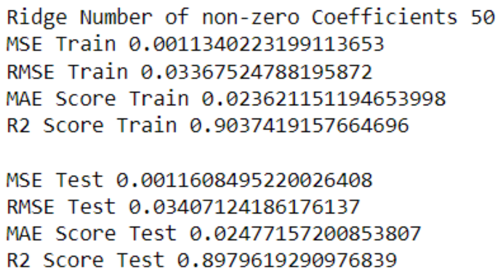
Question 1

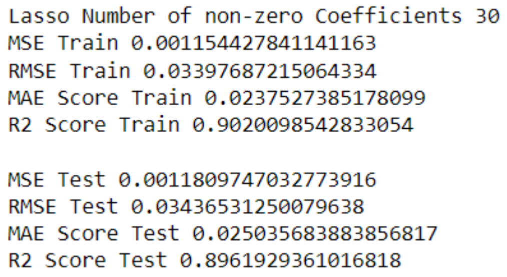
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

- The Optimum alpha Value (regularization parameter) for Ridge regression is 2
- The Optimum alpha Value (regularization parameter) for Lasso regression is 0.0001



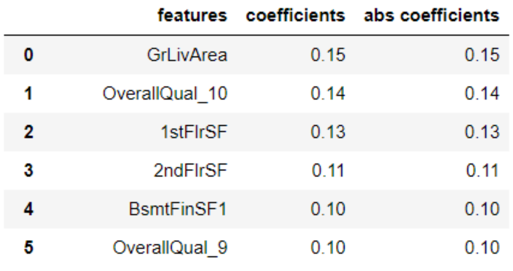
```
Ridge Number of non-zero Coefficients 50
MSE Train 0.0011340223199113653
RMSE Train 0.03367524788195872
MAE Score Train 0.023621151194653998
R2 Score Train 0.9037419157664696

MSE Test 0.0011608495220026408
RMSE Test 0.03407124186176137
MAE Score Test 0.02477157200853807
R2 Score Test 0.8979619290976839
```

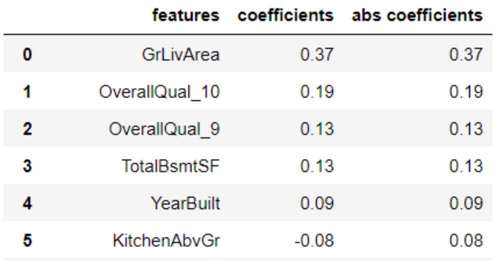


```
Lasso Number of non-zero Coefficients 30
MSE Train 0.001154427841141163
RMSE Train 0.03397687215064334
MAE Score Train 0.0237527385178099
R2 Score Train 0.9020098542833054

MSE Test 0.0011809747032773916
RMSE Test 0.03436531250079638
MAE Score Test 0.025035683883856817
R2 Score Test 0.8961929361016818
```

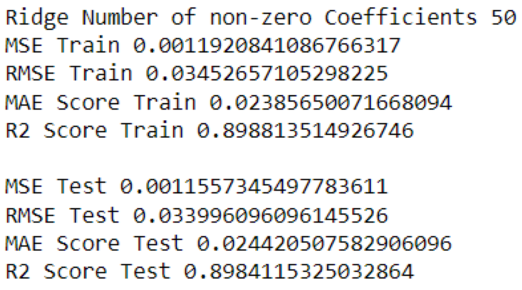


	features	coefficients	abs coefficients
0	GrLivArea	0.15	0.15
1	OverallQual_10	0.14	0.14
2	1stFlrSF	0.13	0.13
3	2ndFlrSF	0.11	0.11
4	BsmtFinSF1	0.10	0.10
5	OverallQual_9	0.10	0.10



