**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. The optimum (α) value of ridge and lasso for deriving the best possible model, for the house prediction the optimum value of ridge regression is 10 and for the lasso the alpha value is 0.0001.

For the ridge model with alpha as 10 the r-squared values for test and train data:

R2 score of Training Data: 0.9452686397556161

R2 score of Testing Data: 0.898830963641904

For the lasso model with alpha as 0.0001 the r-squared values for test and train data:

R2 score of Training Data: 0.9544338157467334

R2 score of Testing Data: 0.8921047640675391

Doubling the alpha value doesn’t bring any big difference in r-squared values of priority variables coefficients

**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. From the assignment of house price prediction, there are so many columns were derived from the categorical variables. In ridge regression it is difficult to remove the columns with less priority where as in lasso it can minimize the coefficient to zero with the column is not relevant. Here in the assignment both the models have given almost a similar R-squared values but ridge regression is performing better with just 1% difference. If accuracy is more important then, we have to go for the ridge regression because that gives the less difference between training and test R-squared values

**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 following features will be the top 5 features if the incoming data doesn’t have the first five important features:

1. LotFrontage
2. OverallCondition
3. YearBuilt
4. Condition1\_Norm
5. BsmtExposure\_Gd

**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?

Ans. To know whether the model is robust we can check the difference between train and test results the higher the difference the lesser the robust and at the same time it is generalizable when the model is predicting almost all of them correctly and for the accuracy we can see the R-squared values for the test set and train set. If the difference between the test set and train set values of R- squared will show how accurate our model is.