

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:

For Ridge, it is 1.0. For Lasso, it is 0.0001.

When we double the Lasso and Ridge regression best alpha-> the most important predictors are as follows

Ridge:

1. TotalBsmtSF
2. OverallQual_10
3. OverallQual_9
4. LotArea
5. FullBath

Lasso:

1. TotalBsmtSF
2. OverallQual_10
3. OverallQual_9
4. LotArea
5. FullBath

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 choose Lasso- as the R squared value varies less for train and test data in case of Lasso regression

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: We drop the top 5 predictor variables, and then rebuild the model. Here are the next top 5, after rebuilding the lasso regression

1. LotFrontage
2. Exterior2nd_ImStucc
3. GarageCars
4. MasVnrArea
5. Neighborhood_StoneBr

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: Simpler models are considered more robust and generalizable. However- the implication of a simpler model is underfitting, resulting in high bias, but low variance. Simpler models has high bias, but low variance. Whereas complex models have low bias and high variance. So, it is always the tradeoff between Variance and Bias