

### 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 Lasso and ridge are

Lasso – 0.001

Ridge - 9

After doubling the values there is decrease in the score

Top 5 predictor variables after change are – 1stFlrSF, 2ndFlrSF, OverallQual, OverallCond, SaleCondition\_Partial

### 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?

After changing the values and rebuilding the model with ridge and lasso regression,

I choose Lasso regression since the lasso score is slightly higher than the ridge.

### 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?

The model is built again after dropping the most important 5 predictor variables and the final most important 5 variables are

FullBath, KitchenQual, GarageArea, LotArea, HalfBath

#### Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

The model is more robust and generalizable when it is neither overfitting nor underfitting

In case of overfitting, there will be a high variance and low bias. And in case of Linear regression Overfitting scenario, we may get a very high accuracy in training data, but model will not perform well in test data and we can see a deviation in train and test score.

If the model is not robust and generalized and robust, accuracy of the model won't be consistent and we cannot rely on such model.

Hence, we use Regularization, so that it will reduce the variance by some compromise on bias and avoid the overfitting model. Also, with this train and test data scores would be almost near to each other.