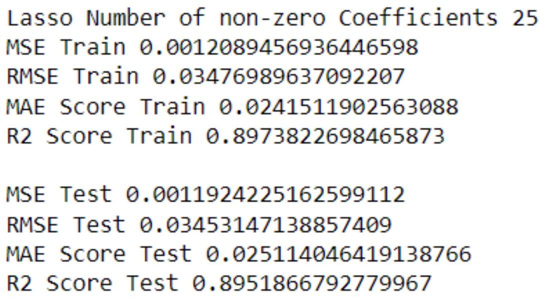
	features	coefficients	abs coefficients
0	GrLivArea	0.37	0.37
1	OverallQual_10	0.19	0.19
2	OverallQual_9	0.13	0.13
3	TotalBsmtSF	0.13	0.13
4	YearBuilt	0.09	0.09
5	KitchenAbvGr	-0.08	0.08

- After increasing the alpha Value to 4 for the Ridge regression
- After increasing the alpha Value to 0.0002 for the Lasso regression



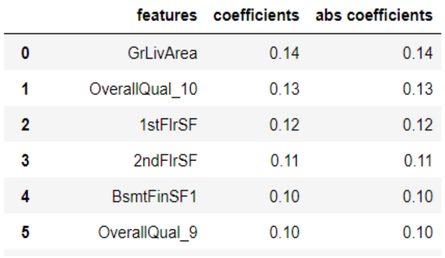
```
Ridge Number of non-zero Coefficients 50
MSE Train 0.0011920841086766317
RMSE Train 0.03452657105298225
MAE Score Train 0.02385650071668094
R2 Score Train 0.898813514926746

MSE Test 0.0011557345497783611
RMSE Test 0.033996096096145526
MAE Score Test 0.024420507582906096
R2 Score Test 0.8984115325032864
```

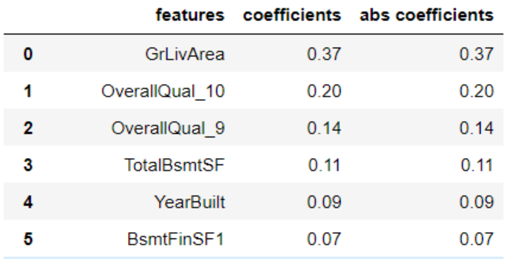


```
Lasso Number of non-zero Coefficients 25
MSE Train 0.0012089456936446598
RMSE Train 0.03476989637092207
MAE Score Train 0.0241511902563088
R2 Score Train 0.8973822698465873

MSE Test 0.0011924225162599112
RMSE Test 0.03453147138857409
MAE Score Test 0.025114046419138766
R2 Score Test 0.8951866792779967
```



	features	coefficients	abs coefficients
0	GrLivArea	0.14	0.14
1	OverallQual_10	0.13	0.13
2	1stFlrSF	0.12	0.12
3	2ndFlrSF	0.11	0.11
4	BsmtFinSF1	0.10	0.10
5	OverallQual_9	0.10	0.10



	features	coefficients	abs coefficients
0	GrLivArea	0.37	0.37
1	OverallQual_10	0.20	0.20
2	OverallQual_9	0.14	0.14
3	TotalBsmtSF	0.11	0.11
4	YearBuilt	0.09	0.09
5	BsmtFinSF1	0.07	0.07

- After doubling the alpha values, a very small value of r2 score decreased on the Train and Test dataset, with small variation in MSE,RMSE,MAE values, but this can be a negligible effect.
- Top five features of Ridge and Lasso regressions do not change but the sixth feature in Lasso regression is changed.
- Few coefficients of top five features are changed.

Question 2

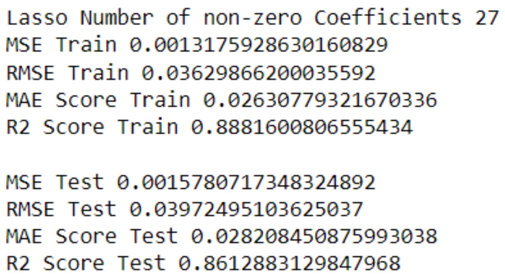
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

- The Optimum alpha Value (regularization parameter) for Ridge regression is 2.
- The Optimum alpha Value (regularization parameter) for Lasso regression is 0.0001.
- Both Ridge and Lasso regressions performed similarly on the Train and Test datasets, however, I would like to take Lasso as the best-performed model because of fewer number of predictor variables.
- 30 predictor variables using Lasso regression can achieve the same result as Ridge regression with 50 predictor variables.
- A model should be as simple as possible and it should be robust, Lasso is performing best on both train and test datasets using fewer predictor variables.

