Question-1:

Rahul built a logistic regression model having a training accuracy of 97% while the test accuracy was 48%. What could be the reason for the seeming gulf between test and train accuracy and how can this problem be solved.

Ans: Feature selection could be a reason. Also if the data clean up and EDA process if not done properly can cause this issue, It is also advised the processing of categorical variable is a necessity in making the model quite accurate.

Question-2:

List at least 4 differences in detail between L1 and L2 regularization in regression.

A regression model that uses L1 regularization technique is called Lasso Regression and model which uses L2 is called Ridge Regression. The key difference between these two is the penalty term. Ridge regression adds "squared magnitude" of coefficient as penalty term to the loss function.

L1-norm loss function is also known as least absolute deviations (LAD), least absolute errors (LAE). It is basically minimizing the sum of the absolute differences **(S)** between the target value (**Yi**) and the estimated values (**f(xi)**):

$$S = \sum_{i=1}^{n} |y_i - f(x_i)|.$$

L2-norm loss function is also known as least squares error (LSE). It is basically minimizing the sum of the square of the differences **(S)** between the target value (**Yi**) and the estimated values (**f(xi):**

$$S = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

L1 is Robust while L2 is not very robust. L1 is unstable and L2 is stable solution. L1 possibly leads multiple solutions whereas L2 has always one solution,

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 dataset, which one would you prefer and why?

Ans: It is clear than L2 equation is much simpler and if your Linear equation is simple you model would be simpler. Easier to compute.

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?

Ans: to make a model robust and generalizable it should be trained on a good amount and diverse amount of data, since diversity in data would help model to be more give more acceptance to test data inputs and model would behave and perform well. Log scale transformation method is often used to reduce the variability of data including outlying observation.

Question-5:

As you have determined the optimal value of lambda for ridge and lasso regression during the assignment, which one would you choose to apply and why?

Ans: Yes I have choosed the optimal value of lambda for ridge and lasso regression, since the optimal value was clear from having the plot between alphas and validation in Ridge. Also lasso performs variable selection and I have choosed the tuning parameter lambda by cross validation.