Lab 4: Regression Analysis with Regularization Techniques

Assignment Instructions

Overview:

In this lab, you will explore multiple regression techniques and enhance your understanding by incorporating regularization methods. Specifically, you will:

- Implement Linear Regression, Multiple Regression, and Polynomial Regression models.
- Apply Ridge Regression and Lasso Regression to observe how regularization helps prevent overfitting and improve model performance.
- Evaluate the performance of your models using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).
- Visualize the regression results to better understand model performance and data patterns.

Lab Instructions:

Dataset
For this lab, use the *Diabetes* Dataset from *sklearn.datasets* to perform regression analysis. This dataset contains various health measurements used to predict disease progression, making it suitable for demonstrating regression techniques.

Step 1: Data Preparation

- Create a new Jupyter Notebook and include a Markdown cell at the top with your name, course title, and lab assignment title.
- Load the *Diabetes* dataset from *sklearn.datasets*.
- Explore the dataset's features, target values, and data distribution.
- Perform any necessary data cleaning steps (e.g., handling missing values).

Step 2: Linear Regression

- Implement a Simple Linear Regression model using one feature as the independent variable and the target variable as the dependent variable.
- Split the data into training and testing sets.
- Train the model and evaluate its performance using MAE, MSE, RMSE, and R^2.
- Visualize the model's predictions against the actual data.

Step 3: Multiple Regression

- Implement a Multiple Regression model using multiple independent variables to predict the target variable.
- Train the model and evaluate its performance using MAE, MSE, RMSE, and R^2.
- Visualize the predicted values versus the actual values.

Step 4: Polynomial Regression

- Implement a Polynomial Regression model by extending the linear regression model with polynomial features.
- Train and test the model while comparing the results with those from the linear and multiple regression models.
- Demonstrate how increasing the polynomial degree can affect overfitting and underfitting.

Step 5: Regularization with Ridge and Lasso Regression

- Implement Ridge Regression and Lasso Regression models.
- Train the models and compare their performance to the earlier models using the same evaluation metrics.
- Explain how regularization parameters (alpha values) influence the model's behavior and results.
- Visualize the model predictions and highlight how Ridge and Lasso differ in performance.

Step 6: Model Comparison and Analysis

- Summarize and compare the performance of the different regression models.
- Discuss key observations, including:
- How well each model performed.
- Which models handled overfitting or improved performance.
- Insights gained about the *Diabetes* dataset.

Submission Instructions:

- Create a GitHub repository for this lab (e.g. MSCS_634_Lab_4).
    - Add the following files to your repository:
        - Your Jupyter Notebook file (.ipynb).
        - A README.md file that briefly explains:
            - The purpose of your lab work.
            - Key insights gained from the regression analysis.
            - Any challenges faced or decisions made during the lab.
- Submit your GitHub repository link and ensure the repository is public or that you have granted instructor access.

Submission

Drag and drop files here or click to add text

Word count: 0

Save and Close **Submit**