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:

Optimal value of alpha for ridge and lasso regression are

Ridge, Alpha : 100

```
R-Square value for Train dataset:  0.8902726109073386
R-Square value for Test dataset:  0.8768554645234838
```

Lasso, Alpha: 0.001

```
R-Square value for Train dataset:  0.8920304982202351
R-Square value for Test dataset:  0.8725563102964339
```

After doubling Alpha value:

Ridge, Alpha = 200

```
R-Square value for Train dataset:  0.8880723098606176
R-Square value for Test dataset:  0.8755126366228568
```

Lasso: 0.002

```
R-Square value for Train dataset:  0.8869453265324618
R-Square value for Test dataset:  0.8783393628989038
```

Below are the most important variables after change implemented:

LotArea, 1stFlrSF, 2ndFlrSF, GarageArea, BsmtFinSF1

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:

As many variables available, Ridge fails to do feature selection. I always prefer Lasso to eliminate features which are not required.

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:

1stFlrSF, Exterior1st, GarageArea, TotalBsmtSF, WoodDeckSF

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:

Main criteria for a model to be robust is there should not be huge difference in Training and test R-square values. Ensure there should not be outliers and correlated variables in the training data.