Student name; Devendran Mani

Date: 24th-Oct-2023

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?
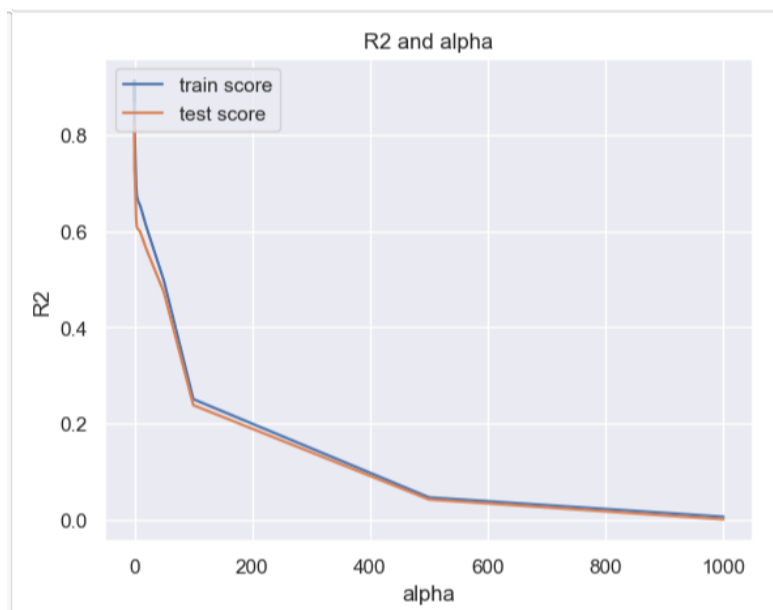
[Ans]:

Lasso optimal alpha = 0.0001, below are the top 10 co-efficients

Out[430]:

	coeffs1
15	0.062156
16	0.048624
45	0.045863
54	0.039446
47	0.032609
61	0.030228
59	0.029638
48	0.029300
13	0.020977
4	0.017803

Out[431]:

	coeffs1
70	-1.479146
3	-0.182884
50	-0.080688
12	-0.077831
39	-0.045695
7	-0.037844
77	-0.025356
53	-0.022518
26	-0.018084
37	-0.016296

R2 and alpha

Lasso: 2 x optimal alpha = 0.0002, below are the top 10 co-efficients

Out[434]:

	coeffs2
15	0.062961
16	0.048375
45	0.045391
54	0.039578
47	0.031546
59	0.030190
61	0.029229
48	0.028223
13	0.019843
4	0.016847

Out[437]:

	coeffs2
70	-1.414817
3	-0.162127
12	-0.075699
50	-0.074579
39	-0.044259
77	-0.024961
53	-0.022234
26	-0.017407
37	-0.015019
63	-0.013090

Most important predictor variables with alpha = 0.0001

70   -1.479146 -->   PoolArea

3   -0.182884 -->   LotArea

50   -0.080688 -->   BedroomAbvGr

12	-0.077831 -->	Condition1
15	0.062156 -->	HouseStyle
16	0.048624 -->	OverallQual
45	0.045863 -->	GrLivArea

What will be the most important predictor variables after the change is implemented?

[Ans]: PoolArea

What will be the most important predictor variables after the change is implemented?

[Ans]: Below are the most important predictor variables after the change is implemented

Most important predictor variables with alpha = 0.0002

70	-1.414817 -->	PoolArea
3	-0.162127 -->	LotArea
12	-0.075699 -->	Condition1
50	-0.074579 -->	BedroomAbvGr
15	0.062961 -->	HouseStyle
16	0.048375 -->	OverallQual
45	0.045391 -->	GrLivArea

Ridge optimal alpha = 0.241, below are the top 10 co-efficients

	coeffs1
15	0.063191
45	0.050402
16	0.048150
54	0.038912
47	0.035506
59	0.034229
61	0.032702
48	0.030065
13	0.021830
4	0.019478

	coeffs1
70	-1.307846
3	-0.195702
7	-0.128879
50	-0.084101
12	-0.077983
39	-0.048490
77	-0.024939
53	-0.022808
26	-0.017905
37	-0.016851

Lasso: 2 x optimal alpha = 0.482, below are the top 10 co-efficients

	coeffs1
15	0.064522
45	0.053176
16	0.047551
54	0.038598
59	0.037683
47	0.036882
61	0.033840
48	0.029897
13	0.021527
4	0.019943

	coeffs1
70	-1.134983
3	-0.187648
7	-0.110040
50	-0.081702
12	-0.076540
39	-0.049427
77	-0.024643
53	-0.022759
26	-0.017330
63	-0.016629

Most important predictor variables with alpha = 0.241

70	-1.307846	--> PoolArea
3	-0.195702	--> LotArea
7	-0.128879	--> LandContour
50	-0.084101	--> BedroomAbvGr
12	-0.077983	--> Condition1

15   0.063191   --> HouseStyle

45   0.050402   --> GrLivArea

What will be the most important predictor variables after the change is implemented?

