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 of alpha for ridge regression is 10

And the optimal value of alpha for lasso regression is 100.

If we choose double the value of alpha for both ridge and lasso then the smoothness of the constraint would be much higher.

The alpha value need not be small but for a larger value the flexibility would be much difficult.

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?

Answer: Lasso regression would be a better option to choose and apply as the model will help in feature elimination.

Lasso regression will produce many solutions to the same problem.

It also has built-in 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?

Answer:

	Featuere	Coef
0	MSSubClass	150859.635253
112	Exterior1st_AsphShn	39053.034464
65	Neighborhood_NridgHt	35028.492536
66	Neighborhood_OldTown	30217.175477
72	Neighborhood_Timber	27211.565810

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?

Answer: A model is considered to be robust if its output dependent variable is consistently accurate even if one or more of the independent variables are changed .

The model should also be generalizable such that the test accuracy is not less than the training score.

The model should be accurate for datasets rather the one used during training.

Outlier analysis need to be done and only those relevant to the dataset must be retained.

This helps in increasing the accuracy of the model.