Introduction to Ridge Regression

1. Introduction

Ridge Regression is a regularized version of linear regression. It adds a penalty term to the ordinary least squares (OLS) loss function to reduce model complexity and prevent overfitting. The loss function is:

penalty term

$$\text{Loss} = \underbrace{\sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)\right]^2}_{\text{OLS Loss function}} + \lambda \sum_{j=1}^{p} \beta_j^2$$

if λ is small, very lite penalty

Where:

- λ is the regularization parameter
- β_j are the model coefficients
- Larger λ values imply stronger penalty (more shrinkage)

2. Why Use Ridge Regression?

When features are highly correlated or when the number of features is large, OLS may produce overfit models. Ridge regression reduces this by shrinking coefficients toward zero (but not exactly zero).

3. Applications of Ridge Regression

- **Healthcare:** Predicting disease progression using high-dimensional medical records
- **Marketing:** Modeling customer behavior with many demographic and behavioral variables
- **Finance:** Portfolio prediction with correlated asset returns

4. Pros of Ridge Regression

- **Reduces overfitting** through regularization
- **Works well with multicollinearity**
- **Retains all features** while shrinking their coefficients
- **Computationally efficient**

5. Cons of Ridge Regression

- **Coefficients are harder to interpret** due to shrinkage
- **Not suitable for feature selection** (unlike Lasso)
- **Still assumes linearity** and independence between errors
- **Hyperparameter tuning (λ) is required**

8. Summary

Ridge Regression is a powerful tool for handling high-dimensional, multicollinear datasets where standard linear regression fails. It introduces bias to reduce variance and improve generalization.