Question 3

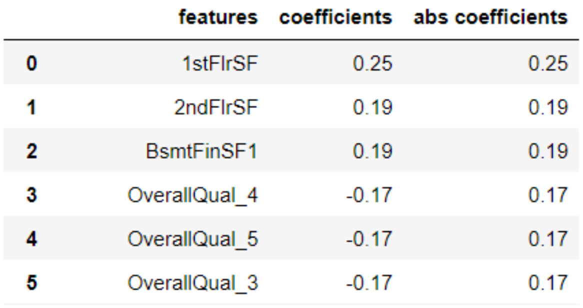
After building the model, you realized that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

- The Optimum alpha Value (regularization parameter) for Lasso regression is 0.0001.
- GrLivArea, OverallQual_10, OverallQual_9, TotalBsmtSF, YearBuilt are the top five features of Lasso regression.
- After removing the top five features, The r2 score decreased from 0.90 to 0.88 on the training dataset and decreased from 0.89 to 0.86 on the test dataset.
- 1stFlrSF, 2ndFlrSF, BsmtFinSF1, OverallQual_4, OverallQual_5 are the top five features of Lasso regression.



```
Lasso Number of non-zero Coefficients 27
MSE Train 0.0013175928630160829
RMSE Train 0.03629866200035592
MAE Score Train 0.02630779321670336
R2 Score Train 0.8881600806555434

MSE Test 0.0015780717348324892
RMSE Test 0.03972495103625037
MAE Score Test 0.028208450875993038
R2 Score Test 0.8612883129847968
```

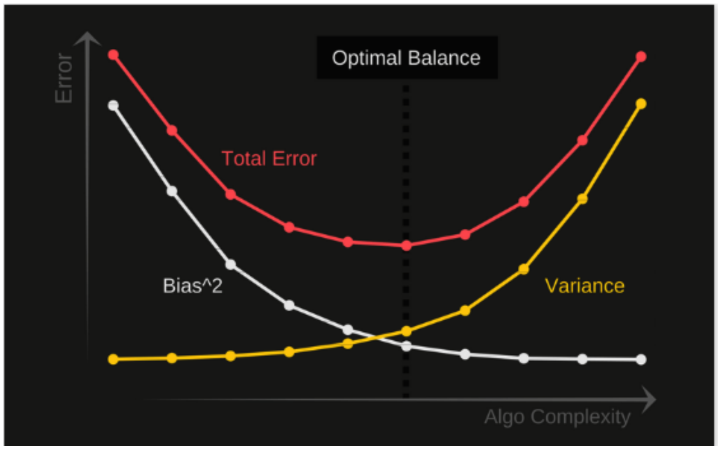


	features	coefficients	abs coefficients
0	1stFlrSF	0.25	0.25
1	2ndFlrSF	0.19	0.19
2	BsmtFinSF1	0.19	0.19
3	OverallQual_4	-0.17	0.17
4	OverallQual_5	-0.17	0.17
5	OverallQual_3	-0.17	0.17

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?

- A model should be as simple as possible and it should be robust.
- Considering bias and variance tradeoff, underfit model has high bias and low variance, and overfit model has high variance and low bias.
- A good model should have low bias and low variance, it should not be complex and it should utilize a minimum number of variables to make predictions.
- A robust model should perform well on both train and test datasets, so test data is very important for analyzing the robustness of the model.
- Regularization helps to overcome overfitting and a robust model performs well on the unseen dataset.
- Sometimes due to the regularization effect the accuracy on the training dataset may decrease a little bit but the model will do well on the unseen dataset.
- So, it is very important to remember that model should not overfit on the train dataset and it should be flexible to make predictions if we make any changes to the training dataset.
- As complexity increases, bias reduces and variance increases, and we aim to find the optimal point where the total error is least.


Error
Optimal Balance
Total Error
Bias^2
Variance
Algo Complexity