Answers of Assignment-based Subjective Questions

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 of alpha for Ridge regression is 6.0 and the optimal value of lambda for Lasso regression is 0.0001.

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 LASSO regression is used for feature selection, while the Ridge regression is used to reduce the coefficients of features. For the given model LASSO removes 91 features, but at a significant cost. That is a significant drop in model accuracy. Ridge regression therefore provides the best prediction accuracy for the given 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?**

Answer: After excluding the five most important predictor variables, the present five most important predictor variables (in descending order) are as follows: 'BsmtFullBath_2', 'TotRmsAbvGrd_14', 'GarageQual_Gd', 'BsmtFullBath_3' and 'GarageCars_4'

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 said to be robust when variances in the data have little impact on how well it performs. On the other hand, a generalised model is capable of adapting to new unknown data that comes from the same distribution as the data that was used to build the model. In other words, we should be careful that a model doesn't overfit the training data in order to ensure that it is robust and generalisable.

- From the perspective of accuracy, a complex model has very high accuracy, but it fails to generalize to new unknown data. On the other hand, a simple model has less accuracy. In order to make the model more robust and generalisable, we have to strike a balance between model accuracy and complexity, i.e., the trade-off between bias and variance. This bias-variance trade-off can be achieved by regularization methods such as Ridge and LASSO regressions.