

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 of alpha for:

- a. Ridge: 0.1
- b. Lasso: 0.0001

If we double the alpha for ridge, the model has performed slightly better with improved R2 value and lower RMSE on test data. Same is the case for Lasso regression. There is no change in most important variables: 'LotFrontage', 'LotArea', 'OverallQual', 'OverallCond'

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:

Theoretically we would choose Lasso as it does variable selection making the model easy to interpret. But in this case, ridge regression has performed better.

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 next most important variables were the same ones which were after the earlier variables. Basically, the variables which were important in the order from 6-10, have now become top 5 important variables.

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 use Lasso regularization to reduce the number of variables and increase interpretability. However, this comes at a cost of slightly reduced accuracy in this case, as we have observed ridge to be better performing.