

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?

- The optimal value of alpha for Ridge & Lasso regression is 1000, which is the highest value for the alpha in both the regression models.
- If we choose to double the value of alpha then the model will start to get underfit & we might miss some important features from the given features.
- After doubling the value for alpha, the optimal value for alpha is still 1000 we can see that r2 score & RSS values decreased for
- Also, the coefficient values decreased.
- Ridge Regression.

Before:

- r2 : 0.8375304519305928
- RSS : 0.8395020105989172

After:

- r2 : 0.8089458250862206
- RSS : 0.8276198913190594
- Lasso Regression : There are no changes in the Lasso Regression model after doubling the value for alpha.
- These are important variables after Lasso Regression : **Street, LotArea, Utilities, BldgType, SaleCondition, YrSold, GarageCars, KitchenQual**

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 optimal value of lambda for me is 1000.
- We need to do some feature selection & then again check for the lambda value.
- Because the current value of alpha is the highest which won't help us to remove the overfitting issue in the model
- After doubling the value of alpha we noticed that the alpha value is the same but our coefficients decreased & the value for r2 & RSS increased.

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?

- **MSZoning, LotArea ,BsmtQual,BsmtFinType2, ExterCond**

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?

- Depending upon our r^2 score & RSS value we can determine the model is robust to what level & is generalisable or not.
- After doing some evaluation & doubling the value for alpha we saw that there is some increase in our r^2 score & RSS value.
- The r^2 score & RSS value after doing some important feature selection are
 - r^2 : 0.828700989465371
 - RSS : 0.8481536475172736
- Also we can see that there is not much significant difference between the r^2 & rss value.