***PART II***

***Activity 1. Consider the task of predicting whether children are likely to be hired to play members of the Von Trapp Family in a production of The Sound of Music, based on these data:***

|  |  |  |  |
| --- | --- | --- | --- |
| **height** | **hair** | **eyes** | **hired** |
| short | blond | blue | + |
| tall | red | blue | + |
| tall | blond | blue | + |
| tall | blond | brown | – |
| short | dark | blue | – |
| tall | dark | blue | – |
| tall | dark | brown | – |
| short | blond | brown | – |

***(1) Compute the information (entropy) gain for each of the three attributes (height, hair, eyes) in terms of classifying objects as belonging to the class, + or – .***

***(2) Construct a decision tree based on the minimum entropy principle.***

Diagram

Description automatically generated(1) There are 3 objects in class '+' and 5 in '-', so the entropy is:   
Entropy(parent) = Σi Pi log2Pi = -(3/8)log(3/8) - (5/8)log(5/8) = 0.954   
  
Suppose we split on height:

Of the 3 'short' items, 1 is '+' and 2 are '-', so Entropy(short) = -(1/3)log(1/3) - (2/3)log(2/3) = 0.918  
Of the 5 'tall' items, 2 are '+' and 3 are '-', so Entropy(tall) = -(2/5)log(2/5) - (3/5)log(3/5) = 0.971  
The average entropy after splitting on 'height' is Entropy(height) = (3/8)(0.918) + (5/8)(0.971) = 0.951 The information gained by testing this attribute is: 0.954 - 0.951 = 0.003 (i.e. very little)

Your work:

Information Gain for hair:

Information Gain for eyes:

Part (2):

***Activity 2. The Laplace error estimate for pruning a node in a Decision Tree is given by:***

Diagram

Description automatically generated

Diagram

Description automatically generated with medium confidence

where  is the total number of items,  is the number of items in the majority class and  is the number of classes. Given the following sub-tree, should the

N

n

k

children be pruned or not?   Show your calculations.

Table

Description automatically generated

Text, letter

Description automatically generated

Your work: draw the decision tree and tell us what class is assigned to the instance {D15, Sunny, Hot, High, Weak}?