Problem Statement - Part II

Assignment Part-II

Q1:

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?

A:

The optimal value of alpha for ridge and lasso regression Ridge Alpha 1 lasso Alpha 10 If we double the value of alpha then Ridge_new Alpha =2 and lasso_new Alpha = 20 For changes Alpha we can see the beta coefficients of predictors are changed

Ridge	Ridge_new	Lasso	Lasso_new	
LotArea	59778.431939	55922.640992	63955.064210	63617.887669
OverallQual	115599.252408	110944.014490	119957.483345	121719.072148
OverallCond	35638.745398	33226.593469	37354.981812	36948.765235
YearBuilt	54545.692314	54344.573607	53864.332906	53764.548095
BsmtFinSF1	51586.657410	52663.731203	50216.539701	50458.153814
TotalBsmtSF	76674.754264	74096.707724	78348.099735	78209.333502
1stFlrSF	73061.086063	71476.123090	8832.898863	8244.958141
2ndFlrSF	37149.879346	35224.759353	0.000000	0.000000
GrLivArea	87839.676484	85326.415089	163982.920640	162804.680303
BedroomAbvGr	-52962.603870	-44604.715801	-62831.358381	-61134.170375
TotRmsAbvGrd	52937.952456	53633.210113	51280.023696	50757.774874
Street_Pave	49959.412426	40419.432038	63045.460825	59515.001052
LandSlope_Sev	-27846.862924	-21531.677392	-37188.510825	-29661.614776
Condition2_PosN	-11908.785655	-5843.960364	-21920.323877	-11645.855795
RoofStyle_Shed	11641.731102	7274.217976	17801.452620	1966.058339
RoofMatl_Metal	18201.049929	11164.959608	32845.684073	16580.031007
Exterior1st_Stone	-37132.047065	-23655.805061	-69633.615929	-59674.587283
Exterior2nd_CBlock	-32941.699298	-21223.133721	-60463.906721	-49678.514531
ExterQual_Gd	-54900.543840	-51867.902074	-58459.152105	-57016.336034
ExterQual_TA	-62317.508218	-60497.044122	-64902.622534	-63508.829030
BsmtCond_Po	-2488.039788	-4021.786999	0.000000	-0.000000
KitchenQual_TA	-5437.664855	-6282.925595	-4495.491440	-4450.468043

	Ridge	Ridge_new	Lasso	Lasso_new	
Functional_Maj2	-23574.925049	-15094.639225	-40743.007254	-31654.783158	
SaleType_CWD	-27224.575631	-20812.381122	-35460.118834	-3	

Q2:

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?

A:

Ridge Regression : R2 score(Train)- 0.884 , R2 score(Test)- 0.869 Lasso Regression : R2 score(Train)- 0.885 , R2 score(Test)- 0.864

The R2_score of lasso is higher than R2_score of Ridge Regression, so we will choose lasso regression to solve this problem

Q3:

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?

A:

Previously R2score (train) at =0.885 R2score (test) =0.864

Now R2score (train) = 0.798, R2score (test) = 0.758

R2score of training and testing

Q4:

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?

A:

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. Too much importance should not given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outlier's 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. If the model is not robust, it cannot be trusted for predictive analysis