**Sri Sivasubramaniya Nadar College of Engineering, Chennai**
(An Autonomous Institution Affiliated to Anna University)

| | | |
|---|---|---|
| Degree & Branch | B.E. Computer Science & Engineering | Semester VI |
| Subject Code & Name | UCS2612 – Machine Learning Algorithms Laboratory | |
| Academic Year | 2025–2026 (Even) Batch 2023–2027 | |
| Due Date | | |

**Experiment 3: Regression Analysis using Linear and Regularized Models**

# Objective

To implement linear and regularized regression models for predicting a continuous target variable, evaluate their performance using multiple metrics, visualize model behavior, and analyze overfitting, underfitting, and bias–variance characteristics.

# Dataset

A real-world regression dataset containing numerical and categorical features related to loan applications is used. The target variable is the **loan amount sanctioned**.

Dataset reference:

- Kaggle: Predict Loan Amount Data

# Brief Theory (For Lab Understanding)

## Linear Regression

Linear Regression models the relationship between input features and a continuous target variable. It is simple, interpretable, and serves as a baseline regression model.

## Regularized Regression Models

Regularization techniques are used to control model complexity:

- Ridge Regression reduces coefficient magnitudes
- Lasso Regression performs feature selection
- Elastic Net combines Ridge and Lasso behavior

Regularization helps improve generalization and reduce overfitting.

## Task Description

Students must:

- Implement Linear, Ridge, Lasso, and Elastic Net regression models
- Tune regularization hyperparameters using Grid Search or Randomized Search
- Visualize regression results and errors
- Analyze overfitting, underfitting, and bias–variance trade-off

## Implementation Steps

1. Load the dataset
2. Perform data preprocessing:
    - Handle missing values
    - Encode categorical variables
    - Standardize numerical features
3. Perform Exploratory Data Analysis (EDA)
4. Visualize feature distributions and target distribution
5. Split the dataset into training and testing sets
6. Train baseline Linear Regression
7. Train Ridge, Lasso, and Elastic Net models
8. Perform hyperparameter tuning using 5-Fold Cross-Validation
9. Evaluate all models using regression metrics

## Required Visualizations (During Coding)

Students must generate and include:

- Target variable distribution plot
- Feature vs. target scatter plots
- Predicted vs. actual values plot
- Residual plot
- Training error vs. validation error plot
- Coefficient comparison bar plot

## Performance Metrics to be Reported

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- $R^2$ Score
- Training Time

## Hyperparameter Search Space

- Ridge: $\alpha \in \{0.01, 0.1, 1, 10, 100\}$
- Lasso: $\alpha \in \{0.001, 0.01, 0.1, 1, 10\}$
- Elastic Net:
  - $\alpha \in \{0.01, 0.1, 1, 10\}$
  - $l1\_ratio \in \{0.2, 0.5, 0.8\}$

## Hyperparameter Tuning Results

Table 1: Hyperparameter Tuning Summary

| Model | Search Method | Best Parameters | Best CV $R^2$ |
|---|---|---|---|
| Ridge Regression | Grid / Random | | |
| Lasso Regression | Grid / Random | | |
| Elastic Net Regression | Grid / Random | | |

## Cross-Validation Performance (K = 5)

Table 2: Cross-Validation Performance

| Model | MAE | MSE | RMSE | $R^2$ |
|---|---|---|---|---|
| Linear Regression | | | | |
| Ridge Regression | | | | |
| Lasso Regression | | | | |
| Elastic Net Regression | | | | |

Table 3: Test Set Performance

| Model | MAE | MSE | RMSE | $R^2$ |
|---|---|---|---|---|
| Linear Regression | | | | |
| Ridge Regression | | | | |
| Lasso Regression | | | | |
| Elastic Net Regression | | | | |

## Test Set Performance Comparison

## Effect of Regularization on Coefficients

Table 4: Coefficient Comparison

| Feature | Linear | Ridge | Lasso | Elastic Net |
|---|---|---|---|---|
| Feature 1 | | | | |
| Feature 2 | | | | |
| Feature 3 | | | | |

## Overfitting and Underfitting Analysis

Students must discuss:

- Difference between training and validation errors
- Effect of regularization strength
- Improvement in generalization after tuning

## Bias–Variance Analysis

Students must comment on:

- Bias behavior of Linear Regression
- Variance reduction using Ridge and Elastic Net
- Feature sparsity effect in Lasso

## Conclusion

Summarize the performance of all regression models, justify the choice of optimal hyperparameters, and discuss the trade-off between accuracy and model complexity.

## Report Format (Mandatory)

1. Aim and Objective
2. Dataset Description
3. Preprocessing Steps
4. Implementation Details
5. Visualizations
6. Performance Tables
7. Overfitting and Underfitting Analysis
8. Bias–Variance Analysis
9. Observations and Conclusion

## References


- Scikit-learn: Linear Models
- Scikit-learn: Hyperparameter Optimization
- Loan Amount Dataset