

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:

Please find the optimal values of alpha below :

Ridge = 0.1

Lasso = 20

R² values are increased on test data, when alpha is doubled

Important predictor variables are

GrLivArea

OverallQual_8,9,10

LotArea(log)

PoolArea

GarageArea

OverallQual_10

RoofMatl_Membran

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:

since r² value is better for ridge, will go for ridge.

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:

RoofMatl_WdShng

HouseStyle_1.5Unf

MSZoning_RH

FullBath_3

SaleType_Con

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:

Model should be tuned such that the bias and variance are balanced appropriately. This can be achieved by making model less complex and including the appropriate independent variables while modelling.

Following above steps we might have less accurate model on the test data, but if the variance is reduced, the model have better accuracy on the test data.