

Surprise Housing Assignment

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 optimal value of alpha for ridge and lasso regression are:

Ridge: 10

Lasso: 0.001

If we decide to double the value of alpha for Ridge, It will lower the coefficients.

For Lasso, it will create more features that are less important, making coefficients 0.

The most important predictor variable after implementing the change is the significant variable.

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: Lasso regression can be used because it leads to model parameters with zero coefficients for less important features.

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: Next 5 significant variables for Lasso:

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

Next 5 significant variables for Ridge:

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

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:

- Tree-based models are better because they are immune to outliers. Outliers are affected in regression-based models.
- For Statistical test, use a non-parametric test.
- Using absolute difference instead of mean squared error reduces the effect of outliers.
- Transform your data
- You can remove outliers if there are few of them and not worthy of predicting.