

## 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?

In the model the Optimum Alpha Value for Ridge Regression is 100 and the associated Top 5 Features as follows

- GrLivArea
- OverallQual
- BsmtQual
- FireplaceQu
- GarageArea

When the alpha is value is changed following things can be observed the list of features have changed to following order

- OverallQual
- GrLivArea
- FullBath
- KitchenQual
- FireplaceQu

RSS and MSE have increased (increase in error) and there is a decrease in  $r^2$  score on both training and test set.

In the model the Optimum Alpha Value for Lasso Regression is 20 and the associated Top 5 Features as follows

- GrLivArea
- OverallQual
- LandContour\_Lv1
- OverallCond
- BsmtQual

With Lasso there is no change in importance features but minimal increase in RSS and MSE with slight variation in  $r^2$  score in both test and train

## 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?

The Optimum alpha value for Ridge is 100 while for the lasso its only 20. As a general rule its been taught than as the alpha value increases the model becomes generalised while loosing important features. It uses far less variables to get a good result in the meantime as well this suggests that best regression to apply in this scenario would be lasso regression with just alpha of 20.

### 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?

After removing the feature and re running Lasso Regression it can be seen that top features are now

- FullBath
- FireplaceQu
- KitchenQual
- GarageArea
- OpenPorchSF

The r2 score has dropped from 0.855 to 0.775 on test set and on training set from 0.880 to 0.795.

### 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?

Model needs to be built to robust and generalizable so it will handle

- Make the design more resistant to outliers
- Ensure the model is simple as possible; while considering the trade-off of bias vs variance throughout the design to get the design to a medium mark.
- If the model is generalizable, the accuracy will not fall in respect to training data on new unseen data sets.