

## Assignment: Advanced Regression

### 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?

- Optimal value of alpha for ridge is 0.8 and for lasso is 0.0001
- Double the value of alpha for ridge is 1.6 and for lasso is 0.0002.
- For ridge the value of some coefficients goes increases, while value of some coefficients decreases.
- For lasso the value of LotArea was reduced to 0.
- After change the most important predictor variable for ridge is GrLivArea.
- After change the most important predictor variable for lasso is GrLivArea.

### 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?

- Lasso works well on small number of significant parameters. While ridge works well on many parameters with equal weightage. Since we have small number of parameters i.e. 29 and some coefficients are significant compared to others, I will choose to go with Lasso.

### 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?

- Fireplaces, WoodDeckSF, LotArea, Neighborhood\_\_Edwards, 'Neighborhood\_\_Sawyer]

### 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?

- We can make sure that a model is robust and generalisable by deliberately making the model simple but not naive. This is a tradeoff as simpler models have high bias and low variance and complex models have high variance and low bias. Making a model too simple will make it have a very high bias and it will underfit. Making a model too complex will make it have high variance and thus it will overfit the training data and fail on test. Thus we need a model with low bias and low variance which are a tradeoff. Thus we need to balance these by picking optimally complex model. Regularisation is one way to reduce complexity of a model. The total error is the sum of variance and bias for a model. The goal should be to minimise the total error, which will give us an optimally complex model.