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

Solution:

Ans:-1 Optimal value of alpha for Ridge :- 5

After Making alpha for Ridge = 10, Coeff values are increasing as alpha will increase and R2 score of train data is also drop, or slight change

Top predictor variable after change :- OverallQual_9, Neighborhood_StoneBr, Neighborhood_Crawfor, functional_Typ, BsmtQual_Grd.

Optimal value of alpha for Lasso :- 100.

After Making alpha for Lasso = 200, Coeff values are increasing and more feature removed from model and R2 score for train and test decrease.

Top predictor variable after change :- OverallQual_9, Neighborhood_StoneBr, SaleCondition_Partial, functional_Typ, OverallQual_8.

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?

Solution:

Ans: 2 ① The optimal lambda Value in case of Ridge and Lasso is as below:
* Ridge :- 5
* Lasso :- 100

② The mean square Error of Lasso is slightly lower than of Ridge and R2 for Train and Test of Lasso is better.

③ Also, since Lasso helps in feature reduction (as the coefficient value of one of the feature became 0) so Lasso have 159 feature to zero, whereas Ridge has only 20 feature equal to Zero.

Therefore, the variable predicted by Lasso can be applied to choose significant variables for predicting the price of a house.

Question 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?

Solution:

Ans 3

Current 5 most important predictor variable

1. Overall Qual
2. BsmtQual
3. Exterior1st
4. Mas Vnr Type.
5. Neighborhood.

After Excluding upper 5 important variable given by lasso, here's the next 5 important variable for prediction.

1. Sale Condition.
2. Bsmt Exposure
3. MS Zoning
4. Bsmt Fin Type 1
5. MS Sub Class.

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Solution:

Ans 4: To make model Robust and generalisable 3 feature are required:

1. Model Accuracy should be > 70-75% .
2. P-value of all the features is < 0.05
3. VIF of all features are < 5

Thus we are sure that Model is robust and generalisable

The simpler the Model the more the bias but less variance and more generalizable. Its Implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data.

Bias - variance Trade off