Lasso Regression: Explanation and Formula

Lasso Regression (Least Absolute Shrinkage and Selection Operator) is a linear regression technique that uses **L1 regularization** to improve model performance by adding a penalty to the loss function. The penalty term encourages sparsity in the model by shrinking some coefficients to exactly zero, effectively performing **feature selection**.

Objective Function

The objective function for Lasso Regression is:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\theta_j|$$

Where:

- $\mathcal{L}(\theta)$: Total loss function.
- y_i: Actual target value for the i-th observation.
- $\hat{y}_i = X_i^\top \theta$: Predicted value for the i-th observation.
- θ_j: Coefficients (parameters) of the j-th predictor.
- λ: Regularization hyperparameter (controls the penalty strength).
- n: Number of observations (samples).
- p: Number of predictors (features).

Key Characteristics of Lasso Regression

- **L1 Regularization:** The penalty term, $\lambda \sum_{j=1}^{p} |\theta_j|$, forces some coefficients to become exactly zero.
- **Feature Selection:** By setting some coefficients to zero, Lasso automatically selects the most important predictors.
- **Sparsity:** Lasso regression is particularly effective in high-dimensional datasets where many predictors are irrelevant.

Effect of Regularization Parameter (λ)

- When $\lambda = 0$, Lasso regression reduces to Ordinary Least Squares (OLS) regression.
- As $\lambda \to \infty$, all coefficients shrink towards zero, resulting in an overly simplistic model.
- The optimal λ balances model complexity and generalization ability. It is typically determined using cross-validation.

Soft-Thresholding in Lasso Regression

The coefficients in Lasso regression are estimated using a method called **soft-thresholding**. For each predictor j, the coefficient θ_j is updated as:

$$\theta_j = \text{sign}(z_j) \cdot \max\left(|z_j| - \lambda, 0\right)$$

Where:

- z_j: The correlation between the j-th predictor and the residuals of the model.
- $\text{sign}(z_j)$: The sign of z_j (positive or negative).
- λ: Regularization parameter.

This formula shrinks coefficients with small contributions to zero, while retaining those with large contributions.

Advantages of Lasso Regression

- **Feature Selection:** Identifies and retains only the most relevant predictors.
- **Simplicity:** Produces a sparse model that is easier to interpret.
- **Effective in High-Dimensional Data:** Handles datasets with many irrelevant predictors.

Limitations of Lasso Regression

- **Bias in Coefficients:** The L1 penalty can introduce bias in the coefficient estimates.
- **Multicollinearity:** Lasso may arbitrarily select one predictor from a group of highly correlated predictors.
- **Dependent on Scaling:** Features must be standardized for proper regularization.