# Extended Report on Linear Regression

Introduction

Linear Regression is one of the oldest and most commonly used algorithms in statistics and machine learning. It establishes a relationship between a dependent variable (Y) and one or more independent variables (X). The concept dates back to the early 19th century with contributions by Sir Francis Galton and Karl Pearson.

Mathematical Background

The equation of a linear regression model is: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon$. The most used cost function is Mean Squared Error (MSE). Gradient Descent is often applied to minimize this error.
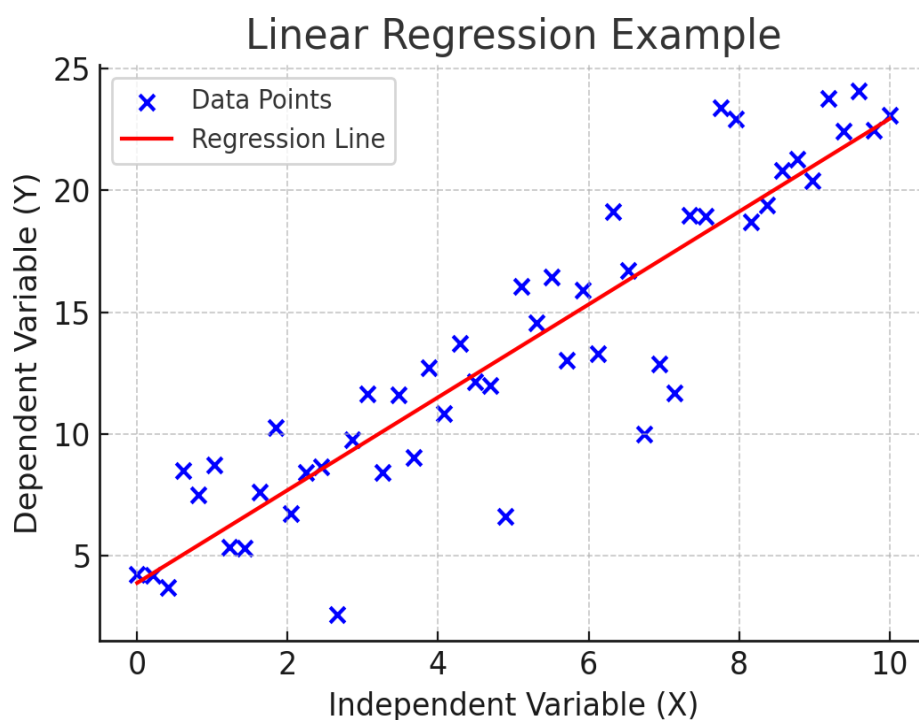

Assumptions of Linear Regression

- Linearity: The relationship between variables is linear. - Independence: Observations must be independent. - Homoscedasticity: Constant variance of errors. - Normality: Errors should follow a normal distribution. - No Multicollinearity: Inputs should not be highly correlated.

Evaluation Metrics

- R²: Explains variance captured by the model. - Adjusted R²: Adjusts for number of predictors. - MSE: Mean Squared Error. - RMSE: Root Mean Squared Error. - MAE: Mean Absolute Error.

Steps in Linear Regression

ollection

ocessing

Training

aluation

rediction

Process Flow
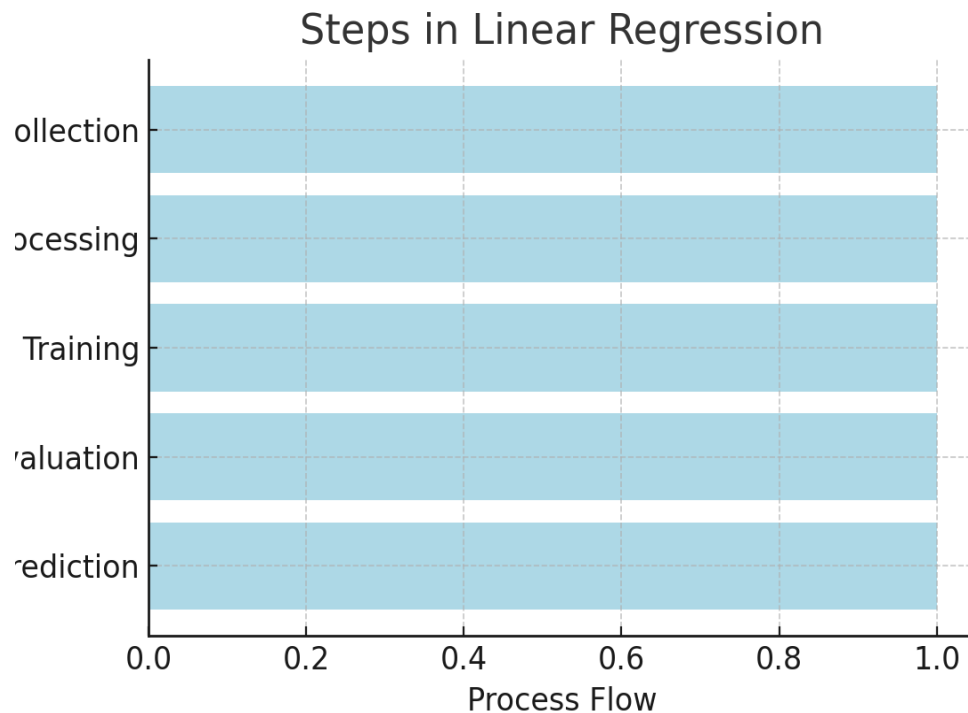
Applications

- Business: Sales forecasting, pricing strategies. - Healthcare: Disease prediction, medical cost estimation. - Education: Student performance analysis. - Finance: Credit risk, stock market trends. - Environment: Weather and pollution forecasting.

Advantages

- Easy to implement and interpret. - Efficient on small and medium datasets. - Provides clear coefficient-based insights.

Limitations

- Sensitive to outliers. - Fails for non-linear data. - Relies on strong assumptions. - Less flexible than modern ML models.

Comparison with Other Models

- Logistic Regression: For categorical outcomes. - Decision Trees: Handle non-linear relationships. - Neural Networks: Capture complex patterns but need more data. - SVM: Suitable for high-dimensional datasets.

Future Scope

Linear Regression still plays an important role as a baseline model, in feature selection, and in advanced methods like Ridge, Lasso, and ElasticNet. It also supports deep learning interpretability.

Conclusion

Linear Regression is a fundamental model that balances simplicity, speed, and interpretability. It remains an essential tool for statistical modeling and machine learning applications.