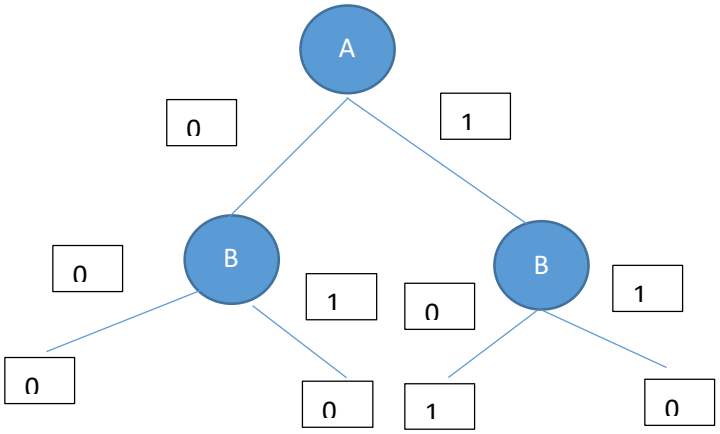
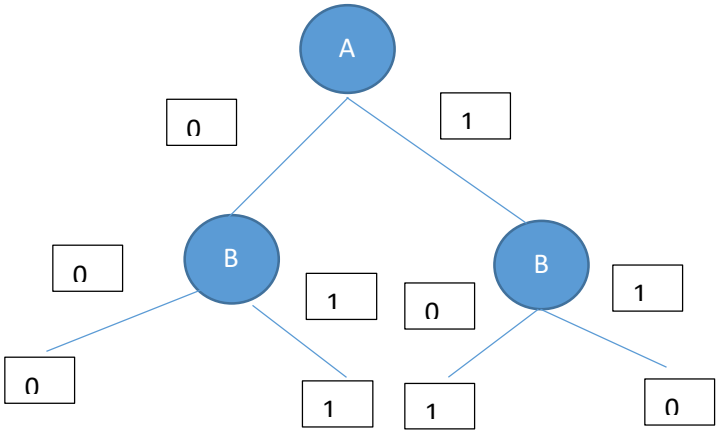
1a)



1b)



2a) w1 = 1 w2 = 1 w3 = -0.9

2b) w1 = -2 w2 = 3 w3 = 1

2c) Not possible

3) Precision = 30/(30+30)

Recall = 30/(30+20)

4)

a. False. For some functions (e.g. XOR) we need to use more nodes than features

b. i) true since the data is linearly separable ii) False since no guarantees on validation set

c. Yes. In decision trees prediction is quite fast (need to traverse the depth of the tree at most)

d. Networks seems the best choice since the features are continuous and the decision boundary is on-linear

5) Price correctly classify all the instances. The decision tree consists of just one node (Price)

6) Since each unit is a linear unit, connecting several will still produce a linear function. Specifically, the linear combination of linear functions is a linear function. This means the output of the neural network (that may consist of several linear units) is still linear. Therefore it is no more expressive than a single unit. However, for sigmoid units, the output of a neural network is a non-linear function and therefore more expressive than a single unit.