Assignment Part -II

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?

This is a normal symptom of over fitting. Model has effectively memorised the training data set.
This can be solved through regularisation. It is a process of deliberately simplifying the models to achieve the correct balance between keeping the simple model and yet not too naive. For Regression this involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.

Question 2

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

A regression model that uses L1 regularization technique is called **Lasso Regression** and model which uses L2 is called **Ridge Regression**.

	L1	**L2**
Penalty term	L1 adds "absolute value of magnitude" of coefficient as penalty term to the loss function.	L2 adds "squared magnitude" of coefficient as penalty term to the loss function.
Computational cost	Lasso is computationally more expensive. Since, it does not convert into a nice invertible function, it is to be solved using an iterative process which has significantly more computational requirement compared to ridge.	Ridge regression demands a simple tweak to the simple linear regression solution and can be converted to an invertible matrix and can be solved using matrix operations and thus has significantly lower computational cost associated to it.
Model parameters	Most important benefits of lasso regression is that in model parameters, such that the lesser important features co-efficient becoming zero. In other words L1 regression indirectly performs feature selection.	Ridge, on the other hand, reduces the co-efficient to arbitrarily low values, though not zero.
Output	Sparse outputs	Non- sparse outputs.

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?

I would **prefer L2 over L1** because L1 is more complex model compared to L2. Though 43.2 is greater than 39.76, there is too many bits of precision in L1 i.e. 32.648628 compared to 19.8.
Model becomes complex when -messy, too many bits of precision, large numbers, etc. are the part of coefficients in the model.
And when simple has major advantages over complex model like – simple model are usually more generic, requires fewer training samples and are robust. It is better to prefer simple model i.e in this case L2 over L1.

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?

By considering simple model over the complex model will make sure that a model is robust and generalizable. There is no specific definition of complexity of a model. However there are few ways of looking the complexity of a model:

1. No. of parameters required to specify the model completely. Lesser the parameters, simpler the model is.
2. The degree of function required to specify the model completely. Lesser the degree of the function, simpler the model is.
3. The depth or size of a decision tree. Lesser the depth, simpler the model is.
4. Lesser the size taken by the best – possible representation of the model, lesser the complexity.

Having said that, simple models have low variance, high bias and complex models have low bias, high variance. “Variance” is how sensitive is the model to input data and “Bias” is the deviation from the expected, ideal behaviour i.e. how much error the model is likely to make in the test dataset.

Although, in practice, we often cannot have a low bias and low variance. As the model complexity goes up, the bias reduces while the variance increases, hence the trade- off is required and the phenomenon is referred to as the bias-variance trade-off.

The implications on the accuracy of the model: Robust and generalizable models are simple models are not very accurate compared to their counterpart – complex model. And that is because simple models have low variance, high bias. Bias qualifies for the accuracy of the model - how accurate is the model likely to be on future (test) data. This is because the simple model doesn’t memorize entire training dataset, so when the future (test) data appears – it is bound to make some error.

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?

The final variables from both lasso and ridge regression models are same for me.
Rsquared score from ridge – Train is 93.07% and test is 89.40%
Rsquared score from lasso- Train is 92.97% and test is 89.34%. These scores are also almost same.

However, lasso gives an additional advantage of feature selection. That is least important and multicollinearity variables are eliminated by co-efficient becoming zero.
In the present model, after rfe with 50 variables, lasso automatically eliminated 10 of them by making co-efficient to zero. So rest of them being the same, I would choose to apply lasso regression , given computational cost is not an issue.