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 double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Optimal value of alpha for Ridge Regression = 2.

Optimal value of alpha for Lasso = 0.01

After doubling Both:

In Lasso Regression R2 score of train set decreased from 0.885 to 0.871

In Lasso Regression R2 score of test set decreased from 0.889 to 0.881

In Ridge Regression R2 score of train set decreased from 0.936 to 0.934

In Ridge Regression R2 score of test set increased from 0.907 to 0.908

The following will be the most important predictor variables after the change:

1. GrLivArea
2. OverallQual_8
3. OverallQual_9
4. Functional_Typ
5. Neighborhood_Crawfor
6. Exterior1st_BrkFace
7. TotalBsmtSF
8. CentralAir_Y

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?

The better option here is lasso regression as it would be helpful to eliminate features which in turn makes the model more robust.

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?

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. GarageArea

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?

The model should be as simple as possible. In Fact the more simple it is the more the model is robust and generalisable .Accuracy is little bit less in simpler models. The accuracy must not change more between the test and train data. This means that the model will perform well on both train and test data.