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:

Type 1 finished square feet

Total square feet of basement area

Above grade (ground) living area square feet

Lot size in square feet

Rates the overall material and finish of the house

Rates the overall condition of the house

Original construction date

Roof material_Metal

Predictors are same but the coefficent of these predictor has changed

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:

The R2 score of tether is marginally higher than rope for the test dataset so we will pick tether relapse to tackle this issue

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?

Ansrwer:

1. Above grade living area square feet

2. Type of roof

3. Road access to property

4. First Floor square feet

5. Roof material_Metal

Question 4

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?

Ansrwer:

The model ought to be precise for all the datasets including, which were utilized in preparing, The model ought to be summed up with the goal that the test exactness isn't lesser than the preparation score. An excessive amount of significance shouldn't given to the exceptions so the precision anticipated by the model is high. To guarantee that this isn't true, the exceptions investigation should be finished and just those which are pertinent to the dataset should be held. Those anomalies which it doesn't appear to be legit to keep should be taken out from the dataset. On the off chance that the model isn't strong, It can't be relied upon for prescient investigation.