

### Question 1

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 values of alpha I got:

Ridge = 0.9

Lasso = 0.0001

	Metric	Ridge regression	Lasso regression
0	R2 Score Train	0.928618	0.932763
1	R2Score Test	0.898725	0.906227
2	RSS Train	1.476826	1.391060
3	RSS Test	0.714483	0.661552
4	MSE Train	0.001391	0.001310
5	MSE Test	0.002013	0.001864

When double the value of alpha:

Ridge = 1.8

Lasso = 0.0002

	Metric	Ridge regression	Lasso regression
0	R2 Score Train	0.927084	0.929167
1	R2Score Test	0.896682	0.905904
2	RSS Train	1.508549	1.465450
3	RSS Test	0.728893	0.663831
4	MSE Train	0.001420	0.001380
5	MSE Test	0.002053	0.001870

Double the value of alpha got slightly change in value of  $R^2$  and RSS.

Important predictors after double the value of alpha are:

- 1] BsmtFullBath
- 2] OverallCond
- 3] MasVnrArea
- 4] HeatingQC
- 5] OverallQual

### 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?

>> Lasso. As it gives robust model as well as difference between  $R^2$  of test and train is less and value of RSS, MSE and RMSE are wise.

### 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?

>> I have drop five most important predictor variables in the lasso model and created another model and got below five most important variables:

- 1] BsmtFinSF1
- 2] OverallQual
- 3] OverallCond
- 4] BsmtExposure
- 5] MSSubClass

#### **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?

- >> 1] The model is robust and genaralisable when the difference between the train and test score is minimum
- 2] We can detect outliers in the dataset using boxplot. treating the outliers will not effect the mean, median,etc. so that we can input correct values to the missing values. Outliers analysis should be done to avoid bias.
- 3] The predicted features should be significant and this significance can be determined by p-values, R-square and adjusted R2 Implications of accuracy of a model.
- 4] The model should not impact by the outliers therefore outlier treatment is important to get the robust model.



