

Regularization

L1 (Lasso), L2 (Ridge), Elastic Net & Weight Decay

Comprehensive Demo with Sklearn Comparison

Seed: 42

From-scratch NumPy implementation
Phase 1: Foundations

Summary of Results

Example 1: Overfitting Demonstration

- No regularization MSE: 0.0193
- Ridge ($L_2=0.1$) MSE: 0.0715
- Lasso ($L_1=0.05$) MSE: 0.0790

Example 2: L1 vs L2 Weight Sparsity

- Ridge near-zero weights: 1/50
- Lasso near-zero weights: 28/50

Example 3: Lambda Sweep

- Best Ridge lambda: 0.0051 (val MSE: 237.09)
- Best Lasso lambda: 0.6210 (val MSE: 221.23)

Example 5: Sklearn Comparison (Ridge)

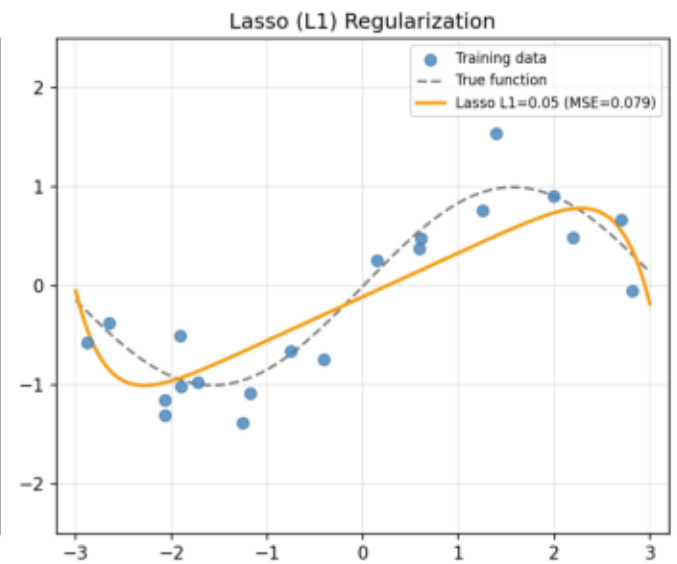
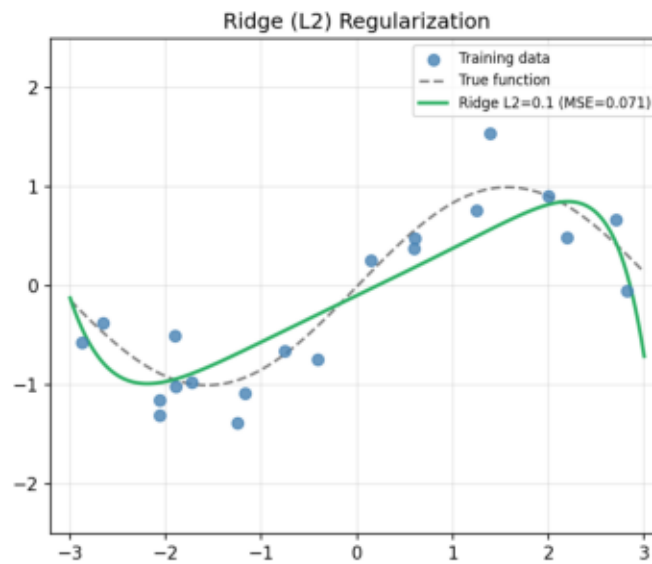
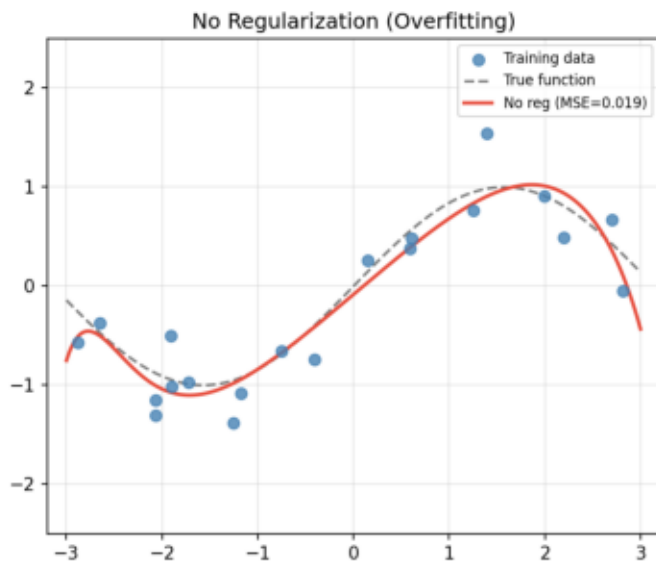
- Max weight diff (closed-form): $5.68e-14$
- Our R^2 : 0.731681
- Sklearn R^2 : 0.731681

