**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: The optimal values of alpha in Ridge is 5 and Lasso is 0.0004. As the alpha values increases or doubles the variance decreases and bias increases and it leads to underfitting. The important predictor is** MSZoning\_RL**.**

**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:** The Mean Squared Error of Lasso is slightly lower than that of Ridge Also, since Lasso helps in feature reduction (as the coefficient value of one of the feature became 0), Lasso has a better edge over Ridge. Hence based on Lasso, the factors that generally affect the price are the Zoning classification, Living area square feet, Overall quality and condition of the house, Foundation type of the house, Number of cars that can be accomodated in the garage, Total basement area in square feet and the Basement finished square feet area.

**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:** MSZoning\_RL, GrLivArea, MSZoning\_RM, OverallQual, MSZoning\_FV are the top 5 variables.

**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:** A model needs to be made robust and generalizable so that there is no implication of the outliers in the training data. The model should also be generalisable so that the test accuracy is not lesser than the training. The model should be accurate for datasets other than the ones which were used during training. Too much weightage should not given to the outliers so that the accuracy predicted by the model will be high.