L1 and L2 Regularisation

1 Overview

The objective of this lab assignment is to implement L1 and L2 regularisation to overcome overfitting of model. To mitigate overfitting in linear regression, we applied L1 (Lasso) and L2 (Ridge) regularRization techniques to the California housing dataset. We first performed exploratory data analysis (EDA) to understand the distribution of data and key patterns. Then, we implemented regularization techniques and trained models using Gradient Descent optimization. Finally, we evaluated the models' performance across different training-test split ratios using relevant metrics to determine the impact of regularization on model generalization.

2 Import Libraries and Load Dataset

To build a regularized linear regression model, we first imported the necessary libraries and loaded the California housing dataset from the `sklearn.datasets` module. This dataset contains various housing attributes and their corresponding median house values, which serve as the target variable.

3 Preprocessing

Before applying regularization techniques, we conducted thorough preprocessing to ensure data quality.

3.1 Exploratory Data Analysis (EDA)

- Performed EDA using statistical visualizations to understand feature distributions.
- Identified and handled missing values in the dataset.
- Checked for outliers that could affect model performance.

3.2 Feature Scaling and Normalization

- Checked for correlation between features using correlation matrix.
- Checked for multicollinearity using variance inflation factor and dropped features with high variance inflation factor.
- Applied standard scaling to standardize numerical attributes.
- Ensured that all features were on a similar scale to facilitate stable gradient descent convergence.
- Split the dataset into 80% training and 20% testing to evaluate the generalizability of the model.

4 Implement Regularization Techniques

To mitigate overfitting, we implemented both L1 and L2 regularization techniques in the cost function:

4.1 L1 Regularization (Lasso)

- Introduced the absolute value penalty term in the cost function.
- The cost function for Lasso regression is given by:

$$J(w) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 + \lambda \sum_{j=1}^{n} |w_j| \tag{1}$$

- The gradient update rule for L1 regularization is:

$$w_j = w_j - \alpha \left(\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i) X_{ij} + \lambda \text{sign}(w_j) \right) \tag{2}$$

4.2 L2 Regularization (Ridge)

- Added a squared penalty term to the cost function.
- The cost function for Ridge regression is given by:

$$J(w) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 + \frac{\lambda}{2} \sum_{j=1}^{n} w_j^2 \tag{3}$$

- The gradient update rule for L2 regularization is:

$$w_j = w_j - \alpha \left(\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i) X_{ij} + \lambda w_j \right) \tag{4}$$

5 Training and Evaluation

After implementing the regularization and optimization techniques, we trained the linear regression models for the L1 and L2 regularization, tested these two models on the testing data and evaluated their performance using standard regression metrics:

5.1 Evaluation Metrics

- Root Mean Squared Error (RMSE) to measure the prediction error.
- Mean Absolute Error (MAE) for the assessment of absolute deviation.
- R-squared (R^2) to determine the model explanatory power.

5.2 Performance Comparison Table

Regularization	RMSE	MAE	R^2
L1	0.7604	0.5677	0.4004
L2	0.7525	0.5467	0.4129

Table 1: Performance metrics for both regularization

5.3 Result Analysis

- In this case L2 regularization is preforming slightly better than L1 regularisation.
- For same number of epochs L2 regularization has a lower cost than L1 regularization.
- L2 has a lower RMSE and MAE than L1 and has a higher R^2 value than L1.

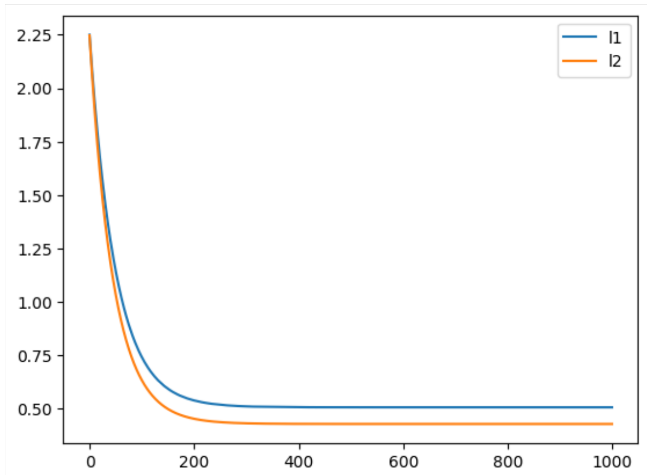

Figure 1: cost history for both regularization

6 Effect of L1 and L2 Regularization

L1 and L2 regularization have different impacts on model performance and feature selection:

6.1 Effect of L1 Regularization (Lasso)

- It introduces the absolute value penalty term in the cost function.
- Encourages sparsity, leading to feature selection by driving some coefficients to zero.
- Helps in high-dimensional datasets where irrelevant features can be ignored.
- Can result in unstable solutions when features are highly correlated.

6.2 Effect of L2 Regularization (Ridge)

- Added a squared penalty term to the cost function.
- Prevents large coefficient values but does not enforce sparsity.
- Works well in multicollinear datasets by distributing weights among correlated features.
- Improves generalization by reducing overfitting without completely removing features.

6.3 Comparison of L1 and L2 Regularization

- L1 is better suited when feature selection is necessary, while L2 is preferable when dealing with multicollinearity.
- L1 can lead to sparse models, whereas L2 produces more stable and smooth solutions.
- A combination of both (Elastic Net) can balance sparsity and stability.

7 Conclusion

In this lab, we explored how L1 (Lasso) and L2 (Ridge) regularization impact linear regression models. By applying these techniques to the California housing dataset, we saw that L1 helps with feature selection by setting some coefficients to zero, while L2 improves stability by preventing large weight values. Using gradient descent for optimization, we evaluated model performance and confirmed that regularization reduces overfitting and improves generalization.

Overall, regularization is a powerful tool for making models more reliable. In the future, experimenting with Elastic Net or adaptive learning rates could further refine performance.