# ADVANCE REGRESSION CASE STUDY

Question Answers

NOVEMBER 2, 2022

IIIT B

# Contents

# Assignment-based Subjective

**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:

- Optimal Value of lambda for Ridge: 100
- Optimal Value of lambda for Lasso: 0.001

We experimented with doubling the value of alpha for Lasso and Ridge. Following were the observations:

1. The r2 score for Ridge reduced by approximately 0.01
2. The r2 score for Lasso increased slightly
3. For both Ridge and Lasso , however, the coefficient value decreased
4. The impact was higher in Lasso and could be observed in the order of the top 10 variable weights which got changed for Lasso but not for Ridge.

**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**: Lasso and Ridge both resulted in good models where metrics of train versus predicted were good. But going by Occam's principle, we will go for Lasso to reduce complexity of model. Ridge is good if we need to reduce coefficient values. RMSE is another important factor to be considered. It was lower for Lasso than for Ridge.

**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:**
Based on code, the new set of variables and their coefficients are as below:

| Coefficients | Value |
|---|---|
| 1stFlrSF | 0.114282 |
| 2ndFlrSF | 0.105415 |
| MSZoning_RL | 0.090025 |
| MSZoning_RM | 0.076300 |
| MSZoning_FV | 0.048366 |
| BsmtFinSF1 | 0.040562 |
| Functional | 0.033513 |

**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:**

There are few thumb rules that can be followed to ensure that our model is generalizable:-

1. Ensure the r2 score of train and test are as close as possible
2. Using k fold validation for small data gives better results,
3. The model should be generic enough to be not impacted by outliers
4. Using Ridge and Lasso, penalize coefficients to keep them as low as possible so that it does not overfit
5. Tree models are another great way to ensure that model does not overfit.
6. As per Occam's principle, sometimes accuracy has to be sometimes sacrificed for simplicity of model.