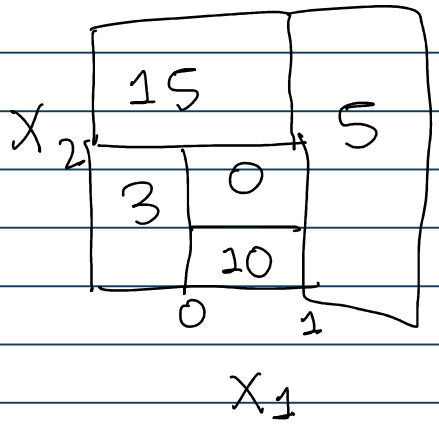
HW 5

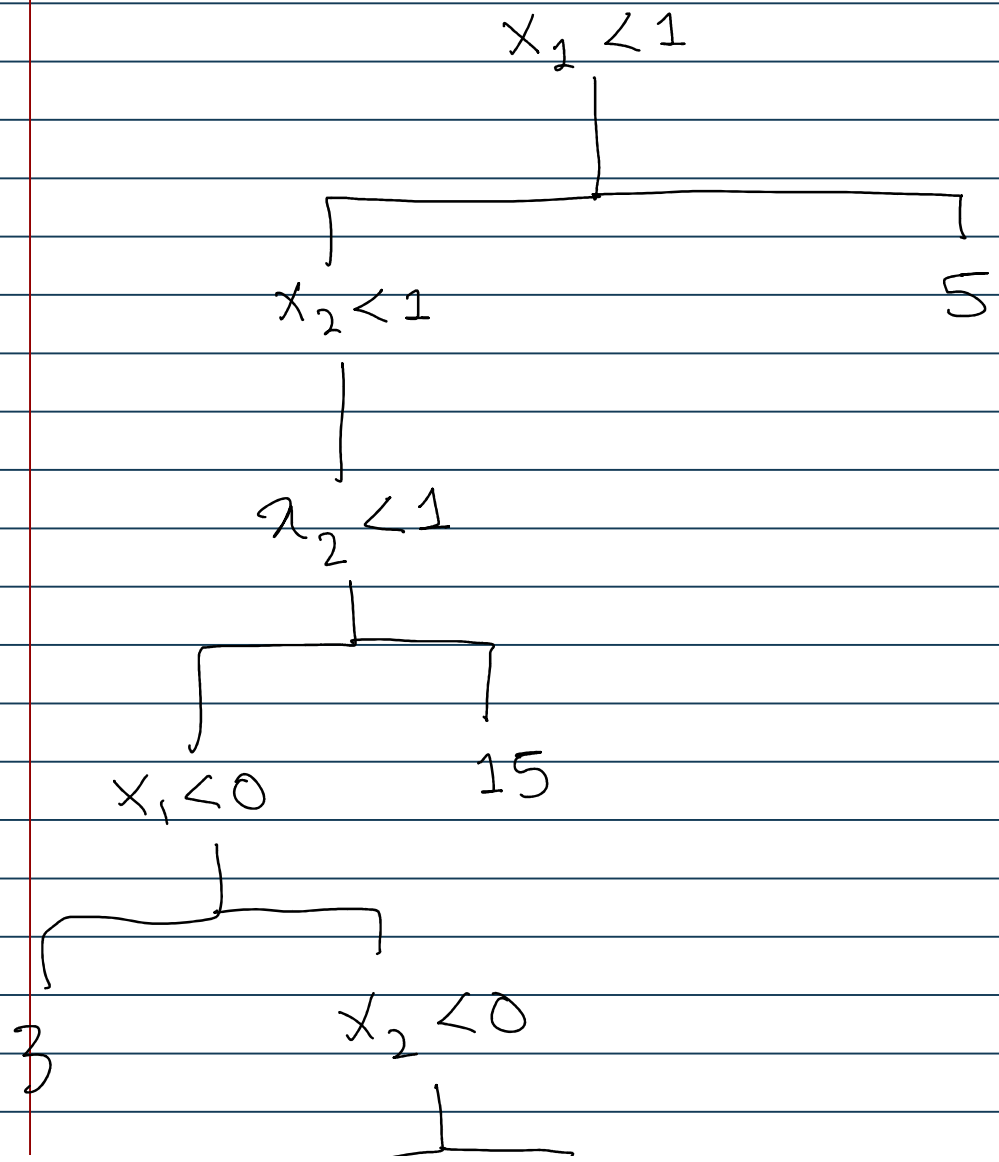
Sunday, October 27, 2024 3:26 PM

Part - A

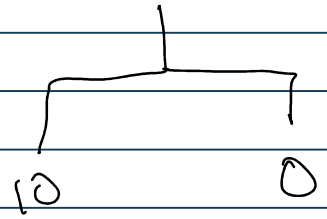
(1)

