

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:

1. alpha value of ridge = 4.0 and lasso its = 100. Model will get under fitting so r^2 will decrease. And important predictor variable co-efficient will be change but order will remain same of predictor variable.

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 chose lasso, because even though both r^2 value are same the complexity of the lasso is very low and as there is not much difference in the efficiency. So it remove most of insignificant variable by assigning it 0 as co efficient but ridge will be assigning those variables to close to 0 not 0

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?

1. OverallQual
2. 2ndFlrSF
3. GrLivArea
4. Neighborhood_NoRidge
5. FullBath

2nd time

1. Neighborhood_StoneBr
2. Exterior2nd_ImStucc
3. Neighborhood_Crawfor
4. Neighborhood_NridgHt
5. BsmtExposure_Gd

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:

We use Cross Validation for increasing the efficiency and Regularization Technique to make the model robust, less complex and to remove under or over fitting of the model. We use Ridge and Lasso regression techniques especially for this issue. The implications is as Bias and variance Trade off. If we try to increase. So both test and train pred accuracy defines the efficiency.