Subjective Questions

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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer: For lasso regression, the optimal value of alpha is 1. The Most important variables would be MSSubClass. RoofMatl_Membran, MSZoning_RL, MSZoning_FV, MSZoning_RH.

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: We would decide the value of alpha where we have good training as well as the test score with the plots. In this case we choose 4 as the value for lambda for Ridge Regression, since it has the best train as well as the test score and we choose 50 as the value for lambda for Lasso Regression, since it has the best train as well as the test score.

Question 3: After building the model, you realized 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: The 5 most important variables are: MSSubClass. RoofMatl_Membran, MSZoning_RL, MSZoning_FV, MSZoning_RH

Question 4: How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer: A model is considered to be robust if the model is stable, i.e., does not change drastically upon changing the training set. The model is considered generalizable if it does not overfits the training data and works well with new data. Its implication in terms of accuracy is that a robust and generalizable model will perform equally well on both training and test data i.e., the accuracy does not change much for training and test data.