Programming Assignment: Linear and Logistic Regression (Regression and Classification)

Objective

This assignment aims to provide hands-on experience with implementing and understanding two fundamental machine learning algorithms from scratch: Linear Regression and Logistic Regression. You will be provided with a Google Colab notebook containing code blocks that you need to complete. This will involve filling in missing code, training models, and evaluating their performance.

Part 1: Linear Regression

Goal: Implement and evaluate a Linear Regression model on a synthetic dataset.

Problem Description

In this section, you will work with a dataset where the target variable has a linear relationship with the features. Your task is to build a Linear Regression model to predict the target variable.

Instructions

Complete the code blocks in the Google Colab notebook for the following steps:

1. **Data Generation:** Generate a synthetic dataset with a linear relationship.
    - Create `X` (features) and `y` (target) such that `y = a*X + b + noise`.
2. **Data Splitting:** Split the dataset into training and testing sets.
    - Use `train_test_split` from `sklearn.model_selection` `(Not needed for this problem)`.
3. **Model Implementation:** Implement Linear Regression.
    - You can either implement it from scratch using NumPy (e.g., normal equation or gradient descent) or use `sklearn.linear_model.LinearRegression`.`(Not needed for this problem)`

4. **Implement cost function:** write function to implement Mean Squared Error(MSE)
5. **Implement Gradient Descent algorithm:** Write function to optimize training parameter using gradient descent algorithm.
6. **Model Training:** Train your Linear Regression model on the training data.
7. **Prediction:** Make predictions on the test set.
8. **Visualization:** Plot the regression line over the data points and predictions.
    - Ensure axes are labeled and the plot has a title.

Part 2: Logistic Regression

Goal: Implement and evaluate a Logistic Regression model for binary classification on a synthetic dataset.

Problem Description

In this section, you will work with a dataset designed for binary classification. Your task is to build a Logistic Regression model to classify data points into one of two categories.

Instructions

Complete the code blocks in the Google Colab notebook for the following steps:

1. **Data Generation:** Generate a synthetic 2D dataset for binary classification.
    - Use `make_blobs` from `sklearn.datasets` to create two distinct clusters.
2. **Data Splitting:** Split the dataset into training and testing sets.
    - Use `train_test_split` from `sklearn.model_selection`. (Not needed for this problem)
3. **Model Implementation:** Implement Logistic Regression.
    - Implement it from scratch.

4. **Implement the cost function:** Implement Binary cross entropy loss function.
5. **Implement gradient descent algorithm for optimization.**
6. **Model Training:** Train your Logistic Regression model on the training data.
7. **Prediction:** Make predictions on the test set.
8. **Evaluation:** Evaluate the model's performance.
    - Calculate and report accuracy, precision, and recall. (Not needed for this problem)
9. **Visualization:** Visualize the decision boundary on the dataset.
    - Clearly show the different classes and the boundary separating them.

Submission Guidelines

- **Google Colab Notebook:** Ensure all your code is implemented within the provided Google Colab notebook.
- **Code Completion:** Fill in all the indicated code blocks.
- **Comments:** Add comments to explain your code, especially for custom implementations or complex logic.
- **Readability:** Ensure your code is well-structured and easy to read.
- **Plots:** For all plots, include appropriate titles, axis labels, and legends where necessary.

Evaluation Criteria

Your assignment will be evaluated based on the following:

- **Correctness:** Proper implementation of Linear and Logistic Regression algorithms.
- **Functionality:** The code runs without errors and produces expected outputs.
- **Visualization Quality:** Clarity and effectiveness of generated plots (regression line, decision boundary).
- **Code Quality:** Readability, comments, and adherence to Python best practices.
- **Understanding:** Demonstrated understanding of the underlying concepts of both algorithms.

Due Date

Please submit your completed Google Colab notebook by
Oct 3, 2025 5:00 PM GMT+5:30.

[Link to Google Colab Notebook for the Assignment]
ML_Regression_and_Classification.ipynb