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:

For Ridge regression optimal value of alpha is 0.3 and for Lasso regression it is 7.0 .
If we double the value for alpha, the coefficients will start moving towards zero.
Important predictors are as follows
[RoofStyle_Shed ,
Exterior1st_BrkFace, Exterior1st_BrkComm, RoofMatl_Metal, RoofMatl_Membran]

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 will go with Ridge regression and optimal value of alpha 0.3 as it has better performance on both train and test data set

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:

'OverallCond_9'(and other dummy classes 8,7,6,5) , OverallQual_10(and other dummy classes 9,7,5) , Exterior1st_CBlock, HouseStyle_2.5Unf, HouseStyle_2.5fin.

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 robust if accuracy remains same even with changes is training data or assumptions made. Adjusted R square in case of regression and ROC curve in case of classification will help us determine how robust and generalized is our model.