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 reason for the gap between the test and train accuracies is because the model is overfitting. Overfitting happens because of building complex models which memorizes the data instead of intelligently learning the underlying trends .The problem can be solved by making the model more generic by reducing the complexity. This can be achieved using Regularization.

Question 2

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

Answer:

L2 Regularization(Ridge)	**L1 Regularization (Lasso)**
Regularization term contains sum of the squares of the coefficients	Regularization term contains sum of the absolute values of the coefficients
Computational cost is less	Computationally more intensive so computational cost is more
Ridge regression only shrinks the coefficients but includes all the coefficients and features in the model	It shrinks the coefficients of redundant variables to zero and helps in feature selection
If there are correlated variables ridge regression is preferred	If there are correlated variable it retains only one variable and shrinks the coefficients of other correlated variables to zero which might lead to loss of information resulting in less accurate model.

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: I would prefer L2 ,as the coefficients are **of** less precision, so the model would be simple.It will save the computational cost

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:

We can make sure that model is robust and generalisable by reducing the complexity.A model should be made as simple as necessary but no simpler because if it is made extremely simple,it will increase the bias,which inturn effects the model.

The implications of the same for the accuracy are , the accuracy of the model might decrease.There are chances of bias increasing by making the model extremely simple.This is because extremely simple models are likely to fail in predicting .

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: I would choose the lambda for lasso regression because I need to do the feature selection.

I need to find out the significant features which effect the sale price of the houses.