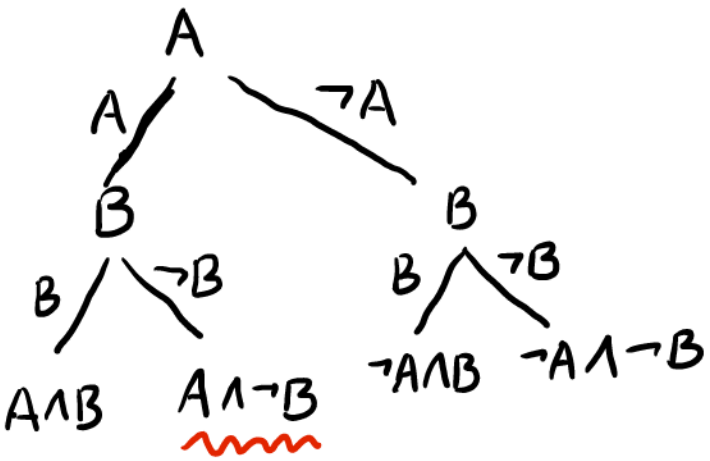
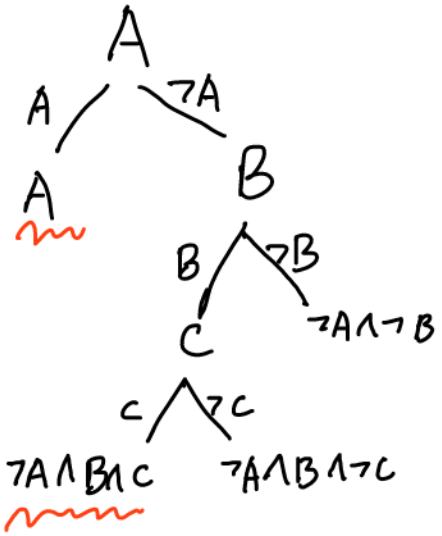
Question 1: (Mitchell, Exercise 3.1, page 77).
Give decision trees to represent the following boolean functions:

