Question 1 :

Rahul built a logistic regression model with a training accuracy of 97% and a test accuracy of 48%. What could be the reason for the gap between the test and train accuracies, and how can this problem be solved?

Answer –

In this scenario, when test accuracy is lower than training accuracy, and have measurable difference , then this problem due to Overfit model. Has learned that some particulars help it perform better in training data that are not applicable to the larger data population and therefore result in worse performance.

To overcome the problem, if we are going to defend any research when such cross-validation methods are highly recommended, with the Regularization too.

Question 2 :

List at least four differences in detail between L1 and L2 regularisation in regression.

Answer –

L1 - Lasso Regression

1) The Lasso regression uses L1 norm for regularization. The main difference between ridge and lasso regression is a shape of the constraint region. The Lasso estimate is an estimate which minimizes the sum of square , it carries with sum of absolute error term.

2) Lasso regression helps for feature selection.

3) L1 penalty for non-sparse data will give you give you a large estimation error.

4) if lambda is zero then we will get back OLS whereas very large value will make coefficients zero hence it will under-fit.

Equation –

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p} |\beta_j|$$

L2 - Ridge Regression

1) The ridge regression uses L2 norm for regularization. The ridge regression gives an estimate which minimize the sum of square error

2) The most important one is that it penalizes the estimates. It doesn't penalize all the feature's estimate arbitrarily.

3) Here, if lambda is zero then imagine we get back OLS. However, if lambda is very large then it will add too much weight and it will lead to under-fitting. Having said that it's important how lambda is chosen. This technique works very well to avoid over-fitting issue.

4) If estimates of BETA value are very large, then the SSE term in the above equation will minimize, but the penalty term will increases. If estimates BETA values are small, then the penalty term in the above equation will minimize, but, the SSE term will increase due to poor generalization.

Equation –

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

Question 3 :

Consider two linear models:

L1: y = 39.76x + 32.648628

And

L2: y = 43.2x + 19.8

Given the fact that both the models perform equally well on the test data set, which one would you prefer and why?

Answer –

Since linear regression gives output as continuous values, so in such case we use mean squared error metric to evaluate the model performance. Also, COD (Coefficient Of Determination) value at higher side can consider.

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 define the quality of a model, it should being accurately measure its predication error is of key importance. However, there are techniques which measures the error.

Normally true predication error identifies with the combination of training error and training optimism.

Overfitting is another concept which need to consider which mainly involves the to identify the model complexity. Here model complexity involves the model predication error, training error, optimism and predication error for new data. So ideally tradeoff between the bias-variance.

To measure the methods, normally Adjusted R Square, AIC,BIC and likelihood, cross validation, resampling etc. are the key points that must be considered along with the above mentioned overview , while to gain confident that model is robust and generalisable.

Question 5:

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 –

They seems to perform similarly for the data, but lambda is not optimal . Below are the few points to compare the ridge and lasso :-

1) Often neither one is overall better.

2) Lasso can set some coefficients to zero, thus performing variable selection, while ridge regression cannot.

3) Both methods allow to use correlated predictors, but they solve multicollinearity issue differently:

4) In ridge regression, the coefficients of correlated predictors are similar;

5) In lasso, one of the correlated predictors has a larger coefficient, while the rest are (nearly) zeroed.

6) Lasso tends to do well if there are a small number of significant parameters and the others are close to zero, when only a few predictors influence the response.

7) Ridge works well if there are many large parameters of about the same value, when most predictors impact the response.