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 1:

The Optimum Value for ALPHA, or RIDGE - 10,

for LASSO - 200

For Double the value of Alpha in Lasso, minor reduction in Train but Test remains same

And for Double the value of Alpha in Ridge, no noticeable change

Predictor variables Ridge –

'Exterior1st_Plywood', 'Heating_GasW', 'Exterior1st_MetalSd', 'Heating_GasA', 'LotFrontage',

'MasVnrArea', 'GarageType_BuiltIn', 'BldgType_2fmCon', 'BsmtFinSF2', '2ndFlrSF'

Predictor variables Lasso -

'Exterior2nd_CmentBd', 'Heating_GasW', 'MasVnrArea', 'GarageType_BuiltIn', 'BldgType_Duplex', 'Condition2_RRNn', 'BsmtFinSF2', 'BldgType_2fmCon', '2ndFlrSF'

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 2:

I will select Lasso regression for the model.

Ridge regression cannot zero out coefficients; they end up including all the coefficients in the model, or none at all. But LASSO on the other hand does both parameter shrinkage and variable selection.

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 3:

'GarageType_BuiltIn', 'Exterior1st_Plywood', 'BldgType_Duplex', 'MasVnrArea', 'Exterior2nd_ImStucc'

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 4:

When the model is able to work with many different kinds of unseen data.
The model should be able to resist outliers; tree model has good resistance towards outliers in the data.
We can add regularisation techniques to make model more generalised.
The ridge regression improves generalization in the data set, by reducing overfit.
Lasso regression can be used to eliminate irrelevant noises and do feature selection and hence regularize a model.