Homework Assignment 5: Regularization

DATA 443: Statistical Machine Learning

Due Date

Friday, April 19, 2025 at 11:59 PM

Objective

In this assignment, you will implement and compare different regularization techniques on a real-world dataset. Your goal is to evaluate how these methods impact model performance and interpret the results using appropriate metrics.

Tasks

1. **Select a Real-World Dataset:** Choose a dataset that is suitable for regression analysis. Ensure it contains both numerical predictors and a continuous target variable.
2. **Data Preparation:**
   - Clean and preprocess the data as needed (e.g., handling missing values, feature scaling).
   - Split the dataset into training and testing sets.
3. **Implement the Following Regularization Techniques:**
   - Ridge Regression (L2 regularization)
   - Lasso Regression (L1 regularization)
   - Elastic Net Regression (combination of L1 and L2)
4. **Model Evaluation:** Evaluate the performance of each model using the following metrics:
   - Root Mean Squared Error (RMSE)
   - R-squared (R^2)
5. **Compare and Interpret the Results:**
   - Create a comparison table summarizing RMSE and R^2 for each method.

- Discuss which regularization technique performed best and why.
- Reflect on the effect of regularization in preventing overfitting and improving generalization.

Deliverables

- Python or R code (well-commented)
- A brief report (1–2 pages) that includes:
  - Description of the dataset and preprocessing steps
  - Model results and comparison table
  - Interpretation of results

Bonus (Optional)

Explore how different values of the regularization parameter (alpha or λ) affect model performance using cross-validation.