

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 :

Optimal Value for Alpha

1. Ridge = 0.9

2. Lasso = 50

Before making the alpha Ridge model was performing better as there RSS is significantly less than other models and delta between R^2 Train and r^2 test was also similar to Lasso.

After changes, delta bw train and test for ridge has increased and However RSS is still less than others and r^2 is better for lasso for both train and test.

OverallQual, Neighborhood=NoRidge, Total rooms above grade, Full bathrooms above grade, GarageCars are the most imp variable after change.

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?

Ans:

Using GridCV I have used Optimal Value of Alpha for Ridge = 0.9 and Lasso = 50. I will use Alpha = 50 for lasso regression. Lasso regression adds the absolute magnitude of the coefficients to the cost function. Here i am looking for predictors which are effecting the Saleprice of the house. As Lasso will remove the unwanted features and we will be left with relevant features which indeed is required.

Considering once we have strong predictors we can see the houses with these features having low price can be bought as of now and will be sold at higher price in future.

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?

Ans:

After removing 5 important features from input model and lasso model

1.Fireplaces, 2.BsmtExposure_Gd, 3.Exterior1st_BrkFace, 4.BsmtFinType1_GLQ, 5.LotConfig_CulDSac

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?

Ans:

Model with low bias and low variance identifies all the patterns so it performs well with unseen data.

we need to manage model complexity: It should neither be too high, which would lead to overfitting, nor too low, which would lead to a model with high bias (a biased model) that does not even identify necessary patterns in the data.

Thus we use regularization so our model to perform well on unseen data and identifies the underlying patterns present in the data.

Regularization doesn't affect the accuracy the model because Regularization can significantly reduce model variance while not increasing bias much.

