Subjective Questions

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?

In Lasso regression, when we increase the value of alpha, the model is penalized more and it tries to get the coefficient value close to 0.

So by making alpha double, we will put more penalty on model and r2 value will also decrease

In ridge regression, when the value of alpha increases then the train error also increases but error term decreases.

If we double the value, then the penalty will also increase. This will in turn try to make the graph more generalized.

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 - Though the model performance by Ridge Regression was better in terms of R2 values of Train and Test, it is better to use Lasso, since it brings and assigns a zero va

lue to insignificant features, enabling us to choose

the predictive variables.

It is always advisable to use simple yet robust model.

Question 3

After building the model, you realized 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?

	Variable	Coeff
0	constant	220544.137
1	MSSubClass	-1448.227
2	LotFrontage	-2916.105
3	LotArea	7830.499
4	OverallQual	11197.794

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Having the model generalizable means that it should not be over fitted and performs equally good on training an test set both

This might decrease the accuracy but in return make the mode much simpler, robust and generalizable

Generalizability is measured by more bias and less variance which also means, the model is simple.

Bias can be defined as error in model. So when bias is high, the model fails to learn all data points and performs poor on m=both training and test set.

On the other hand, high variance means that the model learns all data points in the training set and performs really well on training set but when it is run on test set, it performs really poor as compared to former.