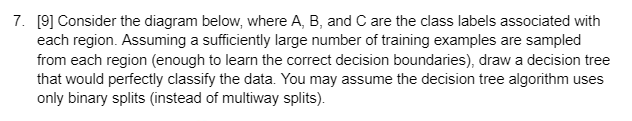
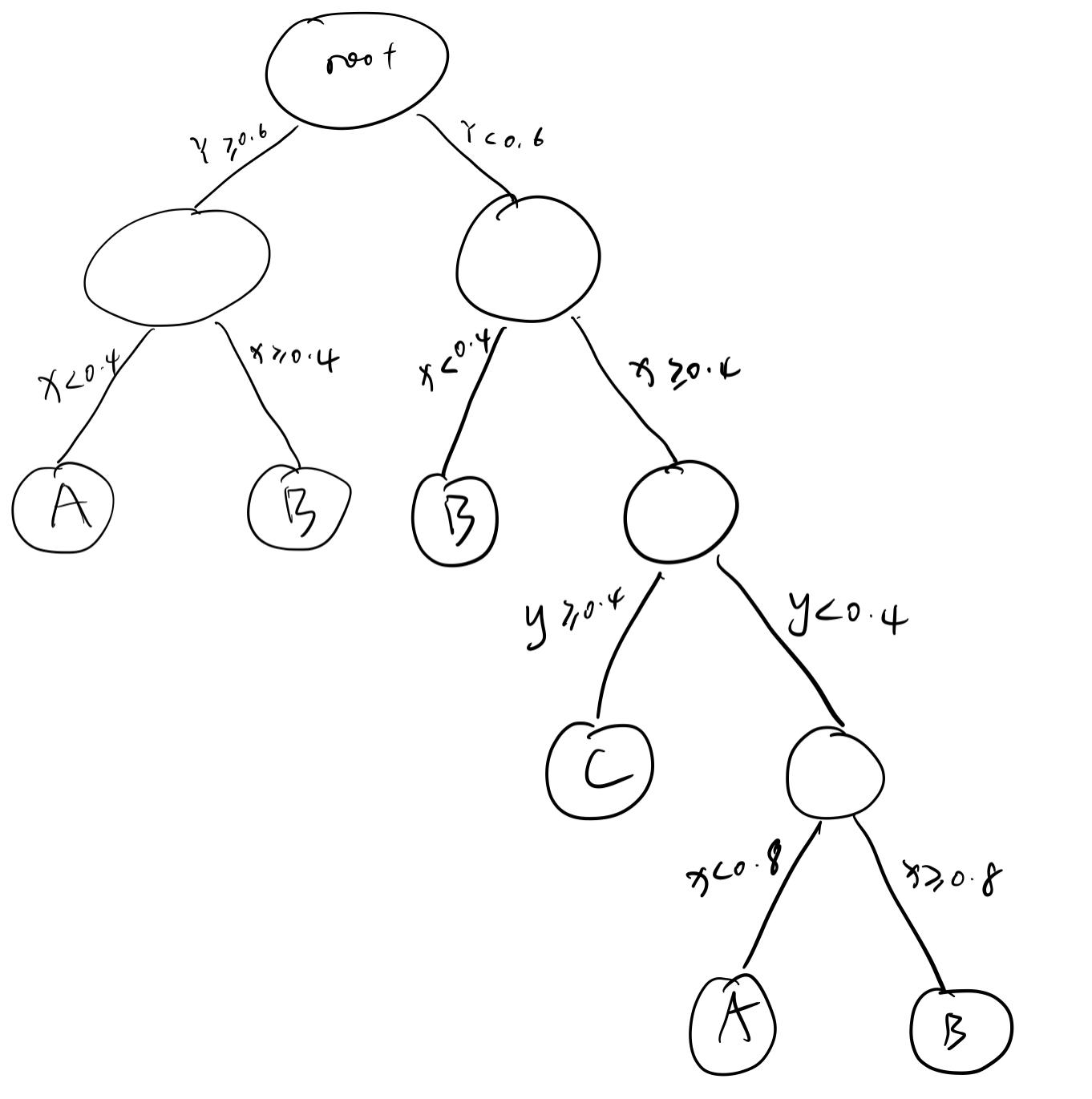
Chapter 3:

4. What is the worst-case time complexity for training a decision tree with N observations and M features (attributes)? Include finding the best splits for each attribute. First, consider only binary attributes. Then, what if you have nominal attributes with three options for each and you use multiway splits? What if you have nominal attributes with three options for each and you try all binary splits (group the options into all possible sets)? Showing a concrete example is helpful. I.e., make up some data and specify N, M, …

* The depth of the tree in the worst case is O(N) rather than O(NlogN) since each split only split one node out.
* We have to loop through N possible places to split on.
* If there are c classes for categorical variables, you have 2^c -1 subsets available to choose, or 2^c -1 ways to split
* If we have continuous features, finding the best splits for features requires a loop through the features, the worst case is O(M)
* So it takes O(N^2\*M) for continuous and O(MN2^c) for categorical.
* This is for binary attributes split.
* If we are dealing with a multiway split, we can combine some similar attributes to make them first a binary split, and then from then on apply the method for the binary split.

Chapter 4:

6.a

P(graduate | smoke) = P(smoke | graduate) \* P(graduate) / P(smoke)

= 0.23 \* 0.2 / (0.15 \* 0.8 + 0.23 \* 0.2) = 0.277

6. b

Undergraduate

6.c

Still undergrad

12. a

x= 5

1NN: +

3NN: -

5NN: +

9NN:-

12. b

1NN: +

3NN: +

5NN: +

9NN:+