PART-2

By – Nikhileshwar Reddy B

Q1) What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if we double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Ans)

- The Optimal values for ridge and lasso regression are 3.0 and 0.0001 respectively.
- Now if we double the values the values will be 6.0 and 0.0002 respectively
- Attaching a picture of what happens when we double the alpha value
- There are no major changes in ridge regression where the changes are in decimals for both R2 value and mean squared error value
- There is a drop of 0.01 in R2 value and slight drop in MSE value
- In both cases 1stFlrSF and 2ndFlrSF are the most important variables.

Q2) You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now which will you choose to apply and why?

Ans)

- Lasso Regression would be a better option when compared to ridge regression
- The reason is it would be a better option in feature elimination.
- And it is more robust when compared to ridge
- And even in the algorithm we have there is not much difference between ridge and lasso, but I would prefer Lasso.

Q3) After building the model, you realized that the five most important predictor variables are not available in the incoming data. You will now have to create another model excluding the five most important variables. What are the five most important predictor variables right now?

Ans)

- The 2nd five most important variables are thus:
    1. LotArea
    2. FullBath
    3. Neighborhood_StoneBr
    4. YearRemodAdd
    5. Fireplaces

Q4) 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?

Ans)

- We should always ensure R2 score of tests is not less than train.
- The MSE (mean square error) should be very less for any data.
- Outliers which are way too much are to be removed and if still there are any outliers present, we should use MinMaxscaler ().
- Always model should be as simple as possible.
- It is important to avoid under and over fitting.