Universität des Saarlandes
Fachrichtung 6.2 – Informatik
J. Siekmann, S. Autexier, K. Fischer, E. Melis, C. Zinn
M. Schiller, E. Schulz, R.Vollmann

12th Theoretical Assignment in Artificial Intelligence (WS 2006/2007)
Solutions

Exercise 12.1 Consider the following set of training examples:

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
3	−	T	F
4	+	F	F
5	−	F	T
6	−	F	T

1. What is the entropy of this collection of training examples with respect to the target function classification?
2. What is the information gain of a_2 relative to these training examples?
3. What is the information gain of a_1 relative to these training examples?
4. Construct a decision tree with the decision-tree-learning algorithm from the lecture for the given set of training examples! Use information gain as the criterion for choosing the "best" attributes for splitting the tree!
5. Assume that *Instance 6* is an error with your measurement, and should in fact be classified as a + and not as a −. Does this change your decision tree? How?

Solution:

1. *The collection contains 3 positive and 3 negative training instances. Therefore,* $Entropy([3,3]) = -\frac{1}{2}log_2\frac{1}{2} - \frac{1}{2}log_2\frac{1}{2} = 1$ *(this means that the collection of examples is very impure).*

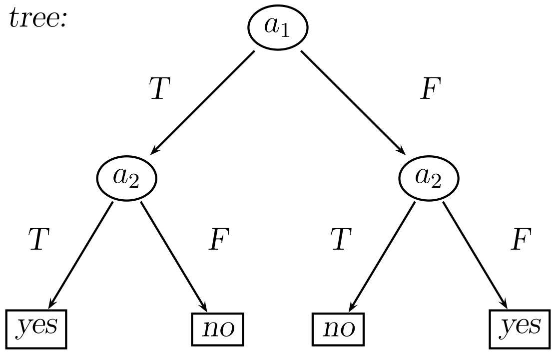
2.

$$\begin{aligned}
Gain(S, a_2) &= Entropy(S) - \sum_{v \in \{T,F\}} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= Entropy(S) - \frac{4}{6} Entropy(S_{a_2=T}) - \frac{2}{6} Entropy(S_{a_2=F}) \\
&= 1 - \frac{2}{3} Entropy([2,2]) - \frac{1}{3} Entropy([1,1]) \\
&= 1 - \frac{2}{3} - \frac{1}{3} \\
&= 0
\end{aligned}$$

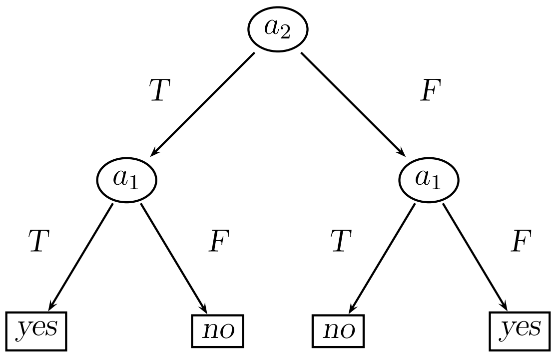
3.

$$\begin{aligned}
Gain(S, a_1) &= Entropy(S) - \sum_{v \in \{T,F\}} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= Entropy(S) - \frac{1}{2} Entropy(S_{a_1=T}) - \frac{1}{2} Entropy(S_{a_1=F}) \\
&= 1 - \frac{1}{2} Entropy([2,1]) - \frac{1}{2} Entropy([1,2]) \\
&= 1 - \frac{1}{2} Entropy([2,1]) - \frac{1}{2} Entropy([1,2]) \\
&= 1 - \frac{1}{2}(-\frac{2}{3} log \frac{2}{3} - \frac{1}{3} log \frac{1}{3}) - \frac{1}{2}(-\frac{1}{3} log \frac{1}{3} - \frac{2}{3} log \frac{2}{3}) \\
&= 1 - \frac{1}{2}(-0.390 - 0.528) - \frac{1}{2}(-0.528 - 0.390) \\
&= 0.082
\end{aligned}$$

