Advanced Regression Assignment Question and Answers.

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?

Ans: the optimal alpha values for the Ridge and Lasso regression are 6 , 20 and then now they are doubled.

Alpha optimal value 6 and 20 :

Ridge : OverallQual (positive slope)and Condition2_PosN (Negative slope)

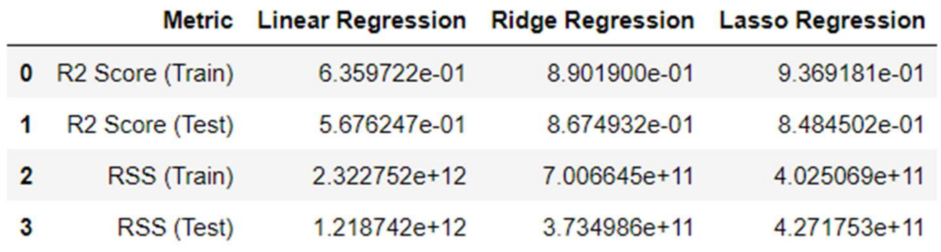
Lasso : RoofMatl_WdShngl (Positive Slope) and Condition2_PosN (Negative slope)

Alpha value doubled 12 and 40:

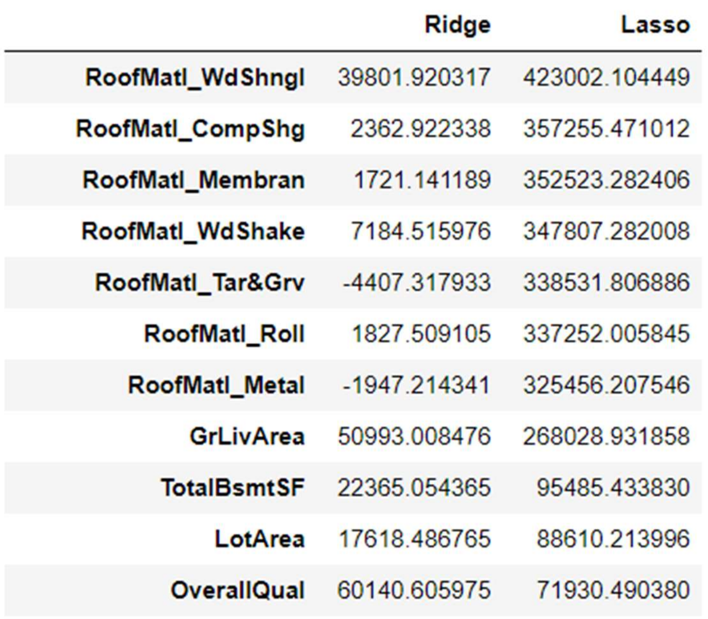
Ridge OverallQual (Positive slope) and KitchenQual_TA (Negative Slope)

Lasso GrLivArea (Positive Slope) , Condition2_PosN (Negative Slope)

Ridge and Lassso Regression with the actual alpha value :

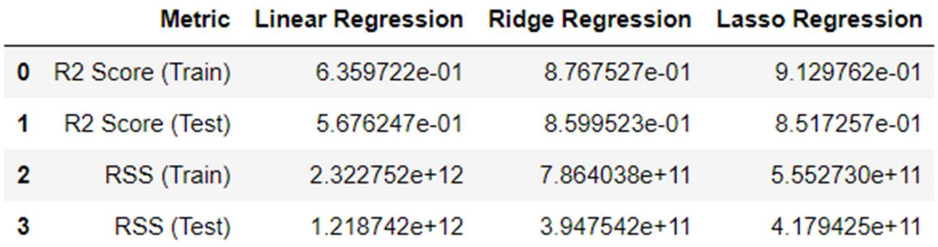


	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	6.359722e-01	8.901900e-01	9.369181e-01
1	R2 Score (Test)	5.676247e-01	8.674932e-01	8.484502e-01
2	RSS (Train)	2.322752e+12	7.006645e+11	4.025069e+11
3	RSS (Test)	1.218742e+12	3.734986e+11	4.271753e+11

