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?

- ✓ The optimal value of alpha for Ridge is 2 and for Lasso it is 0.001. With these alphas the R2 of the model was approximately 0.83.
- ✓ On doubling the alpha values in the Ridge and Lasso, the prediction accuracy remains around 0.82 but there is a small change in the co-efficient values.

	Ridge Co-Efficient
Total_sqr_footage	0.107290
GarageArea	0.060016
TotRmsAbvGrd	0.055534
OverallCond	0.031072
SaleType_CWD	0.026793
CentralAir_Y	0.022028
SaleType_ConLD	0.019839
Total_porch_sf	0.018925
Condition2_Norm	0.018785
RoofMatl_WdShngl	0.017527
LotFrontage	0.016641
HouseStyle_2.5Unf	0.016258
KitchenQual_Ex	0.014301
Heating_GasW	0.014298
LotArea	0.013787
PavedDrive_Y	0.013762
SaleType_Con	0.012998
MSSubClass_70	0.012218
ScreenPorch	0.011916
Neighborhood_Veenker	0.011746

	Ridge Doubled Alpha Co-Efficient
Total_sqr_footage	0.107290
GarageArea	0.060016
TotRmsAbvGrd	0.055534
OverallCond	0.031072
SaleType_CWD	0.026793
CentralAir_Y	0.022028
SaleType_ConLD	0.019839
Total_porch_sf	0.018925
Condition2_Norm	0.018785
RoofMatl_WdShngl	0.017527
LotFrontage	0.016641
HouseStyle_2.5Unf	0.016258
KitchenQual_Ex	0.014301
Heating_GasW	0.014298
LotArea	0.013787
PavedDrive_Y	0.013762
SaleType_Con	0.012998
MSSubClass_70	0.012218
ScreenPorch	0.011916
Neighborhood_Veenker	0.011746

	Lasso Co-Efficient
Total_sqr_footage	0.153426
GarageArea	0.062548
TotRmsAbvGrd	0.047457
OverallCond	0.028820
CentralAir_Y	0.019930
Total_porch_sf	0.017330
KitchenQual_Ex	0.011972
BsmtQual_Ex	0.010626
PavedDrive_Y	0.009753
LandContour_HLS	0.009594
MSSubClass_70	0.009294
BsmtCond_TA	0.008538
Condition1_Norm	0.008249
ScreenPorch	0.007934
Condition2_Norm	0.007129
OpenPorchSF	0.006706
ExterQual_Ex	0.005733
MasVnrType_Stone	0.005724
SaleCondition_Partial	0.005654
HouseStyle_2.5Unf	0.005517

	Lasso Doubled Alpha Co-Efficient
Total_sqr_footage	0.132475
GarageArea	0.058255
TotRmsAbvGrd	0.049791
CentralAir_Y	0.019524
Total_porch_sf	0.017527
OverallCond	0.014271
BsmtQual_Ex	0.012965
KitchenQual_Ex	0.012207
PavedDrive_Y	0.008620
BsmtCond_TA	0.007456
Condition1_Norm	0.006065
MasVnrType_Stone	0.003999
MasVnrType_BrkFace	0.003705
SaleCondition_Partial	0.002879
MSSubClass_70	0.002578
ExterCond_TA	0.001414
HouseStyle_2Story	0.000245
RoofStyle_Hip	0.000147
Exterior2nd_Wd Shng	-0.000000
Foundation_Wood	-0.000000

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?

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 - 0.0018396090787924262 •

Lasso - 0.0018634152629407766

▪ 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?

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

1. Total_sqr_footage 2. GarageArea 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 drops to 0.73 The Mean Squared Error increases to 0.0028575670906482538

The new Top 5 predictors

	Lasso Co-Efficient
LotArea	0.061965
Total_porch_sf	0.040386
LotFrontage	0.033480
KitchenQual_Ex	0.024378
BsmtQual_Ex	0.022575

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?

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.

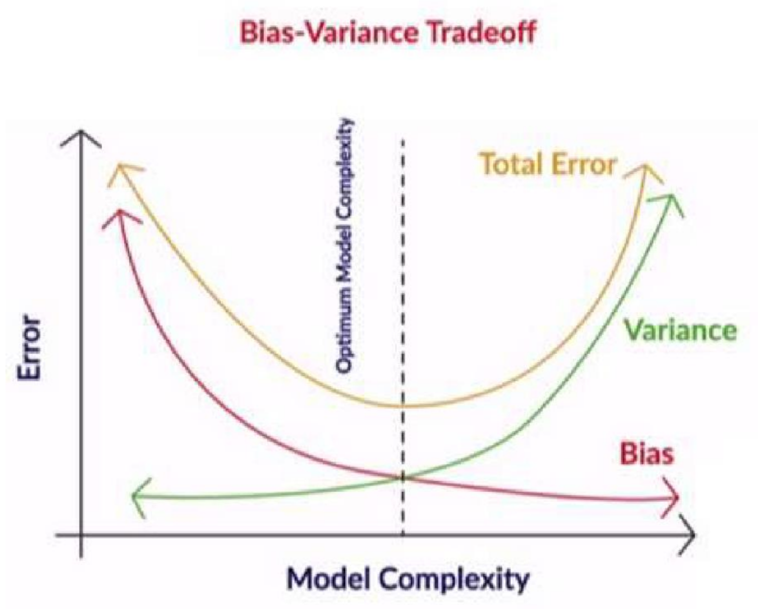
▪ Simpler models are more robust. o Complex models tend to change wildly with changes in the training data set o Simple models have low variance, high bias and complex models have low bias, high variance o Simpler models make more errors in the training set. Complex models lead to overfitting — they 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.


Bias-Variance Tradeoff
Optimum Model Complexity
Total Error
Variance
Bias
Error
Model Complexity