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:

Best alpha value for Lasso : 0.001
Best alpha value for Ridge : 0.1

when doubled, the alpha of both ridge & lasso : 0.2 and .002 respectively.

On compairing both the model we can see that the below Features are best explaining the DataSet

- MSSubClass
- RoofMatl_Membran
- MSZoning_RL
- MSZoning_FV
- MSZoning_RH
- MSZoning_RM
- Condition2_PosA
- RoofMatl_WdShngl

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 :

Lasso. As it gives feature selection option. It removes unwanted features from model without affecting the model accuracy which makes the model generalized, simple and accurate.

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 :

Top 5 are : MSZoning_RL, MSZoning_RH, MSZoning_FV, MSSubClass, RoofMatl_Membran

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 said to be robust and generalisable if the predictor variable is consistently accurate even when one or more of the independent variables are changed drastically.

To make model robust and generalisable 3 features 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.