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 1: The ideal value of alpha in ridge and lasso regression relies on the particular dataset and the underlying issue. In these regression algorithms, the hyperparameter alpha regulates the degree of regularisation.

The strength of regularisation will be increased by doubling alpha for both lasso and ridge regression. As a result, the magnitudes of the regression coefficients will be reduced even more sharply as they approach zero. This implies that the coefficients for ridge regression will be more severely penalised, but none of them will be zero. In the case of lasso regression, the rise in alpha may result in the nullification of more coefficients, creating a sparser model with fewer predictor variables.

After making the update, retraining the models using the new alpha values is necessary to account for the most crucial predictor variables. The magnitude of the related coefficients can be used to gauge a variable's significance. In the case of ridge regression, no coefficients will be absolutely zero even after doubling the amount of alpha, hence all variables will still contribute to the model.

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: Different shrinkage measurements are carried out by Ridge and Lasso depending on the hyperparameter "Lambda" value. Lasso decreases the minor features during the shrinkage process. coefficient to zero, completely eliminating some feature. Therefore, in the event that we have a high number of features, this works well for feature selection. Our data set has a large number of variables, thus we will utilise lasso regression to remove some of the variables entirely by reducing the coefficient of the less significant features to zero. Additionally, we obtained nearly identical r2 scores for lasso and ridge regression with the best alpha value.

The following table lists the values for the ideal alpha and accompanying r2 score for ridge and lasso regression:

The lasso regression's optimal value of alpha is 50, and the r2 score for that value is provided below the train's R2 score, which is 0.9372405328256925.

Test's R2 rating is 0.9254664123086983.

The ridge regression's ideal alpha value is 4, and the r2 score for that value is presented below the train's R2 score, which is 0.9371096095852764.

R2 rating: 0.9253982765709686.

For ridge regression on variables chosen via lasso regression, the ideal value of alpha is 1, and the r2 score for the ideal value of alpha is provided below.

Test R2 score: 0.92319482263163

The r2 score for the ideal value of alpha shows that both the model ridge and lasso give almost same r2 score. Hence, we will choose lasso regression which will do variable

selection as well with r2 score.

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: To decide the five most significant indicator factors in the wake of barring the first top five factors recognized by the tether model, we want to retrain the model on the altered dataset. The refreshed model will think about the excess indicator factors and evaluate their significance. The five most significant predictor variables in the new model are as follows:

Variable A: After removing the top five variables, this predictor variable emerges as the most significant. It has a lot of predictive power and has a strong connection to the target variable.

Variable B: The new model also relies heavily on this predictor variable, which has a significant impact on the target variable and contributes to accurate predictions.

Variable C: exhibits a strong association with the target variable and is a useful feature in the revised model despite the absence of the original top predictors.

Variable D: In the absence of the top five previously identified variables, variable D, another influential predictor, proves to be crucial, significantly improving predictive performance.

Variable E: The updated model places a lot of emphasis on this predictor variable because it helps to identify relevant patterns and variations in the target variable.

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:

To make a model is robust and generalizable,

For training, use a diverse and representative dataset with a wide range of samples and scenarios.

Utilize strategies like cross-approval to evaluate the model's exhibition on various subsets of the information, moderating overfitting and giving a more solid gauge of its speculation capacity.

Regularization techniques like ridge and lasso regression can help the model adapt to new data and avoid overfitting.

Feature engineering and selection should be carried out with care to concentrate on relevant, meaningful, and likely to generalize well features.

Implications: An additional powerful and generalizable model will in general have a superior capacity to perform precisely on concealed information or genuine situations. It increases its reliability by lowering the likelihood of overfitting to noise or particular patterns in the training data. However, because the model focuses on capturing more general patterns rather than specific noise or outliers that are only present in the training set, increasing the model's ability to generalize may occasionally result in slightly reduced accuracy on the training data.