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:

The optimal value of alpha for ridge and lasso regression are 1.0 and 0.0001 respectively. After the alpha value is doubled, the metrics of Ridge are R2 score - 0.886827, MSE - 0.063897 and the same metrics for Lasso are R2 score - 0.898340, MSE - 0.060560. There is no change in the metrics when compared with the metrics when calculated with the optimal value of 1.0 and 0.0001 for Ridge and Lasso regression respectively. The most important predictor variables after doubling the alpha value are as follows:

Ridge

GrLivArea

TotalBsmtSF

Neighborhood_Crawfor

OverallQual_8,

LotArea

Lasso

GrLivArea

Neighborhood_Crawfor

OverallQual_8

TotalBsmtSF

LotArea

There is no change in the predictor variables when the alpha value is doubled for Ridge and Lasso regression.

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:

The optimal value of lambda for ridge and lasso regression are 1.0 and 0.0001 respectively. The R2 scores for Ridge and Lasso are 0.886827 and 0.898340 respectively. The MSE values for Ridge and Lasso are 0.063897 and 0.060560 respectively. I would choose Lasso regression as its R2 score is higher than Ridge regression and its MSE value is lower than that of Ridge regression.

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 significant features from the new model after dropping the initial import predictor variables from Lasso regression are :

BsmtUnfSF

BsmtFinSF1

ExterQual_cat

GarageArea

FullBath

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:

As part of building the model, we use the techniques like feature selection, cross-validation and regularization to build the accurate model. Too many feature selection can make a model complex. The complex model has to change for each small change in the data. Regularization tries to keep the model simple. Trying to make the model too simple will result in large bias and the errors will be very high. Regularization allows only the optimal amount of complexity to the model. The robust or generalisable model keeps a balance between Bias and Variance Trade-offs. The accuracy of the model can be maintained by keeping the balance between Bias and Variance which minimizes the error as shown in the below graph: