**Homework 2 – Assignment**

1. I mentioned in lecture that the number of possible decision trees is very large. How many decision trees exist with n binary attributes? Here is way to think about the problem.

* Suppose you have one binary attribute. Then there are 2^1=2 possible values for the attribute and each of those values can be mapped to 2 outputs, so there are 4 decision trees.
* Suppose you have two binary attributes. Then there are 2^2=4 possible values for the pair of attributes, and each value can be mapped to 2 outputs, so there are 2^4=16 decision trees.
* Now suppose you have n attributes. How many possible decision trees are there? Please justify your answer.

Ans: The Decision Trees that exist with n binary attributes, supposing that we add everything and add everything up we have 2^(2^n) possibility of decisions trees.

**Justification:**

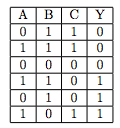
 **2^n:** This represents the number of possible values for the n binary attributes.

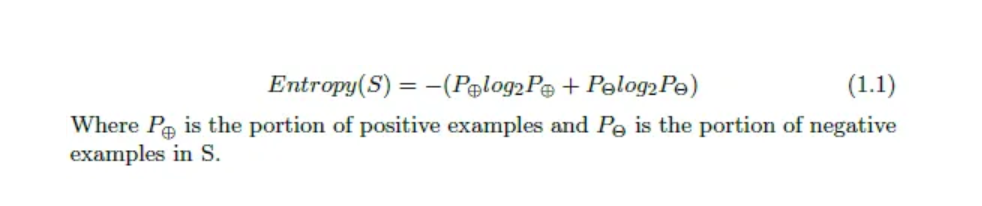
 **2^(2^n):** This represents the number of possible mappings between these values and the two possible outputs (true or false).

Hence, based on scenario we explore 2^1 with 1 binary attribute we have 2^1 = 2 possible values and 4 decision tree if 2^2 we have 4 possible values, and translate to 16 decision trees. Thus, we have 2^(2^n) possible values and creates an infinite number of decision trees. Furthermore, we will not try to study all or explore all decision tree in real-life applications as it is too complex and require huge computing space.

1. Consider the following training set with features A, B, C, and target/label Y.
2. What is the entropy of the output Y?
3. Using the information gain criterion, what is the first node you would split at? Explain clearly why?
4. Using the information gain criterion, complete the learning of the decision tree for this dataset. Draw the decision tree and comment if the tree is unique.

**Questions 2 Answers:**



2a) 

Hence based on this information – dataset we can calculate the entropy of the dataset, which is Positive instance(0.5) and negative instance is (0.5), and hence the entropy value is = 1 max value(Uncertainty).

Ans: Calculation Steps:

P(Y = ‘0’) = 0.5

P(Y = ‘1’) = 0.5

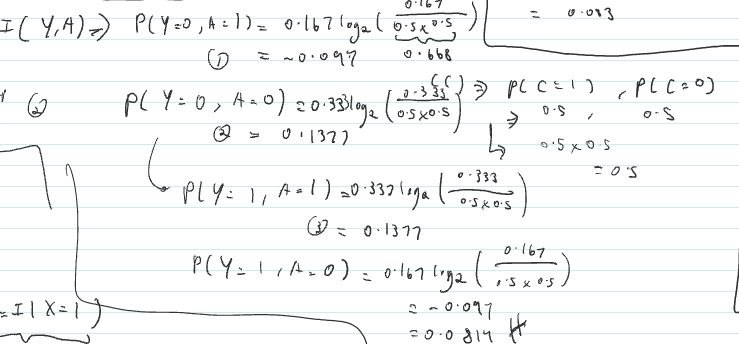
Entropy of output Y = -0.5log20.5 – 0.5log20.5

= 1

2b) The Information gain criterion will reduce entropy after the first node is split at. Given this, compute the information gain value = Entropy – [Average Entropy (Children)]

Ans: The node will split at either **Node A or Node C** as they are the two highest values out of the three nodes based on the information gain criterion. The higher the information gained the more that the attribute contains the label value. Hence, it will split at nodes A and C first.

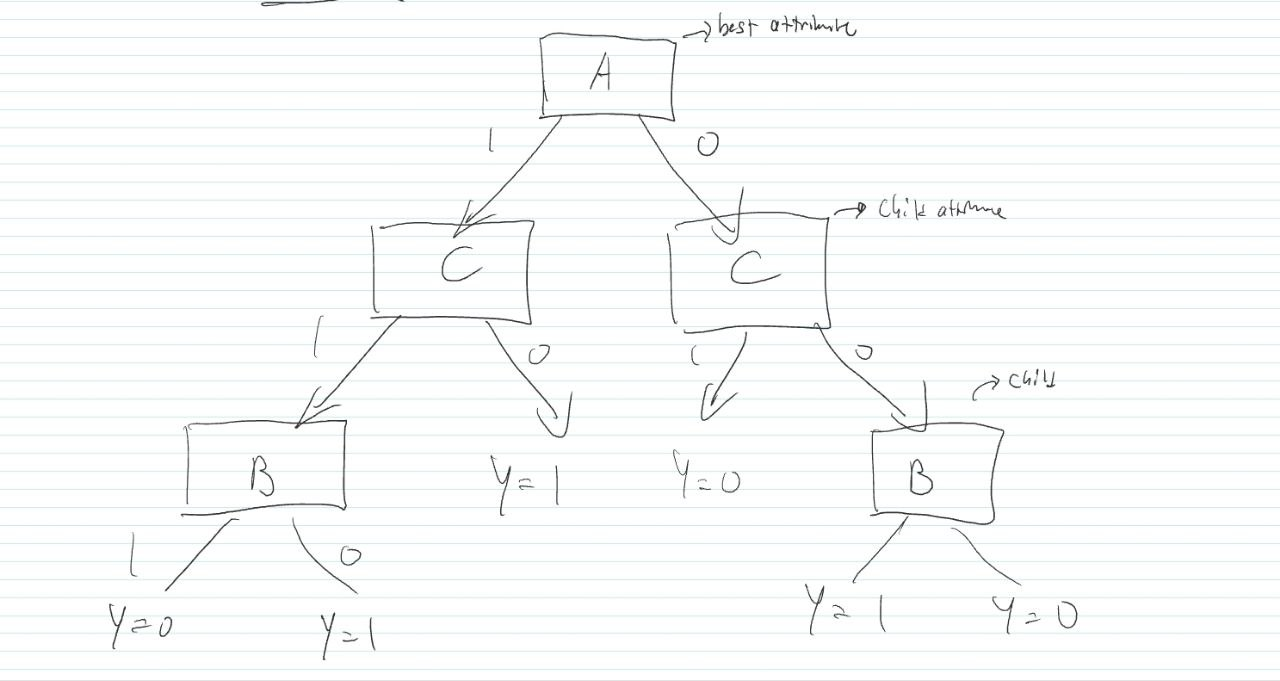
The Calculation is shown below



A close-up of math equations

Description automatically generated A close-up of math equations

Description automatically generated

2c) 

Comment: The Decision Tree is not unique as it can be split either at Node A or Node C with both having an information gain of 0.0813.

**3) Decision Tree Result Values**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Accuracy | | Precision | Recall |
|  | Train | Test |  |  |
| DT1 | 1 | 0.92748538 | 0.91554468 | 0.89219763 |
| DT2 | 0.98366834 | 0.92923977 | 0.91996553 | 0.89221434 |

Comment: DT2 is the decision three set to a max depth of 4 levels to be able to beat the values of DT1.