(a) $A \wedge \neg B$

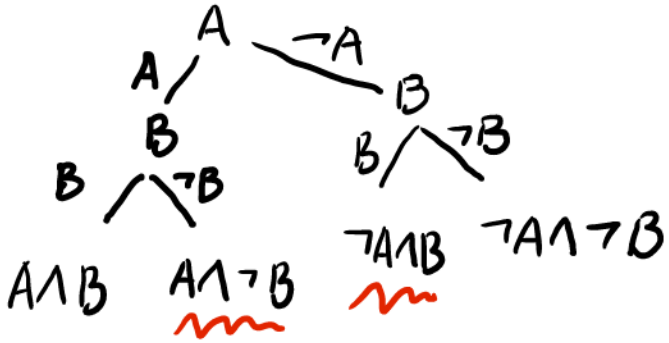


(b) $A \vee [B \wedge C]$

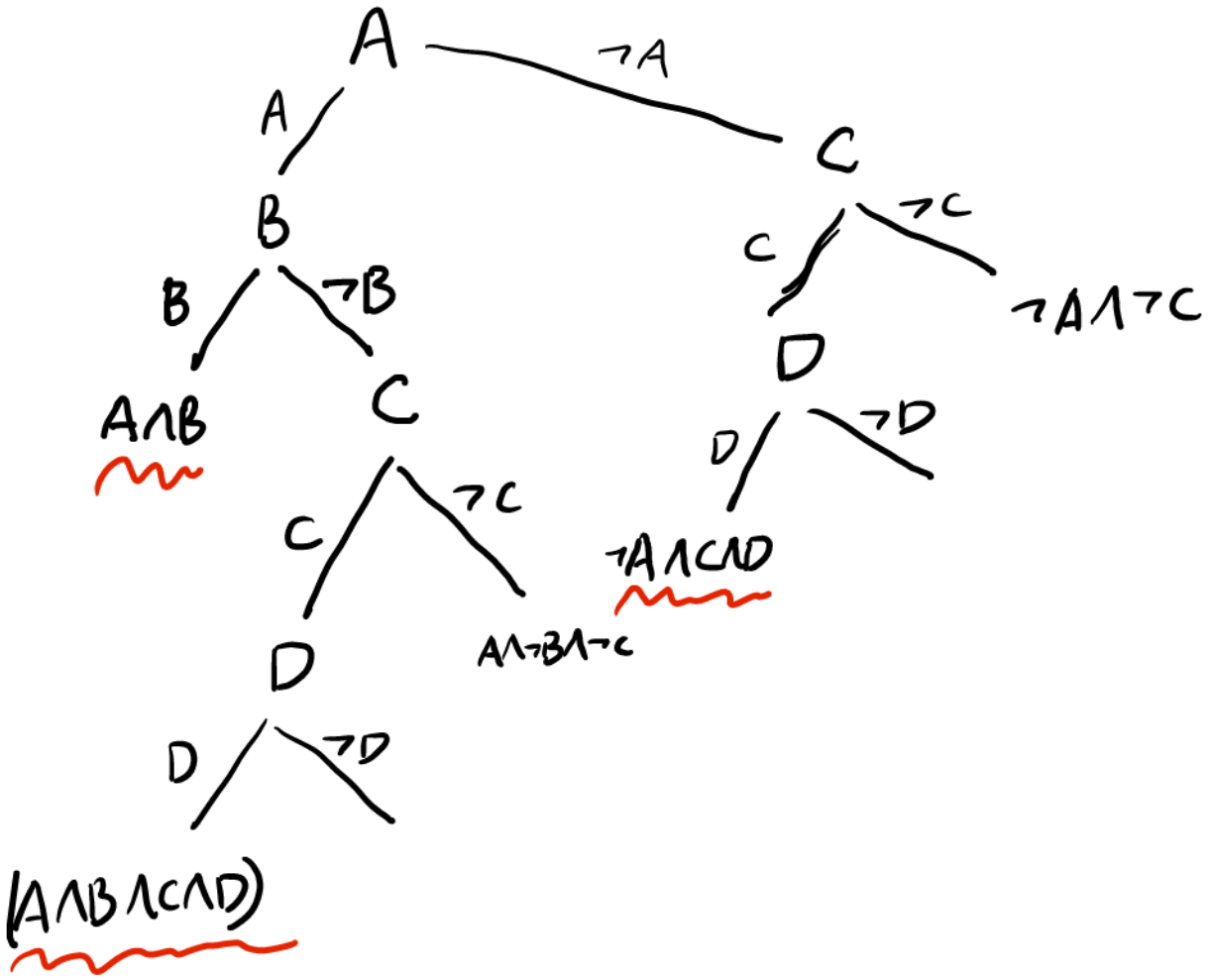


(c) A XOR B

$(\neg A \wedge B) \vee (A \wedge \neg B)$

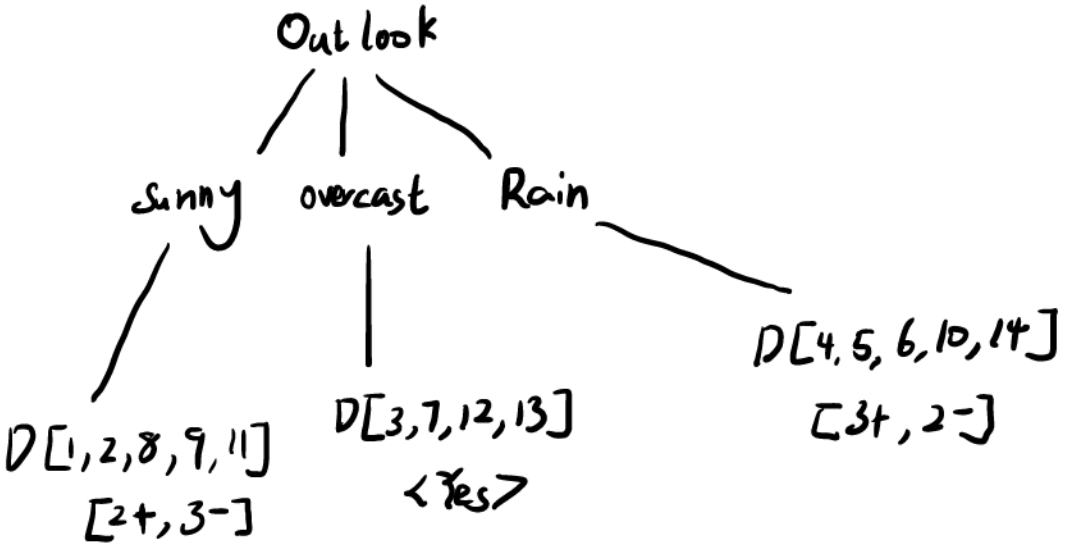


(d) [A∧ B] ∨ [C ∧ D]



Question 2: Consider the samples in the Play-tennis dataset from Table 3.2 in Mitchell's textbook (linked above). If you calculate the information gain for all of the attributes of this set, you will observe that the attribute "Outlook" has the largest information gain, which is equal to 0.246. Therefore, the attribute "Outlook" is the best heuristic choice for the root node.

1. List the labels of the new tree branches below the root node.



For branch outlook = Sunny. The labels is {[D1, D2, D8] =No, [D9, D11] = Yes}

For branch outlook = overcast. The labels is {[D3, D7, D12, D13] = Yes}

For branch outlook = Rain. The labels is {[D4, D5, D10] =Yes, [D6, D14] = No}

2. Which partition of the data will be assigned to each branch by ID3? Please list the sample IDs that will be assigned to each branch.

