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?

1. The optimal value of alpha for Ridge Regression is 10.0 and for Lasso is .001 .

2. When I am doubling the alpha value, technically the regularization effect will decrease and hence the model may become less generic and our scoring parameters should reduce.

For Ridge :

1. The r2_score on the train and test dataset decreases a bit.
   Before : 0.92350 and 0.88956
   After : 0.9198 and 0.8895
2. The beta coefficient for some important features decreases.
3. The mean squared error increases.
   Before : 0.11582 and 0.076499
   After : 0.11584 and 0.8015

For Lasso :

1. The r2_score on the train and test dataset decreases a bit.
   Before : 0.92300 and 0.89129
   After : 0.91700 and 0.89011
2. The beta coefficient for some important features decreases.
3. The mean squared error somewhat remains the same. .
   Before : 0.114011 and 0.07699
   After : 0.11525 and 0.08299

3.

a. Most important features after the alpha change in ridge-

'OverallQual', 'GrLivArea', ‘MSSubClass_30.0’, 'LotArea', ' BsmtFinSF1', 'MSZoning_FV', ‘'MSZoning_RL', Neighborhood_Somerst, Neighborhood_NridgHt

b. Most important features in lasso after doubling alpha are -

'GrLivArea', 'OverallQual', Neighborhood_Edwards, 'OverallCond', MSSub_30.0, 'LotArea', Neighborhood_Somerst

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 optimal lambda value in case of Ridge and Lasso is as below:
Ridge - 10 Lasso - 0.0001

The Mean Squared error in case of Ridge and Lasso are:
Ridge - 0.11582 Lasso - 0.114011 for training.

The r2_score on the train and test dataset in Ridge.
0.9235 and 0.8895

The r2_score on the train and test dataset in Lasso.
0.92300 and 0.8912

The Mean Squared Error of Lasso slightly is less than that of Ridge. Also, since Lasso helps in feature reduction (as the coefficient value of some of the features have become 0), Lasso has a better edge over Ridge even though r2 for Lasso is slightly less than Ridge.
Therefore, the variables predicted by Lasso can be applied to choose significant variables for predicting the price of a house.

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?

Top 5 features after building the first time Lasso models are -
OverallQual, Neighborhood_Somerst, Neighborhood_Edwards, GrLivArea, MSSubClass_30.0

Now after removing these from dataset and building the new lasso model, top 5 are-
Neighborhood_NridgHt, MSZoning_FV, MSZoning_RL, Neighborhood_StoneBr, MSZoning_RH
Also, alpha changes to .004 from .001

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:-

1. Simpler models are usually more 'generic' and are more widely applicable.
2. Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
3. Simpler models are more robust.
    a. Complex models tend to change wildly with changes in the training data set.
    b. Simple models have low variance, high bias and complex models have low bias, high variance.

4. 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. 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.
b. 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 the model is 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.