T-79.5102 **Autumn 2008**
Special Course in Computational Logic
Tutorial 10
Solutions

1. The decision tree for the given data set is constructed iteratively by adding one attribute test at a time. An information theoretic criterion is used to select the best attribute at each step: The attribute A that provides us the highest expected amount of information is selected. Given probabilities $P(v_1), \ldots, P(v_n)$ for the possible values $v_1, \ldots, v_n$ of an attribute A the information content of the actual answer $(A = v_i)$ is calculated as follows:

$$I(P(v_1), \ldots, P(v_n)) = \sum_{i=1}^{n} -P(v_i) \log_2 P(v_i).$$

In the given sample, the objects x_1, x_2, x_6, and x_7 are birds while x_3, x_4, x_5, x_8, and x_9 are not. Thus, the initial information content is:

$$I(\frac{4}{9}, \frac{5}{9}) = \frac{4}{9} \log_2 \frac{9}{4} + \frac{5}{9} \log_2 \frac{9}{5} \approx 0.991.$$

Next, we go through all attributes to find out which gives us the best choice, i.e., is expected to maximise the information gain. The information gain for an attribute A is defined as follows:

$$\mathrm{Gain}(A) = I(\frac{p}{p+n}, \frac{n}{p+n}) - \mathrm{Remainder}(A).$$

The first term, $I(\frac{p}{p+n}, \frac{n}{p+n})$ is the initial information content and the second represents the remaining information content when the value of A is known. Since the goal is to maximise gain and the initial information is independent of the attribute under consideration, the goal is achieved by choosing the attribute with the smallest remainder. The first variable is *Flies* and when we examine the sample data we notice that two birds (x_1 and x_2) fly and two (x_6, x_7) do not. Of the rest, only x_4 and x_5 fly. Thus

$$\mathrm{Remainder}(\mathit{Flies}) = \frac{4}{9} I(\frac{2}{4}, \frac{2}{4}) + \frac{5}{9} I(\frac{2}{5}, \frac{3}{5}) \approx 0.984.$$

Here the left term accounts for the case of flying animals, and the right one the case of non-flying ones. We notice that knowing the value of the attribute *Flies* is not expected to help us much. On the other hand, the number of legs forms a much better choice, since all birds are two-legged:

$$\mathrm{Remainder}(\mathit{Legs}) = \frac{5}{9} I(\frac{4}{5}, \frac{1}{5}) + \frac{4}{9} I(0, 1) \approx 0.401.$$

When we compute remainders for the other attributes, we learn that *Legs* is the best choice for the root of the decision tree. For the sake of completeness, the expected gain of information from the attribute *Legs* is

$$\mathrm{Gain}(\mathit{Legs}) = I(\frac{4}{9}, \frac{5}{9}) - \mathrm{Remainder}(\mathit{Legs}) = 0.991 - 0.401 = 0.590$$

For the moment, our decision tree looks like this:

Legs			
0	2	4	6
|	|	|	|
No	?	No	No

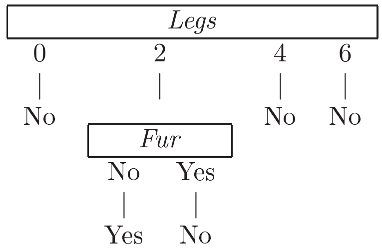
Next we try to find the best attribute for the node marked with "?". For the attribute *Flies*, the remaining examples are divided so that x_1, x_2, and x_5 fly, while x_6 and x_7 do not. Thus we have

$$\text{Remainder}(\textit{Flies}) = \frac{3}{5}I(\frac{2}{3},\frac{1}{3}) + \frac{2}{5}I(1,0) \approx 0.551.$$

It turns out that *Flies* is not the best attribute: By knowing the value of *Flies*, birds can be completely separated from other animals so that

$$\text{Remainder}(\textit{Fur}) = \frac{1}{5}I(0,1) + \frac{4}{5}I(1,0) = 0.$$

By these steps, we have obtained a decision tree that makes correct classifications for the given data set:



The main problem with learning decision trees is that they are only as good as the sample data. If the sample data is not complete enough, the tree may give completely wrong answers. Also, if there is too much data and too many variables, the algorithm can find correlations that are actually only statistical oddities.

