- **Subjective Qs Answered – Advanced Linear Regression Assignment**
- **Submitted by Shailesh Kadre, AI ML C46 Oct 2022 Batch**
- **Submitted on: 21st Feb, 2023 • No of pages: 3**

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:

The optimum value of alpha for Ridge and Lasso are 100 and 0.01. The effect of doubling this alpha for Ridge and Lasso regression model will be on the effect of penalty function. With these doubled values of Alpha, the model coefficients will change and model will have a tendency to be on simple side with more bias.

Ridge first 5 predictors with their coefficient values are as follows for the value of alpha: 200

1. ('tot_sq_ft', 0.262)
2. ('GrLivArea', 0.256)
3. ('OverallQual', 0.219)
4. ('RoofMatl_ClyTile', -0.191)
5. ('LotArea', 0.168)

Lasso first 5 predictors with their coefficient values are as follows for the value of alpha: 0.02

1. ('tot_sq_ft', 0.444)
2. ('GrLivArea', 0.442)
3. ('OverallQual', 0.346)
4. ('RoofMatl_ClyTile', -0.253)
5. ('LotArea', 0.205)

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 optimum value of alpha for Ridge and Lasso are 100 and 0.01. I will choose Lasso regression model over Ridge regression model. Following are the main reasons:

1. Lasso punishes the coefficients of high Xi more compared to Ridge
2. Lasso has another big advantage of making zero coefficients to not so important features. While Ridge regression model tends to put some low but finite values of coefficient ofeach un- important features.

So it can be concluded that Lasso regression model results into less number of key predictors with almost same accuracy on training and test data set. So this critical factor lead me to choose Lasso model.

Question 3

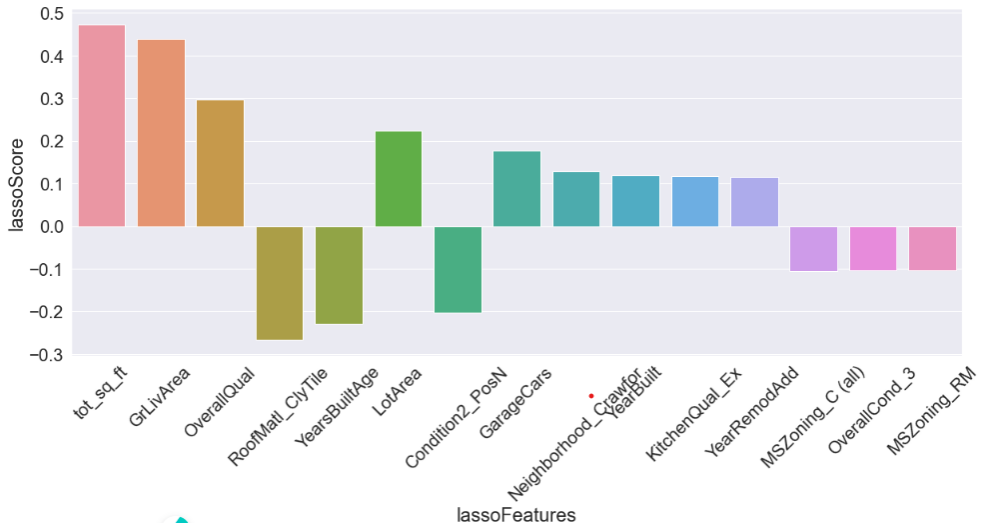
After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer:

Following figure shows the summary of key features from the Lasso regression model

Answer:

Following figure shows the summary of key features from the Lasso regression model



If we don’t have the top five predictor variables, then we have to resort to following next important predictor variables as per above lasso feature plot

1. ('LotArea'),
2. ('Condition2_PosN')
3. ('GarageCars')
4. ('Neighborhood_Crawfor')
5. ('YearBuilt')

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:

By adopting following steps, we make sure that the model is robust and generalisable.

Treatment for Outliers: While building the model, one should not stick very much to outliers. This makes models very specific to that data. So a proper outlier treatment makes regression models more genderized for the new test data and more robust. IQR is one such method followed in current assignment.

Correct Error Metrics- One should use correct error metric to reduce the effect of outliers for evaluating the accuracy of the regression model. Evaluation of linear regression model should be based on mean absolute error instead of mean square sum of error to minimize the effect of outliers.

Feature Transformation – Feature transformation can support in making models robust and generalised. Depending on the data types-logarithmic, exponential and other types of data transformations are adapted for generalization of regression model.

Use of alternate machine learning model: Not all ML models are sensitive to outliers. To name a few such models are K- Means Clustering , Random Forest which are tree based algorithms can be deployed based on the applications.