

3_Regularization

January 28, 2025

1 Exercise: Regularization in Linear Regression

This exercise will guide you through the process of applying and comparing different regularization techniques in linear regression. You will explore Ridge, Lasso, and Elastic Net regression to understand how regularization affects model performance and feature selection.

1.1 Steps to Complete the Exercise:

1.1.1 1. Read and Prepare the Dataset

- Load the dataset from the provided file (`regularization.csv`).
- Separate the dataset into features (\mathbf{X}) and the target variable (\mathbf{y}).
- Split the data into training and test sets.

1.1.2 2. Perform Data Exploration

- Explore the dataset to understand its structure.
 - Check for missing values and handle them appropriately.
 - Visualize the relationships between features and the target variable (e.g., correlation heatmap, scatter plots).
- Standardize or normalize the features if necessary (e.g., for distance-based methods).

1.1.3 3. Apply Linear Regression

- Fit a **linear regression model** to the training data.
- Evaluate its performance using metrics such as:
 - Mean Squared Error (MSE)
 - (R^2) score
- Record the results for comparison.

1.1.4 4. Apply Ridge Regression

- Implement **Ridge regression** and use cross-validation to select the best alpha (α).
- Report:
 - The best alpha value.
 - The corresponding MSE and (R^2) scores on the test set.

1.1.5 5. Apply Lasso Regression

- Implement **Lasso regression** and use cross-validation to select the best alpha (α).

- Report:
 - The best alpha value.
 - The corresponding MSE and (R^2) scores on the test set.
- Analyze feature selection:
 - Plot the **number of features left** (non-zero coefficients) as a function of (λ) .
 - Plot the **MSE** as a function of (λ) .

1.1.6 6. Apply Elastic Net

- Implement **Elastic Net regression** and use cross-validation to select the best alpha (λ) and mixing ratio (α) .
- Report:
 - The best alpha and mixing ratio values.
 - The corresponding MSE and (R^2) scores on the test set.

1.1.7 7. Comparison of Models

- Summarize and compare the results of all models (Linear Regression, Ridge, Lasso, and Elastic Net) using:
 - MSE on the test set.
 - (R^2) score on the test set.
 - For Lasso, analyze the trade-off between the number of features left and model performance.

1.1.8 8. Answer the Following Questions

- Which regularization technique provided the best performance on the test set?
- For Lasso, which alpha value represents a good compromise between simplicity (few features) and accuracy (low MSE)?
- How does Elastic Net compare to Ridge and Lasso in terms of feature selection and accuracy?

[]: