## Assignment Part-II

# Question and answers

### 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

- The optimal value of alpha for ridge – 1.0 and lasso – 50
- Change in the model if we double the value of alpha for both ridge and lasso are: -
    a. In ridge regression, After doubling, the coefficients decreased, the r2 scores decreased, and the rss and mse increased.
    b. The coefficients in lasso regression rose after doubling, r2 scores decreased, and rss and mse increased.
- When the predictor is increased by one unit, the coefficient value indicates the mean change in the response. As a result, it's natural to assume that variables with greater coefficients are more significant since they indicate a larger change in the answer. The following predictors are relevant: -
    a. Lasso regression
        i. Positively influencing predictors
            1. Sale type – contract low down.
            2. Price per square feet (log value).
            3. MS Zoning residential low density.
            4. MS Zoning floating village residential.
            5. Exterior condition of material – fair type.
    b. Ridge regression
        i. Positively influencing predictors – 1. Heating quality and condition – poor, 2. Price per square feet (log value), 3. Exterior condition of material – fair type.
        ii. Negatively influencing predictors – 1. Exterior covering – cement board. 2. MS Zoning residential high density. 3. MS Zoning residntial medium density.

### 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

Because lasso's r2 score for the test dataset is somewhat higher than lasso's, we'll use lasso regression to tackle this problem.

### 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:

- First Floor square feet
- Above grade (ground) living area square feet
- Pave road access to property

- Roof material_Metal
- Type of roof(Shed)

## 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

The model should be generalised to ensure that the test accuracy does not fall below the training score. For datasets other than those used during training, the model should be correct. Outliers should not be given too much weight in order for the model's accuracy to be high. To guarantee that this is not the case, an outliers analysis must be performed, with only those outliers relevant to the dataset being kept. Those outliers in the dataset that don't make sense to keep must be removed. It is impossible to trust a model for predictive analysis if it is not resilient.