Linear Regression with Ridge Regularization

Problem Statement

Consider a dataset where we aim to predict an outcome based on a set of input features. However, we want to avoid overfitting by controlling the complexity of the model. The objective is to formulate an optimization problem that balances the trade-off between fitting the training data well and preventing overfitting caused by excessive model complexity.

Mathematical Formulation

Let $\mathbf{x}_i$ denote the ith input vector and y_i represent the associated target value corresponding to the input vector $\mathbf{x}_i$ in the training dataset. Assuming a linear relationship, the predicted value $\hat{y}_i$ for input vector $\mathbf{x}_i$ is given by $\hat{y}_i = \mathbf{w}^T\mathbf{x}_i + b$, where $\mathbf{w}$ is the coefficient vector and b is the bias term.

Important Note: While the formulation above shows the model for a single observation, Ridge regression typically deals with minimizing the error across all observations in the dataset. However, for a single observation, the summation doesn't strictly apply.

To calculate the total squared error for all observations, we can use the following:

$$\text{Total Error} = \sum_{i=1}^{n} ||\mathbf{w}^T\mathbf{x}_i + b - y_i||^2$$

Here, the summation () iterates over all n observations, and the squared error for each observation (i) is calculated and summed to get the total error.

The goal is to minimize this error to achieve the best fit for the training data. However, overfitting can occur when the model becomes too complex, leading to poor generalization to unseen data. To address overfitting, Ridge regression is employed, which adds a regularization term to the objective function. The regularized objective function becomes:

$$\text{minimize} \sum_{i=1}^{n} ||\mathbf{w}^T\mathbf{x}_i + b - y_i||^2 + \lambda ||\mathbf{w}||^2$$

where λ is the regularization parameter that controls the strength of the penalty. The regularization term $||\mathbf{w}||^2$ penalizes large coefficients in the coef-

ficient vector $\mathbf{w}$, thereby encouraging simpler models that generalize better to unseen data.

Summary

Ridge regularization is a powerful technique in linear regression for preventing overfitting by penalizing large coefficients. By incorporating a regularization term into the objective function, Ridge regression promotes models that strike a balance between fitting the training data well and controlling model complexity. This results in more robust models that generalize effectively to unseen data.