[Ans]: PoolArea

What will be the most important predictor variables after the change is implemented?

[Ans]: There is no change in the most important predictor variables after the change is implemented

Most important predictor variables with alpha = 0.482

70   -1.134983 --> PoolArea

3   -0.187648 --> LotArea

7   -0.110040 --> LandContour

50   -0.081702 --> BedroomAbvGr

12   -0.076540 --> Condition1

15   0.064522   --> HouseStyle

45   0.053176   --> GrLivArea

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?

[Ans]: Will choose Lasso, it has more number of zero co-efficients

In lasso "mean_train_score" = 0.911725 & "mean_test_score" = 0.822013 are higher compare to Ridge regression.

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?

[Ans]

1) MoSold
2) OpenPorchSF
3) TotalBsmtSF
4) YearBuilt
5) RoofMatl

Question 4

How can you make sure that a model is robust and generalisable?

[Ans]

Select an alpha that gives the highest mean_test_score & and mean_train_score values, also the model needs to trained for more combination of training value. Here we used "GridSearchCV()" to search for the best Alpha value. The most generalized model shall perform equally well in test and train data sets.

What are the implications of the same for the accuracy of the model and why?

[Ans]

The best alpha value is the one that best regularizes the model and uses most impactful/essential predictors for the model building. Using non-essential predictors to build the model will reduce the model performance and model performance accuracy.

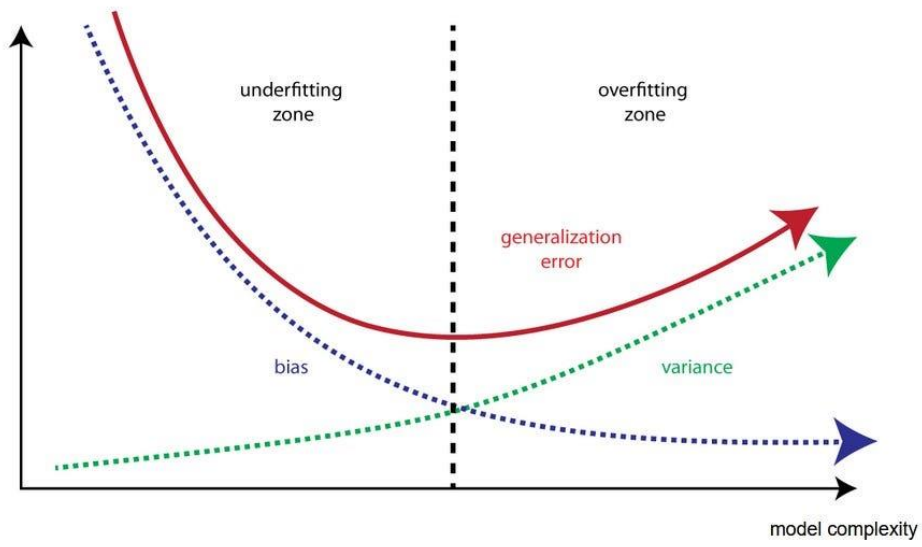The below figure shows the typical relationship between alpha and R2 values,

