Linear Regression

From-Scratch NumPy Implementation

Demonstration & Analysis Report

Seed: 42

Summary

This report demonstrates a from-scratch linear regression implementation using only NumPy. The implementation includes:

- Two solving methods:
  - Normal Equation (closed-form solution using np.linalg.lstsq)
  - Gradient Descent (iterative optimization)

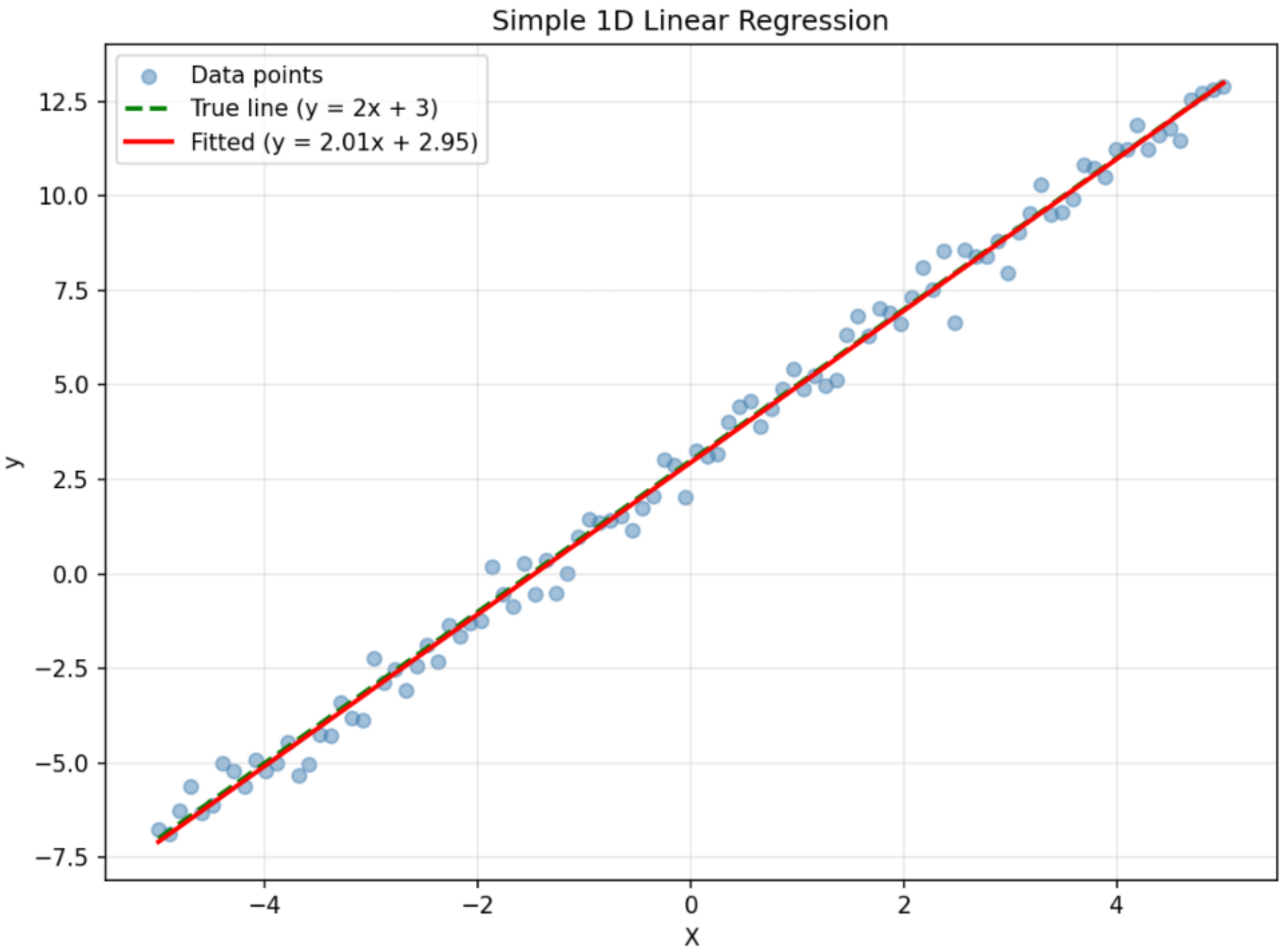
- Features:
  - Supports 1D and 2D input arrays
  - Convergence checking with tolerance
  - Training history tracking
  - R^2 score computation

