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?

Solution:

The optimal alpha value chosen for Ridge and Lasso is as below:

- Ridge - 8
- Lasso - 0.001

If we change the value of alpha to its double, the magnitude of the coefficients decreases, and it would attain values tending to zero but not absolute zero in case of Ridge.
If we double the value of alpha for Lasso, the coefficients of most of the features would reduce to absolute zeroes. This is because lasso selects the only some feature while reduces the coefficients of others to zero to do feature selection. The most important predictor variables would remain the same in both.
I have built a model with double the alpha rate for both Lasso and Ridge. One important observation is that the number of total features selected by lasso decreased from 130 to 109.

RIDGE

alpha = 16

8	OverallQual	0.3140
83	Neighborhood_Crawfor	0.2901
27	GrLivArea	0.2660
138	Exterior1st_BrkFace	0.2584
203	SaleCondition_Partial	0.2499
202	SaleCondition_Normal	0.2382
103	Condition1_Norm	0.2221
99	Neighborhood_StoneBr	0.2204
196	SaleType_New	0.2007

alpha =8

83	Neighborhood_Crawfor	0.3614
138	Exterior1st_BrkFace	0.3237
99	Neighborhood_StoneBr	0.3206
8	OverallQual	0.3097
203	SaleCondition_Partial	0.3093
202	SaleCondition_Normal	0.2722
27	GrLivArea	0.2691
103	Condition1_Norm	0.2550
170	Foundation_Slab	0.2328

LASSO

alpha = 0.002

203	SaleCondition_Partial	0.5407
27	GrLivArea	0.4625
83	Neighborhood_Crawfor	0.4442
99	Neighborhood_StoneBr	0.4184
138	Exterior1st_BrkFace	0.3913
8	OverallQual	0.3228
202	SaleCondition_Normal	0.2880
170	Foundation_Slab	0.2631
103	Condition1_Norm	0.2564

alpha = 0.001

	Features	Coefficient
119	SaleCondition_Partial	0.5812
78	Neighborhood_StoneBr	0.5319
64	Neighborhood_Crawfor	0.4910
23	GrLivArea	0.4576
91	Exterior1st_BrkFace	0.4018
104	Foundation_Slab	0.3941
114	SaleType_ConLD	0.3373
106	Heating_GasA	0.3292
7	OverallQual	0.3072
113	SaleType_CWD	0.3067

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?

Solution:

I am going with lasso regression because house dataset has large number of columns. Ridge regression uses all the features in set making the model very complex. It can cause over fitting, which is seen here by the drop in R2 score from 93% on train data to 83% on test data. In case of lasso, the 215 feature are cut short into 130 by reducing the coefficient to 0. These makes a less complex model and reducing chances of overfitting.

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?

Solution:

After creating a model with lasso, again selected features using RFE. The top 5 features are

0	Neighborhood_StoneBr	True	1	0.531890
1	Neighborhood_Crawfor	True	1	0.490971
2	GrLivArea	True	1	0.457607
3	Heating_GasA	True	1	0.329250
4	OverallQual	True	1	0.307214
5	MSZoning_FV	True	1	0.289408

After removing them, Top features include:

Neighborhood_NridgHt	True	1	0.243340
MSZoning_RL	True	1	0.144828
Neighborhood_Veenker	True	1	0.133904
Heating_Wall	True	1	0.081140
Electrical_FuseF	True	1	-0.016538
MSSubClass_30	True	1	-0.280152

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?

Occam's Razor states that given two models that show similar performance in the finite training or test data, we should pick the simple one because

1. Simpler models are usually more 'generic' and are more widely applicable
2. Simpler models require less training samples for effective training than the more complex ones which makes them easier to train
3. Simpler models are more robust. Complex models are very sensitive to change in dataset but has low bias but Simple models have low variance and high bias
4. Even though simple models make more errors in the training set, Complex models lead to overfitting

Therefore to make the model more robust and generalizable, make the model simple but not naïve.

Regularization helps to make the model simpler. It strives to find the delicate balance between simple model and naïve, useless models. Regularization can be done by adding a regularization term to the cost function which can be either one that adds up the absolute values or the squares of the parameters of the model.

Simple model can lead to Bias-Variance Trade-off:

1. A complex model has low bias and high variance
2. A simpler model has high bias and low variance

Bias quantifies how accurate is the model predicts on test data. Variance refers to the degree of changes or sensitivity in the model with respect to change in the training data. A complex model can do an accurate job prediction provided there is enough training data but simple models that are too naïve has a very large bias and expected error for any test inputs are very high.

We can conclude that the robustness of a model can be attained by finding a balance between Bias and Variance.