Example 7: Weight Decay vs L2 Equivalence

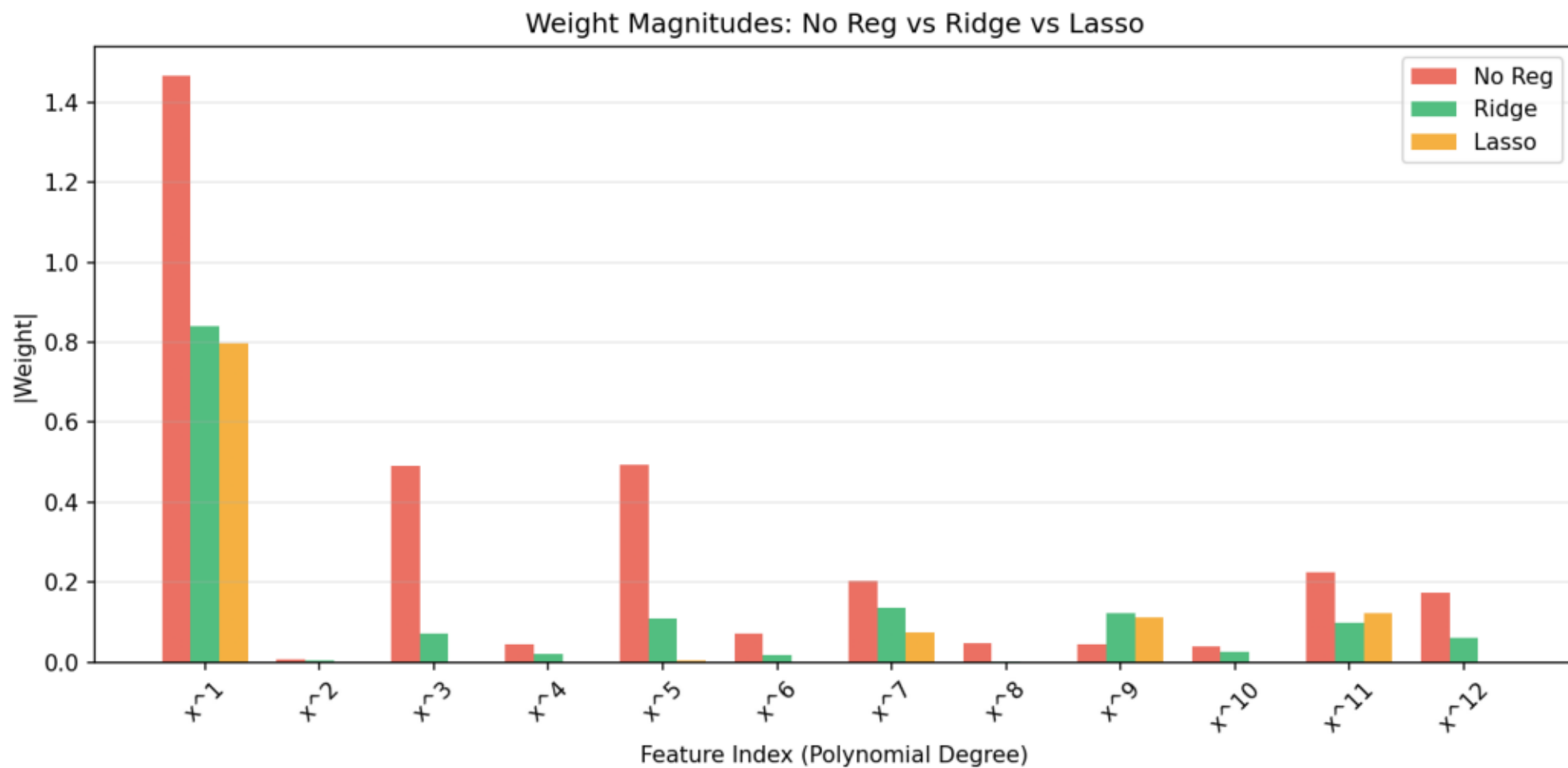
- Final max diff: $4.97e-14$
- Confirms identical trajectories for vanilla SGD