The tree above has too little sample data. Since a human has two legs and no fur, the tree classifies a person as a bird.

2. When learning a decision tree, it is often sufficient to check the classification of examples for each attribute in turn and exact gain values need not be computed (except to exclude doubts in borderline cases).

(a) Let us check how candidates get classified by different attributes:

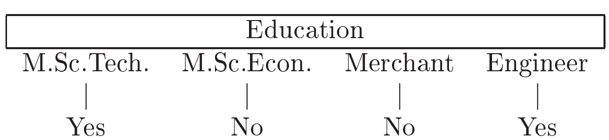
Language proficiency		
English	German	French
+1,+4 -2,-5	+3	+6

Programming skills	
Yes	No
+1,+3,+6	+4 -2, -5

Working experience	
Yes	No
+3,+4 -2	+1,+6 -5

Education			
M.Sc.Tech.	M.Sc.Econ.	Merchant	Engineer
+1,+4,+6	-2	-5	+3

Thus *Education* seems to be the best attribute. In fact, all examples are correctly classified. The following decision tree results:



Given the fact that employees are hired for an IT company, it is slightly surprising that programming skills are not acknowledged.

(b) Next we compute remainder and gain (in bits) for the attribute:
Remainder(*Language proficiency*) =
$\frac{4}{6}I(\frac{2}{4},\frac{2}{4}) + \frac{2}{6}I(1,0) = \frac{2}{3}I(\frac{1}{2},\frac{1}{2}) = \frac{2}{3}$.
Gain(*Language proficiency*) =
$I(\frac{4}{6},\frac{2}{6})$ − Remainder(*Language proficiency*) =
$I(\frac{2}{3},\frac{1}{3}) - \frac{2}{3} = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3} \approx 0.252$(bits).

(c) Let us then test the decision tree obtained above:

	Education	Decision tree	Classification	Result
7	Merchant	No	T	Wrong
8	M.Sc.Tech.	Yes	F	Wrong
9	M.Sc.Tech.	Yes	T	Correct
10	M.Sc.Econ.	No	F	Correct

Only 50 percent of examples are classified correctly by the decision tree. It seems that the classification merchants and M.Sc.Tech.s should be refined. Let us revise the tree using all examples:

Language proficiency		
English	German	French
+1,+4,+7,+9	+3	+6
-2,-5	-8	-10

Programming skills	
Yes	No
+1,+3,+6,+9	+4,+7
-10	-2,-5,-8

Working experience	
Yes	No
+3,+4,+8,+9	+1,+6
-2	-5,-8,-10

Education			
M.Sc.Tech.	M.Sc.Econ.	Merchant	Engineer
+1,+4,+6,+9		+7	+3
-8	-2,-10	-5	

Education is still the best attribute. The classification of Masters of Science in Technology is refined as follows:

Language proficiency		
English	German	French
+1,+4,+9		+6
	-8	

Programming skills	
Yes	No
+1,+6,+9	+4
	-8

Working experience	
Yes	No
+4,+9	+1,+6
	-8

Out of these attributes, language proficiency is already sufficient for complete classification. An analysis of merchants follows:

Language proficiency		
English	German	French
+7		
-5		

Programming skills	
Yes	No
	+7
	-5

Working experience	
Yes	No
+7	
	-5

The resulting decision tree is given in Figure 1.

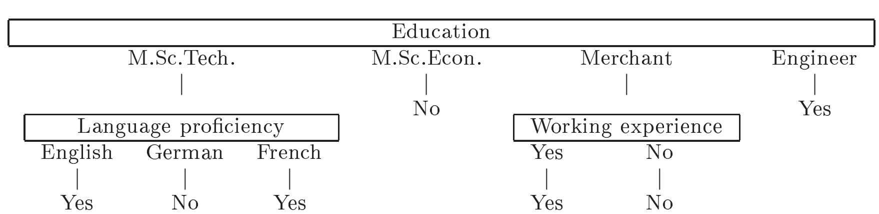


Figure 1: The decision tree obtained for all examples