Advanced Regression Assignment – Part ||

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?

Ans: Optimal value for alpha –

- Ridge = 2
- Lasso = 0.01

Coefficients will be increase as the value of alpha increases in case of Ridge along with R2 score decrease

For Lasso, with the increase of alpha there will be more feature elimination i.e. more features with coefficients becoming zero along with drop in R2 score

The value of alpha chosen should be such that there is a trade off between bias and variance

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?

Ans: We see that in the model, Ridge gives a better R2 score; however, it is preferred to go with Lasso given that it eliminates features that are noisy and unwanted. If we choose the correct value of alpha, Lasso will give us a much more generalized and less complex model which will perform better with unseen data.

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?

Ans: The most important variables are –

1. GrLivAreaAbove
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. GarageArea

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?

Ans: To make model robust and generalisable 3 features are required:

1. Model accuracy should be > 70-75%: In this case its > 80%

2. P-value of all the features is < 0.05

3. Choose appropriate value of alpha to have the right balance between bias and variance

The above 3 important consideration will make the model simple and robust.