Bayesian Networks and Decision Graphs

Exercises with Answers

May 2005

Exercise 1

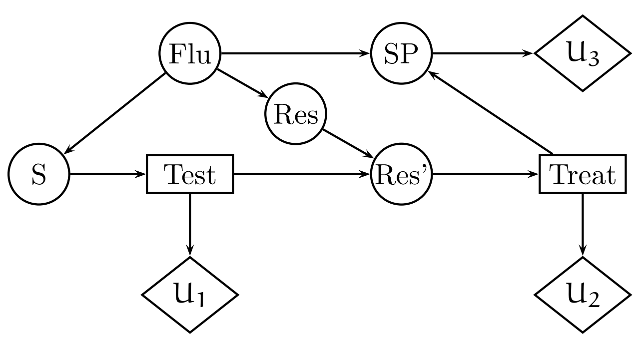
A person, lets call him Frank, goes to the doctor because he believes that he has the flu. At this particular time of the year, the doctor estimates that one out of 1000 persons suffers from the flu. The first thing the doctor checks is whether Frank appears to have the standard symptoms of the flu; if Frank suffers from the flu, then he will exhibit these symptoms with probability 0.9, but if he doesn't have the flu he may still have these symptoms with probability 0.05. After checking whether or not Frank has the symptoms, the doctor can decide to have a test performed which may reveal more information about whether or not Frank suffers from the flu; the cost of performing the test is 40 kr. The test can either give a positive or a negative result, and the frequency of false positives and false negatives is 0.05 and 0.1, respectively. After observing the test result (if any) the doctor can decide to administer a drug that with probability 0.6 may shorten the sickness period if Frank suffers from the flu (if he hasn't got the flu, then the drug has no effect). The cost of administering the drug is 100 kr., and if the sickness period is shortened the doctor estimates that this is worth 1000 kr.

a) Construct an influence diagram for the doctor from the description above.

b) Specify the probability distributions and the utility functions for the influence diagram.

Answers

The influence diagram can be structured in a few different ways depending on how you treat the test, but one approach could be:



The utility functions are $U_1(\text{Test}) = (-40, 0)$, $U_2(\text{Treat}) = (-100, 0)$ and $U_1(\text{SP}) = (1000, 0)$.

For the probabilities we have $P(\text{Flu}) = (0.001, 0.999)$ together with the tables:

		Flu	
		flu	¬flu
Symp.	y	0.9	0.05
	n	0.1	0.95

$P(\text{Symp.}|\text{Flu})$

		Flu	
		flu	¬flu
Res.	pos	0.9	0.05
	neg	0.1	0.95

$P(\text{Res.}|\text{Flu})$

		Res.	
		pos	neg
Test	y	$(1, 0, 0)$	$(0, 1, 0)$
	n	$(0, 0, 1)$	$(0, 0, 1)$

$P(\text{Res'}|\text{Test}, \text{Res})$

		Flu	
		flu	¬flu
Treat	y	$(0.6, 0.4)$	$(0, 1)$
	n	$(0, 1)$	$(0, 1)$

$P(\text{SP}|\text{Flu}, \text{Treat})$

Exercise 2

Assume that a guy (lets call him Frank again) is thinking about buying a used car for 20000 kr., and the market price for similar cars with no defects is 23000 kr. The car may, however, have defects which can be repaired at the cost of 5000 kr.; the probability that the car has defects is 0.3. Frank has the option of asking a mechanic to perform (exactly) one out of two different tests on the car. Test1 has three possible outcomes, namely no-defects, defects and inconclusive. For Test2 there are only two possible outcomes (no-defects and defects). If Frank chooses to have a test performed on the car, the mechanic will report the result back to Frank who then decides whether of not to buy the car; the cost of Test1 is 300 kr. and the cost of Test2 is 1000 kr.

a) Construct a decision tree for Frank's decision problem.

b) Calculate the maximum expected utility and the optimal strategy for the decision tree; calculate the required probabilities from the joint probability table (over the variables Test1, Test2 and StateOfCar) specified below.

