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

Alpha for Ridge = 2.0 and Lasso =0.0001

When alpha is doubled Ridge = 4.0, r2 score of test data changes from 0.8969 to 0.8965

When alpha is doubled Lasso = 0.0002, r2 score of test data changes from 0.8986 to 0.9011

After change is implemented, the most important predictor variable is Total _Home_Quality as marked by eli5 metric

```
import eli5
eli5.show_weights(lasso_new,
```

y top features

Weight?	Feature
+0.194	Total_Home_Quality
+0.178	Total_House_SF
+0.120	1stFlrSF
+0.110	LotArea
+0.079	YearBuilt
+0.077	2ndFlrSF
+0.068	GrLivArea
+0.045	GarageArea
+0.038	Total_Bathrooms
+0.035	SaleCondition_Partial
+0.035	OverallQual
+0.033	BsmtFinSF1
+0.033	Neighborhood_Crawfor
+0.029	Functional_Typ
+0.028	Neighborhood_StoneBr
+0.026	Exterior1st_BrkFace
+0.023	CentralAir_Y
+0.022	WoodDeckSF
+0.018	YearRemodAdd

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

I would choose Lasso, as it performs Feature Selections and eliminates features by making the coefficients equal to 0, therefore making the model less complex and also reducing the number of features from 287 to 144

```
In [70]: betas[betas['Lasso']!=0].shape
Out[70]: (144, 1)
```

```
In [71]: betas[betas['Lasso']!=0]
```

Out[71]:

	Lasso
LotArea	0.121230
OverallQual	0.016780
YearBuilt	0.097676
YearRemodAdd	0.015975
BsmtFinSF1	0.029392
BsmtFinSF2	0.000207
TotalBsmtSF	0.010462
1stFlrSF	0.122522
2ndFlrSF	0.083666
GrLivArea	0.090201
BsmtFullBath	0.002662

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?

Answer 3

After removing the top five features which were predicted before by Lasso Regression, the new features after creating a new model of Lasso Regression with 282 features instead of 287, will be different and are GrLivArea, OverallQual, TotalBsmtSF, OverallCond, GarageArea

```
In [92]: import eli5
         eli5.show_weights(lasso_new_q3, top=-1,
```

Out[92]: y top features

Weight[?]	Feature
+0.313	GrLivArea
+0.132	OverallQual
+0.110	TotalBsmtSF
+0.080	OverallCond
+0.065	<BIAS>
+0.049	GarageArea
+0.041	SaleCondition_Partial
+0.035	Neighborhood_StoneBr

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

A model could be more robust and generalizable when there are when the cost function or RSS is not too high, or when the dependent variable is not changing a lot. A robust model is when the model is not over fitting or underfitting the data. Ideally has right variance and bias and therefore has less model complexity.