Linear Regression: OLS, Gradient Descent, and Regularization

Machine Learning Foundations

November 24, 2025

Abstract

This document presents a comprehensive analysis of linear regression methods including ordinary least squares (OLS), gradient descent optimization, and regularization techniques (Ridge and Lasso). We examine model diagnostics, multicollinearity detection, and cross-validation for hyperparameter tuning.

1 Introduction

Linear regression models the relationship between features X and target y:

$$y = X\beta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I) \tag{1}$$

OLS Solution:

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y \tag{2}$$

Ridge Regression:

$$\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T y \tag{3}$$

Lasso Regression:

$$\hat{\beta}_{Lasso} = \arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \tag{4}$$

2 Computational Environment

3 Data Generation

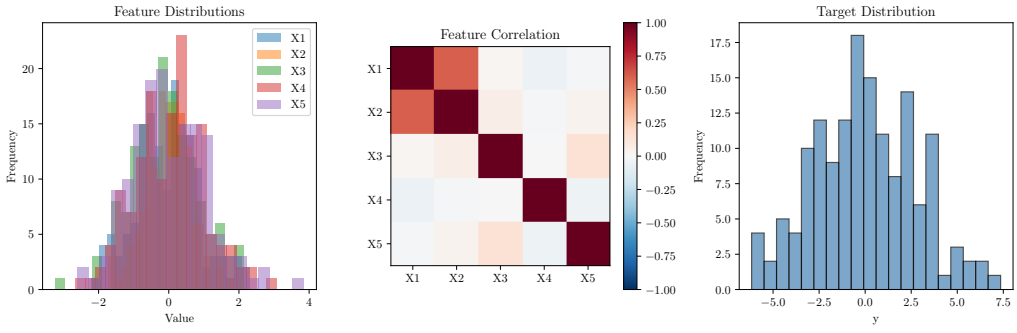

Figure 1: Dataset overview showing feature distributions and correlations.

Dataset: $n = 200$ samples, $p = 5$ features.

