

Advanced Regression Assignment

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 : Optimal Value of alpha for ridge and lasso regression are:

- Optimal Value of lambda for Ridge: 10
- Optimal Value of lambda for Lasso: 0.001

If we choose to double the value of alpha for both ridge and lasso:

In case of ridge that will lower the coefficients and in case of Lasso there would be more less important features coefficients turning 0.

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 :

Optimal Value of alpha for ridge and lasso regression:

- Optimal Value of lambda for ridge: 10
- Optimal Value of lambda for Lasso: 0.001

Got Good score for both Lasso and Ridge. Since lasso does feature selection also would go with Lasso.

Ridge: Train :90.9 Test :87.4 and Lasso : Train :89.8 Test :86.4

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 : On running the same notebook and after removing the top 5 significant

variables : Below variables as next 5 significant.

-Lasso

```
('GarageType_BuiltIn', 0.089),  
( 'GarageType_Detchd', 0.094),  
( 'GarageType_No Garage', 0.101),  
( 'GarageType_Others', 0.12),  
( 'GarageFinish_No Garage', 0.195)]
```

-Ridge

```
(('GarageType_BuiltIn', 0.089),  
(('GarageType_Detchd', 0.093),  
(('GarageType_No Garage', 0.096),  
(('GarageType_Others', 0.103),  
(('GarageFinish_No Garage', 0.14)])
```

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 : Here are some changes you can make to your model:

- Tree-based models are generally not as affected by outliers, while regression-based models are. If you're performing a statistical test, try a non-parametric test instead of a parametric one.
- switching from mean squared error to mean absolute difference (or something like Huber Loss) reduces the influence of outliers. I explain a bit about why this is the case at Why is the median a measure of central tendency? It doesn't have anything to do with any other values of the data set, so how does it "describe" the data set?