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?

Response:

Optimum value of Ridge and Lasso Regression are

1. Ridge Regression: .25

2. Lasso Regression: 100

If we choose to double the value of Ridge and Lasso Regression, then there is no significant difference in the accuracy.

Modified Ridge and Lasso Regression

. Ridge Regression: .50

2. Lasso Regression: 200

	Ridge Regression Alpha value = 0.25	Ridge Regression Alpha value = 0.5	Lasso Regression Alpha value = 100	Lasso Regression Alpha value = 200
Train Data Score	0.6884	0.6883	0.6871	0.6849
Test Data Score	0.6943	0.6947	0.6965	0.6967

Answer 1:

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?

Response: Comparing the K-fold test scope for the optimum values for Ridge (alpha = 0.25, mean test score = 0.662884363) and Lasso regression(alpha=100, mean test score=0.665487002736963), Lasso Regression is more suited.

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?

Response:

After removing the 5 most important predictor variables(

1. MSSubClass_1-STORY W/FINISHED ATTIC ALL AGES
2. MSSubClass_2 FAMILY CONVERSION - ALL STYLES AND AGES
3. MSSubClass_2-1/2 STORY ALL AGES
4. MSSubClass_2-STORY 1945 OLDER
5. MSSubClass_2-STORY 1946 & NEWER

from the data, the five most important variables are

1. LotConfig_Corner - 9633.253
2. LotConfig_Inside - 12493.689
3. LotConfig_Others - 17306.339
4. Neighborhood_Blmngtn - 20431.731
5. Neighborhood_Blueste - 49735.189

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?

Techniques for making the models robust and generalizable

1. Cleanup of data using EDA techniques.
2. Eliminating variables using RFE.
3. Identifying muti-collinearity within attributes and eliminating them.
4. Ensuing homoscedasticity of errors.
5. Using Ridge and Lasso regression in lieu of Linear regression.