4 Ordinary Least Squares

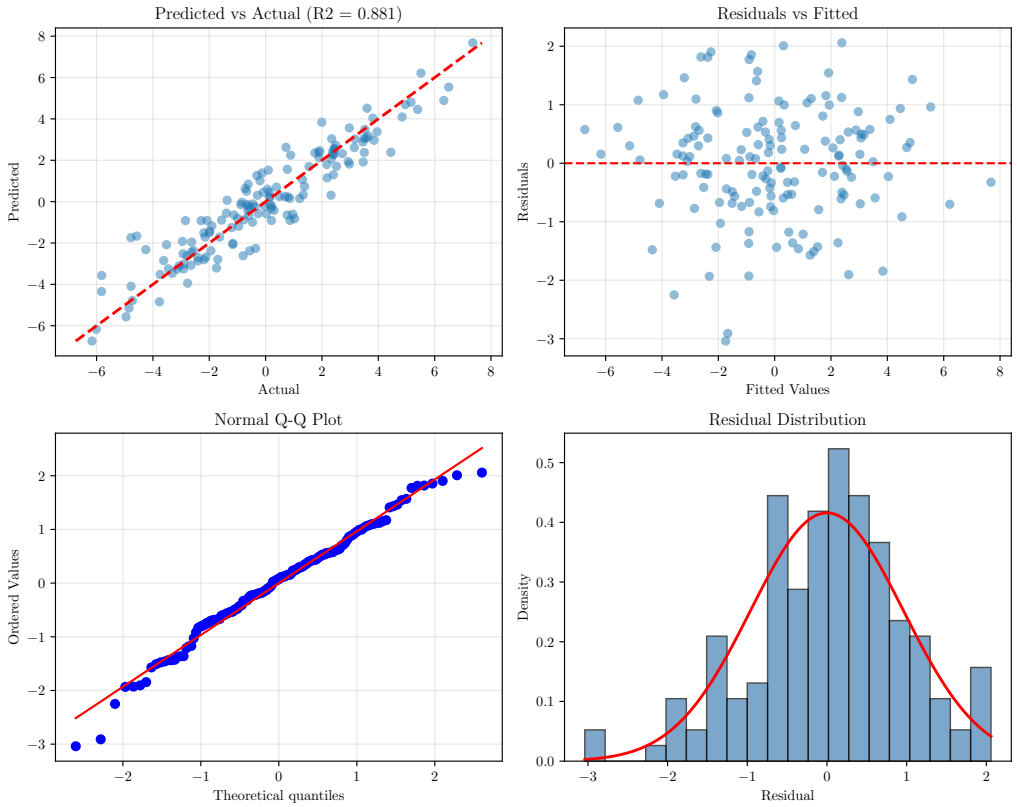

Figure 2: OLS regression diagnostics showing residual analysis.

OLS Performance: Train $R^2 = 0.881$, Test $R^2 = 0.797$.

5 Gradient Descent

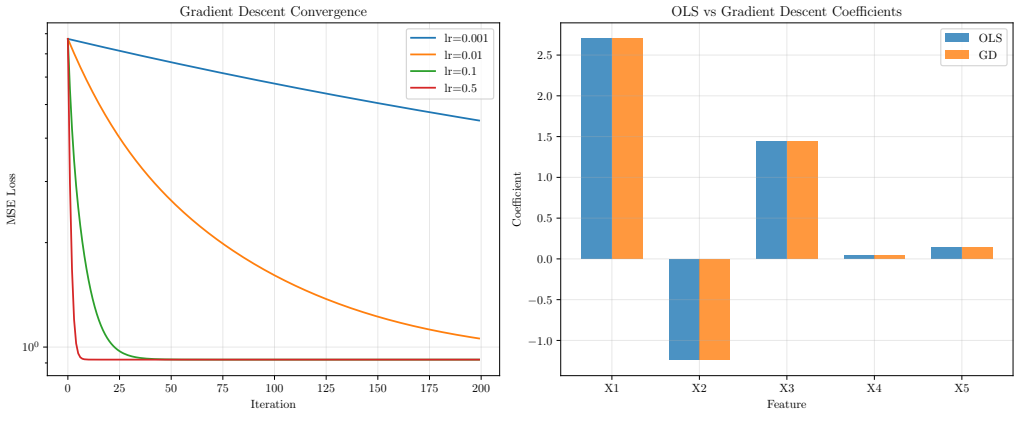

Figure 3: Gradient descent convergence and coefficient comparison with OLS.

6 Ridge Regression

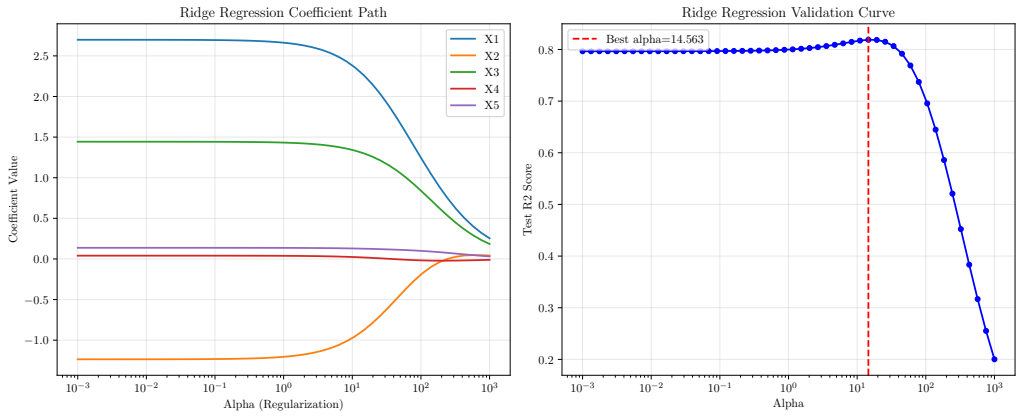

Figure 4: Ridge regression coefficient path and validation curve.

Best Ridge alpha: 14.563 with test $R^2 = 0.819$.