Figure :

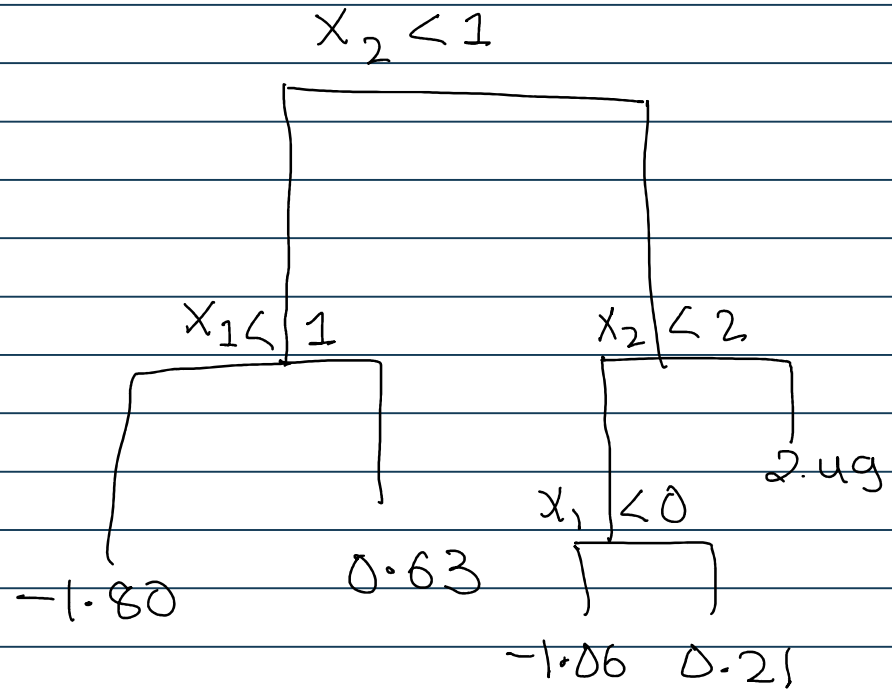


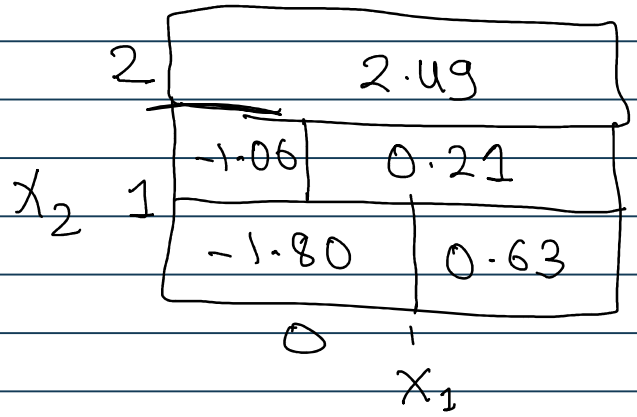


1



2





Q.3) Yes, It is a good idea to try decreasing max_depth if a Decision Tree is overfitting. Doing so, limits the number of splits it can make which helps to reduce the complexity of model.

4) The computational capacity of training a decision tree is given by:

$$O(n \times m \log m)$$

Therefore, if we were to multiply the training set by 10, the training time will be multiplied by

$$K = \frac{n \times 10m \times \log 10m}{n \times m \times \log m}$$

$$= \frac{10 \log 10m}{\log m}$$

if $m = 2 \times 10^6$

$$K = \frac{10 \times \log(10 \times 2 \times 10^6)}{\log(2 \times 10^6)} = 11.59$$

∴ The time taken by 20 million instances will be $2 \times 11.59 = 23.17$ hours