Assignment 5 – Artificial Intelligence I

Name : *Saranya Balasubramaniyan*
Net ID : *sxb9644*
Course and Section : *CSE5360 - 900*

Task 1

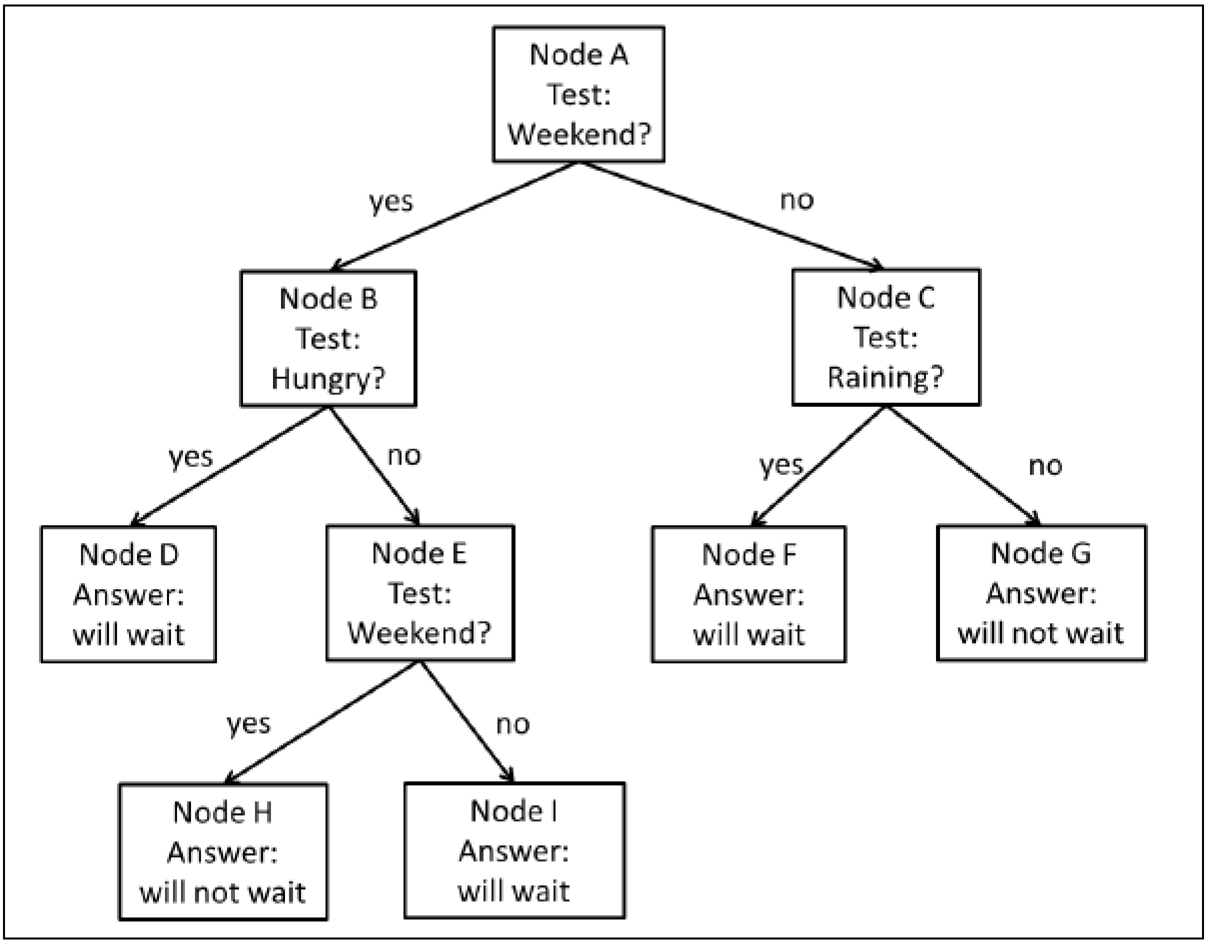


Figure 1: A decision tree for estimating whether the patron will be willing to wait for a table at a restaurant.

Part a:

Suppose that, on the entire set of training samples available for constructing the decision tree of Figure 1, 65 people decided to wait, and 35 people decided not to wait. What is the initial entropy at node A (before the test is applied)?

