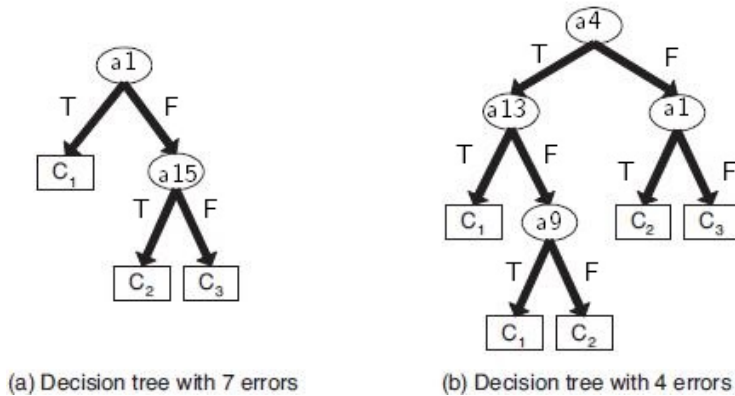


Chapter 3

Loosely based on Exercise 10 [difference between prior points and 3] Consider the decision trees shown in Figure 4.3. Assume they are generated from a data set that contains 16 binary attributes and 3 classes, C_1 , C_2 , and C_3 . Let the data contain total of $n = 200$ data points classified by the trees. Compute the total description length of each decision tree according to the minimum description length principle.



- Each tree is described by tree parameter prefix(number of nodes, number of attributes that includes classes, branching degree, sample size), and description of nodes given in lexicographic order (=BFS order)
- Each node is described by tuple (attribute number, number of children (denoted b), splitting value $_1, \dots$, splitting value $_b$)
- All items in the tree description are encoded using gamma encoding (i.e. if k -bit binary number represents a decimal number that needs to be encoded then gamma code of this number contains $2k + 1$ bits of gamma encoding that will be generated).

1 pt Show left tree cost as function of sample size n

1 pt Show right tree cost as function of sample size n

3 pts Which one of these trees is MDL-better for $\delta = 0.99$ at given sample size?

Hint: Use MDL formulas slide 2 lecture 7-2. If necessary, solve equations (use R, Mathematica or Wolfram alpha).