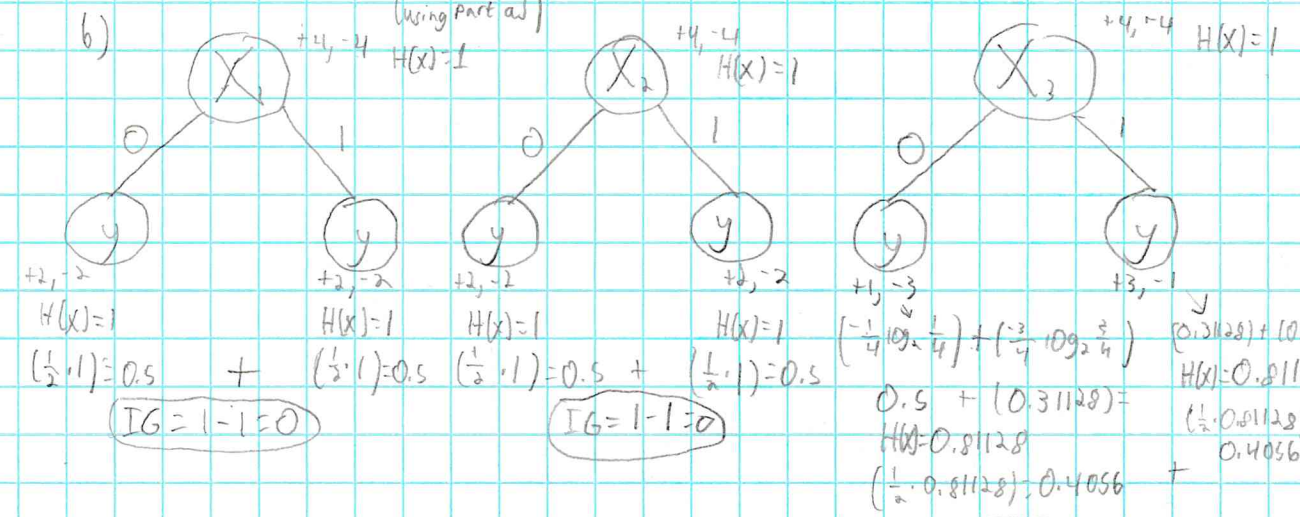
1) a) $\left(\frac{1}{2}\log_2\frac{1}{2}\right)+\left(\frac{1}{2}\log_2\frac{1}{2}\right)$

$-0.5(-1)+(-0.5(-1))$

$0.5+0.5=1$

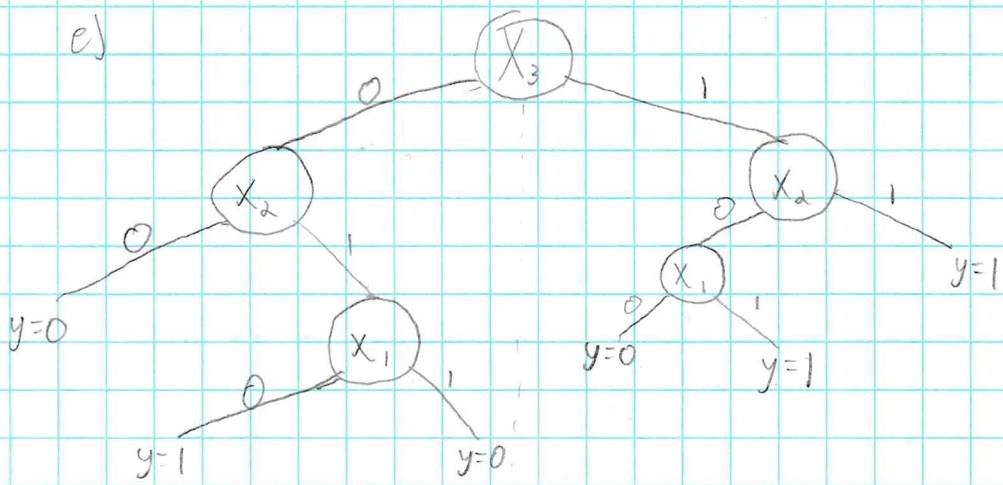
b)



c) Split the decision tree with root X_3 because it has the highest information gain

d) The stopping criteria is all branches have the node have the same result or there are no training examples for a particular node

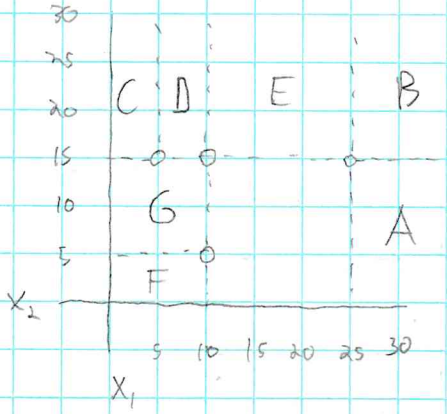
e)



f) 0%

g)/h) Based on the decision tree, those values can't be predicted accurately, therefore the model is overfitting because it doesn't have data for all cases

2) a)



3) b) 7

c) 26

d) PPE

e)

$$0.782 - \left(\left(\frac{42}{185}(0.845)\right) + \left(\frac{141}{185}(0.395)\right)\right) = 0.2891$$

f) 12 attributes were tested. This means there's no training data for the unincluded attributes

g) [[2, 2
3, 3]]

It splits about 50-50. I would rather be given a false positive and then find out I didn't have it then be told I don't have it and find out I do later. It also allows people to get better treatment that way.

h) It does overfit because the testing data isn't perfect but the training data is.