4. *The decision tree:*

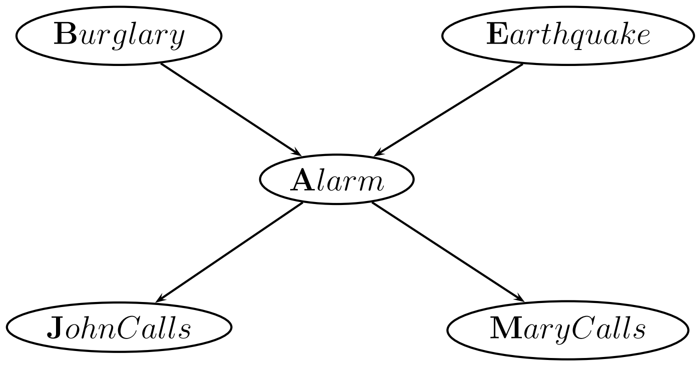


5. *Yes, it does. Now Gain(a_1) $= 0$ and Gain(a_2) $= 0.044$. Therefore, a_2 should be selected as the root of the tree, and not a_1:*



Because the tree is so symmetric, the difference is not very obvious. But in practice the appearance of the tree can radically change due to just some change in the input.

Exercise 12.2 Consider the example from exercise 11.4 on the last exercise sheet.



The *a priory* probabilities and conditional probabilities are:

P(B)
.001

P(E)
.002

B	**E**	**P(A)**
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	**P(J)**
T	.90
F	.05

A	**P(M)**
T	.70
F	.01

In the lecture, we have seen the ENUMERATION-ASK algorithm (reproduced from Russell/Norvig: AIMA, 2003):

function ENUMERATION-ASK(X,e,*bn*) returns a distribution over X
    inputs: *X*, the query variable
        e, observed values for variables E
        bn, a Bayesian network with variables $\{X\} \cup E \cup Y$

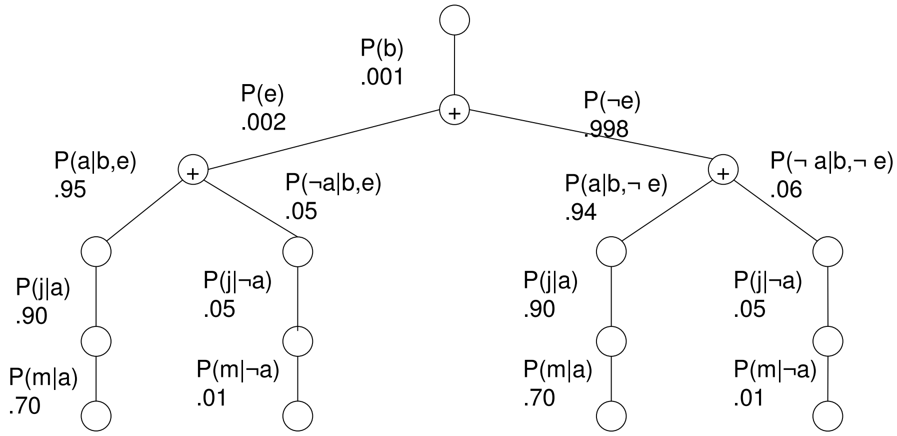
    Q(*X*) ← a distribution over X, initially empty
    for each value x_i of X do
        extend e with value x_i for *X*
        Q(x_i) ← ENUMERATE-ALL(VARS[*bn*],e)
    return NORMALIZE (Q(*X*))

function ENUMERATE-ALL(*vars*,e) returns a real number
    if EMPTY?(*vars*) then return 1.0
    Y ← FIRST(*vars*)
    if Y has value *y* in e
    then return $P(y \mid Parents(Y)) \times ENUMERATE-ALL(REST(\text{vars}), e)$
    else return $\sum_y P(y \mid Parents(Y)) \times ENUMERATE-ALL(REST(\text{vars}), e_y)$

        where e_y is e extended with Y = y

1. Consider the query $P(B|j,m)$. When calling ENUMERATION-ASK (where $X := B$, $e := \{j,m\}$, bn is the given network), ENUMERATE-ALL is called to compute $Q(b)$. This evaluation will produce a tree, which is reproduced below.

