

## Problem Statement 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 value of alpha for ridge and lasso regression are 250 and 0.001 respectively.

In case of Ridge Regression, Coeff values are increasing as alpha increases. r2score of train (0.86 from 0.90) and test data (0.84 from 0.87) also dropped.

In case of Lasso regression, As, alpha value increased more features removed from model. But r2score is also dropped in both train (0.87 from 0.91) and test (0.84 from 0.88) data

Top features are: LotArea, 2ndFlrSF, LowQualFinSF, Neighborhood\_OldTown, MasVnrArea.

### 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:

I will Choose Lasso regression as it gives features selection also ie, eliminating features which are not significant to model without affecting the model accuracy. Also, Accuracy of Lasso regression is more. All these factors make it generalized, simple and accurate.

### 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:

Five most important predictor variables now after removing the five most important predictor variables in the lasso model are: OverallQual, WoodDeckSF, Fireplaces, GrLivArea, BsmtQual

#### 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:

we should consider following Characteristics in order to make a model Robust and Generalised.

- Model should perform well both with the trained and unseen test data. Model accuracy should be good for both trained and test data
- Model should have significant predictor variables i.e., P- values should be less than 0.05 for all the variables contributing to model
- There should not be multicollinearity between predictive variables.
- Model should not be impacted by outliers present in data