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 & lasso? what will be the most important predictor variables after the change is implemented?

Answer-

The optimal value of alpha for ridge is 2 and for lasso it is 0.001. with these alphas the R2 of the model is 0.82.

After doubling the alpha values in the ridge & lasso, the prediction accuracy remains around 0.82 but there is a small change in the coefficient values. The new model is created & demonstrated in the python notebook. Below are the changes in coefficients.

Ridge Regression model-

	Ridge Co-Efficient
Total_sqr_footage	0.169122
GarageArea	0.101585
TotRmsAbvGrd	0.067348
OverallCond	0.047652
LotArea	0.043941
CentralAir_Y	0.032034
LotFrontage	0.031772
Total_porch_sf	0.031639
Neighborhood_StoneBr	0.029093
Alley_Pave	0.024270
OpenPorchSF	0.023148
MSSubClass_70	0.022995
RoofMatl_WdShngl	0.022586
Neighborhood_Veenker	0.022410
SaleType_Con	0.022293
HouseStyle_2.5Unf	0.021873
PavedDrive_P	0.020160
KitchenQual_Ex	0.019378
LandContour_HLS	0.018595
SaleType_Oth	0.018123

	Ridge Doubled Alpha Co-Efficient
Total_sqr_footage	0.149028
GarageArea	0.091803
TotRmsAbvGrd	0.068283
OverallCond	0.043303
LotArea	0.038824
Total_porch_sf	0.033870
CentralAir_Y	0.031832
LotFrontage	0.027526
Neighborhood_StoneBr	0.026581
OpenPorchSF	0.022713
MSSubClass_70	0.022189
Alley_Pave	0.021672
Neighborhood_Veenker	0.020098
BsmtQual_Ex	0.019949
KitchenQual_Ex	0.019787
HouseStyle_2.5Unf	0.018952
MasVnrType_Stone	0.018388
PavedDrive_P	0.017973
RoofMatl_WdShngl	0.017856
PavedDrive_Y	0.016840

Lasso regression model-

	Lasso Co-Efficient
Total_sqr_footage	0.202244
GarageArea	0.110863
TotRmsAbvGrd	0.063161
OverallCond	0.046686
LotArea	0.044597
CentralAir_Y	0.033294
Total_porch_sf	0.028923
Neighborhood_StoneBr	0.023370
Alley_Pave	0.020848
OpenPorchSF	0.020776
MSSubClass_70	0.018898
LandContour_HLS	0.017279
KitchenQual_Ex	0.016795
BsmtQual_Ex	0.016710
Condition1_Norm	0.015551
Neighborhood_Veenker	0.014707
MasVnrType_Stone	0.014389
PavedDrive_P	0.013578
LotFrontage	0.013377
PavedDrive_Y	0.012363

	Lasso Doubled Alpha Co-Efficient
Total_sqr_footage	0.204642
GarageArea	0.103822
TotRmsAbvGrd	0.064902
OverallCond	0.042168
CentralAir_Y	0.033113
Total_porch_sf	0.030659
LotArea	0.025909
BsmtQual_Ex	0.018128
Neighborhood_StoneBr	0.017152
Alley_Pave	0.016628
OpenPorchSF	0.016490
KitchenQual_Ex	0.016359
LandContour_HLS	0.014793
MSSubClass_70	0.014495
MasVnrType_Stone	0.013292
Condition1_Norm	0.012674
BsmtCond_TA	0.011677
SaleCondition_Partial	0.011236
LotConfig_CulDSac	0.008776
PavedDrive_Y	0.008685

Alpha values are small, we did not see the huge change in the model after doubling the alpha value

---

Question 2

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

Answer-

●The optimum lambda value in case of Ridge and lasso is as follows:

Ridge-2

Lasso-0.001

●The mean squared error in case of ridge and lasso are,

Ridge-0.0018396090787924256

Lasso-0.0018634152629407748

●The mean square error of the models is almost same.

●Since lasso helps in feature reduction (as coefficient value of some of the feature becomes zero). Lasso has a better edge over ridge & should be used as the final model.

Question 3

After building the model, you realized 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 variable now?

Answer-

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

1.The_square_footage
2.Garage area
3.TotRmsAbvGrd
4.Overallcond
5.LotArea
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 drop to 0.73
The mean squared error increases to 0.0028575670906482538

The new top five predictors 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 the model is robust & generalizable? What are the implications of the same for the accuracy of the model & why?

Answer-

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 reason—

●Simple models are usually more generic & are more widely applicable

●Simple models require fewer training samples for effective training than the more complex one & are easier to train.

●Simple model is more robust.

- Complex model tends to change widely with changes in the training data set.

- Simple models have low variance, high bias & 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 training samples, fail measurably when applied to other test samples.

Therefore, to make the model more robust & 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 & not making it too naïve 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 lead to Bias- Variance trade off,

● A complete model will need to change for every little change in the dataset & hence is very unstable & 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 widely 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 & makes no discriminations whatsoever has a very large bias as its expected errors across all test inputs are very high.

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

Thus, accuracy of model can be maintained by keeping the balance between bias & variance as it minimizes the total error as shown in the below graph.