Assignment-1: Comprehensive Study of Linear Regression Models

Tools: Python, Google Colab / Jupyter Notebook
Group Activity

OBJECTIVE

To implement, analyze, and compare various Linear Regression models including Simple Linear Regression, Multiple Linear Regression, Polynomial Regression, and Regularization techniques.

PROBLEM STATEMENT

Given a dataset with multiple input features and a continuous target variable, perform end-to-end analysis using linear regression techniques and evaluate model performance.

DATASET REQUIREMENT

- Dataset must contain at least three input features and one continuous target variable.

- Dataset source may be Kaggle, UCI Repository, or synthetically generated.

PROCEDURE

Part A: Exploratory Data Analysis (EDA)

1. Load the dataset and display basic information.

2. Compute summary statistics.

3. Check missing values and outliers.

4. Plot feature distributions and correlation heatmap.

Part B: Simple Linear Regression

5. Select one feature and build a simple linear regression model.

6. Plot regression line and interpret slope and intercept.

Part C: Multiple Linear Regression

7. Build a multiple linear regression model using two or more features.

8. Interpret coefficients and evaluate using MSE, RMSE, and R^2.

Part D: Polynomial Regression

9. Apply polynomial regression to capture non-linear relationships.

10. Compare linear and polynomial regression performance.

Part E: Regularization

11. Apply Ridge and Lasso regression.

12. Compare coefficients and discuss feature selection.

Part F: Model Diagnostics

13. Plot residuals vs predicted values.

14. Validate regression assumptions.

EVALUATION METRICS

- Mean Squared Error (MSE)

- Root Mean Squared Error (RMSE)

- R^2 Score

EXPECTED OUTCOME

Students will gain practical understanding of linear regression models, diagnostics, and model selection.

ASSESSMENT (20 Marks)

EDA and Visualization – 5 Marks

Model Implementation – 7 Marks

Evaluation and Comparison – 5 Marks

Interpretation and Conclusion – 3 Marks

RESULT

To be written by the student after successful completion of the experiment.

As part of the **Linear Regression Lab Assignment**, all students are required to **create a GitHub account and a dedicated GitHub repository** for this activity.

Each student must upload the following to their repository:

- Python Notebook (.ipynb) containing the complete lab work
- Dataset used (CSV/Excel)
- Output plots/figures

- Brief documentation in the **README.md** file explaining the objective, methodology, and observations

The repository should be **public** and properly structured.
You will be required to **submit the GitHub repository link** as part of the lab evaluation.

This activity aims to help you learn **version control, professional documentation, and industry-standard practices**, which are essential skills for data analytics and machine learning professionals.

Please complete the GitHub setup and uploads **before the next lab submission deadline**.