Number of people who decided to wait : A1: 65
Number of people who decided not to wait : A2: 35
Total number of people : A = A1+ A2 = 100

We know that Entropy is defined by the following formula,,

$$H\left(\frac{K_1}{K},\frac{K_2}{K}\right) = \sum_{i=1}^{N} -\frac{K_i}{K}\left(\log_2\frac{K_i}{K}\right)$$

Initial entropy at node A would be,

H(A) =$-\frac{A1}{A}\log_2\frac{A1}{A}$ - $\frac{A2}{A}\log_2\frac{A2}{A}$

H(A) = - $\frac{65}{100}\log_2\frac{65}{100}$ - $\frac{35}{100}\log_2\frac{35}{100}$

H(A) = -0.65($\log_2 0.65$) – 0.35($\log_2 0.35$)

H(A) = -0.65 (-0.6214) – 0.35(-1.5145)

H(A) = 0.4039 + 0.5300

H(A) = 0.9339

Part b: As mentioned in the previous part, at node A, 65 people decided to wait and 35 people decided not to wait.

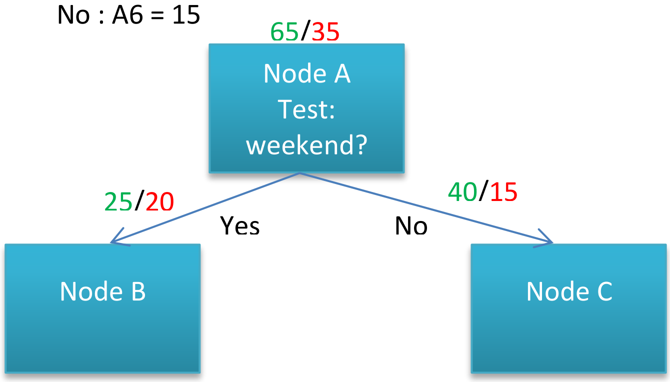
- **Out of the cases where people decided to wait, in 25 cases it was weekend and in 40 cases it was not weekend.**
- **Out of the cases where people decided not to wait, in 20 cases it was weekend and in 15 cases it was not weekend.**

What is the information gain for the weekend test at node A?

Number of people who decided to wait : A1 = 65
Number of people who decided not to wait : A2 = 35

From the graph,

Among the number of people who decided to wait (A1),
Weekend?
Yes : A3 = 25
No : A5 = 40
Among the number of people who decided not to wait(A2),
Weekend?
Yes : A4 = 20
No : A6 = 15



Information gain can be calculated based on the formula,

$$I(E,L) = H(E) - \sum_{i=1}^{L} \frac{K_i}{K} H(E_i)$$

Information Gain = H(A) – $\left(\left[\frac{A3+A4}{A}\right] * H\left(\frac{A3}{A3+A4}, \frac{A4}{A3+A4}\right)\right)$ - $\left(\left[\frac{A5+A6}{A}\right] * H\left(\frac{A5}{A5+A6}, \frac{A6}{A5+A6}\right)\right)$

H(A) = 0.9339 (Calculated in Part a)

$H\left(\frac{A3}{A3+A4}, \frac{A4}{A3+A4}\right)$ =- $\frac{25}{25+20} \log_2 \frac{25}{25+20}$ - $\frac{20}{25+20} \log_2 \frac{20}{25+20}$

$H\left(\frac{A3}{A3+A4}, \frac{A4}{A3+A4}\right)$ =- 0.5555 (-0.8481) -0.4444 (-1.1700)

$H\left(\frac{A3}{A3+A4}, \frac{A4}{A3+A4}\right)$ =0.4711 + 0.5199

$\boldsymbol{H\left(\frac{A3}{A3+A4}, \frac{A4}{A3+A4}\right)}$ **=0.991**

$H\left(\frac{A5}{A5+A6}, \frac{A6}{A5+A6}\right)$ =- $\frac{40}{40+15}\log_2\frac{40}{40+15}$ - $\frac{15}{40+15}\log_2\frac{15}{40+15}$

$H\left(\frac{A5}{A5+A6}, \frac{A6}{A5+A6}\right)$ =- 0.7272 (-0.4595) -0.2727 (-1.8746)

