

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:

Ridge=20

Lasso=0.01

Initial Ridge=20, Lasso=0.01, but after doubling the alpha for both Ridge and Lasso, training score reduced a bit.

Most important predictor variables are - { OverallQual, 2ndFlrSF, BsmtFinType1, MSSubClass }

Ridge Alpha = 20	Ridge Alpha X 2 = 40
R2 Train Score: 0.7401593873925626 R2 Test Score: 0.7520034601196846 RSS Training: 247.20306874068837 RSS Test : 125.97628983624826 MSE Train : 0.24211857859029223 MSE Test : 0.2876171000827586	R2 Train Score: 0.7396373556003266 R2 Test Score: 0.7495907908887729 RSS Training: 247.69971112359366 RSS Test : 127.20186789656722 MSE Train : 0.24260500599764315 MSE Test : 0.29041522350814436

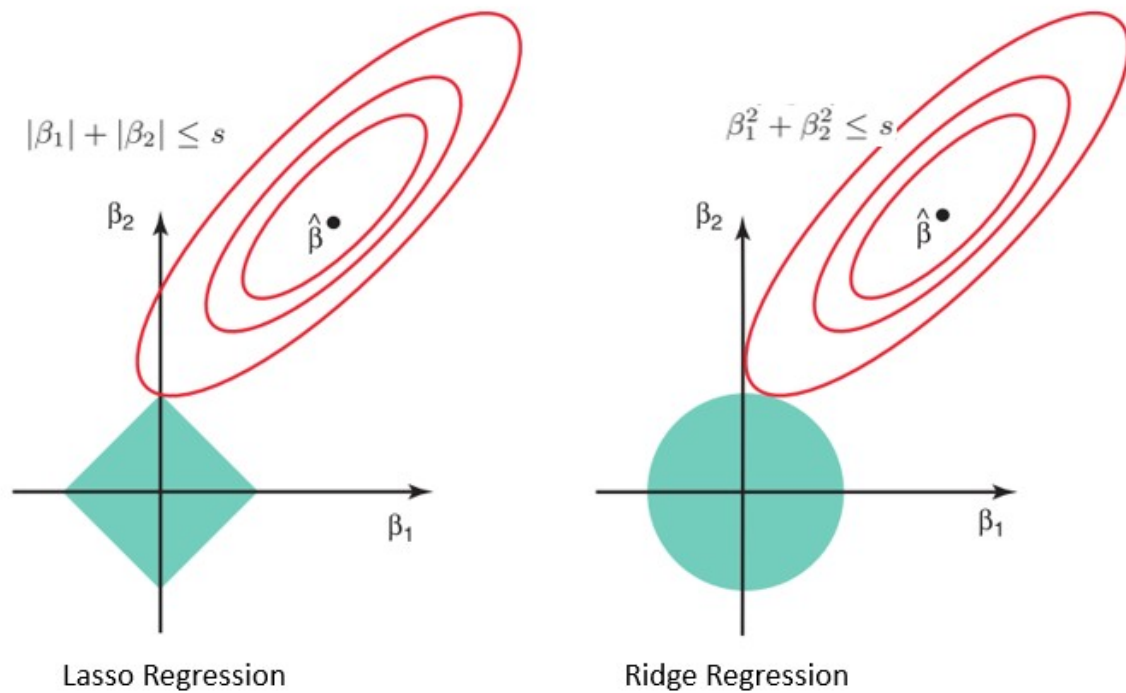
Lasso Alpha = 0.01	Lasso Alpha X2 = 0.02
R2 Train Score: 0.7377541123687783 R2 Test Score: 0.7520437858484106 RSS Training: 249.4913614024021 RSS Test : 125.9558053339552 MSE Train : 0.2443598054871715 MSE Test : 0.2875703318126831	R2 Train Score: 0.7312537095208462 R2 Test Score: 0.7452740096978627 RSS Training: 255.67561226270632 RSS Test : 129.3946891299933 MSE Train : 0.2504168582396732 MSE Test : 0.29542166468035

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 choose Lasso because it helps in reducing the features in the model and helps to create a simple model when compared with Ridge. This is important for creating generalizable models. But, Based on some research I understand that, Lasso works well if the model is impacted by fewer parameters from a lot of variables. Ridge works well if there are many large parameters of about the same value.



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:

```
Lasso OverallQual 0.545237
2ndFlrSF 0.160863
BsmtFinType1 0.147387
Fireplaces 0.107896
BsmtExposure 0.094011
```

After removing them

```
Lasso OverallQual 0.545237
2ndFlrSF 0.160863
BsmtFinType1 0.147387
Fireplaces 0.107896
BsmtExposure 0.094011
```

Predictor variables are same but with different order

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:

To get robust and generalizable, we should make sure that test accuracy needs to be higher than training. However, this can not be too high. Model should generalise during the training, scale if possible.

If model memorizes the data, then it will impact the test data. It will lead to overfitting and failure with actual test data. Hence, regularization is important.