Assignment: Advanced 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 –
The optimal value of alpha for ridge regression is 100
The optimal value of alpha for ridge regression is 0.001

When we double the value of alpha for both ridge and lasso, we can see that coefficients reduce and in a way, start reducing to zero.

The predictor variables will still be the same i.e. according to my model the most important predictor variables are –

- HouseStyle
- HouseStyle
- Exterior2nd
- TotRmsAbvGrd
- GarageFinish
- Neighborhood
- GrLivArea
- ExterQual

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 –
I chose lasso regression. When lambda is small, the result is essentially the least squares estimates. As lambda increases, shrinkage occurs so that variables that are at zero can be thrown away. So, a major advantage of lasso is that it is a combination of both shrinkage and selection of variables.

Question 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 –

- TotRmsAbvGrd
- Functional
- GrLivArea
- ExterQual
- KitchenQual
- Neighborhood

Question 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 –

A model is considered robust and generalizable if it labels variable or the predictor variable is accurate even if any of the input variables or the independent variables are changed drastically. For regression, 'Adjusted R squared' is used for explanation. It tells how well the independent variables selected will explain the variability of the dependent or predictor variable. R-squared is a better metrics compared to RMSE, MSE and MAE.