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             0.072398
GarageCars              0.043039
GrLivArea               0.042976
YearBuilt              -0.040634
OverallCond             0.039261
Condition1_Norm         0.036494
Neighborhood_Crawfor    0.034462
Neighborhood_Edwards   -0.032454
FullBath                0.031004
Fireplaces              0.030687
dtype: float64
```

Lasso:

```
GrLivArea               0.099013
OverallQual             0.083057
Neighborhood_Crawfor    0.069001
GarageCars              0.066603
YearBuilt              -0.058408
MSZoning_RM            -0.057251
Condition1_Norm         0.050784
OverallCond             0.042701
Neighborhood_Edwards   -0.041901
Neighborhood_NridgHt    0.036433
dtype: float64
```

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 - 0.9005964036406708
Test - 0.9009092792340212
```

Ridge= 100

Ridge

```
train - 0.8982613829352001
Test - 0.9030651456275124
```

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 - 0.8553128091344547
Test -0.8591334553805214
```

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
```

Answer 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?

- 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)