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?

Ans :

Optimal Value of Alpha:

• The computed optimal value of alpha for Ridge Regression (Original Model): 4

• The computed optimal value of alpha for Lasso Regression (Original Model): 0.0005

Ridge :

The R2 Score of the model on the test dataset for optimum alpha is 0.816

The MSE of the model on the test dataset for optimum alpha is 0.0068

Lasso :

The R2 Score of the model on the test dataset for 0.0001 alpha is 0.803

The MSE of the model on the test dataset for optimum alpha is 0.0073

Changes in the model, if you choose double the value of alpha for both ridge and lasso regression:

Ridge:

Alpha : 8

The R2 Score of the model on the test dataset for doubled alpha is 0.809

The MSE of the model on the test dataset for doubled alpha is 0.0071

Lasso :

Alpha – 0.001

The R2 Score of the model on the test dataset for doubled alpha is 0.803

The MSE of the model on the test dataset for doubled alpha is 0.0073

The most important predictor variables are as follows:

TotRmsAbvGrd, GarageArea, CentralAir_Y, BsmtFullBath, KitchenQual_Ex

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?

Ans :

Ridge have good R2 value and MSE on test data compared to Lasso. So we chose Ridge for this data set.

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?

Ans : Top five features in Lasso model after before :

TotRmsAbvGrd, GarageArea, CentralAir_Y, BsmtFullBath, KitchenQual_Ex

Top five features in Lasso model after removing top five features:

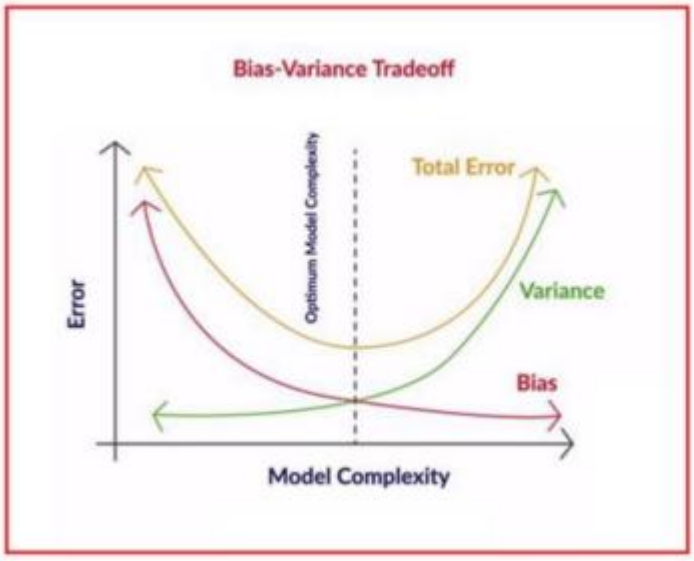
LotArea, BsmtQual_Ex, HouseStyle_2.5Fin, ScreenPorch, LotFrontage

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?

Ans : Robustness of a model implies, either the testing error of the model is consistent with the training error, the model performs well with enough stability even after adding some noise to the dataset. Thus, the robustness (or generalizability) of a model is a measure of its successful application to data sets other than the one used for training and testing. By the implementing regularization techniques, we can control the trade-off between model complexity and bias which is directly connected the robustness of the model. Regularization, helps in penalizing the coefficients for making the model too complex; thereby allowing only the optimal amount of complexity to the model. It helps in controlling the robustness of the model by making the model optimal simpler. Therefore, in order to make the model more robust and generalizable, one need to make sure that there is a delicate balance between keeping the model simple and not making it too naive to be of any use. Also, making a model simple leads to BiasVariance

Trade-off:

➢ A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
➢ A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed. Bias helps you quantify, how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there has to be enough training data. Models that are too naïve, for e.g., one that gives same results for all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high. Variance is the degree of changes in the model itself with respect to changes in the training data. Thus, accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.



Thus, accuracy and robustness may be at the odds to each other as too much accurate model can be prey to over fitting hence it can

be too much accurate on train data but fails when it faces the actual data or vice versa.