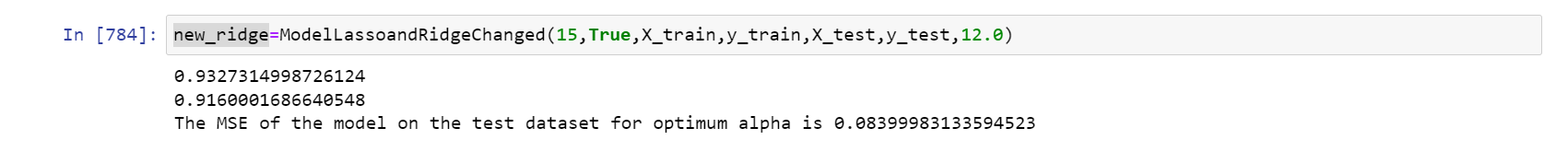
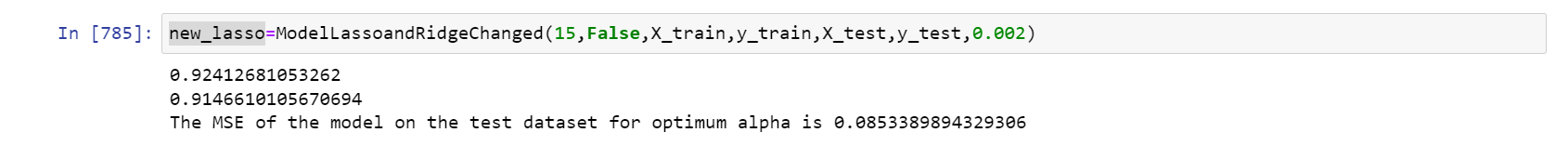
Subjective Questions

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:**



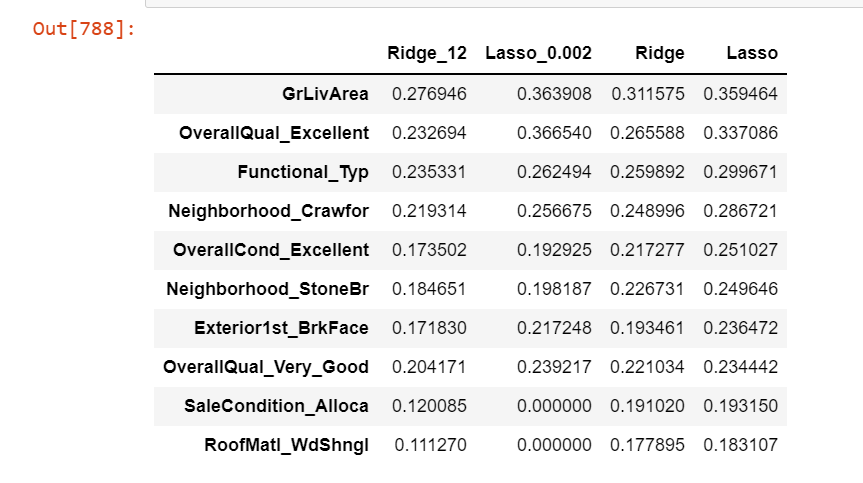


For better reference please check notebook

Here model was trained on Ridge and Lasso for double value .The optimum value obtained was Ridge alpha =6 and Lasso alpha =0.001 .

After doubling (Ridge alpha=12 and Lasso alpha =0.002)

* We can see improvement in R2 test score and reduction of MSE in ridge
* We can see the MSE for Lasso has gone up but very less variation can be seen.But yes increase of MSE indicates model can be better
* We can see in both Ridge and Lasso the R2 scores on train reduced with double of alpha.
* We can clearly see that feature 'SaleCondition\_Alloca' and 'RoofMatl\_WdShngl' is penalised and coeffiecent is very very less .And lasso has removed both of them



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:

**RIDGE METRICS**

* r2 of train 0.9378379406798446
* r2 test 0.914664719097858
* Adjusted r2 of test 0.7984242283554807
* Adjusted r2 of train 0.9174410149654186
* The MSE of the model on the train dataset for optimum alpha is 0.062162059320155366
* The MSE of the model on the test dataset for optimum alpha is 0.08533528090214204

**LASSO METRICS**

* r2 of train 0.9328993288182069
* r2 test 0.9154027790933243
* Adjusted r2 of test 0.800167645750177
* Adjusted r2 of train 0.9108819210866811
* The MSE of the model on the train dataset for optimum alpha is 0.06710067118179312
* The MSE of the model on the test dataset for optimum alpha is 0.08459722090667571

***Lasso is better model***

***Reason***

* r2 score of train and test of Lasso is 1.74 but ridge is 2.32.Hence the difference between the r2 score of train and test should be less
* r2 score for test of Lasso is better than ridge. In regression we consider r2 score test to be better
* MSE on test data for lasso is lower than MSE on test data for ridge.Lower MSE indicates lower error in prediction

1. **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 top predictors were



We remove these and train and create a new model.

X\_train and X\_test dropped these columns shown in notebook and Lasso model with alpha=0.001 was used to build it again.

This time new set of coefficients have come in model.Among the top predictors were

