

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:

For Ridge alpha value is 0.2

```
r2_train_lr 0.9540566495981394
r2_test_lr 0.7792334083702577
```

For Lasso it is 50.0

```
r2_train_lr 0.9473866269176798
r2_test_lr 0.8493410610409731
```

When we double the best alpha interestingly for both Ridge and Lasso Train R2 decreased and at the same time Test R2 increased.

Ridge alpha 0.4

```
r2_train_lr 0.9490652705308924
r2_test_lr 0.8122729204242319
```

Lasso alpha 100

```
r2_train_lr 0.935044428058271
r2_test_lr 0.856603857264153
```

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:

Will be using the lasso regression, as lasso makes certain coefficient to zero to ridge makes close to zero, which helps in lasso to select features automatically. Ridge does the feature selection also, with reduced features the model becomes simple compared to Ridge model and will be having less variance compared to ridge

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:

Top predictors for Lasso are

```
['RoofMatl_WdShngl',  
 'TotalBsmtSF',  
 'GrLivArea',  
 'OverallQual_9',  
 'OverallQual_10']
```

Top predictors for Ridge are

```
['2ndFlrSF', 'RoofMatl_WdShngl', '1stFlrSF', 'PoolQC_Ex', 'GrLivArea']
```

After removing the top 5 predictor variables and rerunning the lasso model, the best alpha is still 50 and the r2 scores are

```
r2_train_lr 0.9420285204905677  
r2_test_lr 0.837493519374845
```

now the top 5 predictors are

```
['LotArea', 'BsmtFinSF1', '1stFlrSF', 'PoolQC_Ex', '2ndFlrSF']
```

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:

Use regularization to penalize the coefficients and make the model simple and with little variance