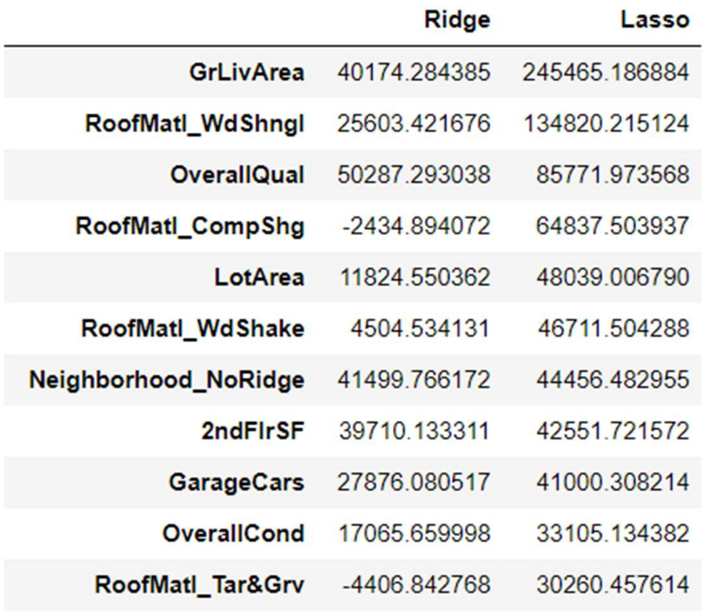


	Ridge	Lasso
RoofMatl_WdShngl	39801.920317	423002.104449
RoofMatl_CompShg	2362.922338	357255.471012
RoofMatl_Membran	1721.141189	352523.282406
RoofMatl_WdShake	7184.515976	347807.282008
RoofMatl_Tar&Grv	-4407.317933	338531.806886
RoofMatl_Roll	1827.509105	337252.005845
RoofMatl_Metal	-1947.214341	325456.207546
GrLivArea	50993.008476	268028.931858
TotalBsmtSF	22365.054365	95485.433830
LotArea	17618.486765	88610.213996
OverallQual	60140.605975	71930.490380

Ridge and Lasso Regression with the doubled alpha value:



	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	6.359722e-01	8.767527e-01	9.129762e-01
1	R2 Score (Test)	5.676247e-01	8.599523e-01	8.517257e-01
2	RSS (Train)	2.322752e+12	7.864038e+11	5.552730e+11
3	RSS (Test)	1.218742e+12	3.947542e+11	4.179425e+11



	Ridge	Lasso
GrLivArea	40174.284385	245465.186884
RoofMatl_WdShngl	25603.421676	134820.215124
OverallQual	50287.293038	85771.973568
RoofMatl_CompShg	-2434.894072	64837.503937
LotArea	11824.550362	48039.006790
RoofMatl_WdShake	4504.534131	46711.504288
Neighborhood_NoRidge	41499.766172	44456.482955
2ndFlrSF	39710.133311	42551.721572
GarageCars	27876.080517	41000.308214
OverallCond	17065.659998	33105.134382
RoofMatl_Tar&Grv	-4406.842768	30260.457614

After the alpha value doubled the only difference is the coefficeints have more difference but the r2 square , RSS value has slight difference.

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?

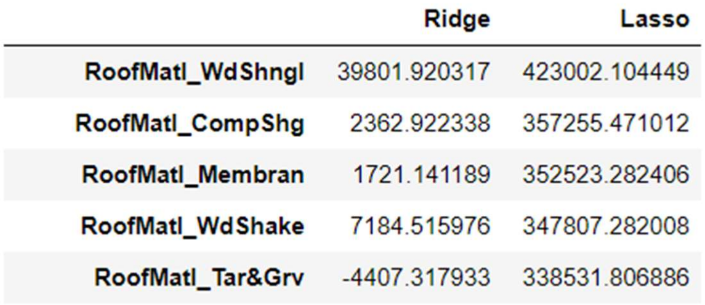
Ans: Alpha values for Ridge and Lasso are 6 and 20 .

Penalty in Lasso forces the some of the coefficinets to estimates to "0". This performs the variable selection.

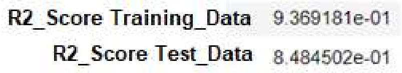
Models generated from the Lasso are generally easier to interpret than those produced by the ridge regression.

