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?

Answer:

The optimal value of alpha for Ridge regression is 1.0

The optimal value of alpha for Lasso regression is 0.0001

Top 10 predictor variables of ridge:
1. LotFrontage
2. OverallCond
3. BsmtFinSF2
4. BsmtFinType1
5. BsmtFinSF1
6. KitchenAbvGr
7. OverallQual
8. MSSubClass_190
9. BsmtCond
10. TotRmsAbvGrd

Top 10 predictor variables of lasso:
1. LotFrontage
2. BsmtFullBath
3. OverallCond
4. Exterior1st_BrkFace
5. OverallQual
6. MSZoning_RH
7. CentralAir
8. MSZoning_RM
9. Street_Pave
10. GarageArea

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?

We will choose Lasso has it is giving features selection option and also it has removed unwanted features from the model without affecting the model accuracy.

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 top 5 important predictor are :

LotFrontage, BsmtFullBath ,OverallCond , Exterior1st_BrkFace,OverallQual.

After dropping these we get

LotArea, FullBath , Exterior1st_CBlock,MSZoning_RH, BsmtFinSF1

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:

To make model robust and generalisable 3 features are required those are:

Model accuracy should be greater than 70-75% in this case its coming

- 92% on train and 73% on test.

- P -value of all the features is < 0.05

- VIF of all the features are < 5

This proves that model is robust and generalizable.