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 Regression Model

The optimal value of alpha for Ridge is 4,

R-Squared for y_train=94.26% and y_test= 92.80%

The most inportant predictor here is Neighborhood_StoneBr,

When alpha is doubles to 8 the diffrence of

R-Squared for y_train=93.85% and y_test= 92.70%

The most inportant predictor here is Neighborhood_StoneBr,

The coefficient and the Intercept value has been changed

Lasso Regression Model

The optimal value of alpha for Lasso is 50

R-Squared for y_train=94.09% and y_test= 92.79%

The most inportant predictor here is SaleCondition_Alloca,

When alpha is doubles to 100 the

R-Squared for y_train=93.48% and y_test= 92.79%

The most inportant predictor here is Neighborhood_StoneBr

The coefficient and the Intercept value has been 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

In this assignment we would select the Lasso model because here the no. of features are large and lasso is good for feature selection.

Lasso Shrinks some of variable coefficient to 0 and hence perform the feature selection.

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 Five most predictor variables are:

1. SaleCondition_Alloca
2. Neighborhood_StoneBr
3. Neighborhood_Crawfor
4. KitchenQual_Gd
5. Functional_Typ

After Removing the above five variables the top 5 variables now are:

1. Street_Pave
2. Functional_Maj2
3. ExterQual_Fa
4. MasVnrType_Stone
5. Neighborhood_Crawfor

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

A model is considered to be robust if the model is stable, i.e. does not change drastically upon changing the training set. Model is considered generalisable if it does not overfits the training data, and works well with new data. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data.