

Question-1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in 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-1

For Ridge regression – We have chosen the value of alpha as 4, because when the value of alpha is 4 test error is minimum. So we decided to go with value of alpha as 4 for our ridge regression model.

For Lasso regression – I have taken the value of alpha as 50, because if we see into the plot when value of alpha is 50, it penalizes the model more and coefficients reduce to zero. And value of r^2 square also reduces.

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-2

Based on Lasso regression plot, we will choose 50 as the value of lambda, as it works best with training and test data.

Based on Ridge regression plot, we will choose value of lambda as 4, as it works well with both training and test data.

Question-3

After building the model, you realized 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-3

After looking into the lasso model parameters, we can ensure following five are the most important predictor variables.

1. MSSubClass
2. MSZoning_RL
3. MSZoning_FV
4. MSZoning_RH
5. MSZoning_RM

After rebuilding the model we can confirm, following 5 are now most important predictor variables –

1. OverallCond
2. SaleCondition

3. Neighborhood
4. Functional
5. KitchenQual

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-4

If a model does not change drastically after changing the training data, then such model is considered ROBUST Model. And if the model works well with new data, then it is considered as generalizable.

So we can conclude that a robust and generalizable model is that model whose accuracy doesn't change much with new data.