# Ridge Regression

January 26, 2025

## 1 Introduction

Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where independent variables are highly correlated. When the independent variables are highly correlated, the coefficients of the regression model are unstable. Ridge regression adds a small bias factor to the diagonal of the correlation matrix to stabilize the model. The bias factor is called the ridge parameter, which is a hyperparameter that needs to be tuned. Ridge regression is also known as Tikhonov regularization, named after Andrey Tikhonov, a Russian mathematician who first proposed the method in 1943.

## 2 Why Ridge Regression?

The most common regression method is Ordinary Least Squares (OLS), which aims to find a line (or hyperplane) that minimizes the error between predicted and actual values.

The goal of OLS is to minimize the sum of squared errors:

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2$$

where: X is the matrix of independent variables. y is the vector of the dependent variable. $\beta$ is the vector of regression coefficients.

The solution is:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

The problem is:

- If the independent variables are highly correlated (e.g. house size and number of rooms are strongly related), the matrix $X^T X$ becomes nearly singular (its determinant is close to zero), causing $(X^T X)^{-1}$ to blow up.
- This makes the estimates of $\beta$ very unstable, with high variance, leading to poor prediction performance.

## 3 Ridge Regression to the Rescue

Ridge Regression solves the multicollinearity problem by adding a regularization term ($\lambda\|\beta\|^2$) to the OLS objective function. The new objective function is:

$$\hat{\beta} = \arg\min_{\beta}(\|y - X\beta\|^2 + \lambda\|\beta\|^2)$$

The solution becomes:

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$$

where: $\lambda$ is the ridge parameter. I is the identity matrix.

To conclude:

- The regularization term ($\lambda\|\beta\|^2$) penalizes large regression coefficients, reducing the model's complexity.
- As $\lambda$ increases, the regression coefficients shrink, reducing the model's variance at the cost of introducing some bias.
- By choosing an appropriate $\lambda$, Ridge Regression achieves a balance between bias and variance, improving the model's prediction performance.

Ridge regression adds the least squares of the second-order regular term to the loss function, also called L2 parametrization, which has the effect of dimensionality reduction, and also limits the matching of the model parameters to the abnormal samples and deals with the highly correlated data sets, thus improving the fitting accuracy of the model to most normal samples. Our team used RidgeCV to adjust the regularization strength alpha to achieve a better fit at alpha 14.

Ridge regression is often the preferred method for handling multicollinearity among features due to its use of L2 regularization, which shrinks coefficients without eliminating variables. This approach maintains all features in the model, allowing for a more comprehensive analysis and easier interpretation. Unlike Lasso regression, which uses L1 regularization to potentially remove features by driving coefficients to zero, Ridge regression ensures a more stable and consistent model by balancing the influence of all predictors. This makes it particularly valuable when all features are considered important for the model's explanatory power.