

### 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 :

Value of alpha for Ridge : 0.9

Value of alpha for Lasso: 0.001

After Doubling the value of alpha

**For Ridge:** Coeff value are increasing the data dropped from 89-45

**For Lasso:** Alpha value increase as more features are removed from the model the decrease was 1%

**Imp Predictors:** MSSubClass, BsmtFullBath, OverallCond, GarageFinish, GarageQual

### 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:**

As we all know LASSO allows us to do Feature selection and helps us remove unwanted features and also have an influence on the Response variable that are affecting our model accuracy

As we can do the variable selection this makes the model interpretation very easy for us.

**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:**

Neighborhood\_Timber, Neighborhood\_OldTown, Neighborhood\_Edwards, Exterior1st\_CBlock, BsmtFullBath are the 5 predictors that were dropped and the model accuracy decreases from 91% and 89% to 45% and 46%. And the next important predictors are MSSubClass, BsmtFullBath, OverallCond, GarageFinish, GarageQual

**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?

**Answer:**

We can say the model is robust when any changes in data do not have any effect on the performance of the model. Generalized model perform very well on seen, unseen data to create model

If our model is both Robust and generalized it will make sure the model doesn't over fit. Otherwise if over fitting happens we get too see very high variance in the data which directly affect the model if we change the slightest of the data

So if we can manage to keep the variance very low then the bias will increase a bit but the model will become robust

The model should be simple and not too complex to be robust .

Factors like i) Model Accuracy >75% , ii) P-value for all the vars<0.05, iii) VIF for all the vars <5