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:

The optimal value for Ridge is 10.0 and for lasso it is 0.001 as suggested by the algorithm.

Alpha changed from 0.01 → 0.02

For Lasso I have choose the value 0.01 and when we double the value of alpha from 0.02 the lasso algo try to penalize more to the coefficients and coefficients further reduces to 19 from 23.

R^2 Square value decreases for both training and test data and Root Mean Squared Error also increases slightly.

Alpha changed from 0.10 → 0.20

For Ridge when I change alpha from 0.10 to 0.20 the efficiency increased for both the training and test data. Also, R^2 Square gets improvement and proved to more efficient in case Ridge alpha increment. The Root Mean Squared Error also reduces and features also changed from 253 to 251 (very slight change).

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 2:

The Optimal value for lasso is 0.001 and for Ridge it is 10.0. Since the lasso helps in feature reduction with alpha 0.01 hence I will choose lasso with alpha 0.01. Also, the mean squared error for lasso is less than ridge hence choosing the lasso.

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 3: The five most important predictor variables that we can assume in other model are as follows:

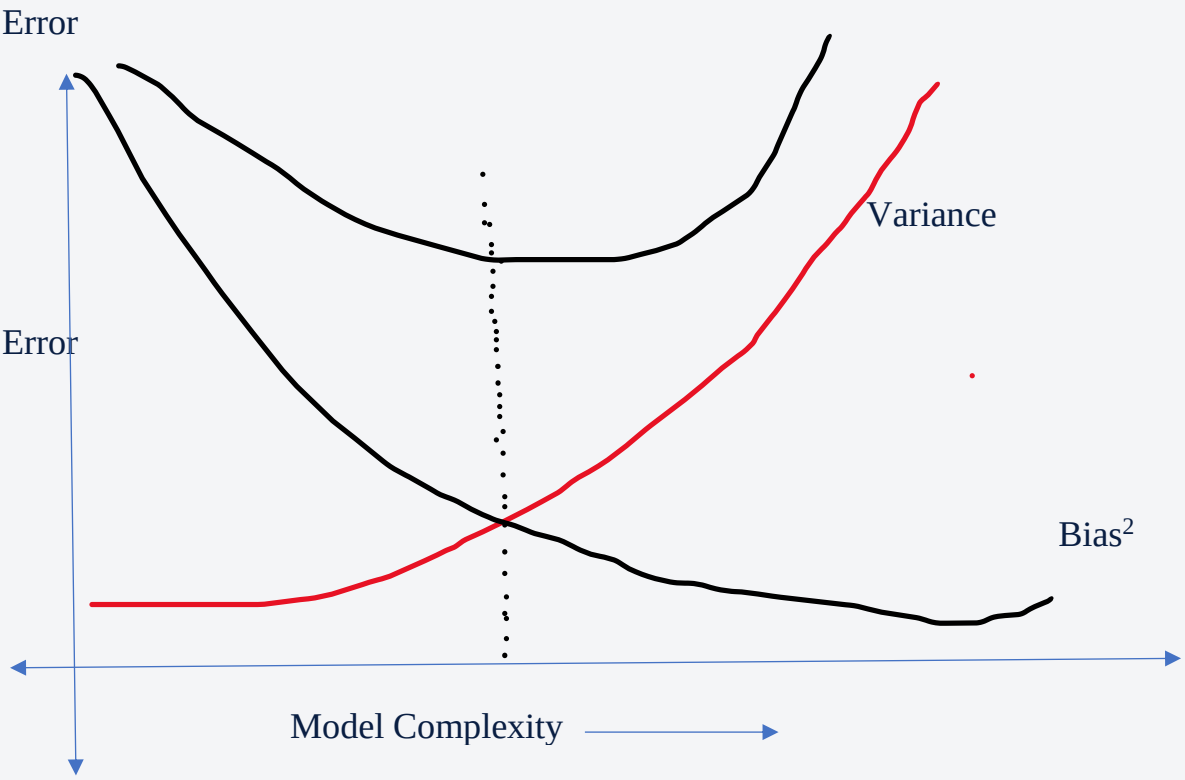
1. OverallQual
2. 1stFlrSF
3. YearBuilt
4. 2ndFlrSF
5. GarageArea

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 4:

For model to be more robust and generalisable we need to make sure that our model should not be too simple and not to be too complex. We can understand more from the bais and variance trade off .



The simpler models will lead to high biasness means model itself on training data and complex model will lead to high variance meaning the model will fail on test data. So there is trade-off between bias and variance with respect to complexity. Now, in either case we will have high error so to choose good model we need to have a low variance and low biasness in model.