**Redo Example 3.1 assuming that a coin is tossed six times with the following results: {0, 1, 0, .0, 1, 0}.**

Suppose that a coin is tossed in the air six times with the following results ( 1 indicates a head and 0 indicates a tail): {0, 1, 0 , 0,1, 0}. If we assume that the coin toss follows the Bernoulli distribution, we know that





Likelihood (when probability of head is 0.8 )

* L(P|0,1,0,0,1,0) = 0.2 \*0 .8 \* 0.2\* 0.2 \*0 .8 \* 0.2 = 0.001024



By this formula

If P is the(Probability of getting head) value to find Maximum likelihood

* P
* 2 /6
* 0.33333

Likelihood (when probability of getting head = 0.3333)

* L(P|0,1,0,0,1,0)
* 0.66 \* 0.33 \* 0.66\* 0.66\*0.33\*0.66
* 0.02195

Probability to get this maximum likelihood is 1/3 & 2/3

2) page 122 exercise 11

Generate a DT for the height example in Table 4.1 using the ID3 algorithm and the training classifications shown in the Output2 column of that table.

**Repeat exercise 11 using the gain ratio instead of gain.**



The beginning state of the training data in Table 4.1 (with the Output2 classification) is that (10/15) are medium, and (5/15) are tall. Thus, the entropy of the starting set is

10/15 log(15/10) + 6/15 log(15/6) = 0.2764

Choosing the gender as the splitting attribute, there are nine tuples that are F and 6 that are M. The entropy of the subset that are F is

3/9 log(9/3) + 6/9 log(9/6) = 0.2764

whereas that for the M subset is

4/6 log (6/4) + 2/6 log (6/2) = 0.2764

The ID3 algorithm must determine what the gain in information is by using this split. To do this, we calculate the weighted sum of these last two entropies to get

((9/15) 0.2764) + ((6/15) 0.2764) = 0.2764

The gain in entropy by using the gender attribute is thus

= 0.2764 - 0.2764 = 0 -----------------------🡪value(1)



 --------------🡪 value(2)

This gives the GainRatio value for the gender attribute as

= value(1) / value(2)

= 0 /0.292 = 0

|  |  |
| --- | --- |
|  |  |
| 0 -1.6 | M,M |
| 1.6 -1.7 | M,M |
| 1.7- 1.8 | M,M,M |
| 1.8-1.9 | T,T,M,T |
| 1.9 – 2 | M , M |
| 2 - INF | T , T |
|  |  |

(0, 1.6], (1.6, 1.7], (1.7, 1.8], (1.8, 1.9], (1.9, 2.0], (2.0, inf]

0 with entropy = (0+ 0 + 0) = 0

0 IN (1.6 ,1.7] with entropy = (0 +0+0) = 0

0 IN (1.7,1.8] with entropy = (0 + 0 +0) = 0

0 IN (1.8,1.9] with entropy = 0+1/4 log (4) + ¾ log(4/3)   = 0.2442

2 in (1.9, 2.0] with entropy = 0 + 2/2 (0) + 0 = 0

two in the last with entropy = (0+0+2/2(0))  = 0

Information gain = 0.2764 – 4/15(0.2442) = 0.2128 -----------------🡪 Value (1)

This has more information gain than gender. So we choose height over gender as splitting attribute.

According to the formula gain ratio 

To calculate the GainRatio for the height split, we first find the entropy associated with the split ignoring classes H

H (2/15 , 2/15,3/15,4/15,2/15,2/15)

= 2/15 (log (15/2) + 2/15 (log (15/2 ) + 3/15 (log (15/3 ) + 4/15 (log (15/4 ) + 2/15 (log (15/2 ) +2/15 (log (15/2 )

= 48\*2\*0.87 + 3\*0.7 +4\*0.574 / 15

= 6.48 + 2.1+2.296 /15

= 0.7251 -----------------🡪 Value(2)

GainRatio with height

= Value (1) / Value(2)

= 0.2128 / 0.7251

**= 0.2934**