Advanced Regression Assignment – Subjective Question

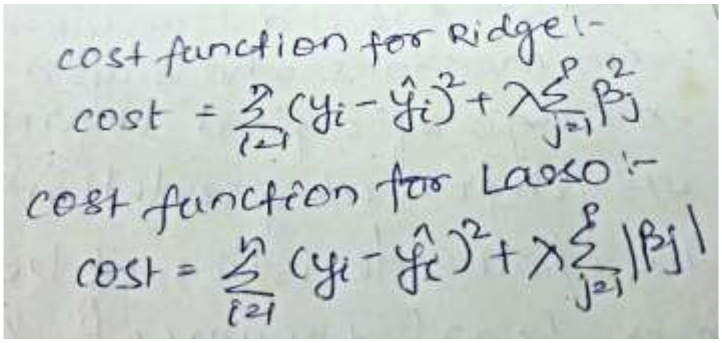
Question 1:

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 obtained optimal value of alpha for ridge and lasso regression is 7.0 and 0.001 respectively. If we double the value of alpha, the new value for ridge and lasso would be 14.0 and 0.002 respectively.

When we use regularization, we add a penalty term to the model's cost function. Here, the cost function would be **Cost = RSS + Penalty**. The formula for cost function: -



cost function for Ridge:-

$$\text{cost} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

cost function for Lasso:-

$$\text{cost} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

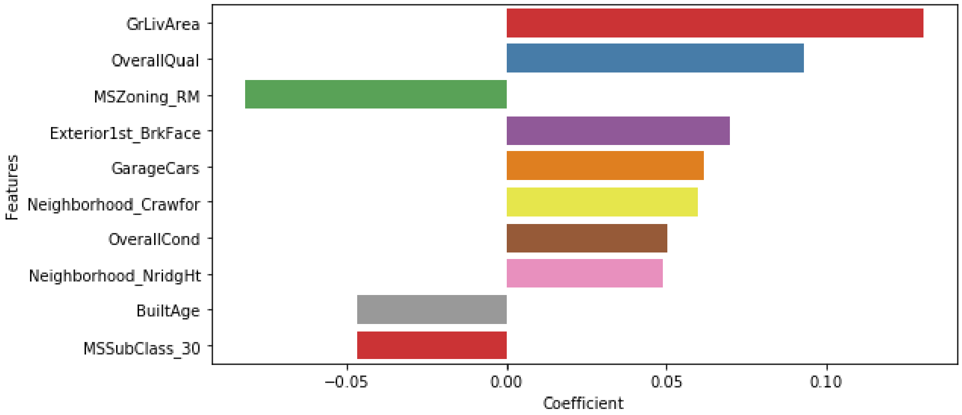
If lambda is 0, then the cost function would not contain the penalty term and there will be no shrinkage of the model coefficients. However, since lambda moves towards higher values, the shrinkage penalty increases, pushing the coefficients further towards 0, which may lead to model **underfitting**.

The changes observed are as follow: -

- If the penalty is very large it means model is less complex, therefore the bias would be high
- In lasso most of the coefficient value became zero, but in case of Ridge, the coefficients became close to zero but not zero. Hence, with the new alpha the coefficient of predictor variables changed accordingly
- The train and test r2 score for both the models reduced slightly. The new r2 score recorded with changed alpha are: -

	Alpha = 7.0	Alpha = 14.0
Ridge R2 (Train)	89.1	88.73
Ridge R2 (Test)	89.6	89.6
	Alpha = 0.001	Alpha = 0.002
Lasso R2 (Train)	88.3	87.2
Lasso R2 (Test)	89.7	89.00

- As per Lasso the most important predictor variables after the change are:



- The change in Old Predictor Variables vs New Predictor Variables: -

Alpha = 0.001

Features	Coefficient
GrLivArea	0.127264
Neighborhood_StoneBr	0.112565
Neighborhood_NridgHt	0.101197
Neighborhood_Crawfor	0.100497
Exterior1st_BrkFace	0.097496
BldgType_Twnhs	-0.086605
OverallQual	0.084726
MSSubClass_30	-0.082701
GarageCars	0.058329
BldgType_TwnhsE	-0.057972
Neighborhood_NoRidge	0.057214

Alpha = 0.002

Features	Coefficient
GrLivArea	0.1302
OverallQual	0.0931
MSZoning_RM	-0.0818
Exterior1st_BrkFace	0.0699
GarageCars	0.0616
Neighborhood_Crawfor	0.0598
OverallCond	0.0503
Neighborhood_NridgHt	0.0489
BuiltAge	-0.0469
MSSubClass_30	-0.0466
Neighborhood_Edwards	-0.0396

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:

The optimal lambda value in case of Ridge and Lasso is 7.0 and 0.001 respectively.

