Linear Regression Machine Learning Pipeline in Google Colab

Objective

In this assignment, you will practice downloading a dataset from GitHub and implementing a simple linear regression machine learning pipeline in Google Colab. You will use the Boston Housing dataset for this exercise. After successfully implementing the basic model, you are encouraged to explore regularization techniques, such as Lasso, to compare their performance with the initial model.

Instructions

Set Up Your Environment in Google Colab

- Open Google Colab and create a new notebook.

Download the Dataset

- Use the following code snippet to download the Boston Housing dataset directly into your Colab environment:

  ```
  !wget https://github.com/PawelSmolinski/MachineLearning/raw/main/
  Datasets/Boston.csv
  ```

- This command downloads the dataset and names it `Boston.csv`.

Load and Explore the Dataset

- Use pandas to load the dataset:

  ```
  import pandas as pd
  data = pd.read_csv('Boston.csv')
  ```

- Explore the dataset using methods like `data.head()`, `data.describe()`, etc.

Implement the Linear Regression Model

- Follow the example provided in the uploaded file regression.ipynb.

- Make sure to split the data into training and testing sets.

Evaluate the Model

- Evaluate the model on the test set and note the performance metrics.

Advanced Challenge (Optional)

- Implement Lasso regularization in your linear regression model.

- Compare the results with the basic linear regression model.

- Document any observations on model performance and complexity.

Resources

- Google Colab
- Pandas Documentation
- Scikit-learn Linear Regression
- Scikit-learn Lasso Regression
- ChatGPT