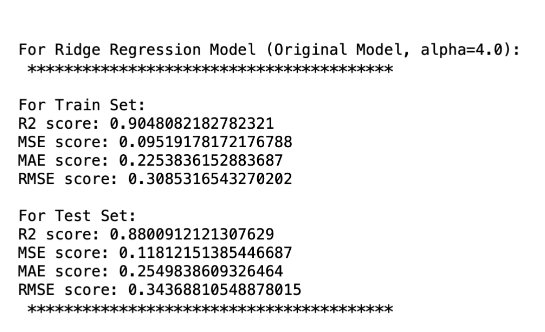
Question 1

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?

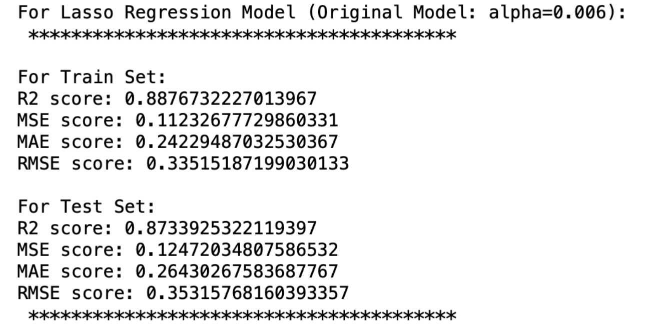
Ans:
Optimal Value of alpha in ridge regression is 4.0
Optimal Value of alpha in lasso regression is 0.006



```
For Ridge Regression Model (Original Model, alpha=4.0):
 ****************************************

For Train Set:
R2 score: 0.9048082182782321
MSE score: 0.09519178172176788
MAE score: 0.2253836152883687
RMSE score: 0.3085316543270202

For Test Set:
R2 score: 0.8800912121307629
MSE score: 0.11812151385446687
MAE score: 0.2549838609326464
RMSE score: 0.34368810548878015
 ****************************************
```

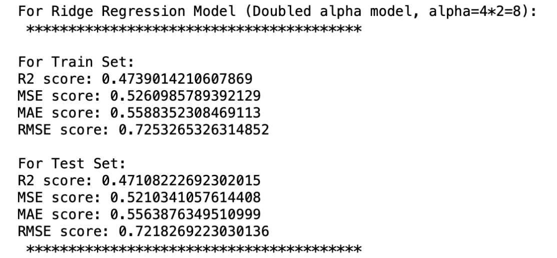


```
For Lasso Regression Model (Original Model: alpha=0.006):
 ******************************************

For Train Set:
R2 score: 0.8876732227013967
MSE score: 0.11232677729860331
MAE score: 0.24229487032530367
RMSE score: 0.33515187199030133

For Test Set:
R2 score: 0.8733925322119397
MSE score: 0.12472034807586532
MAE score: 0.26430267583687767
RMSE score: 0.35315768160393357
 ******************************************
```

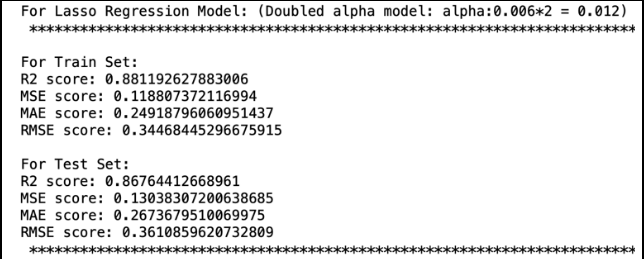
If we double the alpha value



```
For Ridge Regression Model (Doubled alpha model, alpha=4*2=8):
 ****************************************

For Train Set:
R2 score: 0.4739014210607869
MSE score: 0.5260985789392129
MAE score: 0.5588352308469113
RMSE score: 0.7253265326314852

For Test Set:
R2 score: 0.47108222692302015
MSE score: 0.5210341057614408
MAE score: 0.5563876349510999
RMSE score: 0.7218269223030136
 ****************************************
```



```
For Lasso Regression Model: (Doubled alpha model: alpha:0.006*2 = 0.012)
 ************************************************************************

For Train Set:
R2 score: 0.881192627883006
MSE score: 0.118807372116994
MAE score: 0.24918796060951437
RMSE score: 0.34468445296675915

For Test Set:
R2 score: 0.86764412668961
MSE score: 0.13038307200638685
MAE score: 0.2673679510069975
RMSE score: 0.3610859620732809
 ************************************************************************
```

