Subjective Questions

Question1:

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:

For ridge regression, we introduce GridsearchCV. This will allow us to automatically perform 5-fold cross –validation with a range of different regularization parameters in order to find the optimal value of alpha. For lasso, we follow a very similar process to ridge regression.

If alpha is increased, model complexity will have a greater contribution to the cost. Because the minimum cost hypothesis is selected, means higher alpha will bias the selection toward model with lower complexity. If alpha is increased, model will be simple, but with the of risk **underfitting.** There are multiple ways to determine the best predictor, compare coefficients to select the best predictor and Negative mean square error, change in R-squared value.

Question2:

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:

The optimal lambda value in case of Ridge and Lasso is as below:

- Ridge - 20

- Lasso - 0.0002

The Mean Squared error in case of Ridge and Lasso are:

- Ridge - 0.01558

- Lasso - 0.01543

- The Mean Squared Error of Lasso is slightly lower than that of Ridge

-Lasso helps in feature reduction

Lasso regression would be a better option it would help in feature elimination and the model will be more robust

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:

When adding predictors, you should keep all the original predictors in the model, even if they were not significant. Omitted variable bias can cause significant predictors to appear to be insignificant. By adding more variables, key predictors may become significant.

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:

A model needs to be made robust and generalisable so that they are not impacted by outliers in the training data. The model should be generalisable so that the rest accuracy is not lesser than the training score. Outliers which it does not make sense to keep must be removed from the dataset. This would help increase the accuracy of the predictions made by the model.

- Simpler model are usually more generic and more widely applicable.
- Simpler model require fewer training samples for effective training than the more complex ones.
- Simpler models are more robust
- Complex models tend to change widely with changes in training dataset.
- Simpler models have low variance , high bias . complex models have low bias high variance.