**Subjective Questions**

**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?

**Answer 1**

The optimal value of alpha was coming out to be :-

* 0.001 for Lasso Regression
* 5.0 for Ridge Regression

**After doubling the value of alpha for Lasso Regression (Keeping it 0.002) the model’s accuracy come out to be :-**

Train Accuracy is: 0.8891169255232901

Test Accuracy is: 0.8902903805396788

Which is good enough and not significantly different from the optimal Lasso Regression Model.

The 10 most important variable are coming out to be :-

['LotArea', 'OverallCond', 'BsmtFinType1', 'BsmtExposure', 'BsmtFinSF1', 'MasVnrArea', 'BsmtUnfSF', 'BsmtHalfBath', 'OverallQual', 'YearRemodAdd\_Old']

Which are also not much different from the optimal model.

**After doubling the value of alpha for Ridge Regression (Keeping it 10.0) the model’s accuracy come out to be :-**

Train Accuracy is: 0.9133232226647875

Test Accuracy is: 0.900059355563159

Which is good enough and not significantly different from the optimal Ridge Regression Model.

The 10 most important variable are coming out to be :-

['LotArea', 'OverallCond', 'BsmtFullBath', '2ndFlrSF', 'MasVnrArea', 'GarageArea', 'LowQualFinSF', 'Fireplaces', 'GarageFinish', 'HalfBath']

Which are also not much different from the optimal model.

**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?

**Answer 2**

The optimal value of alpha for Ridge is coming out to be 5.0 and for Lasso Regression it’s coming out to be 0.001. For Lasso Regression there is too less compensation on the addition of variables. The R^2 value for Ridge Regression is also coming out to be higher and hence I will go for Ridge Regression over the Lasso Regression.

**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?

**Answer 3**

If the first 5 most important variables are not available in the incoming data and we have to create the model all over again, the 5 most important variables are coming out to be: -

'BsmtQual'

'MasVnrArea'

'ExterQual'

'BsmtFinSF2'

'BsmtFinSF1'

**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?

**Answer 4**

I can say that the model is robust and generalisable as the results do not vary significantly as we make small changes and the difference between the R^2 value for both test an train is good and there is no significant difference between both of them. To really sure of that, we also used cross validation.