

Advanced Regression Assignment

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 values of alpha for Ridge & Lasso are

Ridge: 500

Lasso: 1000

If we double the value of alpha, then below changes will take place:

For Ridge: The coefficients will be reduced

For Lasso: The less important (Significant) features will become 0.

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: The optimal values of lambda for Ridge & Lasso regression are:

Ridge: 500

Lasso: 1000

The accuracy for both the models was nearly same but it's better go for Lasso Regression since it will drag the less important features to 0

	Train	Test
Ridge	88.36	85.6
Lasso	89.24	84.11

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: After dropping the most important predictor variables, the new variables are:

```
('SaleType_ConLI', 6845.151)
('SaleType_ConLw', 7076.102)
('SaleType_New', 7429.085)
('SaleType_Oth', 16743.628)
('SaleType_WD', 27581.124)
```

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: We must use a model that is least affected when few features are changed drastically. We should not over train the model using more complex features as this will overfit the model & will fail when unseen data is fed to the model. So, it is good practice to use regularization methods like Ridge & Lasso.

These regularization methods will regularize the coefficients by reducing them based on the hyper parameter value λ (alpha in sklearn)

We use metrics to evaluate the simplicity & best fit model. The metrics penalize the model for being too complex. Few examples of metrics are Mallows's C_p , Adjusted R^2 (R-Square), AIC & BIC.

As we increase the number of predictors, the penalty in AIC, BIC will increase while the RSS decreases. Lower values of AIC, BIC & C_p and Higher value Adjusted- R^2 , will provide us the best fit model.

By using regularization, the accuracy of the model drops a little since it will generalize the model & give us the option to choose lower bias & lower variance. (Tradeoff between Bias & Variance)