Ridge Regression → L2 regularized model

→ Linear model

Introduction to LASSO Regression

↳ L1 regularized model

Ridge → $\lambda \sum_{j=1}^{p} \beta_j^2$

1. What is Lasso Regression?

Lasso Regression (Least Absolute Shrinkage and Selection Operator) is a regularized version of linear regression that adds an L1 penalty term to the loss function. This regularization allows the model to shrink some coefficients to exactly zero, effectively performing feature selection.

Loss Function

→ $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$

$$\text{Loss} = \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{OLS loss function}} + \underbrace{\lambda \sum_{j=1}^{p} |\beta_j|}_{\text{Penalty term}}$$

Where:

- λ is the regularization strength
- β_j are the model coefficients
- The L1 norm encourages sparsity in the model

2. Why Use Lasso Regression?

Lasso is particularly useful when dealing with high-dimensional datasets or when feature selection is desired. Unlike Ridge, Lasso can eliminate irrelevant features by setting their coefficients to zero.

3. Applications of Lasso Regression

- ✓ **Genomics:** Selecting relevant genes from thousands of features
- ✓ **Marketing:** Identifying key customer attributes driving conversion
- ✓ **Economics:** Modeling predictors of economic performance from many indicators
- ✓ **Text analysis:** Selecting the most important keywords or terms

4. Pros of Lasso Regression

- **Feature selection:** Can reduce model complexity by eliminating irrelevant features
- **Reduces overfitting:** Introduces bias to lower variance
- **Interpretability:** Resulting models are often easier to interpret
- **Useful in high-dimensional settings**

5. Cons of Lasso Regression

- **May exclude important variables:** Especially when predictors are highly correlated
- **Biased estimates:** Coefficients are shrunk toward zero
- **Model performance sensitive to λ:** Requires careful tuning

6. Summary

Lasso regression is an effective tool for both prediction and feature selection in linear models. Its ability to reduce the number of predictors makes it an excellent choice when interpretability and parsimony are desired.