

To compare the two methods first we need to understand what either of these methods are and then choose one which works best.

Because there is a '1' in it we will start by describing the L1 or the Lasso Regression model first.

One would think lasso has something to do with a real lasso but it's just an acronym.

it stands for least absolute shrinkage and selection operator, and lasso regression adds the absolute value of magnitude of the coefficient as a penalty term to the loss function.

$$\sum_{i=1}^n (Y_i - \sum_{j=1}^p X_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Again, if lambda is zero, then the answer will obviously be the OLS (ordinary least squares) whereas if it's a very large value, it will make the coefficients zero, which means our model will become underfit.

Now let's get to the second method the L2 or Ridge Regression.

In the Ridge Regression the squared magnitude of the coefficient is added as the penalty term. The highlighted part below represents the L2 regularization element.

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

Here, just like the L1 model if lambda is zero the value will be equal to the OLS. And, if lambda is very large then it will add too much weight and lead to underfitting. Having said that, how we choose lambda is important. Now as giving lambda a large value, our model will be underfit but if we give it a reasonable value it will help us not overfit the model.

Now just by looking at the two models one can see the only apparent difference between the two is the square on top of beta in the L2 model, but how will this '2' change the two?

When comparing two models there a lot of key factors to take into consideration, robusticity, and the computational price so let's compare the two.

In case of robusticity it is obvious that L1 is more robust since the square will make small and non-important differences seem very large. But in case of computational price even though it would seem L1 is the winner here, because L1 has no closed form solution it is computationally more expensive.

So if I were to choose between the two models I would go with different models based on different situations, if I needed less complicated models I would go with L2 but if robusticity was more important, well then L1 would be my regression model of choice.