- Validated against:
  - Hand-computed examples
  - Scikit-learn LinearRegression
  - California Housing real-world dataset

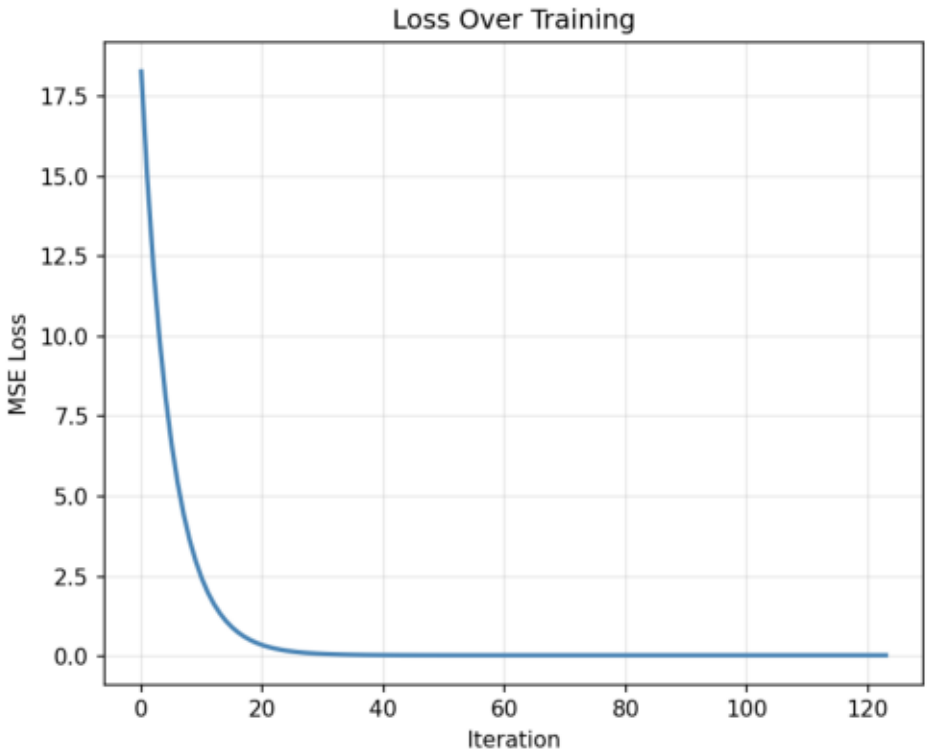
Key Findings:

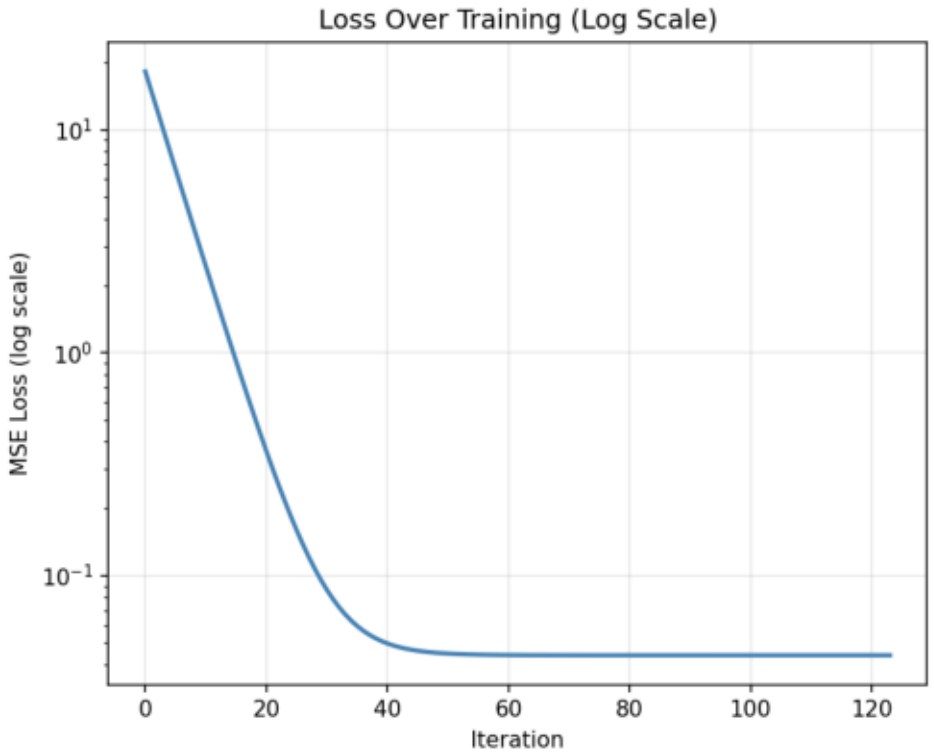
1. Our implementation matches sklearn to machine precision
2. Normal equation is faster for small-medium datasets
3. Gradient descent converges reliably with appropriate learning rate
4. Both methods recover true parameters from noisy data

Example 1: Simple 1D

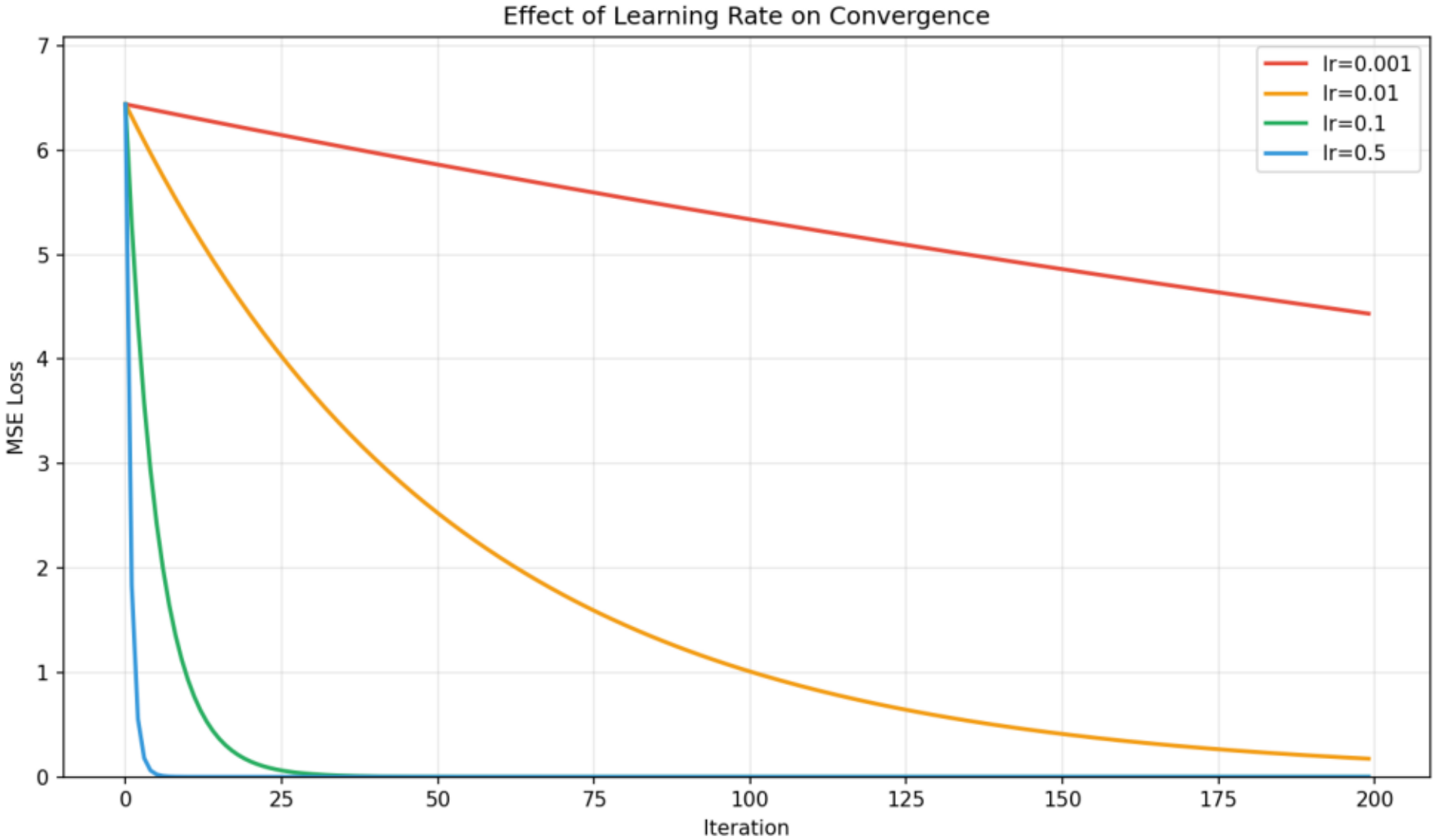


Example 2: Convergence

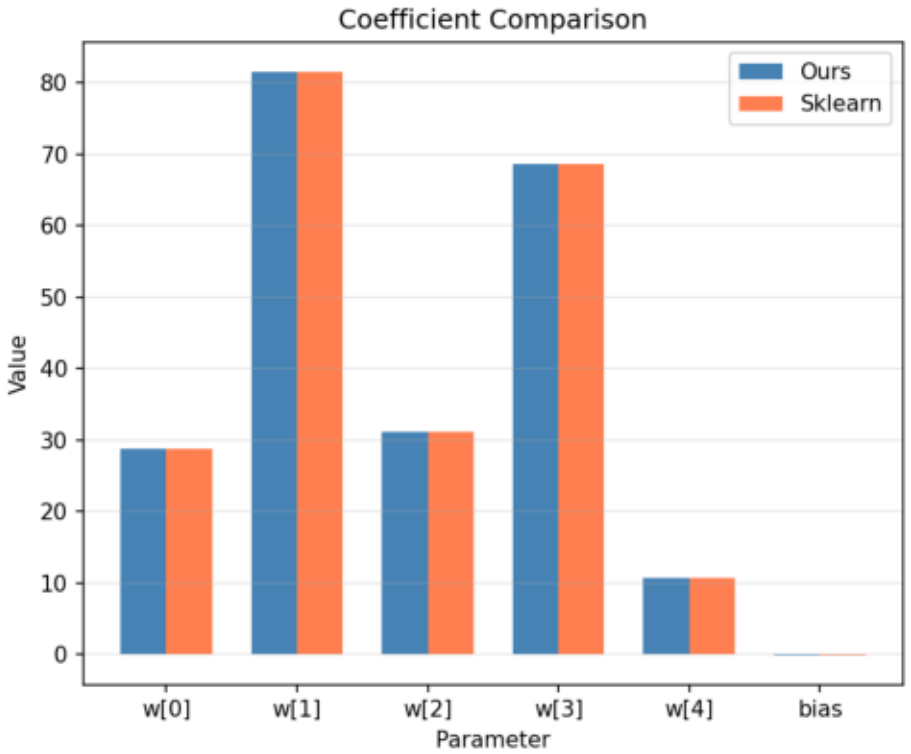


