Assignment Part 2

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?

Answer:

The optimal value of alpha for Ridge and lasso regression is 0.001.
Double means, alpha will be 0.002.
There is no much change in Ridge model but there is considerable decrease in R-Squared value in Lasso Model.
The most important predictor variables after changing alpha =0.002 are :
Ridge Regression : OverallQual, GrLivArea, LotArea, 2ndFlrSF, 1stFlrSF
Lasso Regression: OverallQual, GrLivArea, GarageArea, FirePlaces, Neighborhood_NoRidge

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?

Answer:

I will choose Lasso Regression with alpha = 0.001 .
Because the R-Square value for training data set is 0.812 (I.e 81.2%) which almost equal to the R-Square value for testing data set is 0.793 (I.e 79.3%).
And also Lasso provides feature selection where as Ridge doesn't.

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?

Answer:

The five most important predictor variables now are :
1stFlrSF
2ndFlrSF
YearRemodAdd
BsmtFullBath
Neighborhood_NridgHt

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?

Answer:

The model should be robust and it can be achieved through performing EDA, data cleaning. And the model should also be generalizable which can be achieved by not overfitting the model.
For the above cases Lasso and Ridge regression will help in regularizing the model there by making the model general.

The implications on accuracy are that the test accuracy will be far less than the train accuracy if the model was overfitted or not generalized. Lack of robustness makes model not-trustable while predictions.