

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Best alpha value for Lasso : {'alpha': 0.001}

Best alpha value for Ridge : {'alpha': 0.9}

Important features :

#MiscVal : \$Value of miscellaneous feature

#BsmtHalfBath : Basement half bathrooms

#LowQualFinSF : Low quality finished square feet (all floors)

#BsmtFullBath : Basement full bathrooms

#HalfBath : Half baths above grade

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

I have chose Lasso , as it has the added advantage of feature elimination. . it is always preferred to have a simple model with less features as it genrealises the idea

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Mainly we should make sure that the difference in r^2 score of test set and train set is not too large, basically we want to reduce the value of coefficients of the final model and also eliminate them if possible. I