01 Overfitting Demo

Overfitting Demo: Degree 12 Polynomial, 20 Points

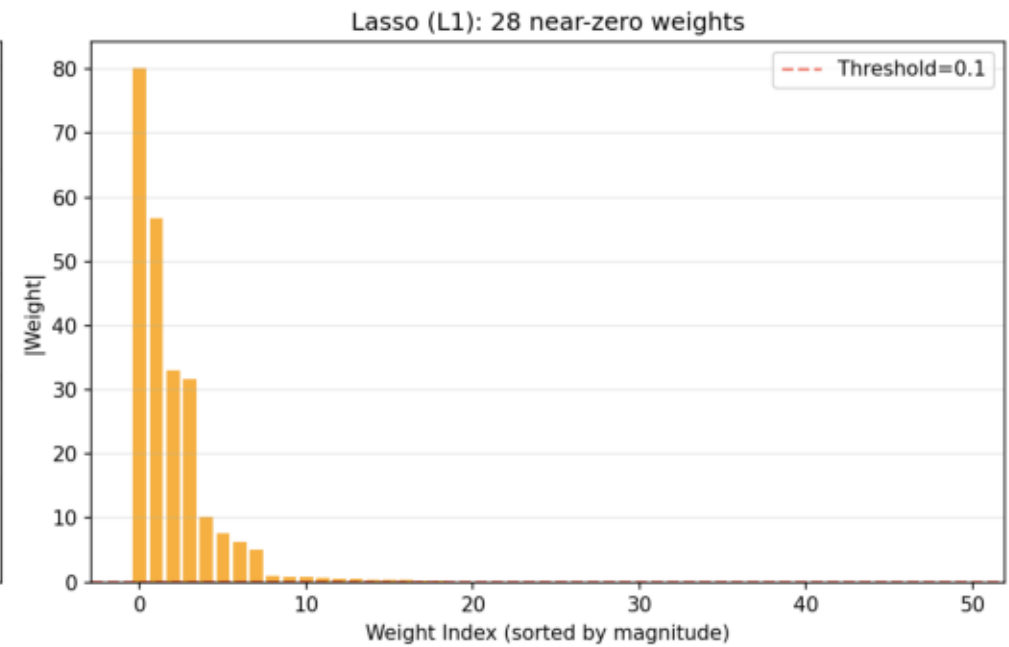
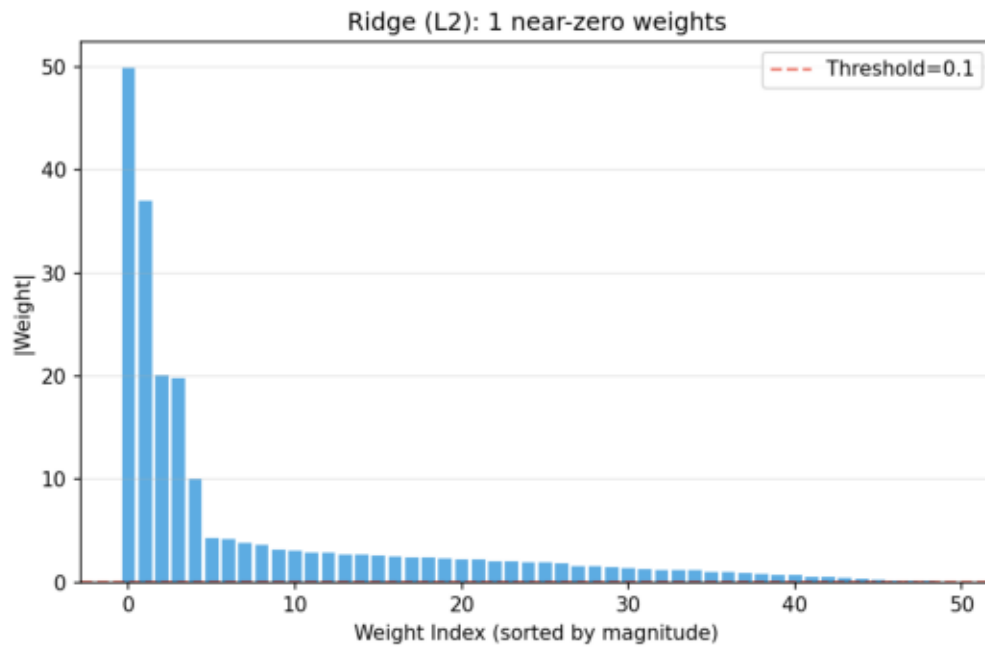


