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:

- Optimal value of alpha ridge is 10 and optimal value of alpha for lasso is 100.
- after make the double alpha for ridge and lasso is 20 and 20

  FOR RIDGE :- Coefficients values are increasing as alpha will inrease.r2 _score of trained data is also drop from .804 to 0.45
- FOR LASSO:- As alpha increased more features removed from model. but r2 score get dropped by 1% in both test and train data.
- Top features - Neighborhood_NoRidge,Neighborhood_NridgeHt, OverallQual,OverallQualNeighborhood_Veenkar

Question-2:

You have determined the optimal value of Lambda for ridge and Lasso during the assignment . Now which one you will choose to apply and why ?

Answer :

We should choose Lasso as it gives an option of feature selection . It removed unwanted features from the model without affecting the accuracy which makes our model generalized, simple and accurate.

Question-3

After Building the model, you realized that the five most important predictor variables in the lasso model are not 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:

5 TOP features are :

Neighborhood_NoRidge,

Neighborhood_NridgeHt,

2ndFlrSF

OverallQual,

OverallQualNeighborhood_Veenker

After dropping them

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

Important features which are important to make mode robust and generalised , we have 3 features that are required:

1. Model accuracy should be > 70-75%: I our case its coming 80%(Train) and 81%(Test) which is correct.
2. P-value of all the features is < 0.05
3. VIF of all the features are < 5

Thus we are sure that model is robust and generalisable

Question-2:
Repeat the above procedure.