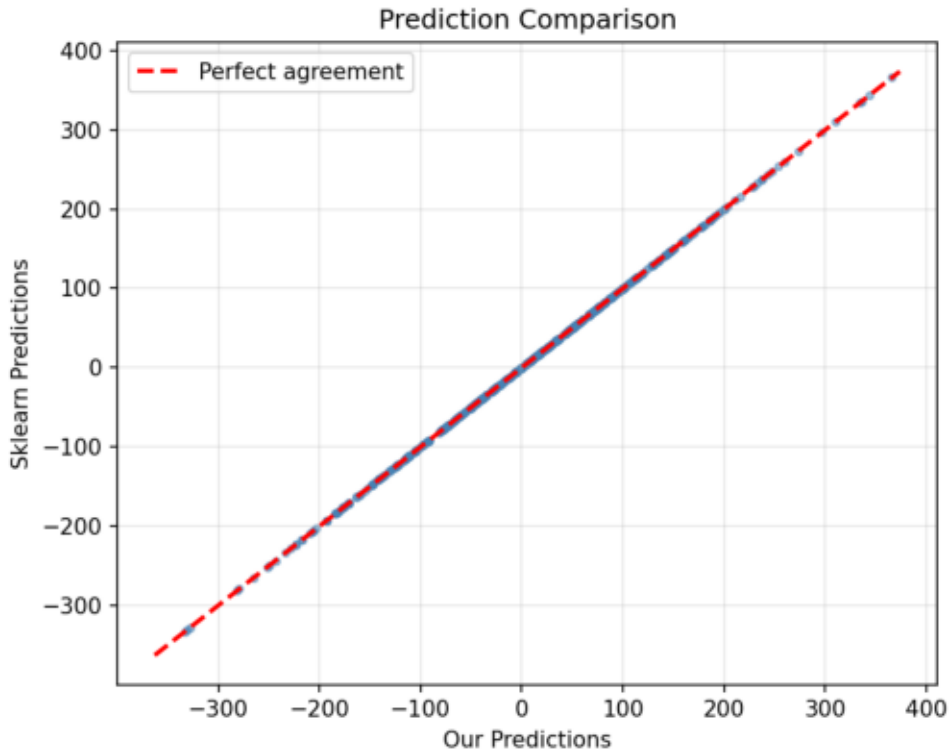


Example 3: Learning Rates

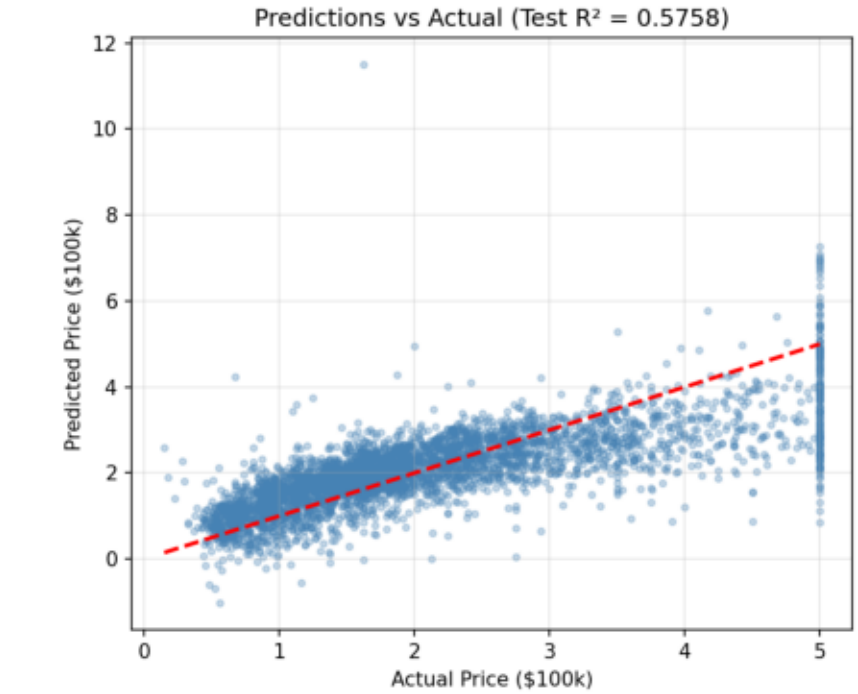