$H\left(\frac{A5}{A5+A6}, \frac{A6}{A5+A6}\right)$=0.3341+0.5112

$\boldsymbol{H\left(\frac{A5}{A5+A6}, \frac{A6}{A5+A6}\right)}$**=0.8453**

Information Gain = H(A) – $\left(\left[\frac{A3+A4}{A}\right] * H\left(\frac{A3}{A3+A4}, \frac{A4}{A3+A4}\right)\right)$ - $\left(\left[\frac{A5+A6}{A}\right] * H\left(\frac{A5}{A5+A6}, \frac{A6}{A5+A6}\right)\right)$

Information Gain = 0.9339 – ([45/100] * 0.991) – ([55/100] * 0.8453)

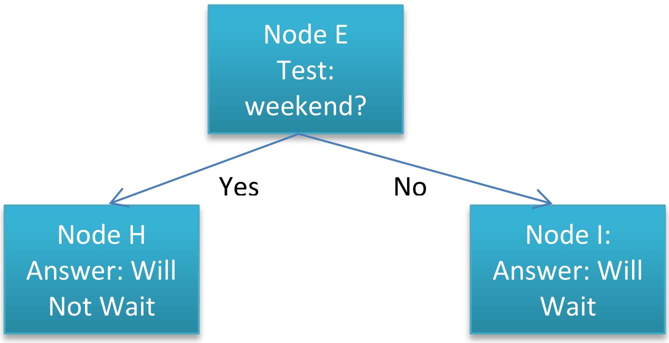
Information Gain = 0.9339 – (0.45 * 0.991) – (0.55 * 0.8453)

Information Gain = 0.9339 – 0.4459 – 0.4649

Information Gain = 0.9339 – 0.4459 – 0.4649

Information Gain = 0.0231

Part c : In the decision tree of Figure 1, node E uses the exact same test (whether it is weekend or not) as node A. What is the information gain, at node E, of using the weekend test?



Node E uses the test that Node A has already used. Node E comes under the “Yes” branch of the Node A which means “Yes for weekend”. The training examples with “Weekend? No” are already filtered at Node A and at Node E, all the training examples would go to the same child Node H.

So, information gain at Node E will be 0.

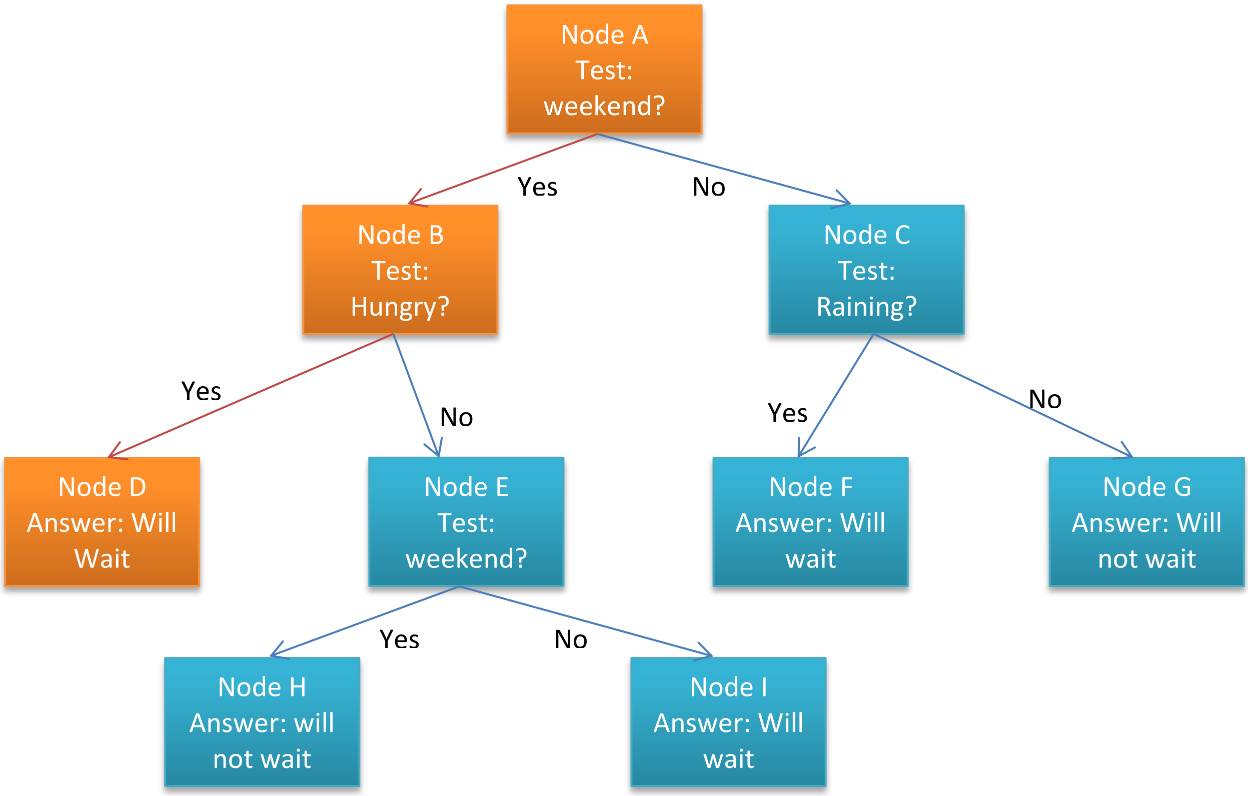
Part d : We have a test case of a hungry patron who came in on a rainy Sunday. Which leaf node does this test case end up in? What does the decision tree output for that case?

