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. Optimal value of Alpha in Ridge and Lasso is 0.18 and 22.30

2. If we choose double the value of the Alpha then the Model will underfit

3. The most important predictor variable is as per my model is Area of the house and Over all Condition of the house.(OverallCond)

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

I will choose Lasso regression model since it gives better set of Feature selections \

Question 3

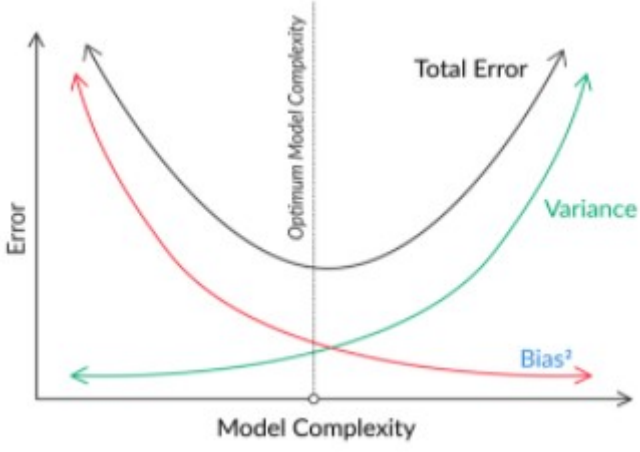
After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Five most imppportant predictor variable are OverallCond, YearBuilt, BsmtCond, KitchenQual_TA and GarageType_Attchd

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?

Generalised model refers to a model whcih works well with any unseen data%
Robust nand generalised model means to have a trade-off between bias and variance with respect to model complexity. You can see in the diagram, a simple model would usually have high bias and low variance, whereas a complex model would have low bias and high variance. In either case, the total error would be high.



To have the Robust and Generalise model we must have Optimum Model Complexity where there is a balance between Variance and Bias.

Accuracy of the model decreases if the model is more generalised but at the same time the model is Robust and gives the optimal result for unseen data