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
The optimal value for Ridge and Lasso transformation is
Ridge : 20
Lasso : 0.001
When we double the alpha the ranking of most important predictors changes
Also we tend to loose biasness and get more variance

After the change most important predictor
Will be

Ridge : Overall Quality
Lasso: GrLivArea

Top 5 Ridge

	Fields	Coeff
0	constant	11.7157
71	Neighborhood_Crawfor	0.0816
4	OverallQual	0.0811
81	Neighborhood_NridgHt	0.0646
23	GrLivArea	0.0605
91	Condition1_Norm	0.0528

Bottom 5 Ridge

47	HouseAge	-0.0357
107	BldgType_Twnhs	-0.0378
101	Condition2_PosN	-0.0507
74	Neighborhood_IDOTRR	-0.0549
72	Neighborhood_Edwards	-0.0643

When alpha is doubled
Top 5 Ridge

0	constant	11.8029
4	OverallQual	0.0802
71	Neighborhood_Crawfor	0.0574
23	GrLivArea	0.0558
91	Condition1_Norm	0.0441
81	Neighborhood_NridgHt	0.0437

107	BldgType_Twnhs	-0.0268
101	Condition2_PosN	-0.0273
47	HouseAge	-0.0342
74	Neighborhood_IDOTRR	-0.0382
72	Neighborhood_Edwards	-0.0509

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:
R2 scores for

Ridge
```
Ridge Regression r2_train:0.9119 , r2_test:0.8888
```

Lasso
Lasso Regression r2_train:0.9059 , r2_test:0.887

No significant change in Lasso and Ridge I would chose **Lasso** as it gives me feature reduction as well

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

	Fields	Coeff	Abs
0	constant	11.8171	11.8171
101	Condition2_PosN	-0.1584	0.1584
23	GrLivArea	0.1111	0.1111
71	Neighborhood_Crawfor	0.1080	0.1080
86	Neighborhood_Somerst	0.0987	0.0987
4	OverallQual	0.0851	0.0851

The Five most important variables are

Condition2_PosN, GrLivArea, Neighborhood_Crawfor, Neighborhood_Somerst, OverallQual

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?

To make model simple and robust we must try to see that we are not overfitting the model

A simplistic model may be low on bias and high on variance whereas a complex model may be high on bias and low on variance . We need to balance the two

To draw an anology we need to make them meet like the demand and supply curve