

## 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?

**Ans:**

- Best alpha value for Lasso : 0.001
- Best alpha value for Ridge : 5.0

There is slight changes in the results when the alpha values are doubled

	Metric	Ridge regression	Lasso regression
0	R2 Score Train	0.869387	0.781764
1	R2 Score Test	0.847304	0.760613

the most important predictor variables after the change is implemented

- OverallCond
- BsmtFullBath
- GrLivArea
- LotConfig\_FR3
- MSSubClass
- 2ndFlrSF
- HalfBath
- Fireplaces
- LotConfig\_Insi
- TotRmsAbvGrd

## 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:**

Based on the alpha values, Ridge regression provides the r2 score more compare to lasso so I am selecting the ridge regression.

### 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?

**Ans:**

Overall condition

Lot area

TotRmsAbvGrd

HouseStyle\_1Story

IsRemodeled.

### 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?

**Ans:**

The model is robust and generalizable when

- No too much difference in the  $r^2$  score of train and test data
- The model should not be impacted by the outliers: Outlier treatment is most important to get the robust model.
- We can detect outliers in the dataset using box plots. This would help standardize the predictions made by the model.
- If the model is not robust, it cannot be trusted for predictive analysis
- Model significance can be determined from the P-values,  $R^2$  and adjusted  $R^2$ .
- Always a simple model can be more robust.

Implications of Accuracy of a model:

- More data leads to a well trained model thereby increasing the accuracy to any type of data
- Fix missing values and outliers in the given data is more important
- After extracting the new data required we
- Can drop the existing features.
- Scaling the values

- Domain knowledge
- Data visualization
- Applying the right algorithm
- Cross validation to get more accurate model