When should Ridge and Lasso Regression be used?

Authored by
stats writer

May 6, 2024

RECOMMENDED CITATION

stats writer (2024). *When should Ridge and Lasso Regression be used?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=143304

Ridge and Lasso Regression are two commonly used statistical techniques in machine learning and data analysis. They are used in situations where there is a need to reduce the complexity of a model and prevent overfitting. These techniques are particularly useful when dealing with high-dimensional data sets or when there is a large number of variables in the model. They are also effective in handling multicollinearity, which occurs when the predictor variables are highly correlated with each other. In such cases, Ridge and Lasso Regression help in selecting the most relevant variables and minimizing the impact of the correlated variables. Overall, Ridge and Lasso Regression are suitable for situations where there is a need to balance model complexity and predictive accuracy.

When to Use Ridge & Lasso Regression

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

Y = β0 + β1X1 + β2X2 + … + βpXp + ε

The values for β0, β1, B2, … , βp are chosen using the least square method, which minimizes the sum of squared residuals (RSS):

RSS = Σ(yi - ?i)2

where:

Σ: A symbol that means "sum"yi: The actual response value for the ith observation?i: The predicted response value for the ith observation

The Problem of Multicollinearity in Regression

One problem that often occurs in practice with multiple linear regression is multicollinearity - when two or more predictor variables are highly correlated to each other, such that they do not provide unique or independent information in the regression model.

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.

Avoiding Multicollinearity: Ridge & Lasso Regression

Two methods we can use to get around this issue of multicollinearity are ridge regression and lasso regression.

Ridge regression seeks to minimize the following:

RSS + λΣβj2

Lasso regression seeks to minimize the following:

RSS + λΣ|βj|

In both equations, the second term is known as a *shrinkage penalty*.

When λ = 0, this penalty term has no effect and both ridge regression and lasso regression produce the same coefficient estimates as least squares.

With Lasso regression, it's possible that some of the coefficients could go *completely to zero* when λ gets sufficiently large.

Pros & Cons of Ridge & Lasso Regression

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

Recall that mean squared error (MSE) is a metric we can use to measure the accuracy of a given model and it is calculated as:

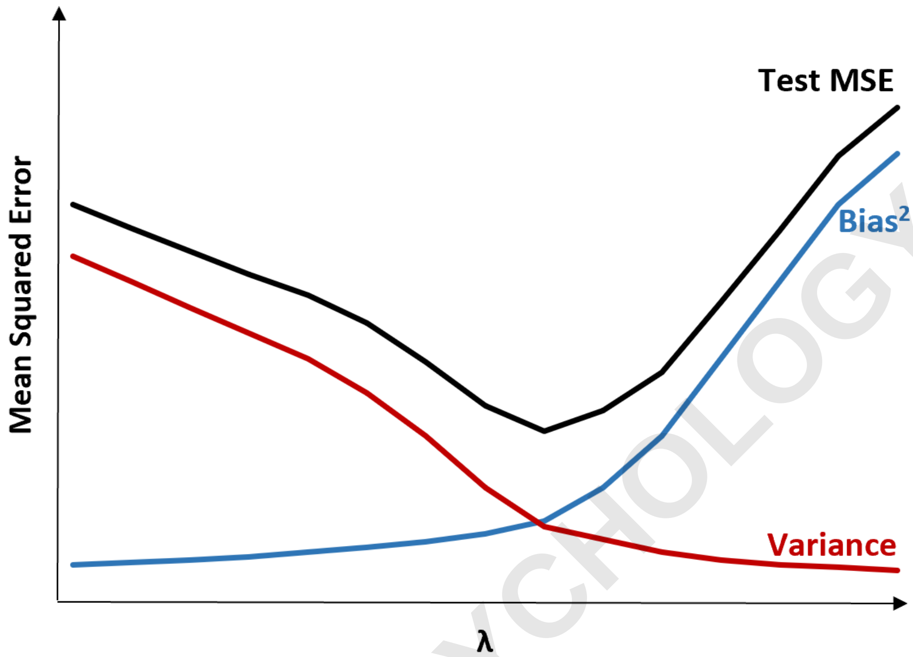
MSE = Var(*f?(*x0)) + 2 + Var(ε)

MSE = Variance + Bias2 + Irreducible error

The basic idea of both ridge and 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:

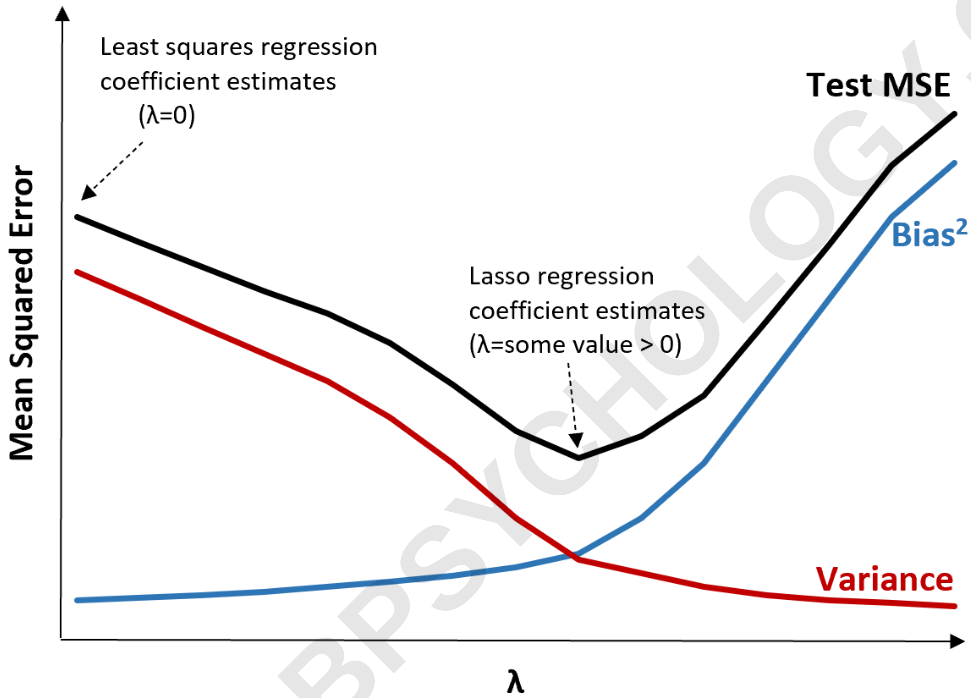


Notice that as λ increases, variance drops substantially with very little increase in bias. Beyond a certain point, though, variance decreases less rapidly and the shrinkage in the coefficients causes them to be significantly underestimated which results in a large increase in bias.

We can see from the chart that the test MSE is lowest when we choose a value for λ that produces an optimal

tradeoff between bias and variance.

When λ = 0, the penalty term in lasso regression has no effect and thus it produces the same coefficient estimates as least squares. However, by increasing λ to a certain point we can reduce the overall test MSE.



This means the model fit by ridge and lasso regression can potentially produce smaller test errors than the model fit by least squares regression.

The drawback of ridge and lasso regression is that it becomes difficult to interpret the coefficients in the final model since they get shrunk towards zero.

Thus, ridge and lasso regression should be used when you're interested in optimizing for predictive ability rather than inference.

Ridge vs. Lasso Regression: When to Use Each

Both lasso regression and ridge regression are 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.

This naturally brings up the question: Is ridge or lasso regression better?

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 typically perform k-fold cross-validation and choose whichever model produces the lowest test mean squared error.

The following tutorials provide an introduction to both Ridge and Lasso Regression:

The following tutorials explain how to perform both types of regression in R and Python: