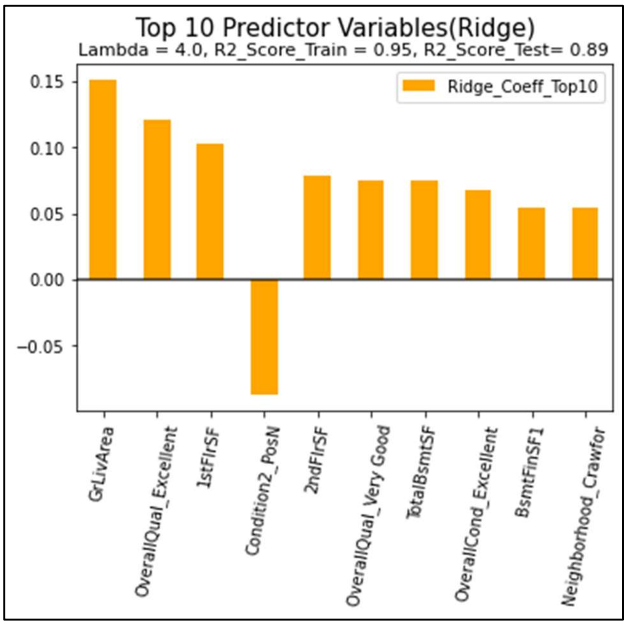
**Question 1**

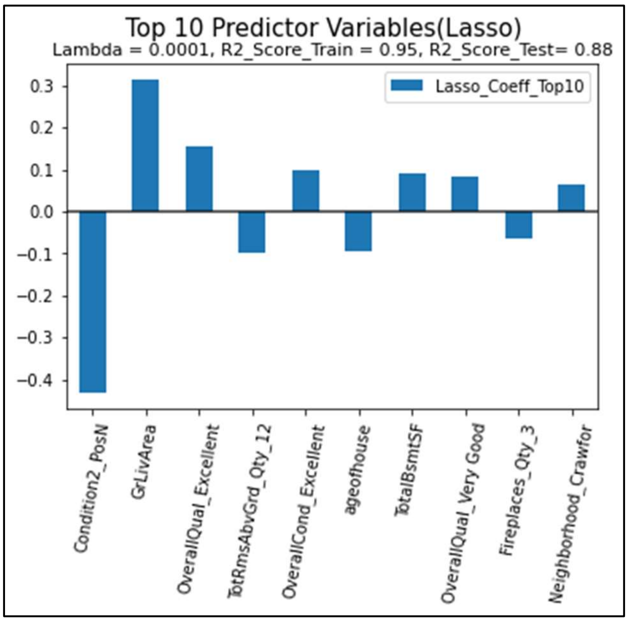
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

➔ The results have been plotted as bellow

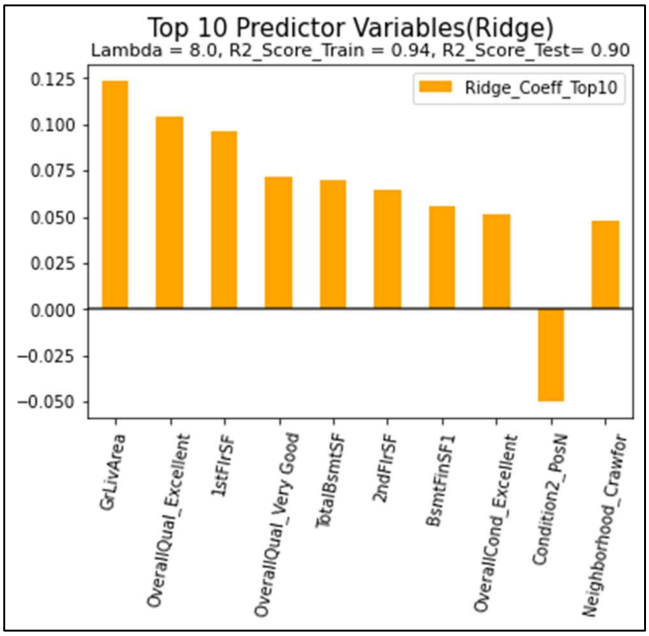
| | Ridge | | | Lasso | | |
|---|---|---|---|---|---|---|
| | Lambda Coeff | R2 Score train | R2 Score test | Lambda Coeff | R2 score train | R2 score test |
| **Optimal Value** | **4** | **94.7%** | **89.4%** | **0.0001** | **95.2%** | **87.7%** |
| Iteration2 -doubling lambda | 8 | 94.1% | 89.7% | 0.0002 | 94.5% | 88.4% |
| Iteration3-again double | 16 | 93.3% | 89.6% | 0.001 | 91.6% | 88.0% |
| Iteration4-again increase | 32 | 92.0% | 89.1% | 0.01 | 75.9% | 75.8% |

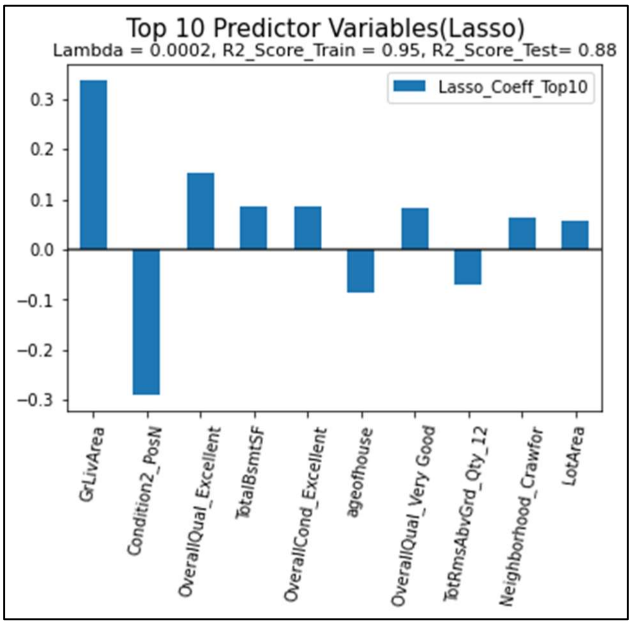
**With optimal coefficients (lambda)**




**After doubling Coefficients…. Following top 10 predictor variables in Ridge & Lasso are important**



**Question 2**

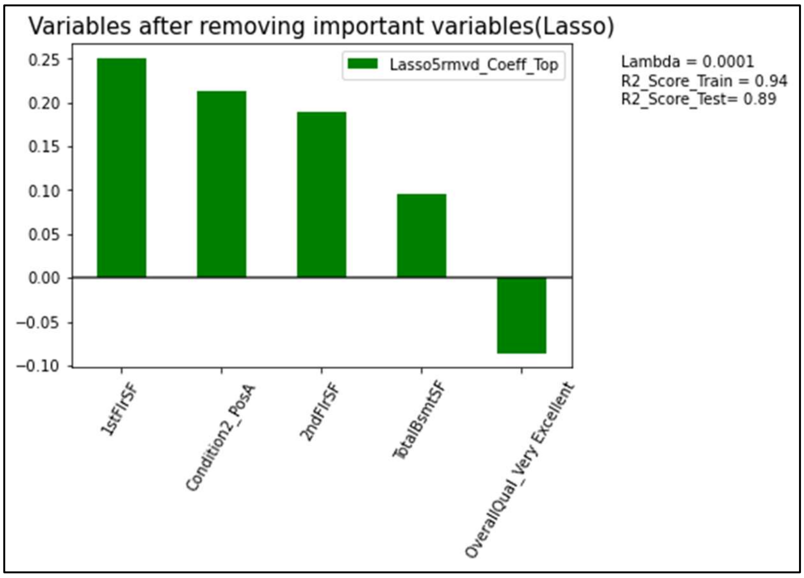
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

- ➔ As stated above optimal value for Ridge coeff is 4. R2 score on train is 94.7% & on test 89.4%
  for Lasso, the coeff is 0.0001. R2 score on train is 95.2% & on test 87.7%
  It can be seen that, for Ridge R2 score on train is comparatively less than Lasso. But the difference between train & test score is lesser.
  **Hence Ridge regression would be first choice, due to less performance variation between train & test data.**

**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?

- ➔ The new variables after removing top 5 important variables are as shown below.
  1. Total basement area in square feet
  2. Condition: Adjacent to positive off-site feature
  3. Second floor square feet
  4. Total square feet of basement area
  5. Overall Quality_Very Excellent'




**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?

→ <u>Robust & generalisable model</u>

The model shall be simple, once developed, which can be used again and again on the different test sets with same performance on training data.

<u>Occam's Razor</u> : The Occam's Razor concept tells what is the criteria of model to be simpler. The model with less number of parameters, or model with small degree of function, depth of size of decision tree.

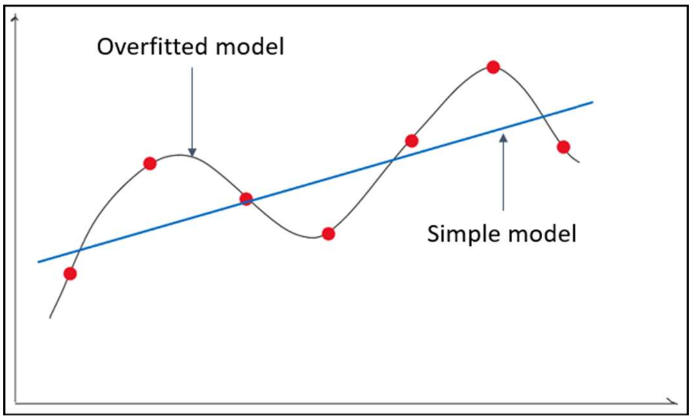
Following ways can be used to make sure model is simple.

Overfitting:

Model is called as overfitted, if it learns each and every aspect from the training data byheart. The overfitted model works maybe 100% on training data, but fails on test data. It means model is not generalisable on the data besides training data. Overfitting can be check by

1. R2 score on train very high, but on test data very less
2. Observing number of parameters, degree of function which shall be on lesser side.

From below fig. it can be seen overfit curve passes through all data points, so error is almost 0. While simple curve doesn't pass through all points, but formation wise is the simple. This is the reason why accuracy of simpler models is less comparatively. Also simple model works effectively on unseen data same as train data.



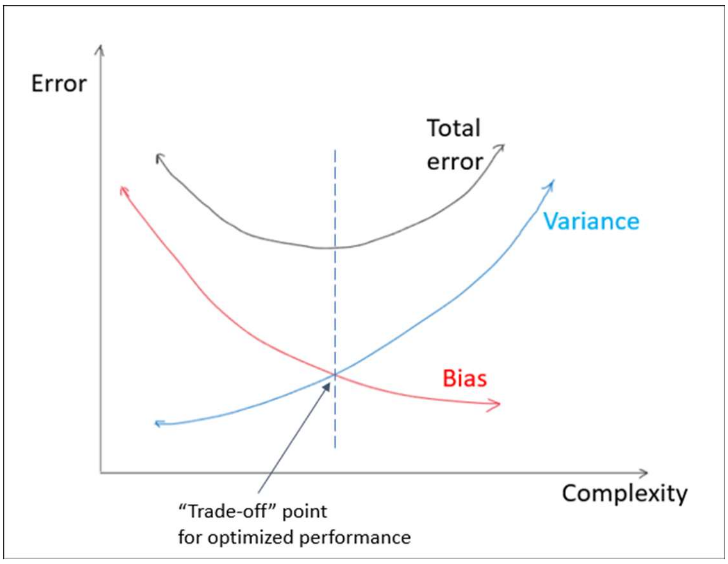

Bias Variance trade-off:

Bias can be considered as accuracy of model

Variance is how model is simple and re-generalizable on new data without error.

As shown below, as model complexity increases Variance increases and Bias reduces. Simplicity may hamper the accuracy of the model. Whereas complex model affects the variance of the model.

Trade-off point shows best possible complexity of model lies.




As a conclusion, accuracy of the model goes on reducing as model becomes simple.