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?

Answer:

If the train Accuracy is more and the test accuracy is very less than the model is the model is said to be over fitted. This occurs if the model shows low bias and high variance. As overfitting reduces the generalizability model will perform very poorly on the unseen data.

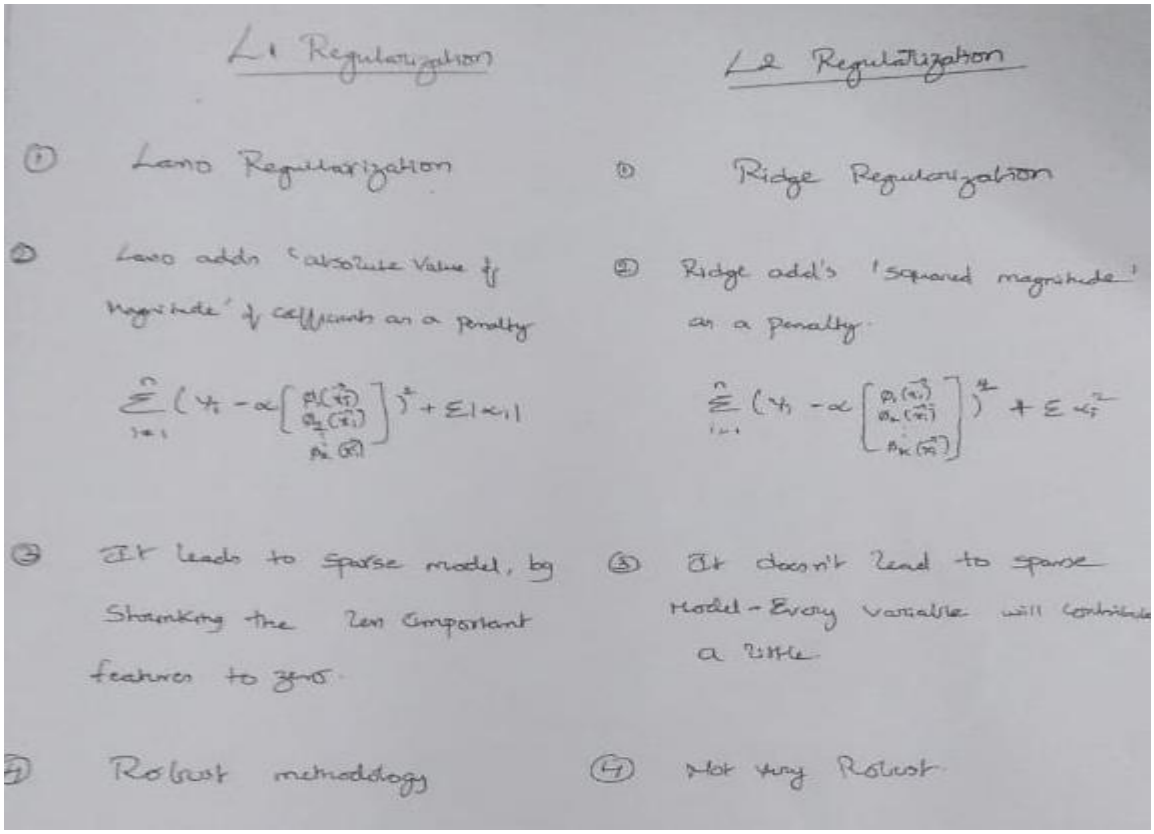
Ways to Overcome the above problem:-

1. **Cross Validation –** this can be a powerful preventive measure to solve overfitting. In K-fold algorithm we partition the data to k- subsets or k folds then algorithm is trained on k-1 folds.
2. **Regularization –** Lasso and Ridge regularization can help in resolving overfit.
3. **Increasing the Train data size –** Increasing the train data size can help in solving the problem to certain extent.

Question-2:

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

Answer:



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?

Answer:

As it is said both the models perform equally well it would be better to choose L2(y = 43.2x + 19.8) model as it is simpler one. It would be easy to work with simpler models as they takes less space and time.

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:

For the model to be more robust and generalizable we should take care that the model should be of low variance and low bias. Bias variance trade-off is a serious problem in machine learning which needed to be taken care by capturing the regularities in the data enough there by making the model more generalizable and accurate.

Note: - Regularization can help in making the model generalizable (i.e. low variance and bias) by imposing a penalty on the cost function there by providing a solution to over fitting.

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?\

Answer:

Lasso will make all the unwanted features to zero.

Ridge on the other hand will reduce the impact of unwanted features by making them close to zero.

Hence Lasso can be used in predicting better on the Linear models.