		Test1		
		no-defects	defects	inconclusive
Test2	no-defects	$(0.448, 0.00375)$	$(0.028, 0.05625)$	$(0.084, 0.015)$
	defects	$(0.112, 0.01125)$	$(0.007, 0.16875)$	$(0.021, 0.045)$

Answers

The decision tree can be seen on the following page (bold lines specify the optimal strategy):

In order to calculate the optimal strategy we need $P(S)$, $P(S|T_1)$ and $P(S|T_2)$.

We have:

$$P(S|T_1) = \frac{P(S, T_1)}{P(T_1)} = \frac{P(T_1|S)P(S)}{\sum_S P(T_1|S)P(S)}$$

Using the tables we get (for $P(S, T_1)$):

		Test1		
		¬d	d	inc.
S	¬d	0.56	0.035	0.105
	d	0.015	0.225	0.06

$P(S, T_1)$

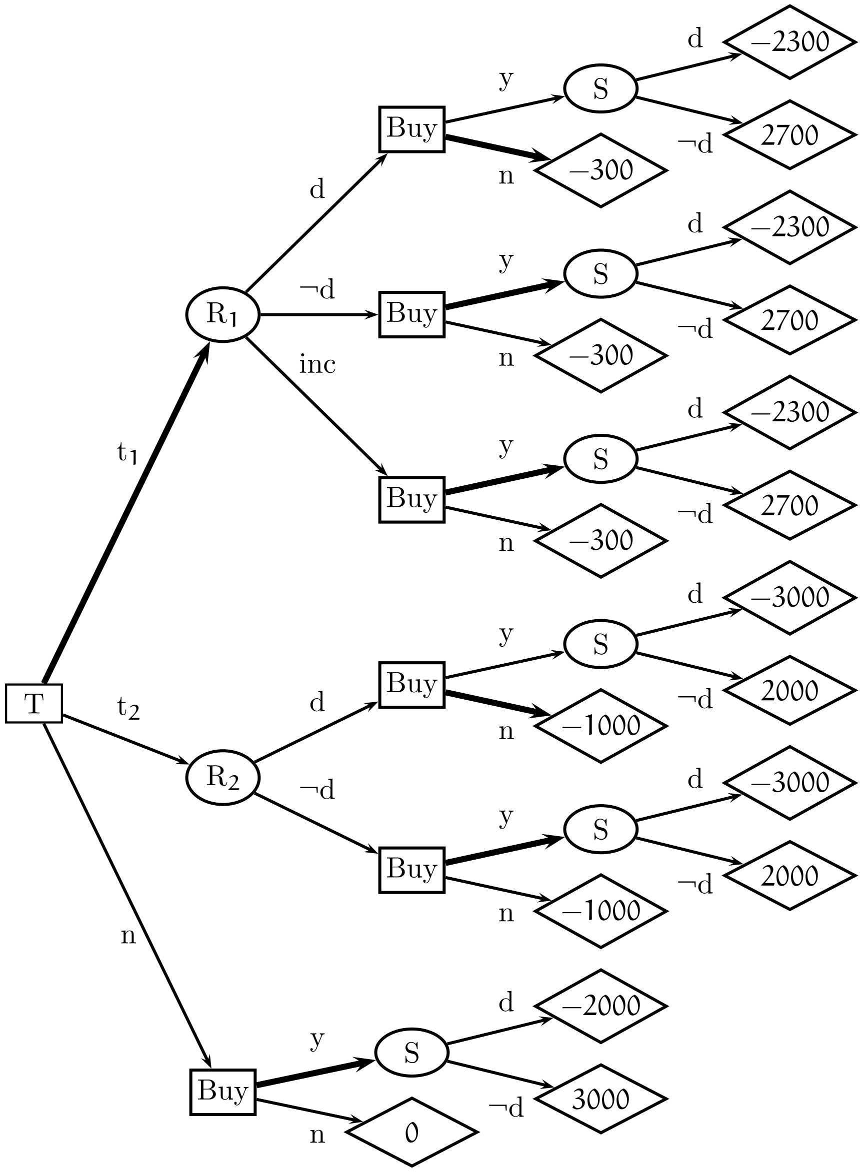
Hence, $P(T_1) = (0.575, 0.26, 0.165)$ and finally:

