

# What is Lasso Regression and how can it be used in data analysis?

Authored by  
**stats writer**

April 22, 2024

## RECOMMENDED CITATION

stats writer (2024). *What is Lasso Regression and how can it be used in data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137960>

Lasso Regression is a type of linear regression that involves adding a penalty to the regression coefficients in order to reduce the complexity of the model and ultimately improve its predictive accuracy. It is commonly used in data analysis as a form of regularization to prevent overfitting and select the most relevant features for prediction. This method is particularly useful when dealing with high-dimensional datasets, where the number of predictors is large compared to the sample size. Lasso Regression has been widely applied in various fields such as finance, medicine, and social sciences, and has proven to be an effective tool for improving the interpretability and generalizability of regression models. By shrinking the coefficients of less important variables to zero, Lasso Regression can help identify the most influential predictors and improve the overall performance of the model.

## Introduction to Lasso Regression

In ordinary multiple linear regression, we use a set of  $p$  predictor variables and a response variable to fit a model of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

where:

$Y$ : The response variable  
 $X_j$ : The  $j$ th predictor variable  
 $\beta_j$ : The average effect on  $Y$  of a one unit increase in  $X_j$ , holding all other predictors fixed  
 $\varepsilon$ : The error term

The values for  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  are chosen using the least square method, which minimizes the sum of squared residuals (RSS):

$$RSS = \sum (y_i - \hat{y}_i)^2$$

where:

$\Sigma$ : A greek symbol that means *sumyi*: The actual response value for the *ith* observation?  
 $i$ : The predicted response value based on the multiple linear regression model

However, when the predictor variables are highly correlated then multicollinearity can become a problem. This can cause the coefficient estimates of the model to be unreliable and have high variance. That is, when the model is applied to a new set of data it hasn't seen before, it's likely to perform poorly.

One way to get around this issue is to use a method known as lasso regression, which instead seeks to minimize the following:

$$RSS + \lambda \sum |\beta_j|$$

where  $j$  ranges from 1 to  $p$  and  $\lambda \geq 0$ .

This second term in the equation is known as a *shrinkage penalty*.

When  $\lambda = 0$ , this penalty term has no effect and lasso

regression produces the same coefficient estimates as least squares.

However, as  $\lambda$  approaches infinity the shrinkage penalty becomes more influential and the predictor variables that aren't important in the model get shrunk towards zero and some even get dropped from the model.

Why Use Lasso Regression?

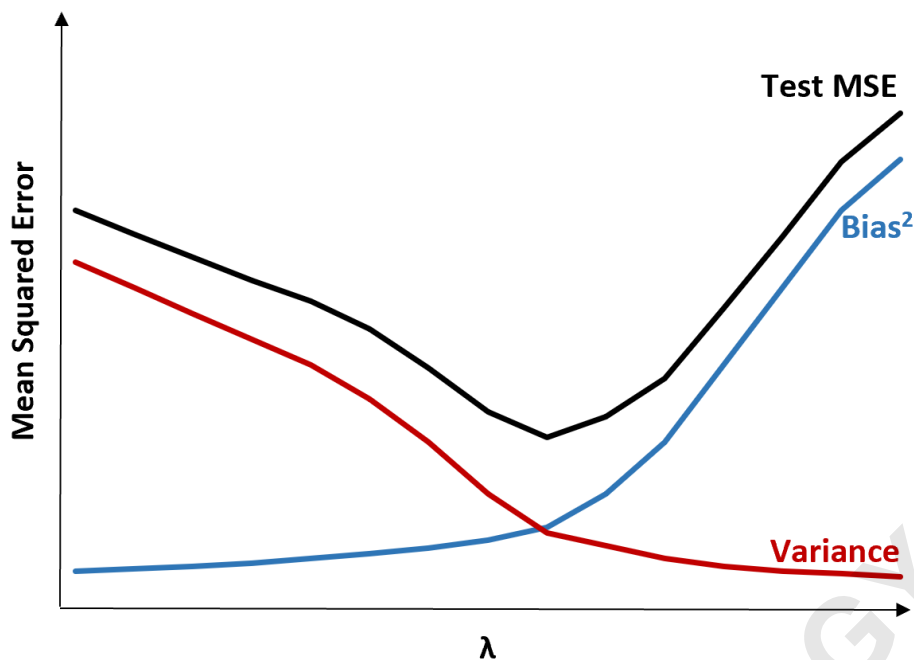
The advantage of lasso regression compared to least squares regression lies in the bias-variance tradeoff.

$$\text{MSE} = \text{Var}(f(x_0)) + \text{Bias}^2 + \text{Var}(\varepsilon)$$

$$\text{MSE} = \text{Variance} + \text{Bias}^2 + \text{Irreducible error}$$

The basic idea of lasso regression is to introduce a little bias so that the variance can be substantially reduced, which leads to a lower overall MSE.

