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: For Ridge and Lasso when we doubled the value of alpha
model started loosing accuracy in predicting the SalePrice because it starts underfitting. The same can be observed in the Jupiter notebook file provided.

Ridge measuements (previous)

```
0.9037696078319818
0.8903282379209618
506026393360.95593
263913642424.4703
647921118.2598667
785457269.1204474
```

Ridge measurements(after this doubling)

```
0.8965394217363346
0.8860667178448332
544046242505.08923
274168545457.26385
696602103.0794996
815977813.8609043
```

Lasso measurements(previous)

```
0.8909835449783725
0.8939741490087668
573261755550.9138
255140138134.0503
734009930.2828602
759345649.208483
```

Lasso measurements(after this doubling)

```
0.8714429477831335
0.8831173205226723
676015757691.5417
281265962107.5724
865577154.5346246
837101077.7011083
```

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: Generally we prefer to apply the lambda which gives better prediction meaning more accurate(which means model is much generalizable) But simpler because if we go with any other values we may face either underfit or overfit problems.

Based on this assignment, regarding regularization model, we can definitely go with lasso as it has penalized many coefficents and improved accuracy.

Because we always wanted to get simpler model as per Occam Razor's definition. Need to go for simpler and generalizable model always.

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: These are most important predictor variables in the same order for the current model which was built ->

MSSubClass, LotFrontage, LotArea, OverallQual, OverallCond

If we need to discard these features we may incur some predicting power but since we don’t have a choice, can consider next set of important varaibles

YearBuilt, YearRemodAdd, MasVnrArea, BsmtFinSF1, BsmtFinSF2

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: We always need to make our model simple enough but not simpler as per Occam Razor's simplified theory on model building.

Because if we make more simpler, it may fail to predict the dependent variable means Bias will be high (It fails to predict in training set itself) or if we make more complex also there is another problem of overfitting.

which will cause model fail in test set miserably.