Logistic Regression Documentation

1. Introduction

Logistic regression is a supervised machine learning algorithm used for binary and multi-class classification. It models the probability of an outcome using a logistic function.

2. Objective

- Develop a web-based interactive application for logistic regression using Streamlit.
- Provide options to upload a dataset, select features, train the model, and evaluate its performance.
- Implement L1, L2, and L1+L2 (ElasticNet) regularization.

3. Dataset Description

- Dataset: User-uploaded CSV file.
- Features: User-selected input variables.
- Target: User-selected target variable.
- Number of records and attributes: Based on the uploaded dataset.

4. Implementation Details

- Frontend: Developed using Streamlit for an interactive UI.
- Backend: Logistic Regression implemented using scikit-learn.
- Steps:
  1. Upload dataset (CSV format).
  2. Select features and target variable.
  3. Choose the regularization type (L1, L2, or ElasticNet).
  4. Adjust the inverse of regularization strength (C value) and L1 ratio (for ElasticNet).
  5. Train the logistic regression model.
  6. Evaluate performance using accuracy, confusion matrix, and classification report.

5. Results and Analysis

- Model accuracy and evaluation metrics are displayed.
- The confusion matrix and classification report provide insights into model performance.
- The impact of different regularization techniques is analyzed.

6. Challenges and Solutions

- Managed data imbalance using appropriate techniques.
- Fine-tuned hyperparameters (C value, L1 ratio) to achieve optimal performance.

7. Conclusion

Logistic regression is a powerful algorithm for classification tasks. The application demonstrates how different regularization methods impact model performance. The interactive Streamlit UI enables users to train and evaluate models efficiently.

8. References

- Scikit-learn Documentation: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- Streamlit Documentation: https://docs.streamlit.io/
- Dataset Source: User-uploaded