Name **Bhushan Date**

Course **Executive PG program in Machine Learning & AI**

Batch **Feb-2022**

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:
Optimal Value of Alpha for Lasso : 0.0001
Optimal Value of Alpha for Lasso : 0.8

After doubling the alpha values for both

Metric	Linear Regression	Ridge Regression	Lasso Regression	Double Ridge Regression	Double Lasso Regression
R2 Score (Train)	0.880246	0.879047	0.879408	0.877700	0.878363
R2 Score (Test)	0.838880	0.842046	0.841584	0.841663	0.843520
RSS (Train)	16.873705	17.042684	16.991772	17.232465	17.139091
RSS (Test)	12.174952	11.935663	11.970574	11.964650	11.824323
MSE (Train)	0.132994	0.133658	0.133458	0.134400	0.134035
MSE (Test)	0.160777	0.159189	0.159422	0.159382	0.158445

Important top predictor variables after change are as follows:

‘OverallQual’ ‘TotRmsAdvGrd’, 'YearBuilt', 'LotArea', 'GarageArea'

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:

I will use Lasso regression to apply for final model, as R2 / MSE & RSS of Lasso regression are better than Ridge regression in this model.

Lasso regression also performs variable elimination & selects only significant variables in the model. Model complexity also reduces as we can remove features with Zero coefficient.

Hence Lasso can be preferred over Ridge regression.

Question 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?

Ans:

After removing the top 5 predictor variables from the data & building model again the following are the new top 5 predictor variables are as follows:

'OverallQual'
'TotRmsAbvGrd'
'YearBuilt'
'LotArea'
'GarageArea'

	Lasso
OverallQual	0.886780
TotRmsAbvGrd	0.552106
YearBuilt	0.522406
LotArea	0.349023
GarageArea	0.327135

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Ans:

A model with testing error & training data error has enough stability even after some noise. This means that an unprecedented change in one or more features does not significantly alter the value of the predicted variable.

Regularization helps in achieving a less complex model by shrinking the model coefficients. Thus, preventing the model from overfitting.

We can control the tradeoff between model complexity and the bias. Regularization helps coefficient for making the model too complex, so to make the module more robust, there should be balance between keeping the model simple, making simple model lead to bias variance trade off.

Accuracy of the model can be maintained by keeping the balance between bias and the variance and minimize the total error.