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?

Optimal values :

Lasso : .01

Ridge : 2

Most important predictor variables after doubling the alpha are

Neighborhood_Crawfor – location Crawford

MSZoning_FV - Floating Village Residential

MSZoning_RL - Residential Low Density

Neighborhood_StoneBr - location Stone Brook

SaleCondition_Partial :- Home was not completed when last assessed (associated with New Homes)

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 :

Ridge uses lambda to penalize the square of coefficient to bring RSS to small value.Variance gets low and bias stays same as we increase lambda,It does not drop the variables.

Lazzo uses lambda to penalize the absolute values of coefficient to bring RSS to small value.Increase in lambda causes drop in coefficient and ultimately to zero. Small lambda model behaves like Linear regression. Variables with 0 value are dropped by model.

It is solely on result we go which regression mode to select. Better R2 value match between train and test result and RSME value

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?

Five most important predictor variable are [which can be excluded]

- GrLivArea
- OverallQual
- OverallQual
- OverallCond
- TotalBsmtSF

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?

To make a model robust and generalizable, it should be simple model. Bias-variance trade off is the best way to analyze the model. Simple model will have more bias, less variance and generaibizable .Accuracy result will be same on train and test data.