01B Weight Magnitudes



02 L1 Vs L2 Weights

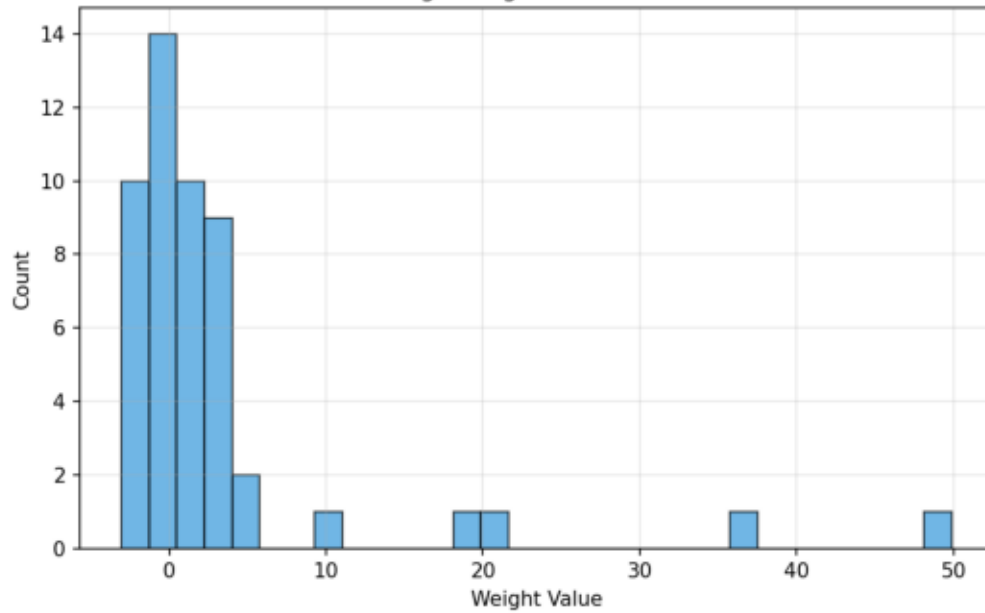
L1 vs L2: Weight Distributions (lambda=0.5, 50 features, 8 informative)



