

Advanced Regression - Assignment QA

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 1:

- Optimal alpha for lasso is `0.001`, and ridge is `10`.
- If we choose double of the alpha value, the `r2_score` of both ridge and lasso will increase a little bit.
- The most important predictor is `MSSubClass`.

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 2:

For Lasso model, the index is:

```
r2_score on train data: 0.9211923846949593
r2_score on test data: 0.9142166005932495
mean squared error: 0.01294206450039103
```

For Ridge model, the index is:

```
train score 0.9315656058788783
Test score 0.9129642074967635
mean squared error: 0.01313101192316349
```

The Lasso model have more small MSE, so we choose **Lasso**.

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 3:

As the output in the code, before it is:

| |
|----------------------|
| MSSubClass |
| BsmtFullBath |
| SaleType_Oth |
| Neighborhood_Edwards |
| OverallCond |

After it is we drop this five variables, and retrain the model, it is:

| |
|--------------------|
| LotArea |
| LotFrontage |
| Neighborhood_NAmes |
| Condition1_PosN |
| BedroomAbvGr |

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 4:

For robust and generalisable, we need to optimize following steps:

- Train-Test Split

- Cross-Validation
- Hyperparameter Tuning
- Outlier Handling
- Feature Selection

For implications for Model Accuracy:

- Robustness
- Generalizability
- Accuracy