Advance Regression Housing Assignment – Subjective Questions

Question 1

a) What is the optimal value of alpha for ridge and lasso regression?

Ans) Solved in the python notebook.

alpha for ridge = 20

alpha for lasso = 0.0001

b) What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

Ans) Solved in the python notebook.

Ridge

Before – Optimum value of aplha

```
R2 Train:  0.913392
R2 Test:  0.887753
RMSE Train:  0.116678
RMSE Test:  0.135900
```

After – Double value of aplha

```
R2 Train:  0.90619500576344 74
R2 Test:  0.8852150362443965
RMSE Train:  0.12142953901397889
RMSE Test:  0.13742808628083644
```

Lasso

Before – Optimum value of aplha

```
R2 Train:  0. 946347
R2 Test:  0.877576
RMSE Train:  0.091835
RMSE Test:  0.141927
```

After – Double value of aplha

```
R2 Train:  0.9320898596820508
R2 Test:  0.8758949461252682
RMSE Train:  0.10331863971554613
RMSE Test:  0.142898520349587 72
```

We can see that the below :-

- R2 Score in Test and Train decreases
- RMSE in Test and Train increases

c) What will be the most important predictor variables after the change is implemented?

Ans) Solved in the python notebook.

Ridge

OverallQual
Neighborhood_Crawfor
Neighborhood_Edwards
GrLivArea
Neighborhood_NridgHt

Lasso

Condition2_PosN
MSZoning_RL
MSZoning_RH
MSZoning_FV
MSZoning_RM

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) Solved in the python notebook.

The R2 score for Ridge is closer between the Train and Test data as compared to Lasso. So, I will choose the Ridge regression in this case.

However had the scores for Both Ridge and Lasso been similar one could choose Lasso as it makes the coefficient of variables to 0.

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) Solved in the python notebook.

Selecting the Ridge model again due to closer R2 scores in Train and Test and lowere RMSE

Top 5 features will now be:-
2ndFlrSF
1stFlrSF
BsmtQual_TA
Condition2_PosN
KitchenQual_TA

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 should be as simple as possible, it should have a fair accuracy. The accuracy may decrease with the model being less complex but as per the Bias-Variance trade-off it would have less variance and hence will be more generalisable but robust.

The implications of a model is that a robust and generalisable model on accuracy will be that it will perform well on training and test sets. Accuracy will not change much on the Training and Test data sets of a robust model.