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 for Ridge Regression is 0.01 and for Lasso Regression is 100

If we double the alpha values for Ridge making it to 0.02 and for Lasso making it to 200 we find significant changes in coefficient values.
**Ridge :** There is a decrease of r2_test to 0.01 %
**Lasso:** There is a significant decrease in r2_train to 2% if we double the alpha value. The features selected on doubling the alpha is :

BedroomAbvGr

Condition2_RRAe

OverallCond

Exterior1st_AsphShn

LotFrontage

BsmtFinSF2

WoodDeckSF

BsmtQual

Neighborhood_Somerst

Neighborhood_NridgHt

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:** We shall use **Lasso Regression Model** as it had better r2 score than Ridge Model.It jas also removed unwanted features from model without affecting the model accuracy. Also the model is more simple and accurate to use.

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:** On dropping the first 5 important predictors below is the observation:
1. The alpha value obtained remains same as before I.e 100
**2.** The important predictors now obtained is
**Exterior1st_BrkComm,LotArea,RoofMatl_Metal,RoofMatl_Roll,RoofMatl_Tar&Grv**

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 robust when any variation in the data does not affect its performance much.
- A generalizable model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.
- To make sure a model is robust and generalizable, we have to take care it doesn't overfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data.
- In other words, the model should not be too complex in order to be robust and generalizable.

- If we look at it from the prespective of Accuracy, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.
- In general, we have to find strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.