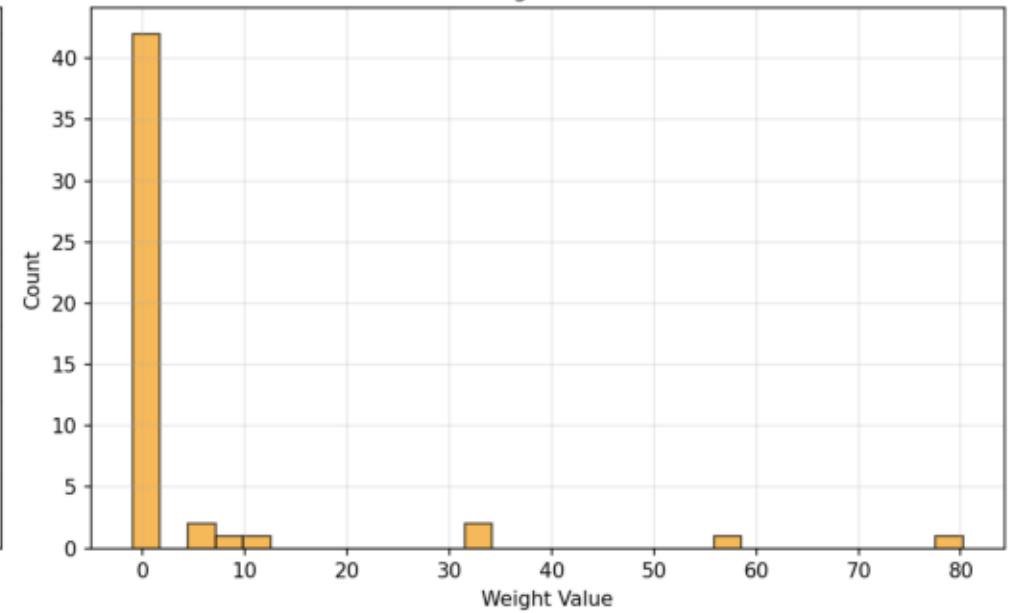
02B Weight Histograms

Weight Value Histograms: Ridge vs Lasso

Ridge Weight Distribution



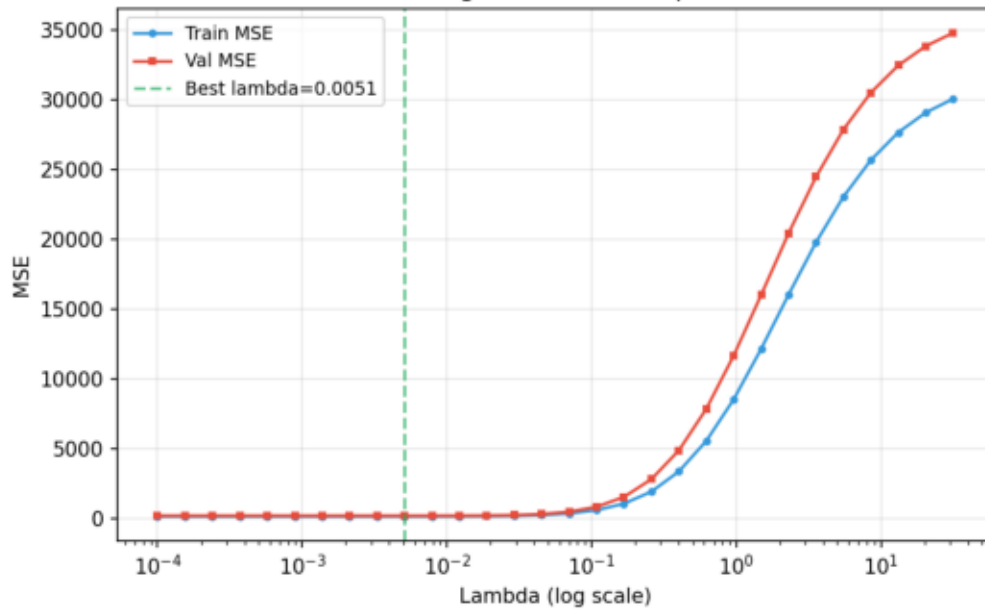
Lasso Weight Distribution