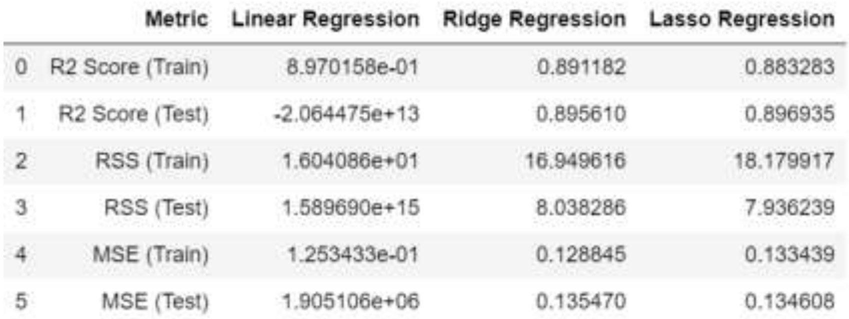
The Mean Squared error in case of Ridge and Lasso are:

- Ridge - 0.018352250695376483 and Lasso - 0.018119267042785627

The R2 score for test data obtained are:

- Ridge - 0.895610 and Lasso - 0.896935

The detailed metric evaluation for ridge and lasso is below: -



	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	8.970158e-01	0.891182	0.883283
1	R2 Score (Test)	-2.064475e+13	0.895610	0.896935
2	RSS (Train)	1.604086e+01	16.949616	18.179917
3	RSS (Test)	1.589690e+15	8.038286	7.936239
4	MSE (Train)	1.253433e-01	0.128845	0.133439
5	MSE (Test)	1.905106e+06	0.135470	0.134608

We built the final model with 70 parameters, chose by RFE, and Lasso helped in feature reduction by making the coefficient of certain features equal to 0. The total number of final predictor variables in lasso turned out to be 37 which is far less than Ridge i.e. 70. Also, the test R2 score, MSE and RSS of Lasso is slightly better than Ridge. Hence, we will opt for Lasso as it has a better edge over Ridge.

Question 3:

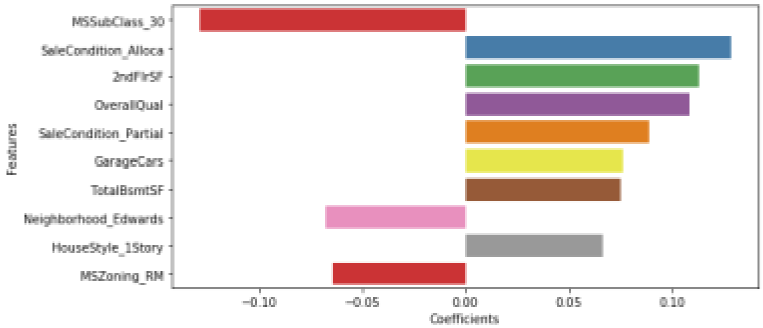
After building the model, you realized that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer:

The steps followed to obtain list of new predictor variables are as follow: -

1. The copy of training and test dataset were created for building new model
2. The top 5 predictor variables were removed from both dataset
3. Executed rfe with 70 in new dataset to obtain new list of features
4. Updated x-train and x-test with new rfe selected cols
5. Built new L1 model with optimal lambda as 0.001
6. Listed out top new predictor variables

Hence, as per the new lasso model the five most important predictor variables are: -



Features	Coefficients
MSSubClass_30	-0.128836
SaleCondition_Alloca	0.128598
2ndFlrSF	0.112997
OverallQual	0.108405
SaleCondition_Partial	0.089003
GarageCars	0.076021

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:

A model needs to be made robust and generalizable so that they are not impacted by outliers in the training data. The model should also be generalizable 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 weightage should not be given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, an outlier analysis needs to be done. Once outliers are detected it is required to treat them accordingly. There are 3 ways to handle them –

1. By completely removing the outliers from the dataset Possible near outliers are identified as observations further than 1.5 x IQR from the quartiles, and possible far outliers as observations further than 3.0 x IQR from the quartiles
2. By performing soft and hard capping within a certain confidence interval. Confidence intervals can be used typically 3-5 standard deviations or Z-score or IQR score

   But the disadvantage of these two methods is that we are restricting the spread of data within a certain range and on that specific range a model is built. Hence, the model will not be able to generalize the pattern beyond the specific range and might not perform well for unseen data. This will degrade the accuracy accordingly.

3. The 3rd and the highly effective way to treat skewness due to extreme outliers is by performing data transformation. There are multiple ways to transform a data and based on the EDA one can choose the required transformation technique

Hence, by doing this the required data can be retained for model building process. This would help to increase the accuracy of the predictions made by the model. It will also help to standardize the predictions made by the model. If the model is not robust, it cannot be trusted for predictive analysis. With a slight change in the spread of data the model will fail drastically resulting into high errors and low accuracy.