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?

Solution :

The optimal value of alpha:

Ridge: 2.0

Lasso: 0.0001

The changes will be as follows, If we double the alpha value :-

1. Ridge: (alpha 4.0)

   R2_score (train) 0.9407

   R2_score (test) 0.8572

The r2_score increases for **train** 0.9277 to 0.9407 and it decreases for **test** - 0.8613 to 0.8572 (alpha = 2.0)

2. Lasso: (alpha – 0.0002)

   R2_score(train) 0.9298

   R2_score(test) 0.8657

The r2_score reduces for **train** - 0.9407 to 0.9298, and r2_score for **test** - 0.8572 increases to 0.8657 (with alpha = 0.0001)

- Most important predictor variable for ridge is :-- **GrLivArea**
- Most important predictor variable from lasso is : **GrLivArea**

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?

Solution:

Considering the alpha/lambda values obtained from the model, opting for Lasso seems more favorable. Lasso effectively shrinks the coefficients of variables, facilitating feature elimination and thereby simplifying the model. This choice is made while acknowledging that the r2_score is similar for both regression models. Additionally, Lasso proves advantageous in addressing issues related to multicollinearity.

**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?

Solution :

*Details on model creation as a part of notebook:*

**Five most important predictor variables after creating another model are:** - > >

| | |
|---|---|
| **OverallQual_9** | (coeff: 0.141) |
| **Neighborhood_Crawfor** | (coeff: 0.136) |
| **CentralAir** | (coeff: 0.110) |
| **OverallQual_8** | (coeff: 0.099) |
| **MSZoning_FV** | (coeff: 0.089) |

**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?

Solution :

Ensuring the robustness and generalizability of a model involves several key steps:

1. Cross-Validation: This process entails dividing the dataset into multiple subsets and training the model on various combinations of these subsets.

2. Regularization: Employing techniques like Lasso and Ridge regression helps prevent overfitting by introducing penalty terms to the loss function. Lasso, in particular, aids in feature elimination, contributing to a simpler model.

3. Feature Scaling: Ensuring that all features are on a similar scale is achieved through feature scaling. Algorithms tend to perform better when numerical input variables share a consistent scale, enhancing learning speed by optimizing algorithm convergence.

4. Hyperparameter Tuning: This step involves adjusting the model's parameters to identify values that yield optimal performance on a validation set.

A robust and generalizable model exhibits higher accuracy on unseen data, having been trained to perform well under diverse conditions without overfitting. Consequently, minor modifications to the test data do not result in a significant change in model accuracy.