

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 of alpha for ridge regression is 10.

If we double the value of alpha for ridge regression i.e. set it to 20 then R2 score becomes 0.7197

The most important predictor variables are:

1. GrLivArea
2. OverallQual
3. MSZoning_RL
4. 2ndFlrSF
5. OverallCond

The optimal value of alpha for lasso regression is 0.001.

If we double the value of alpha for lasso regression i.e. set it to 0.002 then R2 score becomes 0.7194

The most important predictor variables are:

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. SaleType_New

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:

I will use lasso regression because for the optimal value of alpha the R2 score is 0.7211 whereas for ridge regression it is 0.7190.

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 five most important predictor variables after excluding the first five most important predictor variables are:

1. MSZoning_FV
2. MSZoning_RM
3. SaleType_New
4. BsmtFinSF1
5. GarageCars

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:

A model can be robust and generalisable when it does not overfit. It should learn the relations between the predictor variables and the target variable not mug up the mapping of the predictor variables and the target variable.

There is a very little difference in the accuracy score for both the training and the testing set for a robust and generalisable model.