

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?

A) Optimal value of alpha for ridge = 10

Optimal value of alpha for Lasso = 0.001

Ridge: r^2 _score of train data drops because the coefficient values of the model are increases as alpha value will increase.

Lasso: After doing the implementation of the advanced regression. As alpha value increases more features are removed from the model, but r^2 _score also drops in both train and test data.

The most important predictor variables after change in implementation are :

OverallQual_Excellent(0.115), Neighborhood_Crawfor(0.083), Neighborhood_StoneBr(0.073) Functional_Typ (0.069), GarageCars(0.068)

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?

In Ridge Regression because of the alpha value is 10 when i am doing the implementation of surprise housing. which i saw the linearity in the model.

The Lasso varies from point to point from 0.001 to 1 which i saw not in linear.

Lambda is the penalty term. Where λ given here is denoted by an alpha parameter in the ridge function. So, by changing the values of alpha, we are controlling the penalty term in the model

So, I will apply Ridge Regression for doing the model .

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?

A) I realised that the five most important predictor variables in the lasso model are not available in the incoming data then creating another model excluding those five variable after applying the advanced regression same steps the top five predictor variables are

Neighborhood_NridgHt(0.067),GrLivArea(0.066),OverallCond_Excellent(0.062)OverallQual_Very Good
(0.062),SaleCondition_Normal(0.057)

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?

A) Robust : Suppose if the model is robust means does not change in performance if the data having variation .

Accuracy : If the model is Accuracy means does not overfit the data. A too complex model will have a very high accuracy. The model should make as simple. The accuracy might decrease. But it will more robust and generalisable

The simpler the model more the bias and lower the variance, complex the model lower the bias and higher the variance.

Hence the bias and variance have to be at their optimal level. When the model should have robust and generalisable the following features are to be considered.

1. Model Accuracy should be >70- 75% .
2. P-Value of features is < 0.05
3. VIF of all features are < 5