		Test1		
		¬d	d	inc.
S	¬d	0.974	0.135	0.636
	d	0.026	0.865	0.364

$P(S|T_1)$

The latter probability distribution is calculated in a similar way, and is given by:

		Test2	
		¬d	d
S	¬d	0.882	0.384
	d	0.118	0.616



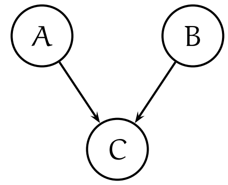
Exercise 3

Consider the Bayesian network shown below (all variables are binary), and assume that C is associated with the conditional probability table $P(C|A, B)$:

		B	
		b_1	b_2
A	a_1	$(0.75, 0.25)$	$(0.7, 0.3)$
	a_2	$(0.4, 0.6)$	$(0.1, 0.9)$

For $P(C|A = a_1, B = b_1)$ we have an initial sample size of 16, and for $P(C|A = a_2, B = b_1)$ we have an initial sample size of 10.

a) Perform fractional updating for the following sequence of cases where $B = b_1$ and where



the states of A and C are:

$$<(a_1,c_1),(a_2,c_2),(a_2,c_1),(a_1,c_1),(a_1,c_2),(a_2,c_2),(a_1,c_1)>.$$

b) Perform fractional updating on the same sequence as above but with fading factor 0.9.

Answers

We start out with $P(C|a_1,b_1) = \left(\frac{12}{16},\frac{4}{16}\right)$ and $P(C|a_2,b_1) = \left(\frac{4}{10},\frac{6}{10}\right)$

For the first part we get the following sequence:

1. $\left(\frac{13}{17},\frac{4}{17}\right)$ and $\left(\frac{4}{10},\frac{6}{10}\right)$
2. $\left(\frac{13}{17},\frac{4}{17}\right)$ and $\left(\frac{4}{11},\frac{7}{11}\right)$
3. $\left(\frac{13}{17},\frac{4}{17}\right)$ and $\left(\frac{5}{12},\frac{7}{12}\right)$
4. $\left(\frac{14}{18},\frac{4}{18}\right)$ and $\left(\frac{5}{12},\frac{7}{12}\right)$
5. $\left(\frac{14}{19},\frac{5}{19}\right)$ and $\left(\frac{5}{12},\frac{7}{12}\right)$
6. $\left(\frac{14}{19},\frac{5}{19}\right)$ and $\left(\frac{5}{13},\frac{8}{13}\right)$
7. $\left(\frac{15}{20},\frac{5}{20}\right)$ and $\left(\frac{5}{13},\frac{8}{13}\right)$

For the second part we get the following two denominators:

$$a_1,b_1 : (((16\cdot 0.9+1)\cdot 0.9+1)\cdot 0.9+1) = 13.9366$$
$$a_2,b_1 : ((10\cdot 0.9+1)\cdot 0.9+1)\cdot 0.9+1 = 10$$

The nominators for a_1,b_1 become:

1. $[12\cdot 0.9+1, 4\cdot 0.9]$
2. $[(12\cdot 0.9+1)\cdot 0.9+1, 4\cdot 0.9^2]$
3. $[((12\cdot 0.9+1)\cdot 0.9+1)\cdot 0.9, 4\cdot 0.9^3+1]$
4. $[((12\cdot 0.9+1)\cdot 0.9+1)\cdot 0.9^2+1, (4\cdot 0.9^3+1)\cdot 0.9] = [10.4122, 3.5244]$

The nominators for a_2,b_1 become:

1. $[4\cdot 0.9, 6\cdot 0.9+1]$
2. $[4\cdot 0.9^2+1, (6\cdot 0.9+1)\cdot 0.9]$
3. $[(4\cdot 0.9^2+1)\cdot 0l.9, (6\cdot 0.9+1)\cdot 0.9^2+1] = (3.816, 6.184)$

