Q1.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?

Ans. This is because of overfitting. The model memorized the training set thoroughly and hence it is not as generalized. It has specialized to the structure in the training dataset. This problem can be solved by decreasing the model complexity so that the accuracy increases. Using cross validation is also better, and using multiple runs of cross validation is better again. This will take time but we will get the best estimate of the models accuracy on unseen data. For regression, Ridge or Lasso regularisation can be used to constrain the complexity (magnitude of the coefficients) during the training process.

Q2.List at least four differences in detail between L1 and L2 regularisation in regression.

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

1. Ridge Regression uses the sum of the square of coefficients as the regularisation term. Lasso Regression uses the sum of the absolute value of coefficients as the regularisation term.
2.Lasso regression can be used for feature selection while ridge regression cannot be.Lasso shrinks the less important feature's coefficient to zero thus, removing some feature.
3.Lasso regression requires more computation power than Ridge since it is to be solved using an iterative process which has significantly more computational requirements compared to ridge regression which demands a simple tweak to the simple linear regression solution.
4.L1(Lasso Regression) can yield sparse models while L2 (Ridge Regression) doesn't. Sparse model is a great property to have when dealing with high-dimensional data.

Q3.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?

Ans. I would prefer L2 as L1 appears to be more complex as it goes up to the 6th decimal.A predictive model has to be as simple as possible, but no simpler.For example the expression (0.552984567 $*$ x^2 + 932.4710001276) could be considered to be more 'complex' than say (2x +3x^2 +1), though the latter has more terms in it.

Q4.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. A model is robust and generabisable if it is not complex but not too simple.Simpler models require fewer training samples for effective training than the more complex ones and are consequently easier to train.
Ideally the model must be immune to the specifics of the training data provided and rather somehow pick out the essential characteristics of the phenomenon that is invariant across any training data set for the problem.
Simple models have low variance, high bias and complex models have low bias, high variance. Here 'variance' refers to the variance in the model and 'bias' is the deviation from the expected ideal behaviour.
Simpler models make more errors in the training set.So training accuracy reduces.

Q5.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?

Ans. I will choose ridge regression because it gave a higher r^2 square than lasso regression so it has a better chance of predicting the sale price accurately. Also ridge is faster as it is computationally less intensive.