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 → The model built by Rahul is showing symptoms of overfitting, where it has test and train accuracy differing by a large amount. This problem of overfitting can be easily fixed by regularization. It can be solved by Ridge and lasso both.

Question 2)

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

Answer → 1) Lasso --- It is computationally intensive.

Ridge – It is not so computationally intensive.

2) Lasso – Regularization term consists sum of absolute values of the coefficients.

Ridge -- Regularization term consists of sum of squares of the coefficients.

3) Lasso – It can be used for feature selection.

Ridge – It can't be used for that purpose.

4) Lasso – It requires iterations to get to the solution.

Ridge – It almost always has a matrix representation for the 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 data set, which one would you prefer and why?

Answer → L2 is preferable because it is simpler and more generalizable and robust.

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 → The model should be as simple as possible ,though its accuracy will decrease but it will be more robust and generalisable. It can be also understood using the Bias-Variance trade-off. The simpler the model the more the bias but less variance and more generalizable.

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→ We should go with lasso because, the model will be more Generalisable with less features.