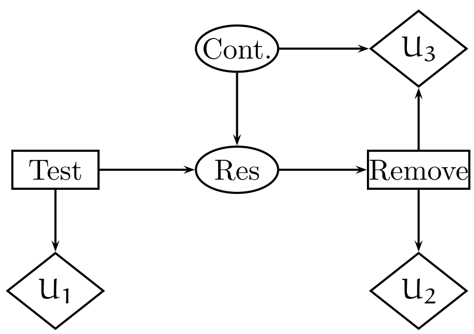
Exercise 4

An environmental department visits a site where a chemical production facility has previously been situated. Based on the department's knowledge about the facility, they estimate that there is a 0.6 risk that chemicals from the facility have contaminated the soil. If the soil is contaminated (and nothing is done about it) all people in the surrounding area will have to undergo a medical examination due to the possible exposure; there are 1000 persons in the area, and the cost of examining/treating one person is \$100. To avoid exposure, the department can decide to remove the top layer of the soil which, in case the ground is contaminated, will completely remove the risk of exposure; the cost of removing the soil is \$30000. Before making the decision of whether or not to remove the top layer of the soil, the department can perform a test which will give a positive result (with probability 0.9) if the ground is contaminated; if the ground is not contaminated the test will give a positive result with probability 0.01. The cost of performing the test is \$1000.

a) Construct an influence diagram for the environmental department from the description above.

b) Specify the probability distributions and the utility functions for the influence diagram.

Answers



The utility functions are: $U_1(\text{Test}) = (-1000, 0)$, $U_2(\text{Remove}) = (-30000, 0)$ and:

		Cont. y	Cont. n
Remove	y	0	0
	n	−100000	0

$U_3(\text{Cont}, \text{Remove})$

For the probabilities we have $P(\text{Cont.}) = (0.6, 0.4)$ and:[1]

		Cont. y	Cont. n
Test	y	(0.9, 0.1, 0)	(0.01, 0.99, 0)
	n	(0, 0, 1)	(0, 0, 1)

$P(\text{Res}|\text{Test}, \text{Cont.})$

[1] Note that the state space for Res consists of three states: (pos, neg, nr), where nr represents no-result.

Exercise 5

An electrical utility firm has build a prototype of some new device. The firm now has to decide whether to make a thorough test of the device or whether to drop it completely. It is estimated that the test will cost \$30k, and past experience indicates that only 30% of all products are successful in the test.

If the test of the device is successful, then the company is faced with a new decision: how many devices should be produced (none, 2000 or 4000)? The cost of producing 2000 devices is \$50k whereas the cost of producing 4000 devices is \$95k. If the test is not successful, then the firm will drop the production of the device.

It is estimated that if 2000 devices are produced there is a 30% chance that a competing company will respond with a similar product; if 4000 devices are produced then the probability is 40%. If a competing company respond the price per unit sold will be as follows (assuming the entire production is sold):

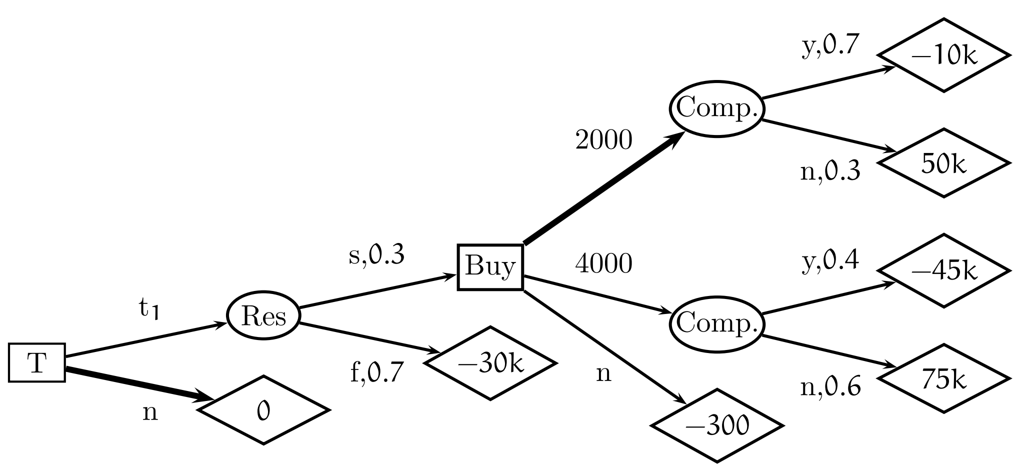
		Production	
		large (4000 devices)	small (2000 devices)
Competition	response	\$20	\$35
	no response	\$50	\$65

a) Construct a decision tree for the company's decision problem.

b) Calculate the maximum expected utility and the optimal strategy for the decision tree.

