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 2 and for Lasso it is 0.0001.
- With above alpha values the R2 of the model was approximately 0.83.
- After doubling the alpha values in the Ridge and Lasso, the prediction accuracy remains around 0.82 - 0.83.
- As we have small change in the co-efficient values, below are the changes in the co-efficients.
- As we have very less alpha values, we don't observe much variation in the co-efficients.

Ridge model:

	Ridge Co-Efficient
Total_sqr_footage	0.171919
GarageArea	0.098997
TotRmsAbvGrd	0.066563
LotArea	0.046725
OverallCond	0.045930
LotFrontage	0.032776
CentralAir_Y	0.030712
Neighborhood_StoneBr	0.027473
Total_porch_sf	0.026740
OpenPorchSF	0.024677
MSSubClass_70	0.024324
Alley_Pave	0.024257
SaleType_Con	0.022986
Neighborhood_Veenker	0.022326
HouseStyle_2.5Unf	0.020355
RoofMatl_WdShngl	0.020207
Condition1_PosN	0.019548
KitchenQual_Ex	0.018904
ExterCond_Ex	0.018745
Condition1_Norm	0.018575

	Ridge Doubled Alpha Co-Efficient
Total_sqr_footage	0.151090
GarageArea	0.089881
TotRmsAbvGrd	0.068068
OverallCond	0.041730
LotArea	0.040630
CentralAir_Y	0.030469
Total_porch_sf	0.029677
LotFrontage	0.027534
Neighborhood_StoneBr	0.025078
OpenPorchSF	0.023767
MSSubClass_70	0.023140
Alley_Pave	0.021636
Neighborhood_Veenker	0.019899
BsmtQual_Ex	0.019847
KitchenQual_Ex	0.019448
MasVnrType_Stone	0.018344
HouseStyle_2.5Unf	0.017626
Condition1_PosN	0.016827
Condition1_Norm	0.016803
PavedDrive_P	0.016416

Lasso model:

	Lasso Co-Efficient
Total_sqr_footage	0.203967
GarageArea	0.107821
TotRmsAbvGrd	0.061172
LotArea	0.047624
OverallCond	0.045982
CentralAir_Y	0.031574
Total_porch_sf	0.025907
Neighborhood_StoneBr	0.022377
OpenPorchSF	0.022141
Alley_Pave	0.020726
LotFrontage	0.018810
MSSubClass_70	0.018789
LandContour_HLS	0.017011
KitchenQual_Ex	0.016782
BsmtQual_Ex	0.016492
Condition1_Norm	0.016184
Neighborhood_Veenker	0.014612
MasVnrType_Stone	0.014475
Condition1_PosN	0.012876
PavedDrive_P	0.011603

	Lasso Doubled Alpha Co-Efficient
Total_sqr_footage	0.205006
GarageArea	0.101267
TotRmsAbvGrd	0.064228
OverallCond	0.042530
CentralAir_Y	0.031833
LotArea	0.029739
Total_porch_sf	0.028033
BsmtQual_Ex	0.017956
OpenPorchSF	0.016946
KitchenQual_Ex	0.016403
Alley_Pave	0.016031
Neighborhood_StoneBr	0.015445
LandContour_HLS	0.014560
MSSubClass_70	0.014421
MasVnrType_Stone	0.013268
Condition1_Norm	0.012901
SaleCondition_Partial	0.010405
LotConfig_CulDSac	0.008337
PavedDrive_Y	0.008100
MasVnrType_BrkFace	0.006483

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 – 2
  - Lasso – 0.0001
- The Mean Squared Error in case of Ridge and Lasso are:
  - Ridge - 0.0017919264106237515
  - Lasso - 0.0018209138725801407
- 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:

The five most important predictor variables in the current lasso model is:

1. Total_sqr_footage

2. GarageArea

3. TotRmsAbvGrd

4. LotArea

5. OverallCond

- After removing above variables R2 of the new model without the top 5 predictors drops to 0.7377266016114529 and MSE increases to 0.0028070626106587587

The new Top 5 predictors are as below:

	Lasso Co-Efficient
LotFrontage	0.130505
Total_porch_sf	0.066769
HouseStyle_2.5Unf	0.059825
HouseStyle_2.5Fin	0.056410
Neighborhood_Veenker	0.042780

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.