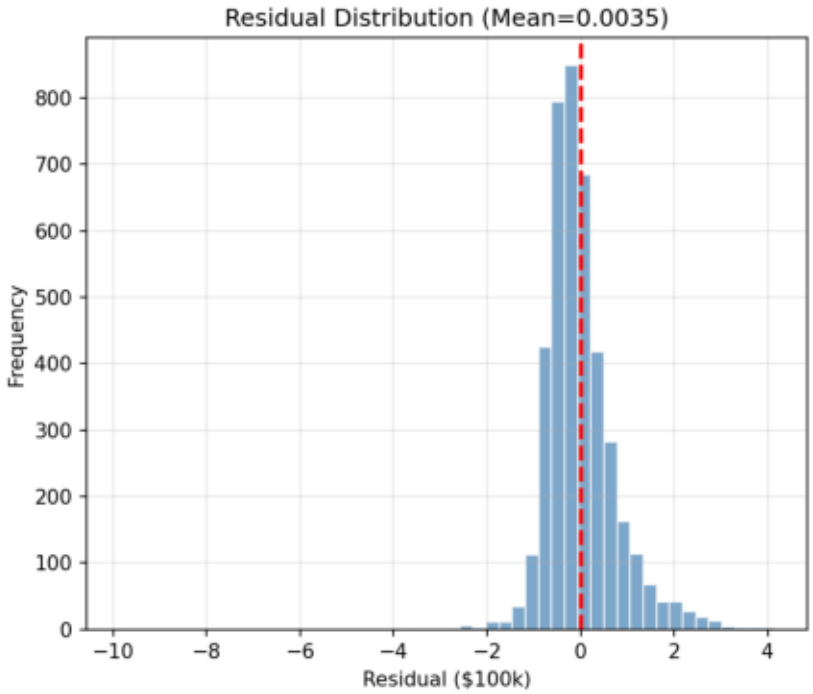
Example 4: Sklearn Comparison

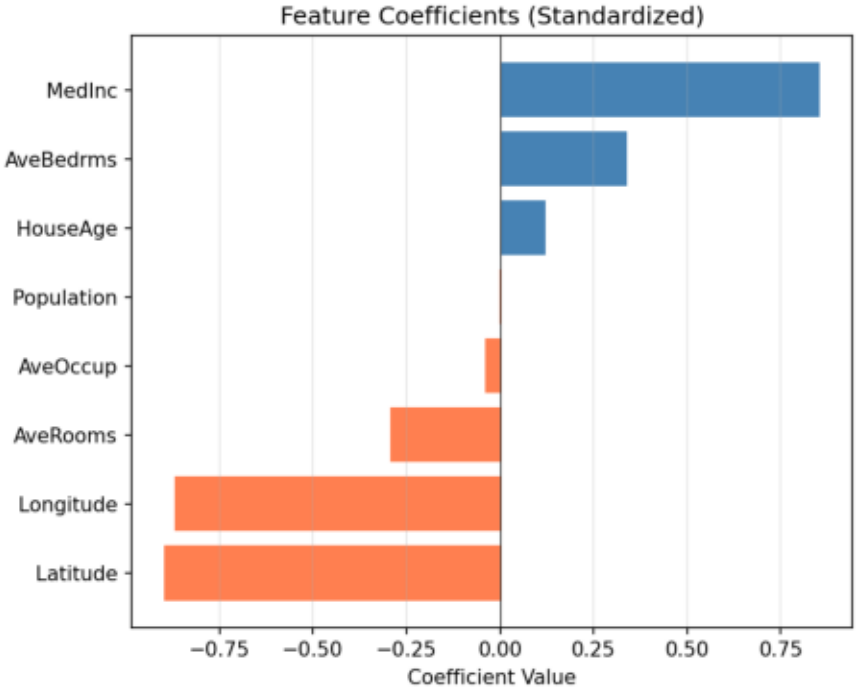


