Surprise Housing Part-II Assignment- Subjective Question Answers - Poonam Attri

Q1. 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?

Ans. The optimal value for Ridge regression is 1400. The Lasso regression could not achieve accuracy and was overfitting the train data with alpha as 0.035. If the alpha values are doubled, the penalty is doubled and R-squared score for train and test data comes down.

Q2. 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?

Ans. I choose to apply Ridge regression model since it has achieved an R-squared score for train and test as 86% and 81% respectively. It is well within a range of 5% variance. It has good Cross validation score of 80% approx.

Q3. 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?

Ans. The five most important predictor variables now are Basement Quality, Sale Condition , Garage Cars, Central Air, Full Bathrooms and Overall Quality.

Q4. 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?

Ans. I made sure that model is robust and generalizable by checking Error terms distribution and variance between train and test scores. The error terms are normally distributed and there’s 5% variance between train and test R-squared scores of the model.