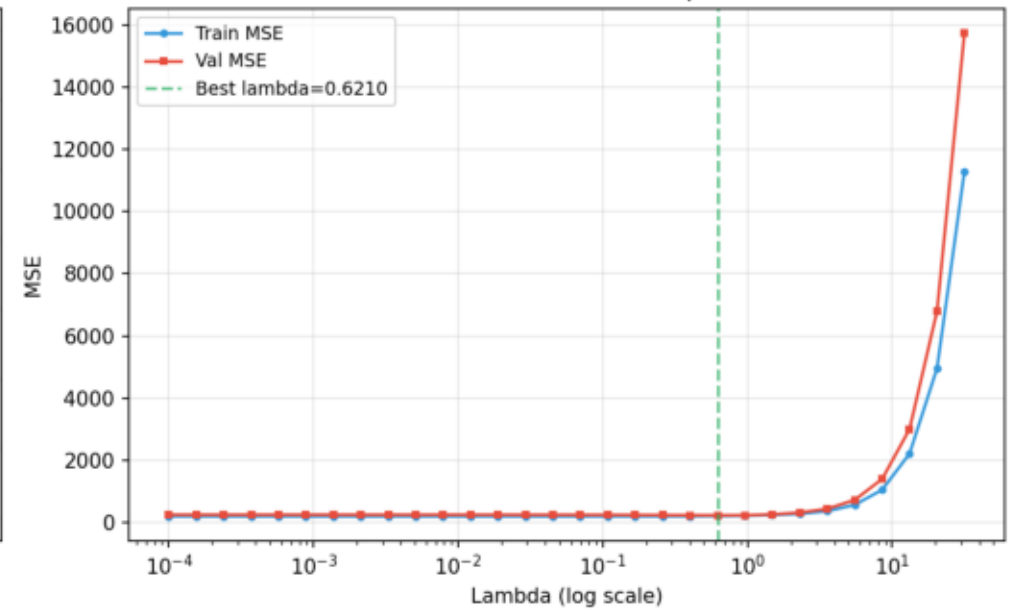
03 Lambda Sweep

Lambda Sweep: Finding the Bias-Variance Sweet Spot

Ridge: Lambda Sweep

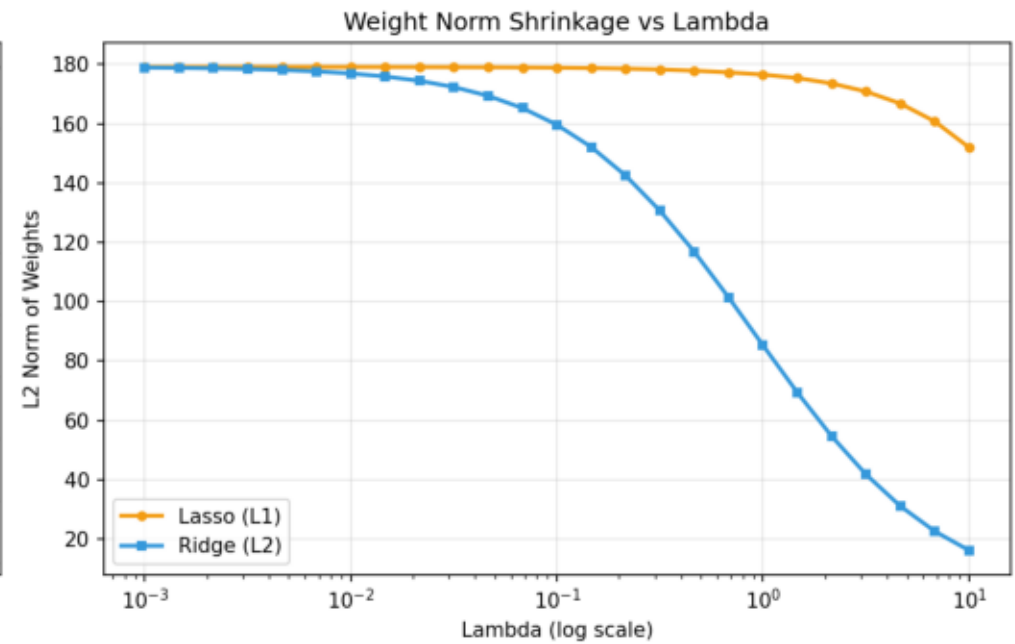
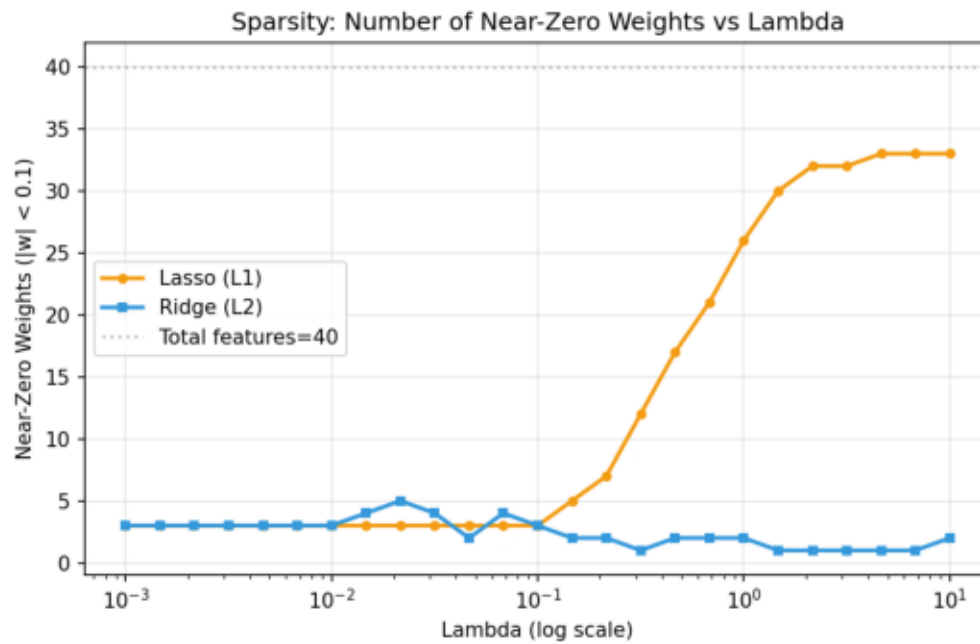