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?

Ans: lasso regression model with five most important variables:

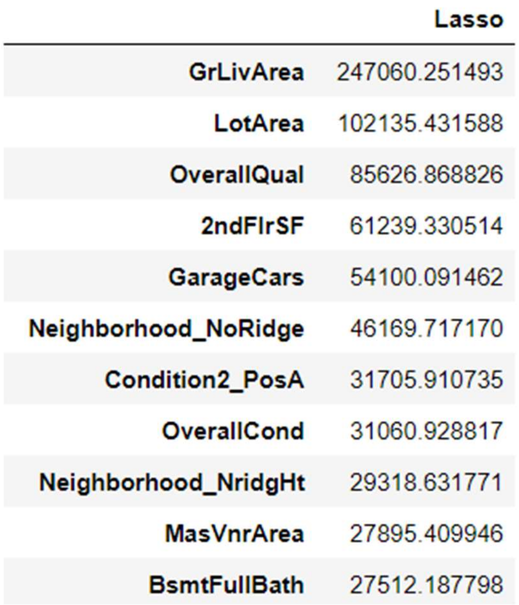


	Ridge	Lasso
RoofMatl_WdShngl	39801.920317	423002.104449
RoofMatl_CompShg	2362.922338	357255.471012
RoofMatl_Membran	1721.141189	352523.282406
RoofMatl_WdShake	7184.515976	347807.282008
RoofMatl_Tar&Grv	-4407.317933	338531.806886

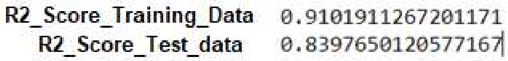


```
R2_Score Training_Data   9.369181e-01
    R2_Score Test_Data   8.484502e-01
```

After removing the above five variables:



	Lasso
GrLivArea	247060.251493
LotArea	102135.431588
OverallQual	85626.868826
2ndFlrSF	61239.330514
GarageCars	54100.091462
Neighborhood_NoRidge	46169.717170
Condition2_PosA	31705.910735
OverallCond	31060.928817
Neighborhood_NridgHt	29318.631771
MasVnrArea	27895.409946
BsmtFullBath	27512.187798



```
R2_Score_Training_Data   0.9101911267201171
   R2_Score_Test_data    0.8397650120577167
```

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?

Ans: As Per, Occam’s Razor— given two models that show similar ’performance’ in the finite training or test

data, we should pick the one that makes fewer on the test data due to following reasons:-

▪ Simpler models are usually more ’generic’ and are more widely applicable

▪ Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.

▪ Simpler models are more robust.

o Complex models tend to change wildly with changes in the training data set

o Simple models have low variance, high bias and complex models have low bias, high variance

o Simpler models make more errors in the training set. Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples

Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use.

Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.

Also, Making a model simple leads to Bias-Variance Trade-off:

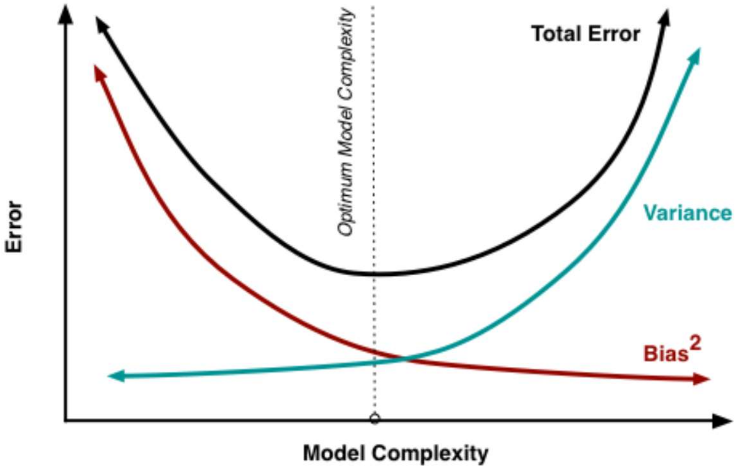
• A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.

• A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed.

Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.

Variance refers to the degree of changes in the model itself with respect to changes in the training data.

Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.


Optimum Model Complexity
Total Error
Variance
Bias^2
Error
Model Complexity