

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:

❑ Ridge - 20

❑ Lasso - 0.0005

If we double the value for ridge and lasso below will be the important predictor variables

Ridge:

OverallQual, GrLivArea, GarageCars, Foundation_PConc, OverallCond

Lasso:

GrLivArea, OverallQual, GarageCars, Foundation_PConc,
OverallCond, MSZoning_RL

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 choose lasso regression, as lasso regularizes the coefficients by making them 0. This helps in reducing the number of features that can predict the model. It uses tuning parameter lambda as penalty for cross validation.

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:

Before excluding the important predictor variables the model had variables as per the solution in question 1. However, post excluding them the model predicts the below variables as important predictor variables.

BsmtFullBath
FullBath
HalfBath
hage2
MSSubClass

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?

Answer:

The model should be as simple as possible. It should have low variance and low bias. It should be able to generalize and robust. Simpler model will have high bias and low variance but it will be more generalizable. So it will be able to perform well on unseen data

and both train and test will be able to provide models that have the metric variables near to each other without much variation.