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: RIDGE REGRESSION: Plotting of curve between Negative Mean Absolute Error and Alpha we see that the value of Alpha increases from 0 and the error term decreases. When the alpha value is 2 the test error term is minimum. Hence we consider Alpha as 2 for the Ridge Regression.

LASSO REGRESSION: We have to keep the value of alpha as small as possible as high value will make most of the coefficient to zero. Initially it came 0.3

When we double the value of alpha in ridge regression it will apply more penalty and we will get more error in train and test .

When we double the value of alpha in lasso regression it will penalize the model decreasing the R2 value.

The most important variables after implementation of changes are :

- Gr Liv area
- Overall quality
- 1st flr and basement sq ft
- Property age

Property age will play an important role as the property grows older the sales price decreases.

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: Regularization and increase of prediction accuracy is very important.

In Ridge regression, the lambda is square of magnitude of coefficient which is identified by cross validation. The penalty is lambda times the square of coefficient. So when we increase the value of lambda the variance is dropped and bias is constant. This regression model includes all variables in final model which is not seen in Lasso regression.

In Lasso regression, the lambda is the absolute value of magnitude of coefficient which is also identified by cross validation. Increase of lambda value will shrink the coefficient to zero. And when lambda is small it performs as a linear regression.

Using both regression models we have to built a simple but robust 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 variables now?

Ans: The five most important predictor variables which are excluded are ;

- GrLivArea
- OverallQual
- 1stFlrSF
- TotalBsmtSF
- PropAge

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: The model should be as simple as possible . It can be explained via BIAS VARIANCE Trade off

The more simple the model more bias it will be and less variance and generalisable which will run equal on train and test data

Bias: More the bias in the error in model it will perform poor on training and test data

Variance: Higher variance performs very good on training data but very poor on test data.

It is important to have balance between Bias and Variance to avoid over fitting and under fitting of data.