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 to 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 for,

1. Ridge Regression: 1
2. Lasso Regression: 0.0001

Change in the model if we double the value of alpha,

1. Ridge Regression: The Mean Squared error remain same
2. Lasso Regression: The Mean Squared error has decreased slightly

The most important predictor variable after the change are,

1. MSZoning_RL
2. GrLivArea
3. MSZoning_RM
4. MSZoning_FV
5. OverallQual
6. TotalBsmtSF
7. Foundation_PConc
8. GarageCars True
9. OverallCond

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:

After our analysis we have found that the optimal value of lambda are as below,

1. Ridge – 1
2. Lasso - 0.0001

And the Mean Squared error are,

1. Ridge - 0.015813
2. Lasso - 0.015777

We can see from above that the Mean Squared error is slightly lower in case of Lasso compared to Ridge.

Also, Lasso helps in feature reduction when we have too many variables as in this case, **Lasso** is **better suited** compared to Ridge.

Question 3:

After building the model, you realized 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:

After dropping the five most important predictors, we get the following top 5 predictors,

1. GarageCars
2. TotalBsmtSF
3. d_FireplaceQu
4. d_KitchenQual
5. LotArea

Question 4:

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answers:

A model is Robust when variation in data does not affect the performance much.

A Generalizable model is one which will adapt to previously unseen data.

To make sure model is Robust and Generalizable we must make sure it does not overfit. We can use Occam Razor principle for selecting the model.

1. Given two model we can choose simpler model, as simpler model are usually more generic
2. Simpler models require fewer training samples than more complex ones
3. Simpler models are more robust

To make the model more robust and generalizable, make model simple but not simpler which will not be of any use.

In general, we must strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.