Advanced Linear Regression Assignment 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 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 Lambda (alpha) for **Ridge = 2.0**, for **Lasso = 0.0001**

After doubling the alpha values the R^2 value for the model increases little bit.

Still 'GrLivArea' (same variable as previous model) is the most significant one.

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:

Alpha value for both Ridge and Lasso Regression is completely different. As mentioned in the above answer.

I will choose **Lasso regression** because the R^2 value for Lasso is better than Ridge regression.

The test and train accuracy for both the regression model is good and not very far away from each other.

Lasso regression is also having the advantage of feature elimination while building the 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 3:

For the current model these are the top five predictors:-

GrLivArea, GarageArea, OverallQual, LotArea, Neighborhood

If these 5 are not available then the next top 5 will be:-

1stFlrSF, OverallCond, YearRemodAdd, MSZoning, GarageFinish

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:

First, Before building and applying the model on any dataset we need to clean and sanitize the dataset i.e. we need to perform EDA and make data consumable by the model.

Second , we can check the difference between R^2 of Train and Test data. If it is significant (very much) in numbers that means that our model is overfitting the Training data and we need to use other techniques like polynomial regression, Non-linear regression or other regression techniques. Also we can use Regularization techniques to improve our model as well.

Third, we need to use out model for some other dataset which have same characteristic and features and if we get acceptable R^2 then that model is robust enough to predict any future data.