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 for ridge Regression – 10

Optimal value for Lasso = 0.0001

After double the value of alpha/lambda,

MSE value reduced from 0.01374 to 0.01366 in Ridge

MSE value reduced from 0.013556 to 0.013331 in Lasso

Important predictor variables are,

GrLivArea

OverallQual_8

OcerallQual_9

Functional_Typ

Neighborhood_Crawfor

Exterior1st_BrkFace

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?

1. Only based on domain knowledge or required of use case we can determine the choose of model.
2. If reduction of co-efficient then will go with Ridge model.
3. If primary goal is feature selections, then we will go with Lasso.

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: 2ndFlrSF, Functional_Typ, 1stFlrSF, MSSubClass_70, Neighborhood_Somerst

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?

1. Model should perform same with what ever data we provide and should have less variance.
2. Sould be generalized for any unseen data and should draw similar results.
3. Model shouldn't overfit.
4. We should decrease variance which will have some bias.
5. Regularization techniques such as Ridge and Lasso will help us to balance between model accuracy and complexity.