

Assignment subjective questions

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?

Ans :

- The optimal value of alpha for ridge regression with RFE is 1.069.
- The optimal value of alpha for lasso regression is 0.001
- If we double the value of alpha the model becomes more generalized. But in this case since alpha is already small, for ridge if we double(alpha=2.138) there is significant difference in the scores.
training r2 score 0.7501645590677039
testing r2 score 0.6942870665991379

If we double the value of alpha for lasso, there is no Difference.

Before : Train r2: 0.7023994213076254 test r2: 0.6653849625073787

After : Train r2: 0.7023994213076254 test r2: 0.6653849625073787

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?

Ans According to the both the models, ridge with RFE selected 60 features perform slightly better than lasso.

Ridge : 0.757861 (train r2), 0.692000(test r2)

Lasso : : Train r2: 0.7023994213076254 test r2: 0.6653849625073787

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?

Ans :

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?

Ans : lambda/alpha value helps to generalize the model keeping in mind the bias – variance trade off. Higher the lambda value greater is the penalty applied to the regression models. This makes the model very simple. It

might cause underfitting, i.e the model will fail to learn on the training data itself. Lower lambda value will cause overfitting, where model performs well on training_data , but performs poorly in test data. Algorithms like gridsearchcv help us obtain the optimal lambda value.