

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?

10 and 100 for ridge and lasso respectively. R2 score reduces by 150 to 200 basis points when we double the alpha values. List of important predictor variables do not change significantly

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?

Both have very similar performance. If feature selection is important, then I will use Lasso. Personally, I will use Ridge as it has a very slight edge over Lasso in test R2_score

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?

FullBath, GarageCars, 1stFlrSF, TotRmsAbvGrd, RoofMatl_WdShngl

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?

Using regularization is one way to ensure the model is generalized. Another way is to remove correlated features which may cause the model to make assumptions about underlying data. Feature scaling and normalizing outliers can also help with generalizing. These steps may result in a model with low variance and slightly high bias. This means the accuracy of training data may be slightly lower, but the performance in test data will be consistently better