Given test case

Hungry?	Weekend?	Raining?
Yes	Yes	Yes

Traversal, starting from Node A

Current Node	**Question**	**Answer**	**Next Node**
Node A	Weekend?	Yes	Node B
Node B	Hungry?	Yes	Node D
Node D	No question : Leaf node	Will Wait	



Based on the above traversal, A Hungry patron on a Rainy Sunday **will wait**.

Task 2

Class	A	B	C
X	1	2	1
X	2	1	2
X	3	2	2
X	1	3	3
X	1	2	2
Y	2	1	1
Y	3	1	1
Y	2	2	2
Y	3	3	1
Y	2	1	1

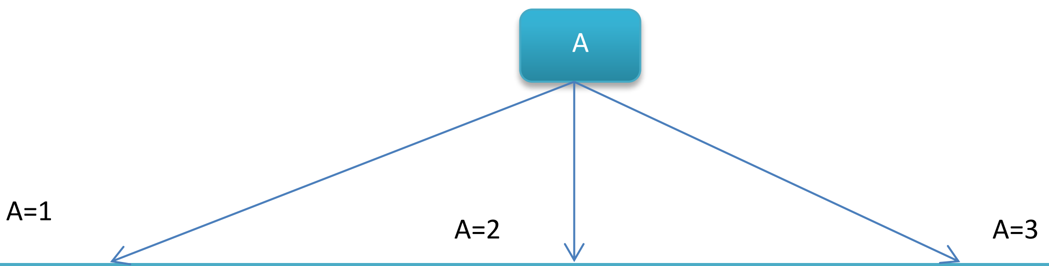
We want to build a decision tree that determines whether a given pattern is class X or class Y. The decision tree can only use tests that are based on attributes A, B, and C. Each attribute has 3 possible values: 1, 2, 3. We have the 10 training examples, shown on the table (each row corresponds to a training example, the first column is the class label and the other 3 columns are the pattern). What is the information gain of each attribute at the root? Which attribute achieves the highest information gain at the root?

To find the attribute with the highest information gain, we need to calculate the Information gain for each of the attribute.

In the table, we can see that the Classes X and Y are equally split. So the entropy would be 1.

H(E) = $-\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10}$
H(E) = -0.5(-1) -0.5(-1)
H(E) = 0.5 + 0.5
H(E) = 1

Calculating the Information gain for Attribute A,



Class	A	B	C	Class	A	B	C	Class	A	B	C
X	1	2	1	**X**	2	1	2	**X**	3	2	2
X	1	3	3	**Y**	2	1	1	**Y**	3	1	1
X	1	2	2	**Y**	2	2	2	**Y**	3	3	1
				Y	2	1	1				

E1
X = 3
Y = 0
Total: 3

E2
X = 1
Y = 3
Total: 4

E3
X = 1
Y = 2
Total: 3

Information gain can be calculated based on the formula,

$$I(E, L) = H(E) - \sum_{i=1}^{L} \frac{K_i}{K} H(E_i)$$

Information Gain = H(E) – $\left(\left[\frac{E1}{E}\right] * H(E1)\right)$ - $\left(\left[\frac{E2}{E}\right] * H(E2)\right)$- $\left(\left[\frac{E3}{E}\right] * H(E3)\right)$

Entropy is defined by the following formula,,

$$H\left(\frac{K_1}{K}, \frac{K_2}{K}\right) = \sum_{i=1}^{N} -\frac{K_i}{K}\left(\log_2 \frac{K_i}{K}\right)$$