   Use the ENUMERATE-ALL algorithm step by step to evaluate ENUMERATE-ALL({B,E,A,J,M},{j,m,b}) – i.e. follow the steps of the algorithm that produce the given tree!



2. Use ENUMERATION-ASK to answer the query $P(B|j,m)$! You can use the solution from part 1., and then do the missing parts.

   One note about normalization: Remember that when you calculate the distribution of some variable, lets say D, where D can take the values d and $\neg d$, for example, then the probabilities of the possible values of D need to add up to 1. Furthermore, $P(d|e_1,...,e_n) + P(\neg d|e_1,...,e_n) = 1$. If the computation involves an unknown normalization constant α, such that $P(d|e_1,...,e_n) = \frac{x}{\alpha}$, and $P(\neg d|e_1,...,e_n) = \frac{y}{\alpha}$, α can be computed easily by knowing that $P(d|e_1,...,e_n) = P(\neg d|e_1,...,e_n) = 1$. This value α is what NORMALIZE is supposed to determine. The result of normalization is the computed probability distribution multiplied by α (look also at p. 476 and p. 505 in Russell/Norvig).

Solution:

1. *ENUMERATION-ASK is called with $X := B$, $e := \{j,m\}$, and bn is the given semantic net.*

2. *B as two possible values, namely $x_1 := b$ and $x_2 := \neg b$. Now ENUMERATE-ALL is called twice.*

3. *ENUMERATE-ALL is called with $vars := \{B,E,A,J,M\}$, $e := \{j,m,b\}$. This produces the tree in the figure above in the following steps:*

   *- $Y := B$. Variable B has value b in e, return P(b) * ENUMERATE-ALL($\{E,A,J,M\}$,$\{j,m,b\}$).*
   ENUMERATE-ALL($\{E,A,J,M\}$,$\{j,m,b\}$) is computed as:
   - $Y := E$. Variable E has no value in e, therefore return the sum of
   *- P(e) * ENUMERATE-ALL($\{A,J,M\}$,$\{j,m,b,e\}$).*
   ENUMERATE-ALL($\{A,J,M\}$,$\{j,m,b,e\}$) is computed as:
   - $Y := A$. Variable A has no value in e, therefore return the sum of
   *- $P(a|b,e)$ * ENUMERATE-ALL($\{J,M\}$,$\{j,m,b,e,a\}$)*
   *- $P(\neg a|b,e)$ * ENUMERATE-ALL($\{J,M\}$,$\{j,m,b,e,\neg a\}$)*
   *- $P(\neg e)$ * ENUMERATE-ALL($\{A,J,M\}$,$\{j,m,b,\neg e\}$)*
   ENUMERATE-ALL($\{A,J,M\}$,$\{j,m,b,\neg e\}$) is computed as:
   - $Y := A$. Variable A has no value in e, therefore return the sum of
   *- $P(a|b,\neg e)$ * ENUMERATE-ALL($\{J,M\}$,$\{j,m,b,\neg e,a\}$)*
   *- $P(\neg a|b,\neg e)$ * ENUMERATE-ALL($\{J,M\}$,$\{j,m,b,\neg e,\neg a\}$)*

