

Question 1

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?

Answer:

The alpha values are as follows

Ridge: 0.01

Lasso: 0.0001.

In this model case when the value of Alpha is doubled there is no major impact on the model but still, there are a few changes.

Ridge

R Square Value

0.01 :

Test: 0.9576468645814743

Train: 0.867086412086461

0.02

Test: 0.9564755320651155

Train: 0.8699426725659947

Mean Square Error

0.01 :

Test: 0.006657455900643789

Train: 0.02186932496523995

0.02 :

Test: 0.006841576733625558

Train: 0.02139936181404636

There is no change in the most important predictor variable. In ridge it is RoofMatl_Membran.

Lasso

R Square Value

0.01

Test: 0.9576468645814743

Train: 0.867086412086461

0.02

Test: 0.9576468645814743

Train: 0.867086412086461

Mean Square Error

0.01

Test: 0.008541250340637265

Train: 0.020111980756661107

0.02

Test: 0.010744774232067118

Train: 0.019979209290849117

The most important predictor variable is GrLivArea

Question 2

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?

Answer

I would prefer Lasso model over the Ridge as the Lasso is performing better on the test data also since it makes the predictor variables which are of less significance to zero the Adjusted R square will also be better.

Question 3

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?

Answer

The Next 5 important variables once the top 5 predictors are removed are below.

1. RoofMatl_WdShake
2. RoofMatl_Membran
3. RoofMatl_Metal
4. OverallQual
5. MSZoning_RL

Question 4

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

Answer

The model can be made robust and generalizable by training it with one set and testing it against various other sets. The performance R square should be almost similar. If the values are varying largely, it might be a case of Overfitting. These can be corrected by Regularization. The accuracy of the test will reduce but will perform better on the test data.

The model will fit better if we keep it simple as possible but should not be too simple which will cause Underfitting.