H(E) = 1

H(E1) = - $\frac{3}{3}(\log_2 \frac{3}{3}) - \frac{0}{3}(\log_2 \frac{0}{3})$
H(E1) = 0

H(E2) = - $\frac{1}{4}(\log_2 \frac{1}{4}) - \frac{3}{4}(\log_2 \frac{3}{4})$
H(E2) = -0.25(-2) – 0.75(-0.4150)
H(E2) = 0.5 + 0.3112
H(E2) = 0.8112

H(E3) = - $\frac{1}{3}(\log_2 \frac{1}{3}) - \frac{2}{3}(\log_2 \frac{2}{3})$
H(E3) = 0.3333(-1.5851)-0.6666(-0.5851)
H(E3) = 0.5283 + 0.3900
H(E3) = 0.9183

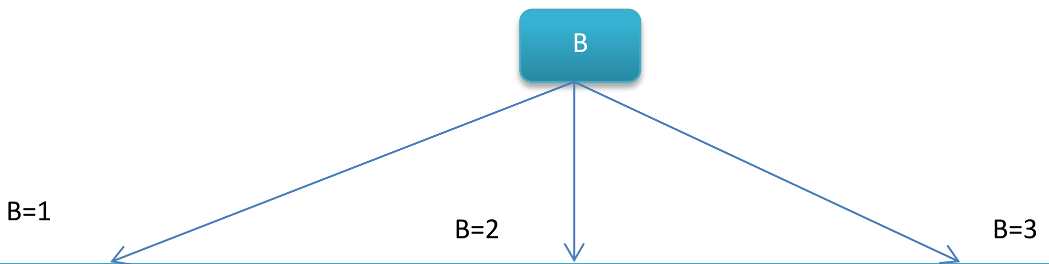
Information Gain for A = 1 – $\left(\frac{3}{10} * 0\right)$– $\left(\frac{4}{10} * 0.8112\right)$ - $\left(\frac{3}{10} * 0.9183\right)$

Information Gain for A = 1 – 0 – (0.4 * 0.8112) – (0.3 * 0.9183)

Information Gain for A = 1-0.3244-0.2754

Information Gain for A = 0.4002

Calculating the Information gain for Attribute B,



Class	A	B	C
X	2	1	2
Y	2	1	1
Y	3	1	1
Y	2	1	1

Class	A	B	C
X	1	2	1
X	3	2	2
X	1	2	2
Y	2	2	2

Class	A	B	C
X	1	3	3
Y	3	3	1

E1
X = 1
Y = 3
Total: 4

E2
X = 3
Y = 1
Total: 4

E3
X = 1
Y = 1
Total: 2

H(E) = 1

H(E1) = $-\frac{1}{4}(\log_2 \frac{1}{4}) - \frac{3}{4}(\log_2 \frac{3}{4})$
H(E1) = -0.25(-2) – 0.75(-0.4150)
H(E1) = 0.5 + 0.3112
H(E1) = 0.8112

H(E2) =$-\frac{3}{4}(\log_2 \frac{3}{4})$ - $\frac{1}{4}(\log_2 \frac{1}{4})$
H(E2) =– 0.75(-0.4150) - 0.25(-2)
H(E2) = 0.3112 + 0.5
H(E2) = 0.8112

H(E3) = $-\frac{1}{2}(\log_2 \frac{1}{2}) - \frac{1}{2}(\log_2 \frac{1}{2})$
H(E3) = -0.5(-1)- 0.5 (-1)
H(E3) = 0.5 + 0.5
H(E3) = 1

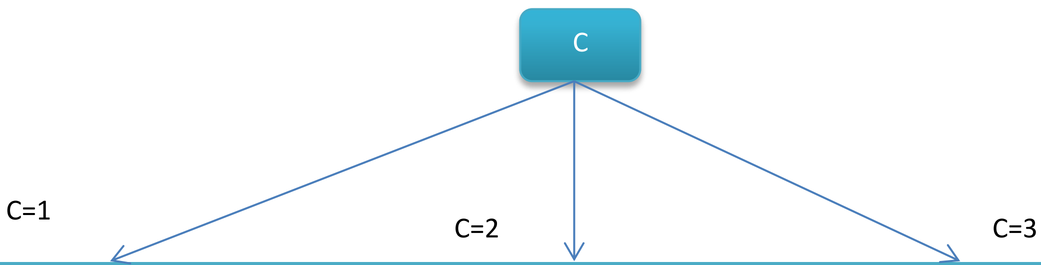
Information Gain for B = 1 – $\left(\frac{4}{10} * 0.8112\right)$– $\left(\frac{4}{10} * 0.8112\right)$ - $\left(\frac{2}{10} * 1\right)$

