Assignment-based 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?

Regression with ridge regularization:

For ridge regression, when the test and train errors for different values of lambda were plotted, we could observe that increase in alpha values decreases the train error, and at lambda value approximately 2, the test errors remain minimum, and on further increasing lambda to a much higher value, increases the test error

Important predictors for Ridge regression:

1. MSZoning_FV
2. SaleType_ConLD
3. MSZoning_RH
4. MSZoning_FV
5. RoofMatl_WdShngl

Regression with ridge regularization:

For Lasso regression, when the when the test and train errors for different values of lambda were plotted, we could observe that increase in alpha value penalizes the model coefficients more and more coefficient of the feature variable will reduced to zero, when we increase the value of lambda to a much higher value r2 square score decreases, and the model tends to underfit.

Important predictors for Lasso regression:

OverallQual (Rates the overall material and finish of the house)
GrLivArea (Above grade (ground) living area square feet)
GarageArea (Size of garage in square feet)
OverallCond (Rates the overall condition of the house)
BsmtFullBath (Basement full bathrooms)

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?

For ridge regression, we could observe that increase in alpha values decreases the train error, and at lambda value approximately 2, the test errors remain minimum, so chosen the optimal value of lambda as 2.

For Lasso regression, the optimal value of chosen lambda is 0.01. Increasing the lambda to a higher value penalizes the model coefficients more and more coefficient of the feature variable will reduced to zero and reduces the r2 score of the model.

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?
After excluding the 5 important predictors in the lasso model, the other 5 most important predictor variables are,

1.1stFlrSF
2.2ndFlrSF
3.Fireplaces
4.TotalBsmtSF
5.BsmtFinType1_GLQ

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?

Addressing the Bias-variance trade-off

If the model is too simple, then it may be on high bias and low variance condition and thus is prone to errors. If model is too complex, then it may be on high variance and low bias. In the latter condition, the model would not generalize well with new entries. We need find a Bias-variance trade-off for the training the model which gives low error in training as well as testing data.
Therefore, would the improve the robustness and generalizability of the model, which is the primary objective of any machine learning.