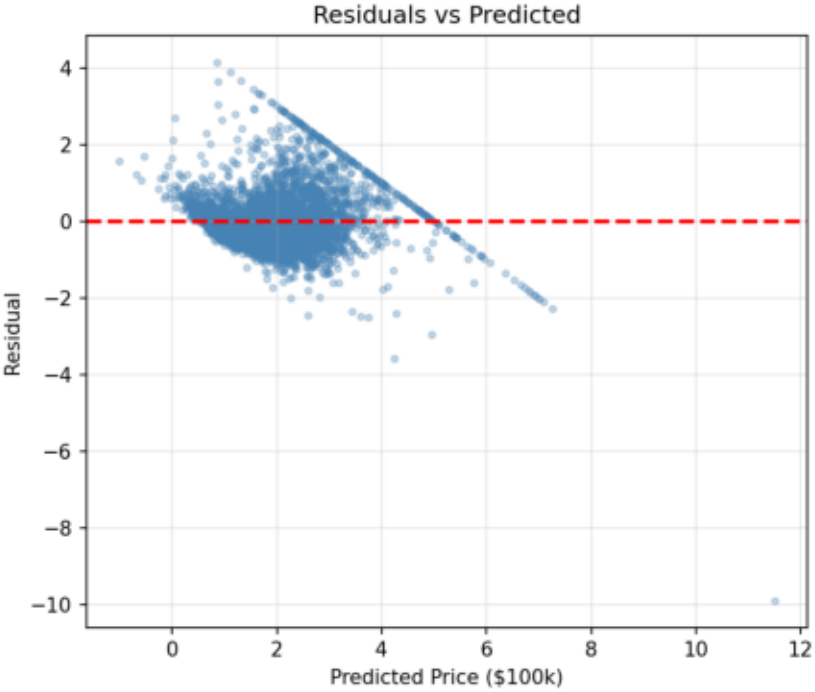


Example 5: California Housing









Example 6: GD vs Normal Eq