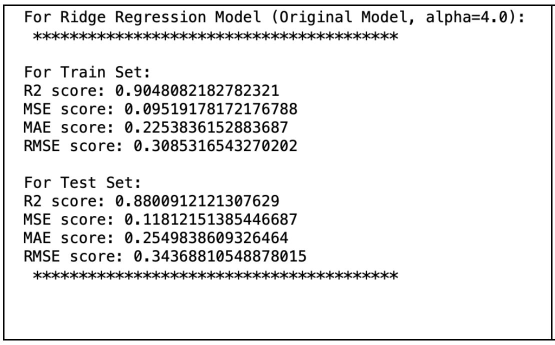
Ridge Model Comparison	**Lasso Model Comparison**
In ridge model, test accuracy there is much difference if we double alpha value .	For lasso regression model , test accuracy is slightly better without double the value
MSE score are slightly low in single alpha value compared to double alpha value	MSE score in both alpha value almost same
Single alpha value seems performing better than double alpha value	Single alpha value seems performing better than double alpha value
Increase in alpha value, decrease R2 score and increase in MSE score .thus single alpha value model is better choice	Increase in alpha value, decrease R2 score and increase in MSE score .thus single alpha value model is better choice

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?

Ans:

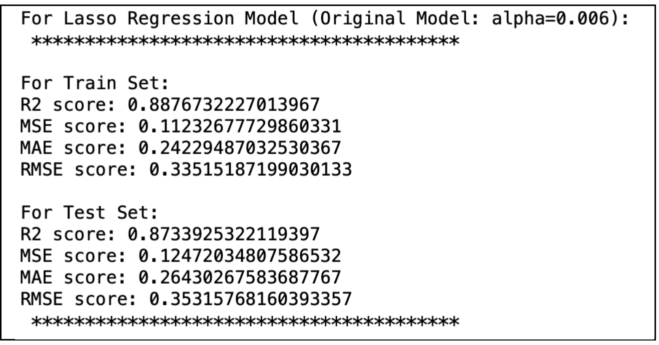
Ridge Model (alpha = 4.0)	Lasso Model(alpha = 0.006)



```
For Ridge Regression Model (Original Model, alpha=4.0):
 ****************************************

For Train Set:
R2 score: 0.9048082182782321
MSE score: 0.09519178172176788
MAE score: 0.2253836152883687
RMSE score: 0.3085316543270202

For Test Set:
R2 score: 0.8800912121307629
MSE score: 0.11812151385446687
MAE score: 0.2549838609326464
RMSE score: 0.34368810548878015
 ****************************************
```



```
For Lasso Regression Model (Original Model: alpha=0.006):
 ****************************************

For Train Set:
R2 score: 0.8876732227013967
MSE score: 0.11232677729860331
MAE score: 0.24229487032530367
RMSE score: 0.33515187199030133

For Test Set:
R2 score: 0.8733925322119397
MSE score: 0.12472034807586532
MAE score: 0.26430267583687767
RMSE score: 0.35315768160393357
 ****************************************
```

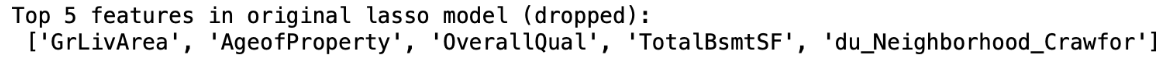
1. R2 score of test data in Ridge model slightly better than R2 score lasso model .But MSE score is lightly high in Lasso model.
2. Lasso model training and testing R2 score has 0.01 difference where as Ridge model has difference 0.03.i.e Lasso model is performing well in unseen data. Though MSE score of training data is slightly decrease in comparison with Ridge model. Since lasso helps in feature selection where some of insignificant variable coefficient will be almost zero, Lasso regression has better over Ridge 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?

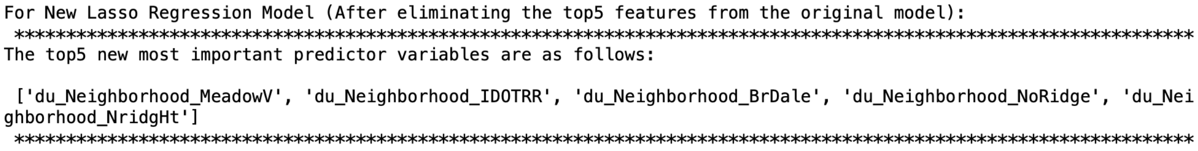
Ans:

Before removing top5 features from model are :



