**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:**

**Optimal value of alpha is 1.0 for Ridge and 10.0 for lasso regression.**

**R2score on training and test data has increased after we double the alpha for Ridge.**

**R2score on training data has increased but it has decreased on testing data after we double the alpha for lasso.**

Important predictor variables after the change:

**LandContour**

**Utilities**

**MasVnrType**

**MasVnrType**

**ExterQual**

**Foundation**

**HeatingQC**

**PavedDrive**

**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:**

The r2\_score of lasso is slightly higher than ridge for the test dataset so we will choose lasso regression to solve this problem

**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:**

**Below are the top 5 predictor variables. So we will drop them and rerun the model.**

'LandContour\_Bnk', 'LandContour\_HLS', 'LandContour\_Low','LandContour\_Lvl', 'Utilities\_AllPub'

**For Lasso, R2score of training and testing data has decreased after removal of above variables. New top 5 variables are mentioned below:**

**‘Utilities\_NoSeWa’ ‘MasVnrType\_BrkCmn’ ‘MasVnrType\_BrkFace’ ‘MasVnrType\_None’ ’ExterQual\_Ex’**

**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 generalizable when test R2 score and train R2 score doesn’t have significant difference. We should remove the outliers from the dataset so that model doesn’t try to learn from outliers. The accuracy of the model can’t be trusted if the data has not been treated for outliers and there will be high variance each time the model is trained.**