The partition sample for each branch is data ID for outlook =Sunny, the ID list is [1, 2, 8, 9, 11].

The partition sample for each branch is data ID for outlook =Overcast, the ID list is [3, 7, 12, 13].

The partition sample for each branch is data ID for outlook =Rain, the ID list is [4, 5, 6, 10, 14].

Q2

3. Calculate the information gain for the remaining attributes in each branch, and determine which attribute will be chosen as the root of the sub-tree in each branch.

1.
For branch where outlook = Sunny:

Gain(Sunny, humidity) =0.97
Gain(Sunny, Temperature)=·0.57
Gain(Sunny, Wind)=0.020

So, for this branch, next split feature is humidity.

2.
For branch where outlook = Overcast:

Because all sample labels in this branch is Yes. So, there is no need to split.

3.
For branch where outlook = Rain:

Gain(Rain, humidity) =0.02
Gain(Rain, Temperature)= 0.02
Gain(Rain, Wind)=0.97

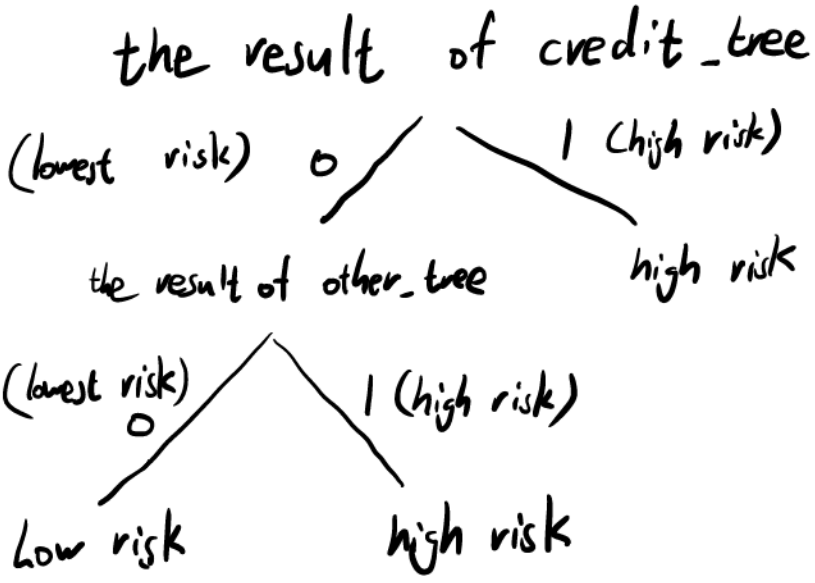
So, for this branch, next split feature is wind.

Question 3: Suppose a bank makes loan decisions using two decision trees, one that uses attributes related to credit history and one that uses other demographic attributes. Each decision tree separately classifies a loan applicant as "High Risk" or "Low Risk." The bank only offers a loan when both decision trees predict "Low Risk."

1. Describe an algorithm for converting this pair of decision trees into a single decision tree that makes the same predictions (that is, it predicts non-risky only when both of the original decision trees would have predicted non-risky).

Assume the decision tree for credit history is credit_tree.
The decision tree for other demographic attributes is other_tree.

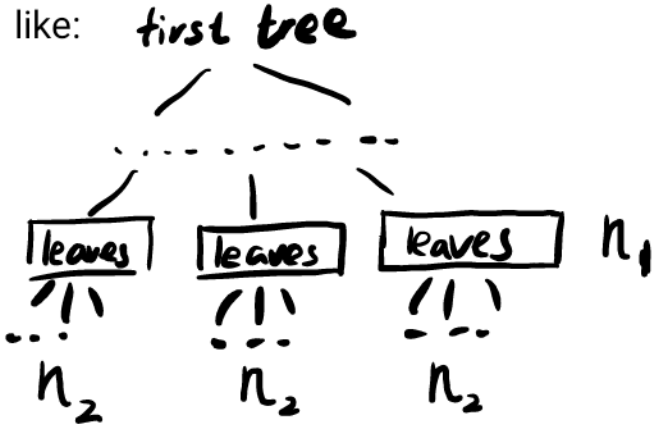
The Algorithm is:



2. Let n_1 and n_2 be the number of leaves in the first and second decision trees, respectively. Provide an upper bound on n, the number of leaves in the single equivalent decision tree, expressed as a function of n_1 and n_2.

Answer:

When we consider the single equivalent decision tree is extreme case. The tree branch of the tree segmentation is based on the feature of the first tree, and then the feature of the second tree. In this condition, the single tree is like:



So, the upper bound n is n1 * n2.