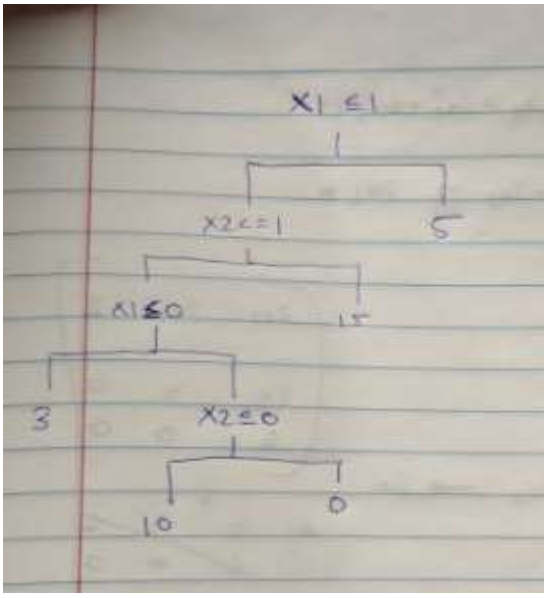
Kunal Mehta
50352960
kunalkam

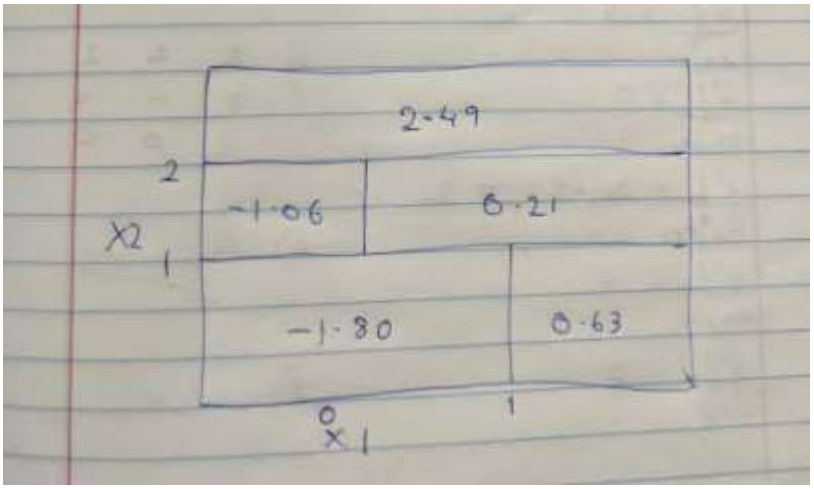
Quiz 5

1. Answers:

a)



b)



2. Answers:

a) The predictors in this case were chosen based on their correlation with the response variable based on all the samples of the data. Moreover, since it is 5-fold cross validation, the data that is used for testing barring the first iteration is data that the algorithm has already been trained on once, therefore, the error rate is underestimated by the model, since error rates over all the folds are averaged out.

b) Chiefly, the selection of predictors on the basis of correlation with response is not a good way to choose predictors over all samples. I'd divide the data into training and test before doing so. Also, I'd make sure that the division into 5 folds is done with random samples entering each fold, not in any order.