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?

Answer 1.

Optimal value of alpha Ridge = 50, lasso = 0.001.

After doubling the alpha the model gets simpler as the coefficients decrease, and might lead to increased bias and reduced variance.

Most important predictors after doubling alpha for ridge and lasso :

Ridge:

```
OverallQual         0.070518
GrLivArea           0.048927
GarageCars          0.043805
OverallCond         0.040632
YearBuilt          -0.039890
2ndFlrSF            0.033105
FullBath            0.031517
YearRemodAdd       -0.030049
1stFlrSF            0.029202
Condition1_Norm     0.029017
```

Lasso:

```
GrLivArea               0.116041
OverallQual             0.080168
GarageCars              0.071765
MSZoning_RM            -0.058499
Neighborhood_Crawfor    0.055171
YearBuilt              -0.054691
OverallCond             0.044424
BsmtFinType1_Unf       -0.043341
Condition1_Norm         0.040963
SaleType_New            0.037317
```

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?

Answer 2

Lasso = 0.002

```
Lasso   train - 0.9091420152176427
        Test - 0.8664160580362419
```

```
        Test - 0.8618922914226035
```

Ridge= 100

```
Ridge train - 0.908105941828082
        Test -  0.8678702952437074
```

Lasso is better to use because it acts as a feature selection.

It's more robust.

And it's a simpler 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?

Answer 3

```
Train - 0.8690337761369153
Test - 0.8198175106217247
```

Most important predictor variables in the lasso model, after model excluding the five most important predictor variables.

```
2ndFlrSF          0.108785
CentralAir_Y      0.100989
1stFlrSF          0.095346
Functional_Typ    0.091539
GarageCars        0.088995
```

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?

Answer 4

- Model should be as simple as possible, as they require less training examples.
- But it should not be too simple such that it under fits the training data.
- Complex models have a tendency to over fit i.e works good on training data but not on test data, as it memorizes the dataset.
- Regularisation is the solution to this problem, as it puts penalty on higher valued parameter.
- Ridge (L2 norm), lasso (L1 norm)