Assignment – Part 2

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 values of alpha for ridge and lasso regression are

Ridge: alpha: 0.9

Lasso: alpha: 0.0001

	RIDGE	LASSO
R2 TRAIN	0.88612	0.88324
R2 TRAIN	0.74116	0.74281

Doubling the alpha variables

Ridge**:** alpha: 1.8

Lasso: alpha: 0.0002

	RIDGE	LASSO
R2 TRAIN	0.88472	0.87972
R2 TEST	0.74598	0.74593

After the values are doubled the important variables are

Ridge :

1. OverallCond_2
2. LotFrontage
3. TotRmsAbvGrd
4. 1stFlrSF
5. smtFinSF2

Lasso:

1. OverallCond_2
2. 1stFlrSF
3. OverallQual_10
4. TotRmsAbvGrd
5. 2ndFlrSF

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:

The optimal values of alpha for ridge and lasso regression from our model are

Ridge: alpha: 0.9

Lasso: alpha: 0.0001

There is not much much difference in the performance of the model as the r2 and RMSE is similar. But in Lasso some coefficients are made zero. Hence the model from lasso is better it is giving similar performance and accuracy with a smaller number of features.

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:

Dropping the 5 important variables of lasso regression

1. OverallCond_2
2. 1stFlrSF
3. OverallQual_10
4. TotRmsAbvGrd
5. 2ndFlrSF

The alpha values we got is

Ridge: alpha = 3.0

Lasso: alpha = 0.0001

The top 5 feature now is

1. LotFrontage
2. BsmtFinSF1
3. Fireplaces
4. Neighborhood_StoneBr
5. LotArea

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:

To ensure that a machine learning model is robust and capable of generalizing well, several factors need to be considered. Firstly, a diverse and sufficient training dataset should be used to expose the model to various scenarios and edge cases. Additionally, careful feature engineering is crucial to select relevant features and eliminate noise. Regularization techniques regularization prevent overfitting, ensuring that the model doesn't become too specialized to the training data. Cross-validation and testing on separate datasets provide a reliable evaluation of the model's generalization ability. Regular monitoring and retraining allow the model to adapt to evolving patterns. Striking the right balance between bias and variance is essential for achieving accurate predictions on unseen data.