Information Gain for B = 1 – (0.4 * 0.8112) – (0.4 * 0.8112) – (0.2 * 1)

Information Gain for B = 1 – 0.3244 – 0.3244 – 0.2

Information Gain for B = 0.1512

Calculating the Information gain for Attribute C,



Class	A	B	C
X	1	2	1
Y	2	1	1
Y	3	1	1
Y	3	3	1

Class	A	B	C
X	2	1	2
X	3	2	2
X	1	2	2
Y	2	2	2

Class	A	B	C
X	1	3	3

Y	2	1	1

E1
X = 1
Y = 4
Total: 5

E2
X = 3
Y = 1
Total: 4

E3
X = 1
Y = 0
Total: 1

H(E) = 1

H(E1) = - $\frac{1}{5}(\log_2 \frac{1}{5}) - \frac{4}{5}(\log_2 \frac{4}{5})$
H(E1) = -0.2(-2.3219) – 0.8(-0.3219)
H(E1) = 0.4643 + 0.2575
H(E1) = 0.7218

H(E2) = $-\frac{3}{4}(\log_2 \frac{3}{4})$ - $\frac{1}{4}(\log_2 \frac{1}{4})$
H(E2) =– 0.75(-0.4150) -0.25(-2)
H(E2) = 0.3112 + 0.5
H(E2) = 0.8112

H(E3) = - $\frac{1}{1}(\log_2 \frac{1}{1}) - \frac{0}{1}(\log_2 \frac{0}{1})$
H(E3) = 0

Information Gain for C = 1 – $\left(\frac{5}{10} * 0.7218\right) - \left(\frac{4}{10} * 0.8112\right)$ - $\left(\frac{1}{10} * 0\right)$

Information Gain for C = 1 – (0.5 * 0.7218) – (0.4 * 0.8112) – 0

Information Gain for C = 1-0.3609-0.3244-0

Information Gain for C = 0.3151

Information Gain for A(0.4002) > Information Gain for C (0.3151) > Information Gain for B (0.1512)

Attribute A has the highest Information gain

Task 3

Suppose that, at a node N of a decision tree, we have 1000 training examples. There are four possible class labels (A, B, C, D) for each of these training examples.
Part a: What is the highest possible and lowest possible entropy value at node N?
Highest entropy means that the training examples are equally split.

Training examples equally split would result in A=250, B= 250, C=250 and D=250 training examples

Hence entropy would be

$H(E_A) + H(E_B) + H(E_C) + H(E_D)$
$= 4 * \left(-\frac{250}{1000} * \log_2 \frac{250}{1000}\right)$
$= 4 * (-0.25 * -2)$
$= 2$

Lowest entropy means that all the examples are in a single class,

Entropy would be
$H(E_A) + H(E_B) + H(E_C) + H(E_D)$
$=\left(-\frac{1000}{1000} * \log_2 \frac{1000}{1000}\right) + 0 + 0 + 0$
$= 0$

Lowest possible entropy at node N is 0
Highest possible entropy at node N is 2

Part b: Suppose that, at node N, we choose an attribute K. What is the highest possible and lowest possible information gain for that attribute?
If the decision tree has less chaotic, more lop sided values, then the lowest possible entropy value is 0.
If the decision tree has more chaotic, least lop sided values, then the highest possible entropy value is $\log_2 N$

Since there are 4 attributes, value of N is 4. So highest possible entropy value is $\log_2 4$ = 2

Lowest possible entropy at node N on choosing attribute K is 0
Highest possible entropy at node N on choosing attribute K is 2

Task 4

Your boss at a software company gives you a binary classifier (i.e., a classifier with only two possible output values) that predicts, for any basketball game, whether the home team will win or not. This classifier has a 28% accuracy, and your boss assigns you the task of improving that classifier, so that you get an accuracy that is better than 60%. How do you achieve that task? Can you guarantee achieving better than 60% accuracy?

Negating the output would result in 72% accuracy of the result. Thus an output with better accuracy of more than 60% can be achieved.