Assignment-Advanced Regression-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?

Ans: The Optimal Values of lambda are:

- Ridge: 10
- Lasso: 100

The r2 score for the optimal values of lambda comes out to be:

	R2_score Train	R2_score Test
Ridge(alpha=10)	0.91	0.82
Lasso(alpha=100)	0.91	0.82

When the values of lambda are doubled:

	R2_score Train	R2_score Test
Ridge(alpha=10)	0.89	0.82
Lasso(alpha=100)	0.90	0.82

When the values of lambda are doubled, there is increase in the r2_score for both ridge and lasso

The significant variables after the value of alpha doubled are:

- OverallQual
- RoofMatl

- Neighbourhood
- SaleCondition
- GarageCars
- Functional

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?

Ans: I will choose Lasso,as it makes some variables coefficients as zero, doing feature selection.

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?

Ans: After the top 5 features are dropped the next important predictors are MSSubClass, OverallQual, KirchenQual, BsmtQual and LotShape.

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?

Ans: The model does not overfit on the training data and should be handle new data.

If the model performs well on training data and testing data, then the model is robust and generalizable.