Lasso: Lambda Sweep



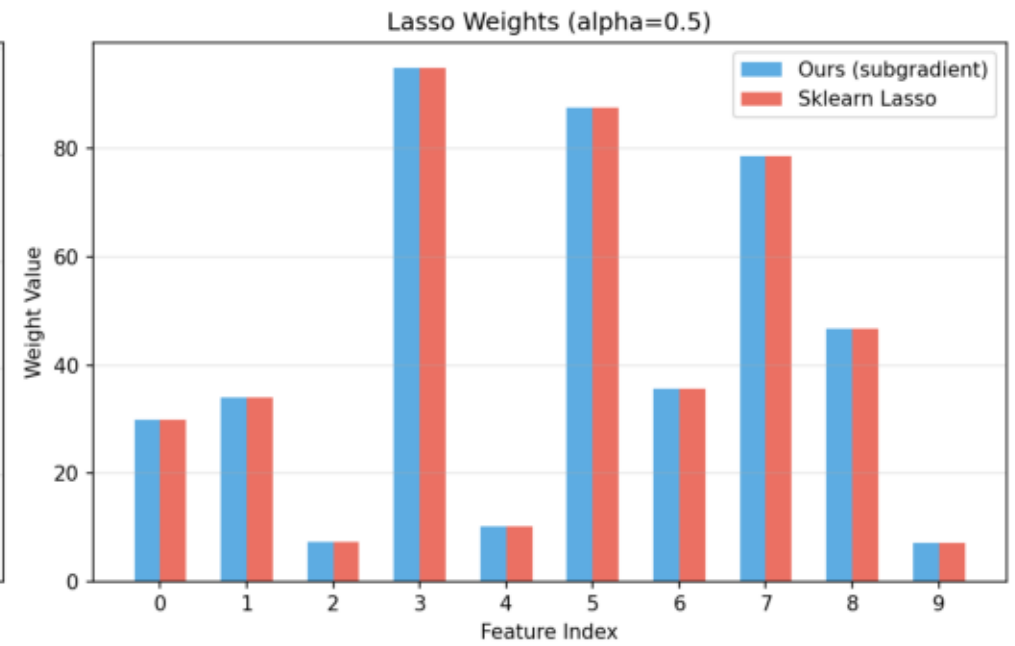
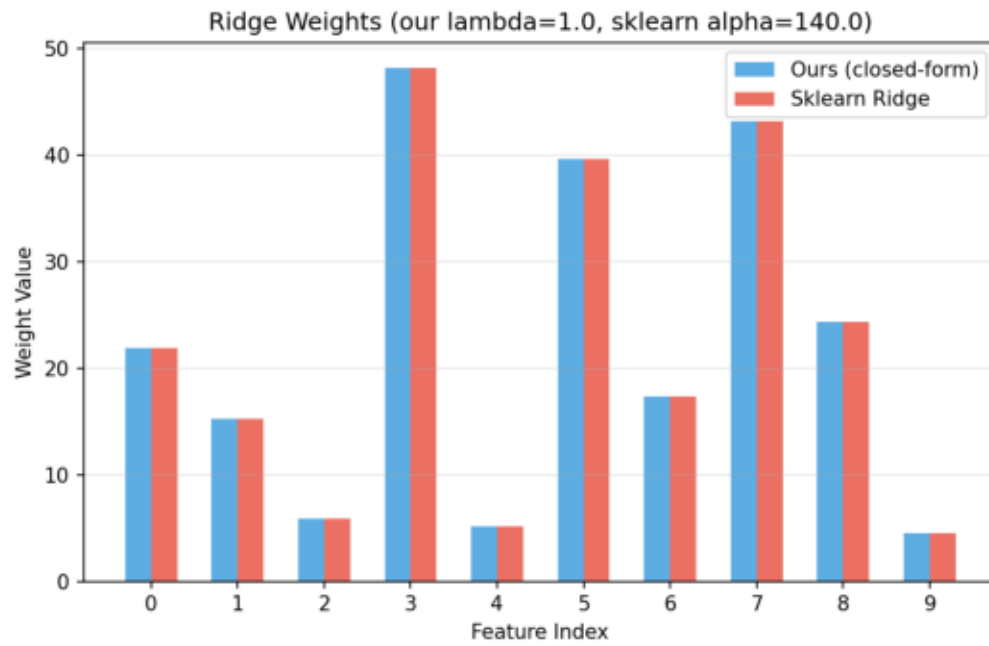
04 Sparsity Vs Lambda