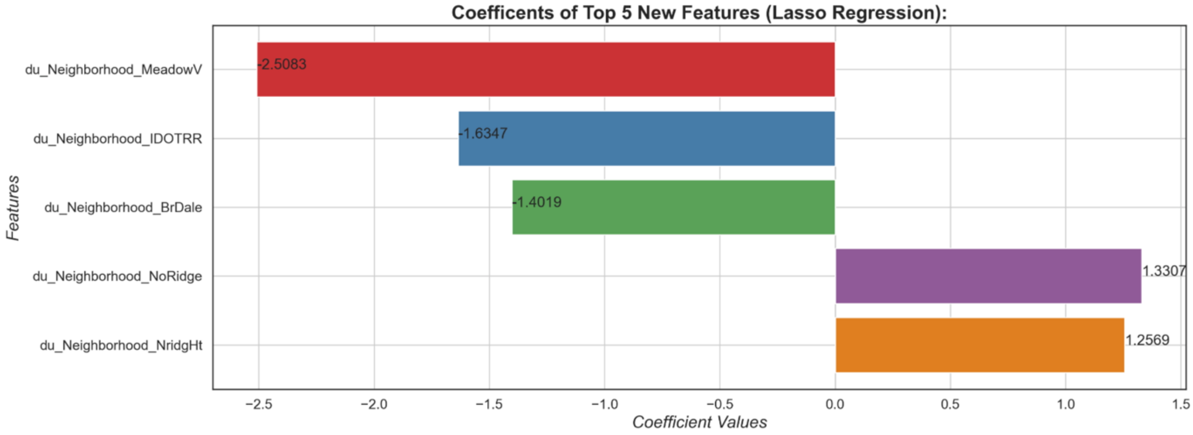
```
 Top 5 features in original lasso model (dropped):
   ['GrLivArea', 'AgeofProperty', 'OverallQual', 'TotalBsmtSF', 'du_Neighborhood_Crawfor']
```

After removing top5 features from model



```
For New Lasso Regression Model (After eliminating the top5 features from the original model):
 ****************************************************************************************************
The top5 new most important predictor variables are as follows:

 ['du_Neighborhood_MeadowV', 'du_Neighborhood_IDOTRR', 'du_Neighborhood_BrDale', 'du_Neighborhood_NoRidge', 'du_Nei
ghborhood_NridgHt']
 ****************************************************************************************************
```



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?

Ans:

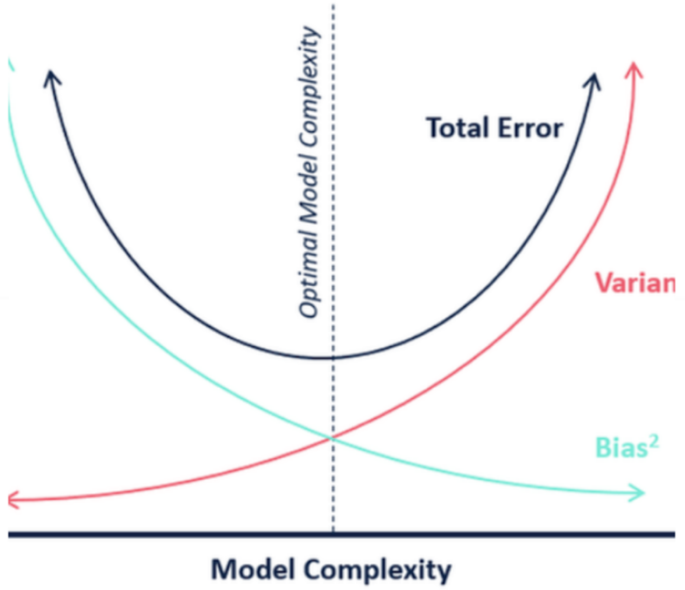
Model robustness refers to the degree that a model's performance changes when using unseen data versus training data. Ideally, statistic measurement should not deviate significantly.

To make model is robust and generalisable, we need to use regularization technique which can control model complexity and bias by penalizing coefficient for making model complex. So, it will allow optimal amount of complexity .

Making model simple leads to bias variance trade off.

1. If model has very less features, then there are possibility to derive pattern then adding new features to model will not consider those variables.
2. If model has so complex then small change the data variable requires remodel the again which is costlier.

Bias helps to know how accurate is likely to be on test data where as complex model can do good prediction provided the sufficient training data is available.
Variance is degree of change in model with respect to change in training data.
So accuracy of the model can be maintained by keeping the balance between Bias-Variance trade off.



As per above figure, we can see, if the bias-variance have the balance, there is very less total error then it will be optimal model complexity.