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:
In our model optimal value of Alpha for Ridge and Lasso Regression are 0.4 for Ridge and 100 for Lasso.

If we double the Alpha value for Both Ridge and Lasso fallowing are the observation .

1. In Ridge regression after increasing the Alpha value from 0.4 to 0.8 r2 score has decreased for train set and slightly increased for test data
2. In Lasso regression after increasing the Alpha values R2score of training & test data has decreased slightly
3. Predictor variables are same but the coefficient are changed a bit

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:

The R2_score , RSS and MSE are better for lasso regression than Ridge , so we will choose Lasso Regression.

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:

In our previous model 'LotArea', 'OverallQual', 'OverallCond', 'YearBuilt', 'BsmtUnfSF' were the top 5 Predators , after dropping them and built a Lasso Model we observed the significant drop in R2 values for both Test and Train data Set.

R2 for Test and train after dropping top 5 Predator variables are
0.8482170044280363
0.8444023878340945

In our previous model R2 values were :
0.8842155622640282
0.8907235903241959

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?

Answer:

If models are built Simpler then it will be more robust but its accuracy will be less , The simpler the model the mode bias it will be but less variance abs become generalisable.

But if our model is very simple it will be underfit , so its important to have a balance in Bias and Variance .

This balance can be achieved with the help of Regularization , Regularization helps in managing the model complexity by shrinking the Co-efficients towards Zero. This avoids the model becoming very complex and reduces the risk of becoming overfitting .

We should have a optimum bias-variance so that it will help in minimizing the Total Error . Model should be Sufficiently simpler to to generalisable and also complex enough to become robust.