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

The Optimal Value of Alpha

Ridge =0.5

Lasso = 0.0001

Double the Alpha Value

Ridge = 0.10

Lasso = 0.0002

After doubling the value

For Ridge, we got an additional Parameter.

1. ExterCond_TA
2. MSZoning_Others

For Lasso after doubling the value, I did not see any change in the parameter. I got the same set of parameters as before doubling the alpha value

GarageArea

LotArea

OverallQual

OverallCond

YearBuilt

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?

Ans

I will choose Lasso. As Lasso has zeros some coefficients while Ridge did not . Lasso has helped with feature elimination.

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?

The top 5 predicator variables I have eliminated are –

"GarageArea", "LotArea", "OverallCond", "OverallQual", "YearBuilt

With the new Lasso model created after eliminating the above 5 parameter, the new list of important predictors are

1. YearRemodAdd.
2. PoolArea.
3. GrLivArea.
4. ExterCond_TA.
5. Neighborhood_NridgHt.

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?

To ensure that a model is robust and generalizable, you can take the following steps:

1. Test -Train split - This helps assess how well the model generalizes to unseen data.
2. Cross-Validation: Perform cross-validation, such as k-fold cross-validation, to assess the model's performance across multiple train-test splits. This helps evaluate the model's consistency and generalizability across different subsets of the data.
3. Feature Engineering:
4. Regularization: Apply regularization techniques like Ridge regression or Lasso regression to prevent overfitting and improve generalization.
5. Hyperparameter Tuning: Optimize the model's hyperparameters using techniques like GridSearch

The implications of ensuring model accuracy are that the model is more likely to perform well on unseen data and new instances