Subjective Assignment – Advanced Regression

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:

1. Optimal value of alpha for Ridge is 50 and for Lasso is 10. When we double these values, the model performance remains same in both the cases.
2. In Ridge the important predictor variables are TotalBsmtSF,GrLivArea , KitchenQual,YearBuilt, OverallQual
3. In Lasso, the important predictor variables are - TotalBsmtSF , 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?

Answer:

Both of them are performing similary, so by ignoring the variables which add noise to the model, we will go with ridge.

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:

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:

Addressing overfitting and making sure the model is simple as possible, would make it more robust and general. The accuracy on the train data would drop because we are **not overfitting it and making it more general.** But Test data accuracy might improve if model is generalized.