ASSIGNMENT-BASED SUBJECTIVE QUESTIONS

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

The optimal value of alpha for ridge and lasso regression as obtained by the model are as below:

1. Ridge – 0.1
2. Lasso – 40

After doubling the value of the alpha for both ridge and lasso, the new alpha values become,

1. Ridge – 0.2
2. Lasso – 80

We see that there is drastic changes in the performance metrics. The R2_score for the train set has been reduced while that of the test set has been increased. We also see that the top predictors have been changed as well.

For ridge regression, we see that the top 05 predictor variables now are,

'RoofMatl_CompShg'
'RoofMatl_Tar&Grv'
'RoofMatl_Membran'
'1stFlrSF'
'RoofMatl_WdShngl'

For lasso regression, we can say that the top 05 predictor variables now are,

'1stFlrSF'
'2ndFlrSF'
'OverallQual_Very_Excellent'
'OverallQual_Excellent'
'LotArea'

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?**

Based on the assignment, we found the optimal value of lambda for ridge regression to be 0.1 and for lasso regression to be 40.
We saw that the R2_score in case of lasso regression for both train set, and test set is better than that in case of ridge regression.

Hence, we will choose to apply lasso regression since it also performed some feature selection for us.

3. **After building the model, you realized 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?**

Initially, when we performed lasso regression, we found the top 05 predictor variables were,

'RoofMatl_Membran'
'RoofMatl_CompShg'
'RoofMatl_Tar&Grv'
'RoofMatl_WdShake'
'RoofMatl_Metal'

However, now, after excluding the above variables and building another model, we see that the top 05 predictor variables have changed and are as follows,

'1stFlrSF'
'2ndFlrSF'
'OverallQual_Very_Excellent'
'OverallQual_Excellent'
'BsmtFinSF1'

4. How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

A model can be made robust and generalizable by making it stable and simple. There should be no drastic change in the model upon changing the train set. In other words, there should be no overfitting in the model, the result of which is that the model performs well on train set but performs poorly on the test data.
In order to avoid the model from being too complex, we may use regularization techniques like ridge and lasso which penalize the insignificant variables trying to make the model too complex.

The implications of a model being robust is that the accuracy of the model on test set is at par with that on the train set. The reason for this being that our model is quite generalizable on any kind of unseen data unlike the case of a complex model which has a very high variance and that changes as soon as the training data is changed.