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 value of alpha for Ridge regression was found to be 2 and for Lasso regression it was 0.001. With the alphas the R2 of the model for both lasso and ridge was approximately 0.83. Doubling the alpha values in the Ridge and Lasso, the prediction accuracy remains as 0.82 with a small change in the co-efficient values.

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?

The optimum lambda value for both Ridge and Lasso regression are as follows: -

- Ridge – 2
- Lasso – 0.0001

▪ The Mean Squared Error for both Ridge and Lasso regression are:

- Ridge -0.001839609078792425
- Lasso - 0.0018634152629407746

▪ The Mean Squared Error of both the models are almost equal.

▪ Since the Lasso regression helps in feature reduction (as the coefficient value of some of the features becomes zero), Lasso has a better edge over Ridge regression and need to be used in the final model.

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) The five most important predictor variables in the current lasso model is:-
1. Total_sqr_footage
2. GarageArea
3. TotRmsAbvGrd
4. OverallCond
5. LotArea

We have finally built Lasso model in the Jupiter notebook after removing these attributes from the dataset. The R2 of the new model without the top 5 predictors drops to .73

The Mean Squared Error increases to 0.002857567090648254

The new top 5 predictors are:

Lasso Co-Efficient	
LotFrontage	0.146535
Total_porch_sf	0.072445
HouseStyle_2.5Unf	0.062900
HouseStyle_2.5Fin	0.050487
Neighborhood_Veenker	0.042532

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?
In order to make the model more robust and generalizable, make the model simple but not simpler.
Regularization can be used to make model simpler.

Regularization helps to strike the right balance between keeping the model simple and not making it too naive to be of any use. In regression, regularization involve adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model. Also, while making a model simple will lead to Bias-Variance Trade-off:

• A complex model will need to change for every little change in the dataset and they are very unstable and extremely sensitive to changes in the training data.

• A simpler model that abstract out some pattern following the data points given is unlikely to change widely even if more points are added or removed.

Bias tell us how accurate is the model likely to be on the test dataset. A complex model can do accurate job prediction given there is enough training data.