ASSIGNMENT PART 2 – SUBJECTIVE QUESTIONS

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 1

Optimal Value of Alpha for **Ridge: 0.3**

Optimal Value of Alpha for **Lasso: 0.0001**

RSS for Train and Test has slightly increased by doubling the Alpha value

```
RSS Score Ridge Train: 21.78233721604314
RSS Score Lasso Test: 16.683684685401246
RSS Score Ridge Train New: 22.121047960583102
RSS Score Ridge Test New: 16.66434825245581


RSS Score Lasso Train: 21.716765961594458
RSS Score Lasso Test: 16.71230422090438
RSS Score Lasso Train New: 22.051970915031227
RSS Score Lasso Test New: 16.72745699930847
```

Coefficients have slightly changed as well

	Beta	Ridge	Lasso	new_ridge	new_lasso
0	TotalBsmtSF	0.859191	0.882212	0.836195	0.880995
1	1stFlrSF	0.402588	0.266127	0.419609	0.192672
2	2ndFlrSF	0.513689	0.412483	0.501143	0.333300
3	GrLivArea	0.626126	0.819832	0.601527	0.919611
4	BedroomAbvGr	-0.554926	-0.579366	-0.507356	-0.543650
5	KitchenAbvGr	-0.629153	-0.648536	-0.555696	-0.573734
6	GarageCars	0.453207	0.451517	0.461661	0.460031
7	MSSubClass_1-STORY 1946 & NEWER ALL STYLES	0.134435	0.130582	0.130048	0.121656
8	MSSubClass_1-STORY PUD (Planned Unit Developme...	0.176002	0.168169	0.177550	0.162958
9	MSSubClass_1-STORY W/FINISHED ATTIC ALL AGES	0.087125	0.057257	0.084529	0.027497
10	MSSubClass_2 FAMILY CONVERSION - ALL STYLES AN...	0.100449	0.099939	0.080057	0.072971
11	MSSubClass_2-STORY 1945 & OLDER	0.096601	0.092743	0.096638	0.088697
12	MSSubClass_2-STORY 1946 & NEWER	0.147784	0.144005	0.150292	0.142599
13	MSSubClass_DUPLEX - ALL STYLES AND AGES	0.066730	0.123452	0.052600	0.090932
14	MSSubClass_PUD - MULTILEVEL - INCL SPLIT LEV/F...	0.092360	0.078535	0.092249	0.066309
15	MSSubClass_SPLIT FOYER	0.156516	0.149615	0.149626	0.135350

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 2

Optimal Value of Alpha for **Ridge: 0.3**

Optimal Value of Alpha for **Lasso: 0.0001**

For optimal values of Alpha, RSS of Ridge is slightly lesser. So I will choose Ridge

	Metrics	Linear Regression	LR with RFE	Ridge	Lasso
0	R2_train	9.641686e-01	8.640582e-01	0.837827	0.839357
1	R2_test	-1.102638e+20	7.516703e-01	0.837827	0.731305
2	MSE_train	5.655146e-03	2.145521e-02	0.021631	0.021566
3	MSE_test	1.726211e+19	3.887671e-02	0.038620	0.038686
4	RSS_Train	5.694732e+00	5.694732e+00	21.782337	21.716766
5	RSS_Test	7.457233e+21	7.457233e+21	16.683685	16.712304

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 3

Top 5 predictor values are

1. TotalBsmtSF
2. GrLivArea
3. KitchenAbvGr
4. MSZoning_FV
5. Functional_Sev

	Beta	Ridge	Lasso
0	TotalBsmtSF	0.859191	0.882212
3	GrLivArea	0.626126	0.819832
5	KitchenAbvGr	-0.629153	-0.648536
17	MSZoning_FV	0.598117	0.633075
29	Functional_Sev	-0.555159	-0.621079

After removing, Top 5 predictor values are

1. 1stFlrSR
2. 2ndFlrSF
3. Exterior1st_BrkComm
4. GarageCars
5. OverallQual_Very Poor

	Beta	New_lasso
0	1stFlrSF	1.276415
1	2ndFlrSF	0.775925
21	Exterior1st_BrkComm	-0.637012
3	GarageCars	0.537319
20	OverallQual_Very Poor	-0.533669

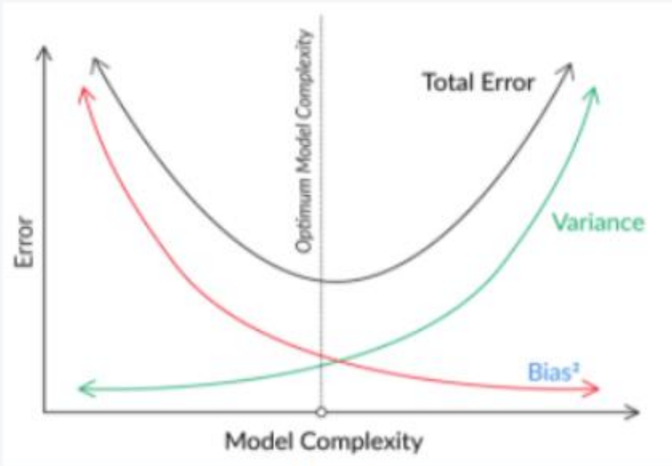
Question 4

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 4

While training a model, we may **Overfit** the model or **Underfit** the model which can lead to inaccurate model. Besides **Multicollinearity** also leads to variability of coefficients.

Regularization helps us with ensuring the model is robust and generalisable. Regularization helps with managing model complexity by essentially shrinking the model coefficient estimates towards 0. This discourages the model from becoming too complex. Below diagram explains this:



Image

We need lowest total error, i.e., low bias and low variance, such that the model identifies all the patterns that it should and is also able to perform well with unseen data. For this, we need to manage model complexity: It should neither be too high, which would lead to overfitting, nor too low, which would lead to a model with high bias (Underfitting) that does not even identify necessary patterns in the data

Implications

When we add penalty and try to get the model parameters that optimise updated cost function (RSS + Penalty), the coefficients that we get given the training data may not be the best in terms of accuracy. Although with this minor compromise in terms of bias, the variance of the model may see a marked reduction. **Essentially, with regularization, we compromise by allowing a little bias for a significant gain in variance**.