Assignment-based Subjective Questions

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 :

- For Ridge regression optimal value of alpha is 1.0
- For Lasso regression optimal value of alpha is 0.0001

For Ridge regression alpha is 1.0 and now doubling it and making it 2.0

For Ridge regression alpha is 0.0001 and now doubling it and making it 0.002

After change, the impact on

Ridge Regression train r2: 0.9238505663513401

Ridge Regression test r2: 0.7526833276310241

After change, the impact on,

Lasso Regression train r2: 0.924842196868562

Lasso Regression test r2: 0.7423837845239192

The most important predictors after the changes would be

[**BsmtFullBath**, **LotFrontage**, **Overall Condition**, **Garage Area**, **CentralAir**, **MSZoning_RH, Overall Quality**, **Exterior1st_CBlock**]

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:

I've opted for Lasso regression due to its comparable performance with Ridge in terms of R2_score. Lasso offers unique advantages such as feature selection and simplicity stemming from a reduced feature set. By eliminating unnecessary features without compromising model accuracy, Lasso promotes model generalization, making it a suitable choice

3. After building the model, you realized 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: Five most important predictor variables now

[**LotArea**, **FullBath**, **1stFlrSF**, **ExterCond**, **MSZoning_RH**]

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:

For a model to be robust and generalisable, it must strike a balance that avoids both overfitting and excessive complexity. Overfitting, characterized by a model that fits the training data too closely, results in high variance, making it sensitive to small data fluctuations. While such a model may excel in training data accuracy, its performance on

unseen test data suffers. To enhance robustness and generalizability, we prioritize reducing model complexity, a task facilitated by regularization techniques.

Bias, on the other hand, represents a model's inability to learn from data effectively, resulting in errors. High bias signifies that the model struggles to capture intricate data patterns, leading to poor performance on both training and testing data