Assignment - Advanced Regression Problem Statement - Part II

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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

The optimal value of lambda for ridge and lasso regression.

- Ridge Regression: 1
- Lasso Regression: 50

When the lambda values is doubled the predictors remain the same but the co-efficient lowers for Ridge Regression and the co-efficient reduces to 0 for Lasso Regression.

Ridge Regression:

	Feaure	Coef
2	BsmtFinSF1	208840.897542
5	2ndFlrSF	116368.231204
7	KitchenAbvGr	107882.284914
9	Condition2_PosN	87373.942955
3	TotalBsmtSF	83903.093989
1	OverallQual	68159.204040
4	1stFlrSF	66731.804690
16	Functional_Sev	54862.003410
11	RoofMatl_Metal	54707.315783
14	RoofMatl_WdShake	47108.357875
6	GrLivArea	45945.459872
15	RoofMatl_WdShngl	30863.072017
12	RoofMatl_Roll	21618.145800
13	RoofMatl_Tar&Grv	19171.684823
23	GarageCond_Po	5990.608411
22	GarageCond_Gd	3788.121506
19	GarageQual_Po	3321.893656
24	GarageCond_TA	218.672166
21	GarageCond_Fa	-6955.263011
20	GarageQual_TA	-7843.778242
18	GarageQual_Gd	-19304.209183
17	GarageQual_Fa	-28959.014744
8	GarageArea	-50432.414089
0	LotArea	-71154.141553
10	RoofMatl_CompShg	-76786.010099

Lasso Regression:

	Feaure	Coef
2	BsmtFinSF1	235053.303915
7	KitchenAbvGr	211188.429057
3	TotalBsmtSF	98467.314393
9	Condition2_PosN	81616.994681
1	OverallQual	79568.536732
5	2ndFlrSF	67046.812269
11	RoofMatl_Metal	34878.070751
16	Functional_Sev	30369.832171
14	RoofMatl_WdShake	21086.132785
4	1stFlrSF	2030.883049
19	GarageQual_Po	345.276121
20	GarageQual_TA	-0.000000
12	RoofMatl_Roll	0.000000
21	GarageCond_Fa	-0.000000
17	GarageQual_Fa	-0.000000
23	GarageCond_Po	0.000000
22	GarageCond_Gd	-0.000000
24	GarageCond_TA	-0.000000
15	RoofMatl_WdShngl	0.000000
13	RoofMatl_Tar&Grv	0.000000
6	GrLivArea	0.000000
18	GarageQual_Gd	-10783.029826
8	GarageArea	-47456.733535
0	LotArea	-61332.482864
10	RoofMatl_CompShg	-110245.415956

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:
Lasso Regression since the R2 score for Train is slightly higher than R2 score of Test.
Also Lasso Regression helps in feature reduction since the co-efficient of certain features are 0 which makes better compared with Ridge Regression.

Question 3

After building the model, you realized 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:

Dropped Predictors

- Functional_Sev
- RoofMatl_Metal
- RoofMatl_WdShake
- RoofMatl_WdShngl
- RoofMatl_Roll

Top 5 predictors after dropping above mentioned 5 predictors and re-building Lasso Regression model

- BsmtFinSF1
- KitchenAbvGr
- 2ndFlrSF
- TotalBsmtSF
- OverallQual

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

Model should Handle Outliners

Too much importance should not give to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outliers analysis needs to be done and only those which are relevant to the dataset need to be retained.
Those outliers which it does not make sense to keep must be removed from the dataset.

Model should be more Robust

The model should be generalized so that the test accuracy is not lesser than the training score.
The model should be accurate for datasets other than the ones which were used during training. If the model is not robust, it cannot be trusted for predictive analysis.

Model should be Simple

Regularization can be used to keep the model simple. Regularization helps to strike balance between to keep model simple and not making it too naive to be of any use.