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 – 0.01 and lasso – 10
- Effect of doubling alpha: -
    a. In the case of ridge regression, the coefficients reduced after doubling, r2 scores reduced and rss and mse increased.
    b. In the case of lasso regression, the coefficients increased after doubling, r2 scores reduced and rss and mse increased.
- Most important predictor variables after implementing change – No changes in most important predictors found. The following predictors are relevant: -
    a. Lasso regression
        i. Positively influencing predictors – 1. LotArea and 2. First foor square feet area.
        ii. Negatively influencing predictors – 1. Price per square feet, 2. Ground living area and 3. Bedroom above ground.
    b. Ridge regression
        i. Positively influencing predictors – 1. Lot area and 2. Lot frontage.
        ii. Negatively influencing predictors – 1. Length width ratio, 2. Price per square feet and 3. Ground living area.

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 with optimal value of lambda of 10. This is because the lasso regression provides a better R2 score and lesser RSS and MSE compared to ridge regression.

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

- Positive influencers
    - OverallQual
    - LotFrontage
    - TotRmsAbvGrd

- Negative influencers
    - LWRatio
    - BsmtQual_TA

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

A model derives its robustness from being capable of providing the same accuracy even with changes in assumptions and independent variables. So, to make a model robust, the features selected should be thoroughly verified and filtered using RFE, VIF and other techniques. In this assignment, business metrics have become predictor variables in the original as well as reduced model selection for lasso regression.

The accuracy of the model varies with the type and number of feature set input to it. A change in the model input feature set can change the different metrics that are used as a measure of efficiency. Adjusted R Squared metric can be used to explain how well independent features explain the variability in the dependent variable/ target variable. For regression models, there are many metrics that provide information on the model's performance. A robust model will show lower susceptibility to changes in the feature sets.