

### 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?

Optimal value of Alpha

- Ridge : 1
- Lasso : 0.0001

After doubling the alpha values for both ridge and lasso, following changes were observed  
Important predictor values after doubling the alpha values for both ridge and lasso

	Beta	Ridge	Lasso	new_ridge	new_lasso
0	TotalBsmtSF	0.79	0.88	0.73	0.87
1	1stFlrSF	0.44	0.41	0.45	0.32
2	2ndFlrSF	0.41	0.46	0.36	0.37
3	GrLivArea	0.51	0.53	0.50	0.67
4	BedroomAbvGr	-0.44	-0.55	-0.36	-0.52

### 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?

Based on metric comparison Ridge and Lasso using the optimal value of lambda. Lasso appears to be performing better than Ridge. R2 score of the Lasso model is also greater than R2 score of Ridge model

	Metrics	Linear Regression	LR with RFE	Ridge	Lasso
0	R2_train	0.96	0.88	0.84	0.85
1	R2_test	-209,305,211,762,087,788,544.00	0.67	0.84	0.69
2	MSE_train	0.01	0.02	0.02	0.02
3	MSE_test	32,767,319,974,553,444,352.00	0.05	0.04	0.05
4	RSS_Train	5.69	5.69	21.32	20.42
5	RSS_Test	14,155,482,229,007,087,304,704.00	14,155,482,229,007,087,304,704.00	17.70	20.01

### 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?

Top 5 Lasso Predictors are:

1. TotalBsmtSF
2. 1stFlrSF
3. 2ndFlrSF
4. GrLivArea
5. BedroomAbvGr

After dropping the above 5 predictors, a new model was created and below are the 5 new predictors

1. GrLivArea
2. 1stFlrSF
3. KitchenAbvGr
4. Exterior1st\_BrkComm
5. GarageCars

#### **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?

A model is considered as robust and generalizable when its performance is not impacted by any variation, outliers etc. It should be adaptable to new data.

Below actions are done to make it robust and generalized.

1. Outlier treatment
2. Regularization to avoid overfitting

Below implications are of the same for the accuracy of the model

1. A thorough outlier analysis shall be done and only relevant columns to the dataset must be retained. Those outliers which are irrelevant must be removed from the dataset. This will help to increase the accuracy of the model.
2. An overfit model will have high accuracy; however, it will also have high variance. This means smallest change in the data will affect accuracy of model. Hence balance between the accuracy and complexity is necessary. Regularization helps in achieving the same