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 alpha for Ridge regression model = 10.0
Optimal value for alpha for Lasso regression mode = 0.001

Below are the changes in the model when the alpha values are doubled for Ridge and Lasso Model.

RSS value has increased slightly for both the models.
The train and test score decrease slightly for both the models.
The coefficients of predictor variables have shrunk to a lower value.

Below are the 5 predictor variables for Lasso model after change is implemented:

1. GrLivArea
2. Neighborhood_Crawfor
3. SaleCondition_Normal
4. Exterior1st_BrkFace
5. SaleCondition_Partial

Below are the 5 predictor variables for Ridge model after change is implemented:

1. GrLivArea
2. SaleCondition_Normal
3. Neighborhood_Crawfor
4. Exterior1st_BrkFace
5. OverallQual

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:
Optimal value for alpha for Ridge regression model = 10.0
Optimal value for alpha for Lasso regression mode = 0.001

After the analysis I have found that both the models performed well on the test data. But to arrive at a conclusion I have selected the Lasso regression model as the best among the two analyses conducted because Lasso Regression have feature selection as compared to Ridge regression model.

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 features drived after removing the actual 5 predictor variables

1. 2ndFlrSF
2. 1stFlrSF
3. MSZoning_FV
4. CentralAir
5. SaleType_New

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:

To make the model more robust and generalisable:

- Make sure that the model is not overfitting.
- Variables should be normally distributed. If it's not normally distributed, uses transformation techniques to make it normally distributed.
- Treat the outlier present in the data properly.

If the model is not generalisable, we might get best accuracy on the train data and the model fails when running on test data.