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: The optimal value of alpha for Ridge is 20 and for Lasso it is 0.004. With these alphas the R2 of the model was approximately 0.83.

After doubling the alpha values in the Ridge and Lasso, the prediction accuracy remains around 0.87 but there is a small change in the co-efficient values. The new model is created and demonstrated in the Jupiter notebook. Below are the changes in the co-efficients.

Ridge Regression Model

```
The R2 Score of the model on the test dataset for doubled alpha is 0.875721597069473
The MSE of the model on the test dataset for doubled alpha is 0.022716441854075516
The most important predictor variables are as follows:
```

Out[113]:

	Ridge Doubled Alpha Co-Efficient
1stFlrSF	0.111950
2ndFlrSF	0.091445
OverallQual	0.077993
OverallCond	0.047482
SaleCondition_Partial	0.032514
LotArea	0.031589
MSZoning_RL	0.029506
BsmtFinSF1	0.027913
SaleCondition_Normal	0.027208
Fireplaces	0.025210
GarageArea	0.025047
Condition1_Norm	0.024611
BsmtQual	0.023195
Neighborhood_Somerst	0.022541
BsmtFullBath	0.017969
FullBath	0.017261
Foundation_PConc	0.017231

Lasso Regression Model

```
The R2 Score of the model on the test dataset for doubled alpha is 0.8763048349500224
The MSE of the model on the test dataset for doubled alpha is 0.022609833713898467
The most important predictor variables are as follows:
```

Out[114]:

	Lasso Doubled Alpha Co-Efficient
1stFlrSF	0.125509
2ndFlrSF	0.105932
OverallQual	0.080077
OverallCond	0.050315
SaleCondition_Partial	0.033873
LotArea	0.030817
SaleCondition_Normal	0.028734
MSZoning_RL	0.028687
BsmtFinSF1	0.027286
Condition1_Norm	0.025008
Fireplaces	0.022895
Neighborhood_Somerst	0.022860
BsmtQual	0.022814
GarageArea	0.022437
BsmtFullBath	0.017345
Foundation_PConc	0.016949
GarageType_Detchd	0.016617

Overall since the alpha values are small, we do not see a huge change in the model after doubling the alpha.

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:

- The optimum lambda value in case of Ridge and Lasso is as follows:-
  - Ridge – 20
  - Lasso – 0.0004
- The Mean Squared Error in case of Ridge and Lasso are:
  - Ridge 0.02288850104034337
  - Lasso 0.022900778195948324
- The Mean Squared Error of both the models are almost same.
- Since Lasso helps in feature reduction (as the coefficient value of some of the features become zero), Lasso has a better edge over Ridge and should be used as the final 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?

Ans:

We build a Lasso model in the Jupiter notebook after removing these attributes from the dataset.
The R2 of the new model without the top 5 predictors drops to .76
The Mean Squared Error increases to 0.028575670906482538

The new Top 5 predictos are:-

	Lasso Co-Efficient
LotFrontage	0.146535
Total_porch_sf	0.072445
HouseStyle_2.5Unf	0.062900
HouseStyle_2.5Fin	0.050487
Neighborhood_Veenker	0.042532

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:

As Per, Occam's Razor— given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:-

- Simpler models are usually more 'generic' and are more widely applicable
- Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
- Simpler models are more robust.
  - Complex models tend to change wildly with changes in the training data set
  - Simple models have low variance, high bias and complex models have low bias, high variance
  - Simpler models make more errors in the training set. Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples

Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use.

Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For

regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.

Also, Making a model simple leads to Bias-Variance Trade-off:
- A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
- A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed.

Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.

Variance refers to the degree of changes in the model itself with respect to changes in the training data.

Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.