

Project Git Link:

<https://github.com/arvindkjangid/HousePricePrediction.git>

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:

The Optimal value lambda for Ridge is 10 & for Lasso is 100.

As we double the value of alpha for both ridge and lasso model become underfit.

For Ridge regression on increasing Lambda from 10 to 20

R-Squared value decreases for train and test data, And Error increases.

Top 5 predictor after change for ridge

Neighborhood_NoRidge	14119.840544
Neighborhood_StoneBr	14222.176741
Neighborhood_NridgHt	19479.063541
BsmtQual_Gd	-12287.891453
LandContour_HLS	11363.346639

Top 5 predictor after change for Lasso

Neighborhood_Crawfor	15014.144898
Exterior1st_BrkFace	15392.805133
Neighborhood_NoRidge	24052.305427
Neighborhood_StoneBr	25171.764175
Neighborhood_NridgHt	32113.707414

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:

The best value of lambda for Ridge is 10 & for Lasso is 100.

It's a bias variance trade for overfitting of model vs underfitting the model. And our goal is to find the optimal value, as we run for different range of alpha for ridge and lasso regression. And we got the best optimal value is for Ridge is 10 & for lasso 100.

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:

After removing the top 5 variable, by using GridsearchCV we found found new optimal value for lambda = 50.

So top 5 predictor after fitting is:

Condition2_PosN -86117.552699

Neighborhood_Mitchel -27633.677269

Neighborhood_Gilbert -23432.906451

RoofMatl_CompShg 22521.906721

RoofMatl_WdShngl 27274.668057

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:

We can say that model is robust and generalised because.

We checked the R-Squared value, the difference between train & test r-squared value is low means model is robust.

And it works well with unseen data.

And model is generalised because, by using GridSearchCV we find out the optimal value of lambda among various test options.

We mean that over model is best fit respect to bias-variance trade off.