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:
- The optimal value of alpha for ridge is 40 and lasso is 900.
- By doubling the optimal value of alpha for ridge test and train accuracy score falls by 1% approx. respectively.
- By doubling the optimal value of alpha for lasso test and train accuracy score falls by less than 1% and 1% approx respectively.

- *Note: Red are the falling ones and green are the rising ones after doubling the optimal value of alpha*
- Before Change
  - LotArea : 4657.836
  - BsmtFinSF1 : 6246.359
  - TotalBsmtSF : 5986.237
  - GrLivArea : 19944.962
  - GarageArea : 5219.217
  - enc_OverallQual : 10609.697
  - enc_OverallCond : 3896.154
  - enc_SaleCondition : 1597.195
- After Doubling value of optimal alpha
  - LotArea : 4084.576
  - BsmtFinSF1 : 6375.205
  - TotalBsmtSF : 4619.111
  - GrLivArea : 18022.586
  - BsmtFullBath : 1260.424
  - Fireplaces : 3982.116
  - GarageArea : 4857.884
  - enc_OverallQual : 11550.633

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:

I will choose to apply lasso model since it drops the features which has less importance and makes the model more simplified unlike 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:

1. GarageArea
2. TotalBsmtSF
3. BsmtFinSF1
4. OverallQual
5. GrLivArea

Next 5 important predictors after dropping five most important predictors:
1. Condition1
2. KitchenAbvGr
3. TotalBsmtSF
4. BsmtUnfSF
5. LotArea

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 :

For a model to be more robust and generalisable the model should not overfit or underfit by keeping low variance and low bias. If variance is high so any change in the data would result in unexpected performance of the model and in case of high bias it would not capture the new patterns in the unseen leading to an underfitting model.

If a model has high variance it would perform very well in training with a very good accuracy but poor on the test data with a very low accuracy whereas in case of high bias it would perform poorly in both the training and test data with a very low accuracy.