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.

A: The gulf between the test and train accuracy is due to the model being:

- Highly over fit to the training data
- Using a very complex and a high degree function to get an exact fit
- The model has memorized the data set

Rahul needs to:

- Create a more generic models to perform better on unseen datasets
- Employ Regularization methods to reduce the complexity
- Employ Ridge and Lasso regularization to get better results

Question-2:

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

A: The Differences are as follows:

- Lasso Reduces redundant feature coefficients completely 0 there for can be used for feature elimination
- For Ridge the term in the cost function is $\lambda \sum \beta^2$ while for lasso $\lambda \sum |\beta|$
- Ridge regression almost always has a matrix representation for the solution while Lasso requires iterations to get to the final solution.
- The ridge and the lasso make different assumptions:
    - In ridge, the coefficients of the linear transformation are normal distributed
    - In lasso they are Laplace distributed
- The ridge is a bit easier to implement and faster to compute,

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?

A: The models are almost the same but we will choose the simpler model L1: *y = 39.76x + 32.648628* as this the simplest model available with the lowest coefficient necessary.

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?

A: we can make a model more robust and generalizable by:

- Ensuring the we eliminate multicollinearity
- Use RFE to reduce features
- Use Ridge or Lasso Regression

There is tradeoff between and making a model more generalizable and accuracy.

- The model will lose some accuracy as we make it more generalizable, but it will not have a radically different score from the training as in the case of a highly complex and over fitted model.

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?

A:

The optimum value of Lambda was found to be

- Ridge:10
  - AIC: 10154
  - BIC: 11420
- Lasso:100
  - AIC: 10135
  - BIC: 11401

The models we have generated are very close so we will combination of lasso and Ridge as follows:

- Lasso will perform the feature reduction
- Ridge being much faster will be able to give us better performance with out a drop in accuracy