Advanced Regression Assignment

Q1) 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 alphas came out as following:
- Ridge Regression alpha = 500
- Lasso Regression alpha =0.01

On doubling the alpha values:
- RidgeRegression : There is a decrease in R2 values for both training and testing dataset. The most significant features remains the same, that is **OverallQual**
- Lasso Regression: There is a decrease in R2 values for both training and testing dataset. The most significant features remains the same, that is **OverallQual** but with higher impact on target after doubling of alpha value

The most import predictor is OverallQual which positively impacts our target variable.

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:

We observe that after finding the optimal alpha values for ridge and lasso regression, we see that ridge shows slightly better R2 scores than lasso. Although ridge shows slightly better score, the alpha value for ridge came out to be 500 which is better high, where as lasso came out to be 0.01 which is quite low. For the sake of better estimation, lasso was chosen as a better model as it eliminates features among 165 features which would help in better generalization.

3)
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) If we want to make our model more robust and generalisable, we can increase the alpha values. This will increase the penalty on features and hence make the model more generalisable.