Assignment Part-II

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?

Answer >> Optimal values of alpha are following:

- Ridge: 10
- Lasso: 100

If we choose double value of alpha for both Ridge and Lasso, then in case of Lasso more feature co efficient will become zero while for Ridge, feature co-efficient will be closer to zero.

Once, change is implemented most important predictors will be following:

- GarageType_Detchd
- GarageType_No Garage
- GarageFinish_No Garage
- GarageFinish_RFn
- GarageFinish_Unf

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 >> Optimal values of Alpha are following:

Model Name	Alpha	Train Score	Test Score
Ridge (Without RFE)	10	90.3	80.6
Lasso	100	91.8	77.3

Both models have performed equally well, we can pick any model.

If we have to pick only one, we will go with Ridge (Without RFE) model as for that model Test score is better than Lasso and Train error is comparable with Lasso error. Model which performs better on unseen data is of more importance.

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?

Answer >> After removing initially selected top 5 most important predictor variables, following are the new top 5 most important predictor variables:

- FireplaceQu_TA
- GarageType_Attchd
- GarageType_Basment
- GarageType_BuiltIn
- GarageType_CarPort

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?

Answer >> For model to be robust and generalized, model should be less complex ie should not be having too many features. Also, it should perform reasonably well on unseen data. To achieve this, we need to use and apply penalties specified by Lasso or Ridge. With addition of penalties, we try to have significant reduction in the variance in the model by allowing some bias.

Also, model's Residual graph should not show any pattern, error should be evenly distributed. We could also try to build and cross check model with more Error metrics like MSE (Mean Square Error) or MAE (Mean Absolute Error).