The result is $Q(b) = P(b) * (P(e) * ((P(a|b,e) * P(j|a) * P(m|a)) + P(\neg a|b,e) * P(j|a) * P(m|a)))) + P(\neg e) * (((P(a|b,\neg e) * P(j|a) * P(m|a)) + P(\neg a|b,\neg e) * P(j|a) * P(m|a))))$ $= 0.001 * (0.002 * ((0.95 * 0.9 * 0.7) + 0.05 * 0.9 * 0.7))) + 0.998 * (((0.94 * 0.9 * 0.7) + 0.06 * 0.9 * 0.7))) = 0.001 * ((0.002 * (0.5985 + 0.000025)) + (0.998 * (0.5922 + 0.00003)))$ $= 0.001 * (0.002 * 0.598525 + 0.998 * 0.59223) =$ *0.00059224259*

4. *ENUMERATE-ALL is called with* $vars := \{B, E, A, J, M\}$, $e := \{j, m, \neg b\}$. *This produces a tree that is similar to the tree given above:*

   - $Y := B$. *Variable B has value $\neg b$ in e, return $P(\neg b)$ * ENUMERATE-ALL($\{E, A, J, M\}$,$\{j, m, \neg b\}$).*
   ENUMERATE-ALL($\{E, A, J, M\}$,$\{j, m, \neg b\}$) is computed as:
     - $Y := E$. *Variable E has no value in e, therefore return the sum of*
       - *P(e) * ENUMERATE-ALL($\{A, J, M\}$,$\{j, m, \neg b, e\}$).*
       ENUMERATE-ALL($\{A, J, M\}$,$\{j, m, \neg b, e\}$) is computed as:
         - $Y := A$. *Variable A has no value in e, therefore return the sum of*
           - *$P(a|\neg b, e)$ * ENUMERATE-ALL($\{J, M\}$,$\{j, m, \neg b, e, a\}$)*
           - *$P(\neg a|\neg b, e)$ * ENUMERATE-ALL($\{J, M\}$,$\{j, m, \neg b, e, \neg a\}$)*
       - *$P(\neg e)$ * ENUMERATE-ALL($\{A, J, M\}$,$\{j, m, \neg b, \neg e\}$)*
       ENUMERATE-ALL($\{A, J, M\}$,$\{j, m, \neg b, \neg e\}$) is computed as:
         - $Y := A$. *Variable A has no value in e, therefore return the sum of*
           - *$P(a|\neg b, \neg e)$ * ENUMERATE-ALL($\{J, M\}$,$\{j, m, \neg b, \neg e, a\}$)*
           - *$P(\neg a|\neg b, \neg e)$ * ENUMERATE-ALL($\{J, M\}$,$\{j, m, \neg b, \neg e, \neg a\}$)*

   The result is $Q(\neg b) = P(\neg b) * (P(e) * ((P(a|\neg b,e) * P(j|a) * P(m|a)) + P(\neg a|\neg b,e) * P(j|a) * P(m|a)))) + P(\neg e) * (((P(a|\neg b,\neg e) * P(j|a) * P(m|a)) + P(\neg a|\neg b,\neg e) * P(j|a) * P(m|a))))$ $= 0.999 * (0.002 * ((0.29 * 0.9 * 0.7) + 0.71 * 0.9 * 0.7))) + 0.998 * (((0.001 * 0.9 * 0.7) + 0.999 * 0.9 * 0.7))) = 0.999 * ((0.002 * (0.1827 + 0.000355)) + (0.998 * (0.00063 + 0.0004995)))$ $= 0.001 * (0.002 * 0.183055 + 0.998 * 0.0011295) =$ *0.001491848*

5. *NORMALIZE(0.00059224259,0.001491848) is called. This results in* $\alpha = 479,83$ *and the distribution* $Q(b : 0.284, \neg b : 0.716)$.

Note: We omitted the computation of the subcalls ENUMERATE-ALL($\{J, M\}$,$\{j, m, b, e, a\}$) and ENUMERATE-ALL($\{J, M\}$,$\{j, m, b, e, \neg a\}$), but in principle it is not different from the other steps of the algorithm.
