

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?

Optimal value of alpha for ridge regression: 10.0

Optimal value of alpha for lasso regression: 0.001

If we choose to double the value of alpha for both ridge and lasso:

Coefficient values of ridge regression becomes lower and more number of coefficients in case of lasso regression is turning 0.

The most important predictor variable after doubling alpha for ridge regression:
1stFlrSF (First Floor square feet)

The most important predictor variable after doubling alpha for lasso regression:
1stFlrSF (First Floor square feet)

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?

Optimal value of alpha for ridge regression: 10.0

Optimal value of alpha for lasso regression: 0.001

We will choose lasso regression model since we got slightly better R2 value and Root Mean Square Error. Details as follows:

Lasso Regression model was able to achieve R2 score of 0.8746 on test data i.e. 87.46% of the variance in test data can be explained by the model.

Root Mean Square Error = 0.1531 on test data, that means the prediction made by the model can off by 0.1531 unit.

Ridge Regression model was able to achieve R^2 score of 0.8712 on test data i.e. 87.12% of the variance in test data can be explained by the model.

Root Mean Square Error = 0.1534 on test data, that means the prediction made by the model can off by 0.1534 unit.

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?

The 5 most important predictor variables from lasso regression (after removing the initial 5 most important variables) are as follows:

FullBath, BsmtFinSF, GarageArea, Fireplaces, HalfBath

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?

Bias-Variance Trade off - A model can be made more robust and generalisable by reducing the variance and thereby sacrificing some bias.

This will decrease the model accuracy of training data but leads to better accuracy of test data.