**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?

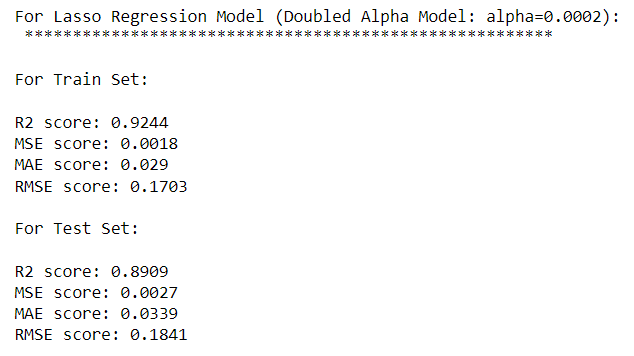
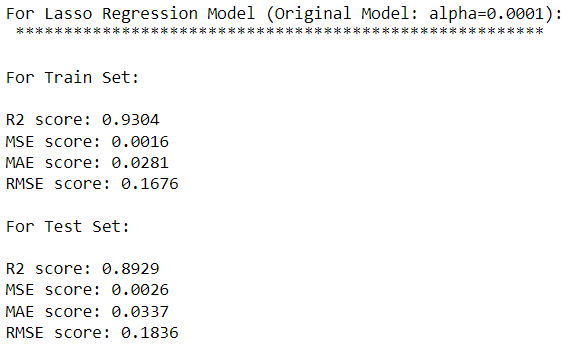
**Answer 1**

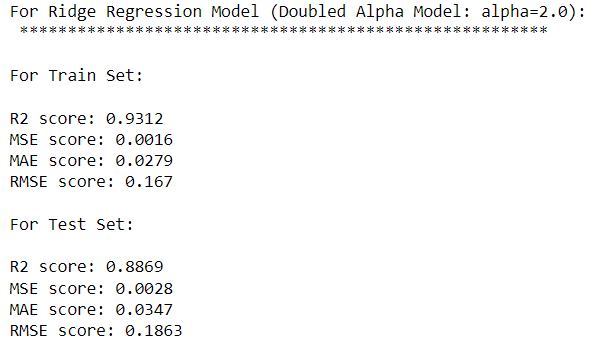
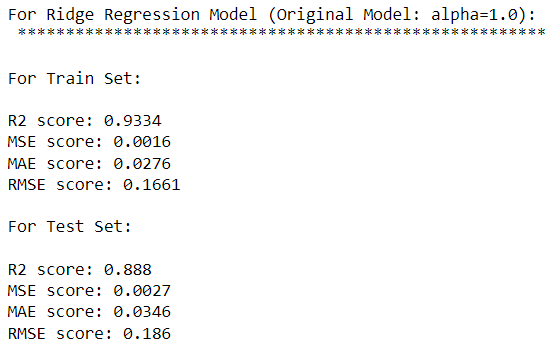
***Part 1*** *– What is the optimal value of alpha for ridge and lasso regression?*

Optimal value of alpha for Ridge = 1.0 and Optimal value of alpha for Lasso = 0.0001

***Part 2*** *– What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?*

The accuracy of train and test scores have reduced slightly and in return the error terms have increased a bit in both the cases of Lasso and Ridge.





***Part 3*** *- What will be the most important predictor variables after the change is implemented?*

Top 10 Predictors for Lasso –

GrLivArea, OverallQual, OverallCond, Age\_Built\_Sold, TotalBsmtSF, LotArea, Neighborhood\_Crawfor, GarageCars, Neighborhood\_Somerst, Neighborhood\_NridgHt

Top 10 Predictors for Ridge –

GrLivArea, OverallQual, OverallCond, TotalBsmtSF, MSZoning\_FV, MSZoning\_RL, MSZoning\_RH, Age\_Built\_Sold, 1stFlrSF, Neighborhood\_StoneBr

(we are considering Lasso predictors over Ridge here as well)

**Question 2**

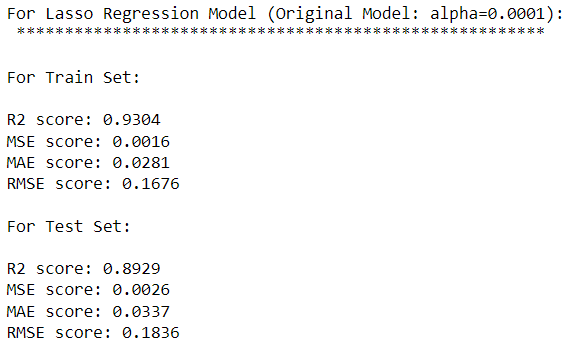
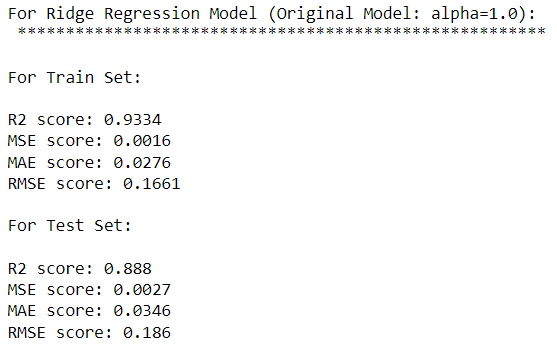
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

**Answer 2**

The test accuracy is better for Lasso and in return the error is less. Also, the train test accuracy difference is less for Lasso.

Lasso helps in feature selection as the non important features coefficients are 0.

Due to above reasons I would select the Lasso model.



**Question 3**

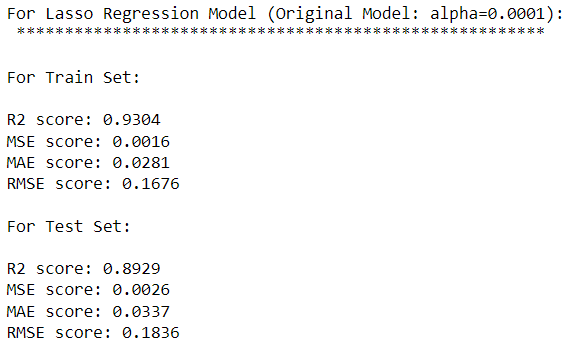
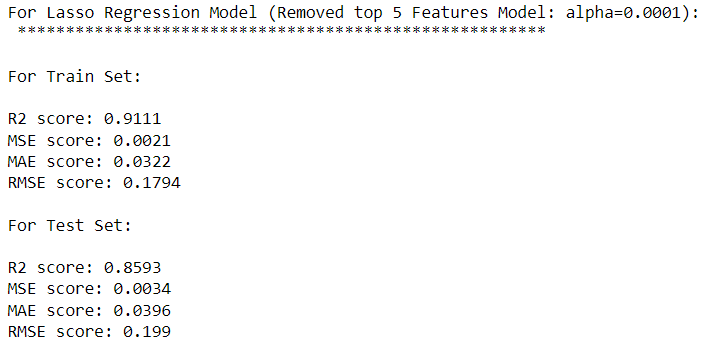
After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

**Answer 3**

The optimal alpha still remains the same i.e. 0.0001. The accuracy for both train and test scores have significantly reduced thus increasing the error’s.

The top 5 predictor variables now are:

1stFlrSF, 2ndFlrSF, Neighborhood\_Somerst, Exterior1st\_BrkComm, Neighborhood\_MeadowV



**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?