

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 optimal value of alpha for ridge and lasso regressions are 0.2 for Ridge and 0.0005 for Lasso. If we choose to double the value of alpha for both ridge and lasso it will produce a "smoother" regression at the expense of a worse fitting of the data, while a small alpha would do the opposite.

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 optimal value of lambda for ridge and lasso regression was determined during the assignment and I would choose Lasso Regression as it creates a good balance between Bias & Variance after it was run on the test data set.

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:

Residential High Density_MSZoning	0.8609
Floating Village Residential_MSZoning	0.856
Residential Low Density_MSZoning	0.8275
Residential Medium Density_MSZoning	0.7215
Excellent_OverallQual	0.5246

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:

1. Dataset must be the same as what's been given
2. EDA – a proper outlier detection for continuous variables. Not every time you have to delete the outliers.
3. Feature Engineering & Scaling
4. The proper measure of train & test split. Fit and transform on Train and fit on Test dataset.
5. Feature selection using multiple methods like RFE or Lasso.
6. K fold for backup test within the train dataset.
7. Using hyperparameter for tuning.
8. Using Open grid CV for accurate hyperparameter selection.