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?

Ans: The optimal value for alpha for ridge and lasso regression is found at 20 and 0.001 respectively.

Ridge Regression:

R2 Score for Training data: 0.91
R2 Score for Test data: 0.89

Top 10 features:

	Feature	Coefficient
2	OverallQual	0.084
148	Neighborhood_Crawfor	0.082
158	Neighborhood_NridgHt	0.070
149	Neighborhood_Edwards	-0.066
41	Exterior1st_BrkFace	0.060
58	BldgType_Twnhs	-0.054
91	Condition1_Norm	0.054
151	Neighborhood_IDOTRR	-0.052
125	CentralAir_Y	0.052
121	BsmtFinType1_Unf	-0.052

Lasso Regression:

R2 Score for Training data: 0.9
R2 Score for Test data: 0.89

Top 10 features:

	Feature	Coefficient
24	MSZoning_RH	0.428
23	MSZoning_FV	0.405
25	MSZoning_RL	0.389
26	MSZoning_RM	0.371
43	Neighborhood_MeadowV	-0.322
40	Neighborhood_Edwards	-0.224
49	Neighborhood_OldTown	-0.188
42	Neighborhood_IDOTRR	-0.180
161	LotShape_IR3	-0.165
44	Neighborhood_Mitchel	-0.149

After doubling alpha:

Ridge Regression:

R2 Score for Training data: 0.9
R2 Score for Test data: 0.89

Top 10 features:

	Feature	Coefficient
2	OverallQual	0.087
43	Neighborhood_Crawfor	0.058
44	Neighborhood_Edwards	-0.052
53	Neighborhood_NridgHt	0.051
3	OverallCond	0.046
113	Condition1_Norm	0.046
19	GarageCars	0.045
188	CentralAir_Y	0.044
11	GrLivArea	0.043
37	BsmtFinType1_Unf	-0.043

Lasso Regression:

R2 Score for Training data: 0.89
R2 Score for Test data: 0.88

Top 10 features:

	Feature	Coefficient
2	OverallQual	0.107
8	GrLivArea	0.077
25	Neighborhood_Crawfor	0.068
27	Neighborhood_NridgHt	0.064
51	CentralAir_Y	0.061
28	Neighborhood_Somerst	0.059
16	GarageCars	0.053
38	Condition1_Norm	0.052
26	Neighborhood_Edwards	-0.051
21	MSZoning_RL	0.047

Doubling of alpha for both Ridge and Lasso decreases the model accuracy. But there is no significant decrease in model accuracy. But there is a significant change in model coefficient values and their relative ranking importance.

Most important predictor variables after the change in Lasso regression is OverallQual, GrLivArea, Neighborhood_Crawfor, Neighborhood_NridgHt, CentralAir_Y, Neighborhood_Somerst, GarageCars, Condition1_Norm, Neighborhood_Edwards, MSZoning_RL

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: Both the models has decent r-square values on train data and test data.

Ridge regression uses a tuning parameter called lambda, which helps to regularise the model. Regularisaion can significantly rescue the reduce the model variance while not increasing bias much. Here we estimate the model coefficients by minimising the cost function which has a penalty term which is equal to lambda multiplied by sum of squared model coefficients.

But disadvantage with Ridge regression is it retains all the variables that are present in the data. When the number of features are very high as in our case, as data can have unrelated and noisy variables, we may not need to keep those variables in the model. Lasso regression helps here by keeping the sum of absolute values of model coefficients instead of squared model coefficients in penalty term of cost function.

This helps Lasso regression to push some of model coefficients to exactly 0 and to perform model selection. So its always better to apply lasso regression because of this added advantage with lasso regression for feature selection.

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?

Ans:

Top 5 predictors of Lasso model are

	Feature	Coefficient
37	Neighborhood_Crawfor	0.111
42	Neighborhood_NridgHt	0.105
2	OverallQual	0.095
9	GrLivArea	0.085
80	BldgType_Twnhs	-0.079

After dropping these 5 predictors from dataset, new top 5 predictors are

	Feature	Coefficient
33	MSZoning_RH	0.428
32	MSZoning_FV	0.405
34	MSZoning_RL	0.389
35	MSZoning_RM	0.371
63	Neighborhood_MeadowV	-0.322

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?

Ans: As per the Occam's razor, when in dilemma between two models, one should always choose simpler model. The model should be simple as necessary but not simpler than that.

It is due to the following reasons:

- Simpler models requires less data to train than that of complex models, which makes simpler models easy to train.
- Complex models tends to change widely with changes in training data set.
- Whereas Simple models are more robust and doesn't change significantly which changes in training data.
- Simpler models are usually more generic and can be widely applicable.
- Although training error on simpler models can be high, it won't be like complex models where training error is very low and testing error would be very high and fails miserably on testing data.

Hence keeping this characteristics in mind, we build a model which is more robust and generalisable.

Regularisation can be used to make the model simpler. It helps to achieve the correct balance between keeping the model simple and not too naive. In case of regression models, regularisation is a process of adding a penalty term to the cost function which adds ups the absolute values of model coefficients or sum of squares of model coefficients to the cost function.

Also, making the model simple may lead to bias-variacne trade off.

- Simple models will have low variance and high bias where as complex models has high variance and low bias.
- Simpler models make more errors in training set while complex models lead towards overfitting. Complex models works well in training but fail miselrablyin testing.
- So perfect model should not have either high bias or high variance, we should find a point where both where there is a balance between bias and variance as it minimised the total error as shown in the below figure, so that model generated will be more robust and generalised.