Lambda 0
MSE Test= 1.04e+01
MSE Train= 4.99e-03

Lambda 0.5
MSE Test= 6.84e-02
MSE Train= 7.59e-02
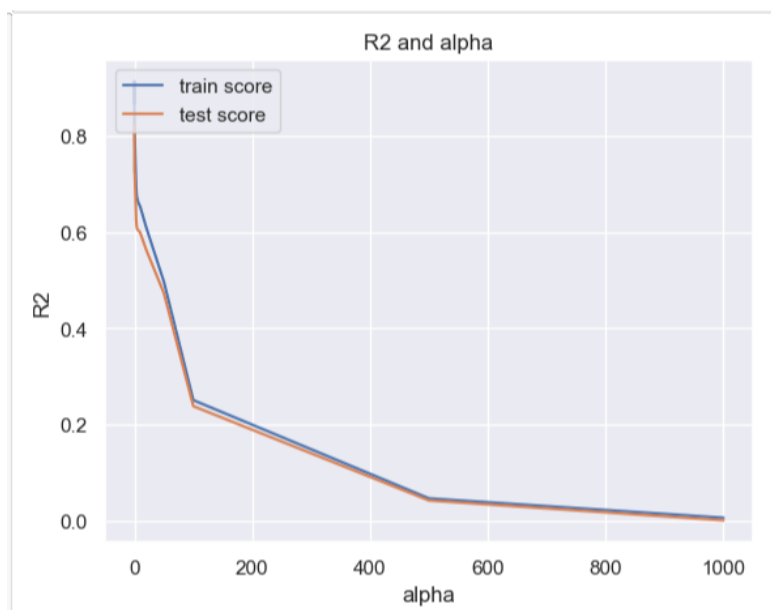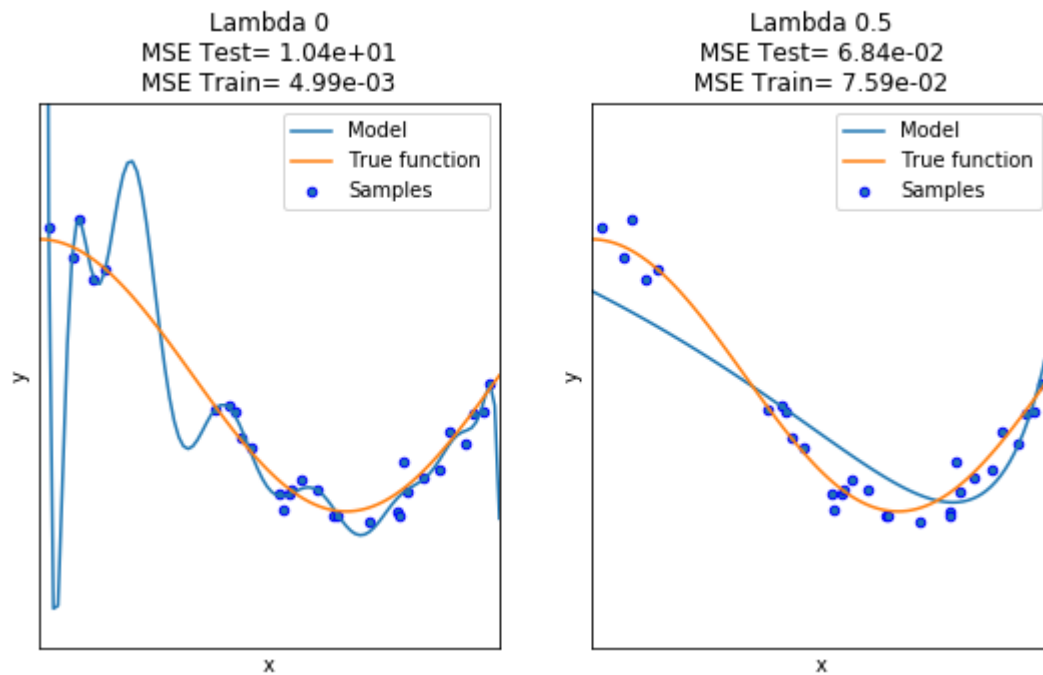
When alpha(lambda) = 0 : model complexity will be high and over fitted model

When alpha(lambda) = High (1000) : model complexity will be low and under fitted model, we need find the optimal alpha(lambda) that give optimal complexity and best accuracy.
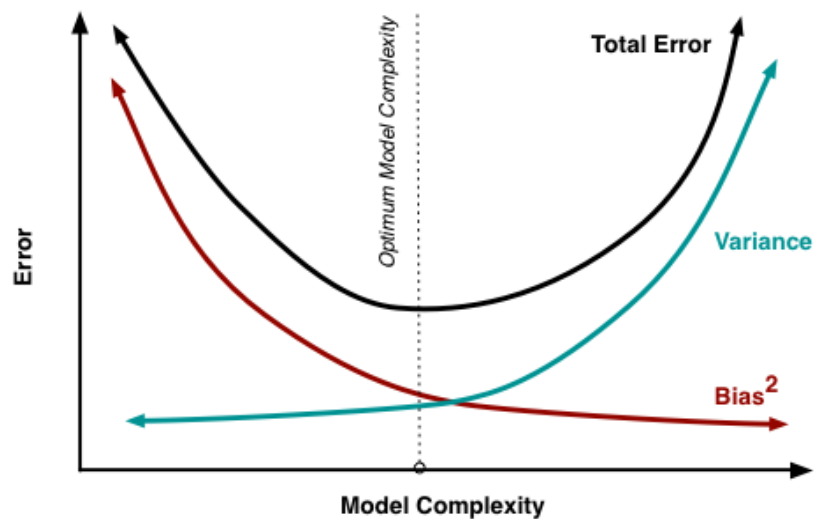
the bias vs. variance trade-off

-End of document -