

### 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:

Optimal alpha values for Ridge and Lasso regression are 1 and 9e-05 respectively.

Doubling the alpha value in Ridge regression i.e.,  $\alpha = 2$ :

The model will be made simple, so we have increased error in the train and test set. R-squared value also increases.

Doubling the alpha value in Lasso regression:

As the coefficients of the variables are not pushed close to zero but are equated to zero, when alpha is increased, the r-squared value decreases.

After the change,

1. LotArea
2. OverallQual\_Very Excellent
3. OverallQual\_Excellent are the most important predictor variables in the 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:

In my model, the alpha value for Ridge is 1 and for Lasso is 9e-05.

I chose Lasso regression for the following reasons:

1. As Lasso has alpha value close to zero, with the error values almost equal to that of Ridge.
2. Lasso explains the SalePrice variable better than Ridge regression with slightly greater R-squared value
3. Though some coefficients are to be shrunked to zero in Lasso, most of them are retained and all the important variables with better coefficient values are present.

### 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:

I'll choose the next 5 important features:

1. WoodDeckSF
2. Neighborhood\_Crawfor
3. OverallQual\_Good
4. Neighborhood\_NridgHt
5. Neighborhood\_NoRidge

### 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:

A General and robust model should be:

1. Simple – Though the accuracy will be low, it will be generalizable.
2. Model to be built on data without outliers, and clean data
3. Model that doesnot overfit on the training data
4. Bias- Variance tradeoff: The model shouldnot have high bias(which means underfit) or high variance(which means overfit)
5. Model that has the hyperparameter tuned to the optimal value.