Sparsity and Shrinkage: L1 vs L2 Across Lambda Values

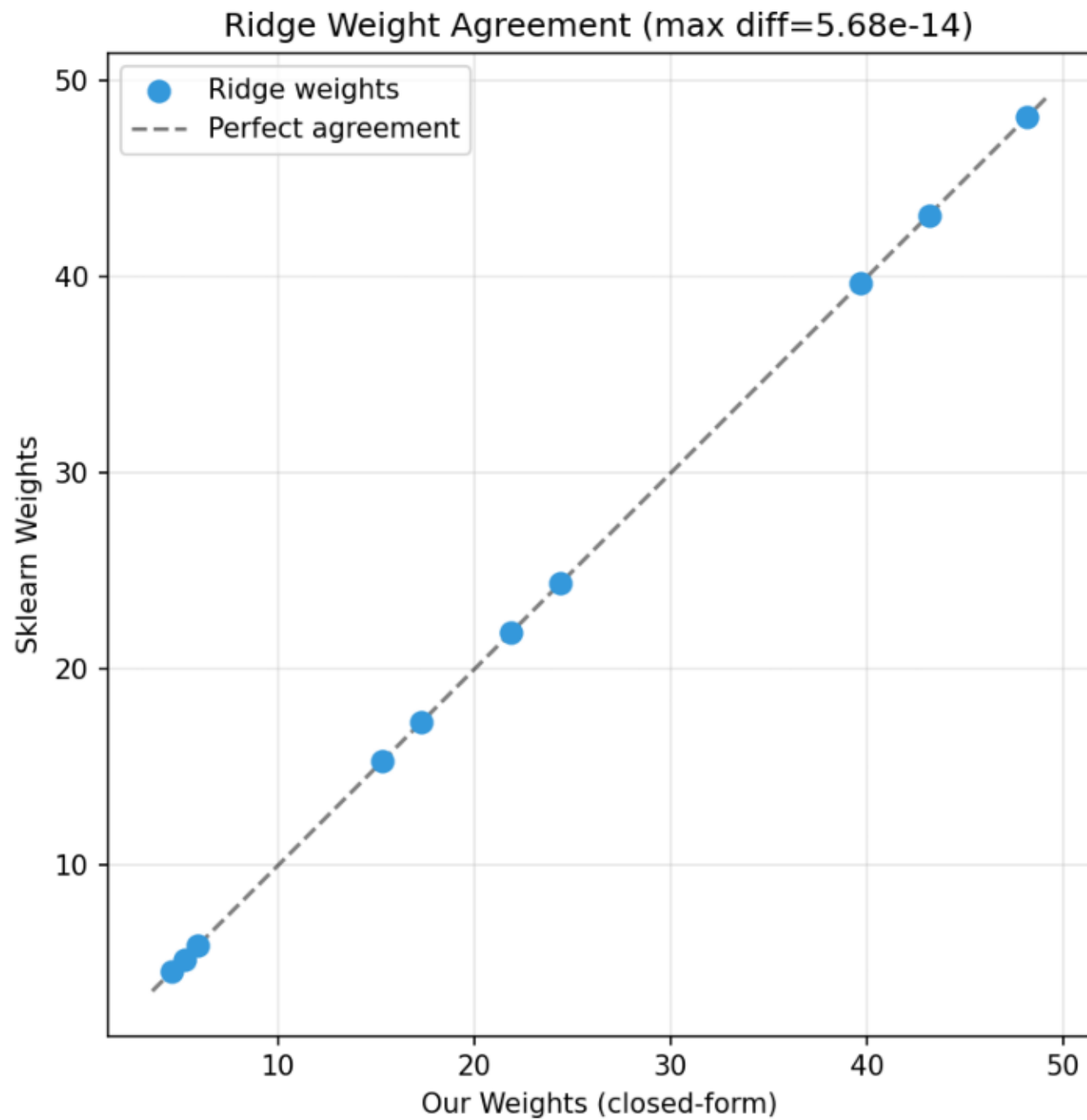


05 Sklearn Comparison

Sklearn Comparison: Our Implementation vs Sklearn

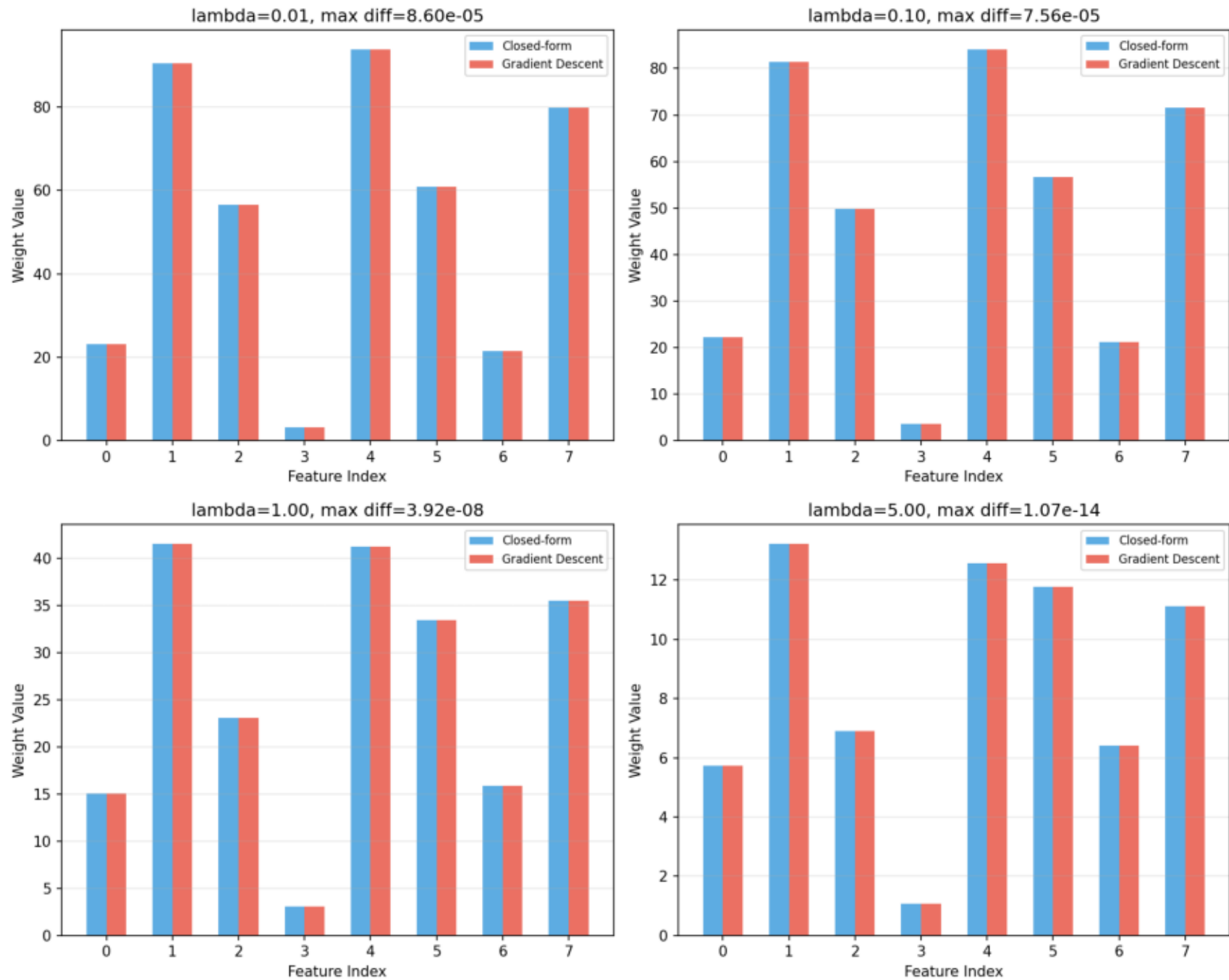


05B Ridge Weight Agreement



06 Closed Form Vs Gd

Ridge: Closed-Form vs Gradient Descent Weights



07 Weight Decay Equivalence

