

# Problem Statement - Part II

## 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?

### Answer :

The optimal value of alpha for Ridge and Lasso regression is a tuning parameter to determine strength of regularization and it depends on the specific dataset. The choice of alpha is made using techniques like cross-validation, grid search, or AIC/BIC criteria where different values are tested, and the one that results in the best model performance is selected.

Increasing the value of alpha in both Ridge and Lasso regression increases the strength of regularization. If we double the value of alpha for both Ridge and Lasso regression, the regularization penalty will be stronger.

Effects of Doubling Alpha are as follows :

Ridge Regression:

- Increased shrinkage
- Less drastic feature elimination
- The model will be more biased, but it may have improved generalization performance

Lasso Regression:

- Enhanced feature selection
- Potentially improved interpretability
- Risk of over-penalization

The most important predictor variables after the change is implemented will be those whose coefficients are not driven to zero. The optimal value of alpha depends on the specific dataset and the problem at hand. The choice of alpha should be made with techniques like cross-validation to make sure good model performance on new, unseen data.

### **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:**

Based on a trade-off between predictive performance, interpretability, and the characteristics of our dataset, Lower RMSE and Higher R-squared both models perform similarly but Lasso will be preferred one.

### **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 :**

The most important predictor variables in Lasso regression model are typically the ones with non-zero coefficients. The coefficients provide a measure of variable importance. We can examine the coefficients of the Lasso model to identify the five most important predictor variables. We can Extract coefficients from the Lasso model, create a DataFrame to associate coefficients with corresponding feature names, sort the DataFrame by absolute coefficient values in descending order and display the top five predictor variables.

The top\_five\_predictors DataFrame will contain the names of the five most important predictor variables based on their absolute coefficient values in the Lasso model.

**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 :**

To make sure that a model is robust and generalizable needs several practices to enhance its performance across different datasets. Some strategies are as below.

1. Cross-Validation
2. Train-Test Split
3. Hyperparameter Tuning
4. Regularization
5. Feature Engineering
6. Ensemble methods
7. Rigorous testing
8. Continuous monitoring
9. Ensemble Methods
10. Data Augmentation
11. External Validation

Also to consider Training Accuracy vs. Test Accuracy, Bias-Variance Tradeoff, Robustness to Variability.