To illustrate this, consider the following chart:

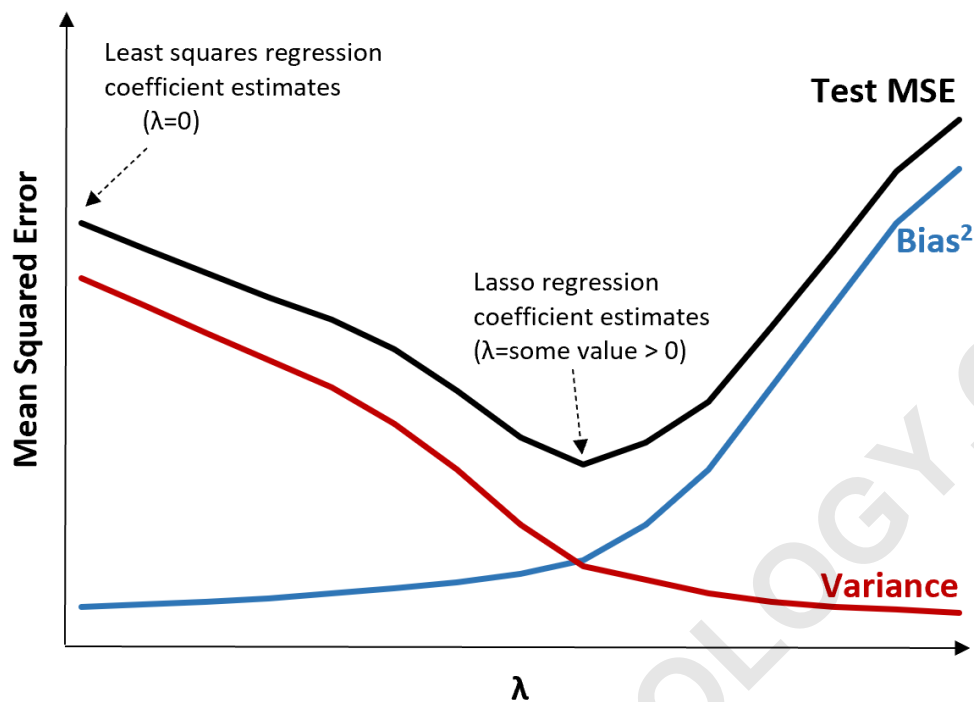


Notice that as  $\lambda$  increases, variance drops substantially with very little increase in bias. Beyond a certain point, though, variance decreases less rapidly and the shrinkage in the coefficients causes them to be significantly underestimated which results in a large increase in bias.

We can see from the chart that the test MSE is lowest when we choose a value for  $\lambda$  that produces an optimal tradeoff between bias and variance.

When  $\lambda = 0$ , the penalty term in lasso regression has no effect and thus it produces the same coefficient estimates as least squares. However, by increasing  $\lambda$  to

a certain point we can reduce the overall test MSE.



This means the model fit by lasso regression will produce smaller test errors than the model fit by least squares regression.

#### Lasso Regression vs. Ridge Regression

Lasso regression and ridge regression are both known as *regularization methods* because they both attempt to minimize the sum of squared residuals (RSS) along with some penalty term.

In other words, they constrain or *regularize* the

**coefficient estimates of the model.**

**However, the penalty terms they use are a bit different:**

**Lasso regression attempts to minimize  $RSS + \lambda \sum |\beta_j|$**   
**Ridge regression attempts to minimize  $RSS + \lambda \sum \beta_j^2$**

**When we use ridge regression, the coefficients of each predictor are shrunken towards zero but none of them can go *completely to zero*.**

**Conversely, when we use lasso regression it's possible that some of the coefficients could go *completely to zero* when  $\lambda$  gets sufficiently large.**

**In technical terms, lasso regression is capable of producing "sparse" models - models that only include a subset of the predictor variables.**

**This begs the question: Is ridge regression or lasso regression better?**

**The answer: It depends!**

**In cases where only a small number of predictor**

variables are significant, lasso regression tends to perform better because it's able to shrink insignificant variables completely to zero and remove them from the model.

However, when many predictor variables are significant in the model and their coefficients are roughly equal then ridge regression tends to perform better because it keeps all of the predictors in the model.

To determine which model is better at making predictions, we perform k-fold cross-validation. Whichever model produces the lowest test mean squared error (MSE) is the preferred model to use.

#### Steps to Perform Lasso Regression in Practice

The following steps can be used to perform lasso regression:

**Step 1: Calculate the correlation matrix and VIF values for the predictor variables.**

First, we should produce a correlation matrix and calculate the VIF (variance inflation factor) values for each predictor variable.

If we detect high correlation between predictor variables and high VIF values (some texts define a "high" VIF value as 5 while others use 10) then lasso regression is likely appropriate to use.

However, if there is no multicollinearity present in the data then there may be no need to perform lasso regression in the first place. Instead, we can perform ordinary least squares regression.

**Step 2: Fit the lasso regression model and choose a value for  $\lambda$ .**

Once we determine that lasso regression is appropriate to use, we can fit the model (using popular programming languages like R or Python) using the optimal value for  $\lambda$ .

To determine the optimal value for  $\lambda$ , we can fit several models using different values for  $\lambda$  and choose  $\lambda$  to be the value that produces the lowest test MSE.

**Step 3: Compare lasso regression to ridge regression and ordinary least squares regression.**

Lastly, we can compare our lasso regression model to a

ridge regression model and least squares regression model to determine which model produces the lowest test MSE by using k-fold cross-validation.

Depending on the relationship between the predictor variables and the response variable, it's entirely possible for one of these three models to outperform the others in different scenarios.

Lasso Regression in R & Python

The following tutorials explain how to perform lasso regression in R and Python:

**[Lasso Regression in R \(Step-by-Step\)](#)**

**[Lasso Regression in Python \(Step-by-Step\)](#)**