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?

The optimal lambda value in case of Ridge and Lasso is as below:

Ridge - 10
Lasso – 0.0004

With above set of values, the Mean Squared Error for Rige and Lasso are
- Ridge - 0.013743
- Lasso - 0.013556

Ater doubling the lambda values, the Ridge and Lasso Mean Squared Error values are
- Ridge - 0.013937
- Lasso – 0.013591

Before doubling the Lambda values for Ridge and Lasso, the most important predicator variables were
- Ridge – MSZoning_RL – Coefficient:0.0877
- Lasso – MSZoning_RL – Coefficient:0.109840

After doubling the Lambda values, the most important predicator variables were
- Ridge – GrLivArea – Coefficient:0.0662
- Lasso – GrLivArea – Coefficient:0.099772

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?

I would choose Lasso over Ridge for the given problem. The Mean Square Error of Lasso is slightly lower than that of Ridge. Also Lasso helps in reducing the features.

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?

The most importan 10 predictor variables and their coefficients are given below

	Features	rfe_support	rfe_ranking	Coefficient
11	MSZoning_RL	True	1	0.109840
5	GrLivArea	True	1	0.107541
12	MSZoning_RM	True	1	0.076511
9	MSZoning_FV	True	1	0.070579
1	OverallQual	True	1	0.069655
4	TotalBsmtSF	True	1	0.046112
2	OverallCond	True	1	0.044889
14	Foundation_PConc	True	1	0.043031
7	GarageCars	True	1	0.036338
3	BsmtFinSF1	True	1	0.033366

If the most important five predictors are removed, then the last five predictor varaibles in the above table can be used for the new model.

4	TotalBsmtSF	True	1	0.046112
2	OverallCond	True	1	0.044889
14	Foundation_PConc	True	1	0.043031
7	GarageCars	True	1	0.036338
3	BsmtFinSF1	True	1	0.033366

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?

A model can be made robust and generlisable by doing the outlier
analysis and retaining only the relevant data sets. The outliers which
are not making sense for the entire set of data should be removed to
increase the accuracy of the model.