**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 values of alpha for ridge and lasso regression are 1.0 and 0.0001 respectively. With double the optimal values of alpha for ridge and lasso resulted in decrease in R2 Score for train data and increase in Test R2 score. This implies over fitting of model was further reduced with double of optimal alpha value. Doubling the alpha values for ridge and lasso has reduced the coefficient values further, but there was no significance difference in the important predictor variables. There was a bit of shuffling in the order of predictor variables observed based on coefficient values.

**Lasso - Top most predictors of Lasso**

RoofMatl_ClyTile

Condition2_PosN

PoolQC_Gd

OverallQual_Excellent

OverallQual_Very Excellent

GrLivArea

**Ridge - Five most predictors of Ridge**

Condition2_PosN

RoofMatl_ClyTile

PoolQC_Gd

PoolQC_Ex

1stFlrSF

GrLivArea

**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:**

Optimal values of alpha for ridge and lasso regression are 1.0 and 0.0001 respectively. R2 scores of Ridge and Lasso was almost same without any significant change in values. Lasso regression is a better option to choose for its feature elimination, which will remove any redundant or unnecessary predictors that are used in the Regression model, this is considered more robust and generalized model.

**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:**

Following predictors are removed from input data based on initial Lasso regression

- GrLivArea
- RoofMatl
- Condition2
- PoolQC
- OverallQual

Considering new predictor variables after removing the five most predictors from input data, the next five most predictors are

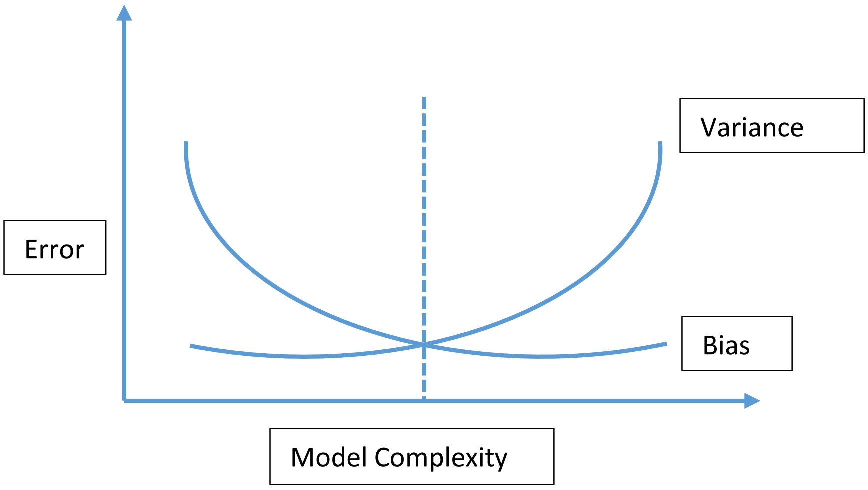
- area in square feet of 1st and 2nd floor has a positive relation with SalesPrice,
- Neighbourhood_NoRidge, Neighbourhood_StoneBr also has positive relation with SalesPrice
- Fireplaces_Three had negative relation with SalesPrice
- MSZoning_C is another feature with negative relation with SalesPrice

## 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:**

The model which is most simple is considered to be robust and general model. However, the accuracy of model might not be as good for a simpler model, since it cannot understand or learn all the underlying conditions or patterns of the training data. This is explained by Bias-Variance trade-off concept.


Variance
Error
Bias
Model Complexity