**I220454 Hunain Maqbool Ai B**

**AI Project Phase 2**

**Report: Manual Linear Regression for Gender Prediction**

**Introduction:**

Linear regression is a fundamental technique in statistics and machine learning used for predicting a continuous target variable based on one or more input features. In this report, we present a manual implementation of linear regression for the purpose of predicting gender based on several acoustic features extracted from audio data.

**Components of the Algorithm:**

1. **Data Preparation:**

- We start with a dataset containing acoustic features such as duration, intensity, pitch, centroid, contrast, rolloff, bandwidth, and formant frequency.

- The dataset is split into features (X) and the target variable (y), where X consists of the acoustic features and y contains the gender labels.

2. **Manual Linear Regression Function:**

- The `manual\_linear\_regression` function is implemented to perform linear regression manually without using any built-in libraries for regression.

- It begins by adding a column of ones to the feature matrix to account for the bias term.

- Then, it calculates the transpose of the feature matrix and multiplies it by the original matrix to obtain \( X^T \cdot X \).

- Next, it calculates the inverse of \( X^T \cdot X \) and multiplies it by \( X^T \) to obtain the parameter matrix.

- Finally, it computes the final parameters (weights and bias) by multiplying the parameter matrix with the target variable.

3. **Model Training and Testing:**

- The dataset is split into training and testing sets using the `train\_test\_split` function from scikit-learn.

- The `manual\_linear\_regression` function is called on the training set to obtain the final parameters.

- The model is then used to predict gender labels for the test set.

4. **Evaluation:**

- Mean Squared Error (MSE) is calculated to evaluate the performance of the model.

- The predicted gender labels are rounded to integers for comparison with the actual gender labels.

- Results are displayed in a DataFrame containing both predicted and actual gender labels.

**Rationale for Techniques:**

-**Manual Linear Regression:** The manual implementation allows for a deeper understanding of the linear regression algorithm and its components. While scikit-learn provides efficient implementations of linear regression, building the algorithm from scratch helps grasp the underlying mathematics and concepts involved.

**- Train-Test Split:** Splitting the dataset into training and testing sets allows for evaluating the model's performance on unseen data. This helps in assessing the model's generalization ability.

- **Mean Squared Error**: MSE is a widely used metric for regression tasks as it provides a measure of the average squared difference between the predicted and actual values. Lower MSE indicates better model performance.

**Conclusion:**

In this report, we demonstrated the manual implementation of linear regression for gender prediction based on acoustic features. By understanding each component of the algorithm and evaluating its performance, we gain insights into the predictive modeling process. This approach can be further extended and optimized for various regression tasks in machine learning.