Nhật Du Ngô

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 value of alpha for ridge and lasso regression:

-   1 and 10 respectively

the changes in the model if you choose double the value of alpha for both ridge and lasso:

-   Predictors are same but the coefficent of these predictor has changed: LotArea, OverallQual, OverallCond, BsmtFinSF1, TotalBsmtSF, GrLivArea, TotRmsAbvGrd, Street_Pave, RoofMatl_Metal

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:

Given that the Lasso regression model demonstrates a slightly higher R2 score compared to the Ridge regression model for the test dataset, we opt to utilize the Lasso regression to address this problem. This choice aligns well with our objectives, especially considering the modest performance difference between the two models.

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:

11stFlrSF, GrLivArea, Street_Pave, RoofMatl_Metal, RoofStyle_Shed

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:

The model's generalization is crucial to ensure that the test accuracy remains on par with the training score. It's imperative that the model performs accurately on datasets beyond those used in training. Additionally, mitigating the influence of outliers is essential for achieving

high predictive accuracy. To achieve this, thorough outlier analysis is necessary, retaining only those outliers relevant to the dataset while discarding outliers that do not contribute meaningfully. A robust model instills trust in its predictive capabilities, making it suitable for reliable predictive analysis.