7 Lasso Regression

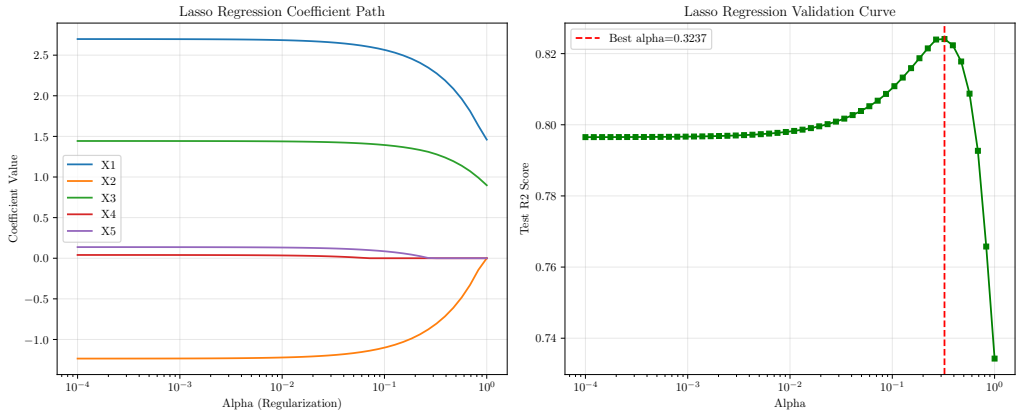

Figure 5: Lasso regression coefficient path showing feature selection.

Best Lasso alpha: 0.3237 with test $R^2 = 0.824$. Selected 3 features.

8 Model Comparison

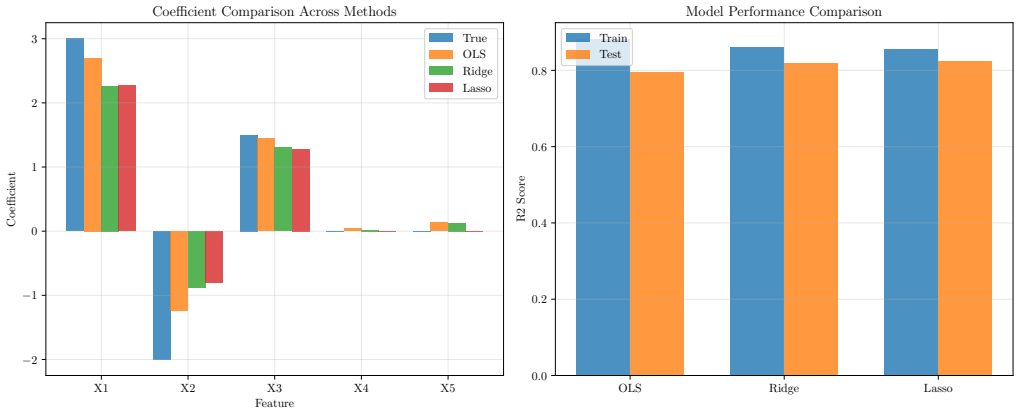

Figure 6: Comparison of OLS, Ridge, and Lasso regression methods.

9 Results Summary

Table 1: Regression Model Performance

Model	Train R^2	Test R^2	Hyperparameter
OLS	0.881	0.797	-
Ridge	0.862	0.819	$\alpha = 14.563$
Lasso	0.857	0.824	$\alpha = 0.3237$

Table 2: Coefficient Estimates

Feature	True	OLS	Ridge	Lasso
X1	3.00	2.70	2.27	2.28
X2	-2.00	-1.24	-0.88	-0.80
X3	1.50	1.44	1.30	1.28
X4	0.00	0.04	0.02	0.00
X5	0.00	0.14	0.13	0.00

10 Conclusion

This analysis demonstrated:

- OLS as the baseline with closed-form solution
- Gradient descent as iterative optimization method
- Ridge regression for handling multicollinearity (L2 penalty)
- Lasso regression for feature selection (L1 penalty)
- Hyperparameter tuning via validation curves

The Lasso method successfully identified the relevant features and set irrelevant coefficients close to zero, demonstrating its utility for sparse solutions.