Linear Regression Documentation

1. Introduction

Linear Regression is a supervised machine learning algorithm used for predicting continuous values. It establishes a relationship between the dependent variable (target) and independent variables (features) using a straight-line equation.

2. Objective

- Develop an interactive web application using Streamlit for Linear Regression.
- Provide users with the ability to upload a dataset, select features, train the model, and evaluate its performance.

3. Dataset Description

- Dataset: **A suitable dataset**
- Features: Describe the input variables
- Target: Describe the variable to predict
- Number of records and attributes

4. Implementation Details

- **Frontend**: Developed using Streamlit for user interaction.
- **Backend**: Linear Regression implemented using `scikit-learn`.
- **Steps**:
  1. Upload the dataset in CSV format.
  2. Select features and target variables.
  3. Train the Linear Regression model.
  4. Evaluate the model using metrics like Mean Squared Error (MSE), R-squared, and Mean

Absolute Error (MAE).

5. Visualize results using scatter plots and regression lines.

5. Results and Analysis

- Provide model performance metrics.

- Plot the actual vs predicted values.

- Discuss how well the model fits the data.

6. Challenges and Solutions

- Managed multicollinearity using correlation analysis.

- Addressed data normalization to improve accuracy.

- Applied feature selection techniques to reduce complexity.

7. Conclusion

Linear Regression is a simple yet effective algorithm for predicting continuous values. The interactive application allows users to explore model behavior and visualize results.

8. References

- Scikit-learn Documentation

- Streamlit Documentation

- Dataset Source