Assignment - Advanced Regression

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?
sol: Optimal value of alpha for Ridge Regression is 0.9 and Optimal value of alpha for Lasso Regression is 0.0001.

If we double the value of alpha then model is not optimal, the test/training R2_score/accuracy of the model will either slightly decrease or it may remain same.

In my model, R2_score of training data decreases slightly while the test R2_score remains same.

Predictor variable remains same after changing the implementation in Ridge and lasso Regression.

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?
Sol: I will go with Lasso Regression model, here it decreases model complexity by feature selection and makes co-efficient to zero.

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?
Sol: 1stFlrSF, 2ndFlrSF, HouseStyle_SFoyer, OverallQual_Rating4, HouseStyle_SLvl

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?
Sol: To make the Model robust and generalisable, we can use Regularization Techniques, use Cross Validation or/and tune the hyper parameters. We generally use Regularization Technique to make the model robust, less complex and to remove under/over fitting of the model. We use Ridge and Lasso regression techniques especially for this issue. The implications of the this is what we call as Bias variance Trade off. If we try to increase one, the other will decrease. So we need to find an optimal model which has acceptable bias and variance i.e. the model has to perform well for both training data as well as unseen test data well. The accuracy of the model will decrease as an implication of the model being robust and generalisable.