custom model achieved 42% accuracy, whereas the Scikit-Learn model achieved 71.5% accuracy. This suggests that further improvements, such as better feature selection, hyperparameter tuning, or optimization techniques, could help improve the custom model’s performance.