Assignment Part-II

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?

A:
1. Ridge Regression:
    a. Optimal Value : 10
2. Lasso Regression:
    a. Optimal Value : 0.001

If we change the model with the double the value of alpha the R2 score of Train Set will be decreased. The R2 score of Test Set will be decreased. The model will become more generalised and penalty on the curve. The model become more simpler. The co-efficient of the variables will be zero.

Metric	Ridge Regression	Lasso Regression
R2 Score (Train)	0.934148	0.905235
R2 Score (Test)	0.927674	0.911638

The important predictor variables after the change is:

1. GrLivArea
2. OverallQual_8
3. OverallQual_9
4. Functional_Typ
5. Neighborhood_Crawfor

6. Exterior1st_BrkFace
7. TotalBsmSF
8. CentralAir_Y

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?

A.

1. We need to reduce the variance to increase the prediction accuracy.
2. If the requirement is feature selection and we have many variables then we can use Lasso. Lasso reduces the coefficient value to zero as the lambda value increases where lambda is the tuning parameter
3. If the requirement is to reduce the coefficient and we do not want the large coefficients than we can use 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?

A . The five most important Lasso predictors are:

1. OverallQual_9
2. GrLivArea
3. OverallQual_8
4. Neighborhood_crawfor
5. Exterior1st_BrkFace

After dropping the five most important predictor variables, the new five most predictors are:

1. 2ndFlrSF
2. Functional_Typ
3. 1stFlrSF
4. MSSubClass_70
5. Neighborhood_Somerst

```
2ndFlrSF              0.098102
Functional_Typ        0.073546
1stFlrSF              0.073456
MSSubClass_70         0.061023
Neighborhood_Somerst  0.056671
```

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?

1. A model is generalized when it is able to adopt any variations in the data set and it should be able to give good output for the unseen data.
2. A model is robust when it does not overfit. It should not have high variance. Small effect on bias is acceptable but it should not affect the model prediction .
3. The model should not be complex in order to be robust and generalisable.
4. A complex model has low bias but variance will be high. There should be balance between accuracy and complexity.
5. As model should not overfit in the same way it should not underfit.
6. To make our model more robust and generalisable it should have low variance and does not effect on the prediction and should not fail to understand the patterns in unseen test data.