Ridge Regression: Explanation and Formula

Ridge regression, also known as **L2 Regularization**, is a linear regression technique that addresses the issue of overfitting and multicollinearity by adding a penalty term to the loss function. It is particularly useful when predictors are highly correlated or when the number of predictors exceeds the number of observations.

Objective Function

The Ridge regression aims to minimize the following objective function:

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

Where:

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

Closed-Form Solution

The Ridge regression coefficients are computed using the following closed-form solution:

$$\theta = \left(X^\top X + \lambda I\right)^{-1} X^\top y$$

Where:

- X: Feature matrix ($n \times p$).
- y: Target vector ($n \times 1$).
- I: Identity matrix ($p \times p$).
- λ: Regularization hyperparameter. A larger λ applies more penalty.

Effect of Regularization

- When $\lambda = 0$, Ridge regression is equivalent to Ordinary Least Squares (OLS).
- As $\lambda \to \infty$, the coefficients θ_j shrink towards zero, simplifying the model but not setting coefficients exactly to zero.

Advantages

- Handles multicollinearity by stabilizing coefficients of correlated predictors.
- Reduces overfitting, improving the model's generalization ability.

Limitations

- Does not perform feature selection; all predictors are retained in the model.
- Requires standardization of features for optimal performance.