Assignment-based Subjective Questions

Question -:
What is the optimal value of 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 in implemented?

Answer -:
The optimal value of Ridge Regression is: 10 whereas the optimal value of Lasso Regression is: 6e-05 i.e., 0.00006.
The changes in the model if choose to double the value of alpha for the ridge regression, it would shrink the coefficient towards zero while in lasso regression it could force some coefficients towards zero.
The most important predictor variable after the change in implemented are those which is significant to model.

Question -:
You have determined the optimal value of lambda of ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer -:
The Optimal Lambda or alpha which the model suggest -: Lasso Regression -: 0.00006 Ridge Regression is 10.

As the Ridge and Lasso regression r-2 score of train and test set-:
Ridge: Train :90.6% Test :86.9%
Lasso: Train :91.8% Test :87.0%

So, both the model has good score. So, I suggest a lasso regression because it show zero coefficients to less significant features.

Question -:

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 -: After the building the model, I realized that the five most important predictor variables in the lasso model are not available in the incoming data and I would create a another model excluding the five most important predictor variable . Then the model suggest me these five variables excluding the five most important variables -:

LASSO -:

Features	Coefficients
'GarageType_CarPort'	0.134
'GarageType_Detchd'	0.135
'GarageFinish_RFn'	0.137
'GarageFinish_Unf'	0.155
'GarageFinish_Unknown'	0.171

RIDGE -:

Features	Coefficients
('GarageType_CarPort'	0.068)
('GarageType_Detchd'	0.069)
('GarageFinish_RFn'	0.069)
('GarageFinish_Unf'	0.071)
('GarageFinish_Unknown'	0.084)

Question -:

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

Answer -:

Ensuring that a machine learning model is robust and generalizable is crucial for its effectiveness in real-world scenarios. Here are several techniques and considerations to make sure a model is robust and generalizable -:

Cross validation -:

Use techniques like k-fold cross-validation to assess the model's performance on different subsets of the data. This provides a more accurate estimate of the model's performance and helps in identifying overfitting.

Train -Test Split -:
Split the dataset into separate training and test sets. The model is trained on the training set and evaluated on the test set. This provides an unbiased evaluation of the model's performance on unseen data.

Feature Selection -:
Choose relevant features and avoid overfitting by selecting features that have a significant impact on the target variable. Feature selection methods like Lasso regression can be employed.

Regularization
Use regularization techniques like L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting by penalizing large coefficients.

Hyperparameter Turing-:
Fine-tune hyperparameters using techniques like grid search or random search. Optimize the model's parameters for the best generalization performance.
