**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:**

Alpha (Hyperparameter) is the co-efficient for the regularization term for Ridge and Lasso regularization. If Alpha is close to 0 it means the model has overfitted and for very high alpha model can be underfitting. The optimal value of alpha is based on the model we are building. In our assignment case for Ridge without RFE, the optimal value was 9 and for Lasso, it was 100. Using RFE it was 2 for Ridge and 50 for Lasso.

If we double the value of alpha, we are telling model to reduce the coefficient of features and in Lasso we will clearly see that more features will be eliminated (coefficients become 0). This reduces the complexity of model but can result in underfitting.

After using double the alpha value, for Ridge(with RFE) as well as Lasso(after RFE) ***OverallQual_10, RoofMatl_WdShngl, OverallQual_9, Neighborhood_StoneBr*** were the top predicator.

**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:**

While using normal Ridge regression we got an optimal value of 9 for lambda which is good but Ridge does not perform feature elimination and we can't use all the feature for predication even if we had excellent r2 of 0.90+ On other hand RFE with 50 feature and then Ridge over it gave us an optimal value of 2 which is ok but it still had 50 features.

For Lasso, we got 100 as lambda value and It also did the feature elimination with good r2 score as well. Further using RFE with lasso gave us alpha of 50 and eliminated more features. So, I will go ahead with Lasso regularization with pre RFE which is 50.

**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:**

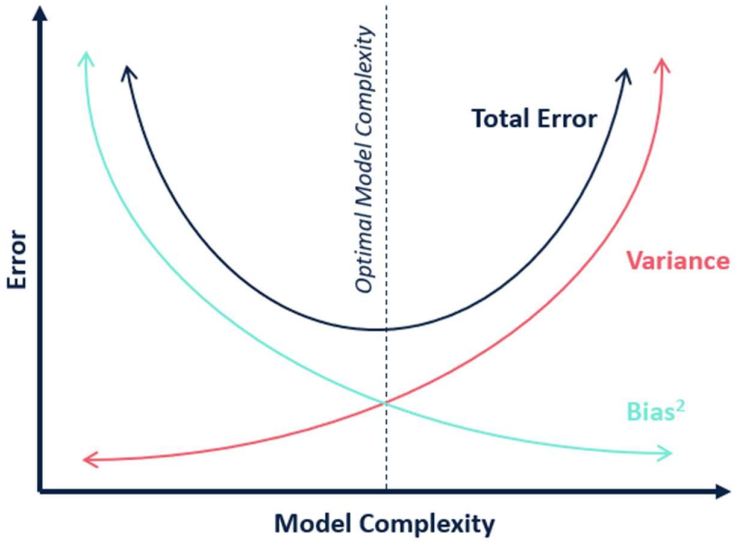
Next 5 most important predictor variables they would be BsmtExposure_Gd , 1stFlrSF, Neighborhood_Crawfor, 2ndFlrSF and Neighborhood_NridgHt.

**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:**

The answer lies in alpha value(lambda). Lower the alpha value more complex and overfitted the model will be and higher alpha value mean model is simpler but can be a underfit. So, selecting right alpha value is important which is similar to striking the perfect balance between variance and bias.




If we choose a very complex model it will have high accuracy on the train set but not on test data and in real-world as it would have memorized the train data set.

A model with low variance and low bias would be the most perfect model with good accuracy on the train and test set as well.