

SUBJECTIVE QUESTIONS:

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 to 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 of Alpha for Ridge and Lasso is **3** and **0.0001**, respectively. There is slight reduction in Train and test R-Squared values if we choose to double the alpha. Even the most important predictor variables also changed.

The new most important predictor variables are **Overall Quality** even if the **alpha doubled** in case of **Ridge Regression**. But in case of **Lasso Regression**, the most important predictor variable is **Above Ground Living Area Square feet** which is changed from **Overall Quality**.

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: I would choose Ridge Regression over Lasso because the optimal lambda value I got is neither too low nor too high in case of Ridge regression which avoids overfitting or underfitting unlike Lasso where the optimal value I got is too low which might overfit my model.

We can also use Lasso, since most of the beta Co-efficient values are zero which might reduce the complexity of the model by compromising little bit on accuracy.

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 next 5 important predictor variables are

- 1). Second Floor Square Feet
- 2). Pool Area

3). On Full bath, that is Full bathrooms above Grade

4). Overall Conditions

5). On height of the basement weather, it is fair, poor or no basement

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: The model needs to be robust and generalizable so that they are not impacted by null values and outliers. Therefore we need to treat null values by replacing them with mean or mode as per the situation and treat Outliers by capping the minimum and maximum quantiles. If the model developed is giving good accuracy for the different data sets that is not at all used for training, then we can say, the model is robust and generalizable, accurate and ready to use.