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?

Ans 1

Optimal value of lambda for Ridge Regression = 10

Optimal value of lambda for Lasso = 0.001

If we choose double the value of alpha in both ridge and lasso model R2 score will decrease by a small bit

Most important predictor variable are:-

MSZoning_RM

Exterior1st_BrkFace

MSZoning_FV

MSZoning_RL

StoneBr 8. GrLivArea

SaleCondition_Normal

Neighborhood_Crawfor

SaleCondition_Partial

MSZoning_RH

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?

Ans 2 Will choose lasso regression as it would help in feature elimination and the model will be more robust , Lasso regression, uses a parameter called lambda ,it is absolute value of magnitude of coefficients . As the lambda value increases Lasso shrinks the coefficient estimates towards zero

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?

Ans 3

Those 5 most important predictor variables that will be excluded :

- GarageArea
- OverallQual
- GrLivArea
- 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?

Ans 4

A model has to be made robust so that they are not impacted by outliers in the training data and generalisable so that the test accuracy is not lesser than the training score. The model should be accurate for unseen datasets.

Outliers has to be analysed and handled , and only those which are relevant to the dataset need to be retained. Unnecessary outliers should be removed , this will in return help in increase accuracy of the predictors by the model