Answers

The decision tree can be seen below (bold lines specify the optimal strategy). Note that the optimal strategy is not to perform the test, which gives a maximum expected utility of 0.



Exercise 6

Consider the Bayesian network shown below (all variables are binary), with the probability distributions $P(B) = (0.1, 0.9)$ and:

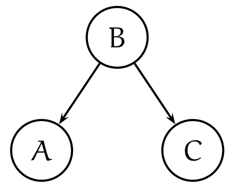
		B: b_1	B: b_2
A	a_1	0.1	0.6
	a_2	0.9	0.4

$P(A|B)$

and

		B: b_1	B: b_2
C	c_1	0.3	0.05
	c_2	0.7	0.95

$P(C|B)$



For $P(A|B = b_1)$ we have an initial sample size of 10, and for $P(A|B = b_2)$ we have an initial sample size of 30.

a) Perform fractional updating of $P(A|B)$ for the following sequence of cases:

$$< (a_1, b_2), (a_2, b_2), (a_1, b_1), (a_1, b_1), (a_2, b_2), (a_2, b_1), (a_1, b_1) > .$$

b) Perform fractional updating on the same sequence as above but with fading factor 0.9.

Answers

We start off with $P(A|B = b_1) = \left(\frac{1}{10}, \frac{9}{10}\right)$ and $P(A|B = b_2) = \left(\frac{18}{30}, \frac{12}{30}\right)$. Note that we can disregard $P(C|B)$.

For the first part we get the following sequence:

1. $\left(\frac{1}{10}, \frac{9}{10}\right)$ and $\left(\frac{19}{31}, \frac{12}{31}\right)$
2. $\left(\frac{1}{10}, \frac{9}{10}\right)$ and $\left(\frac{19}{32}, \frac{13}{32}\right)$
3. $\left(\frac{2}{11}, \frac{9}{11}\right)$ and $\left(\frac{19}{32}, \frac{13}{32}\right)$
4. $\left(\frac{3}{12}, \frac{9}{12}\right)$ and $\left(\frac{19}{32}, \frac{13}{32}\right)$
5. $\left(\frac{3}{12}, \frac{9}{12}\right)$ and $\left(\frac{19}{33}, \frac{14}{33}\right)$
6. $\left(\frac{3}{13}, \frac{10}{13}\right)$ and $\left(\frac{19}{33}, \frac{14}{33}\right)$
7. $\left(\frac{4}{14}, \frac{10}{14}\right)$ and $\left(\frac{19}{33}, \frac{14}{33}\right)$

For the second part we get the following two denominators:

$$b_1 : (((10 \cdot 0.9 + 1) \cdot 0.9 + 1) \cdot 0.9 + 1) \cdot 0.9 + 1 = 10$$
$$b_2 : ((30 \cdot 0.9 + 1) \cdot 0.9 + 1) \cdot 0.9 + 1 = 24.58$$

The nominators for b_1 become:

$$(((1 \cdot 0.9 + 1) \cdot 0.9 + 1) \cdot 0.9) \cdot 0.9 + 1 = 3.1951$$
$$(9 \cdot 0.9^3 + 1) \cdot 0.9 = 6.8049$$

The nominators for $\mathfrak{b}_2$ become:

$$(18 \cdot 0.9 + 1) \cdot 0.9^2 = 13.932$$
$$(12 \cdot 0.9^2 + 1) \cdot 0.9 + 1 = 10.648$$

Exercise 7

Assume that a guy (lets call him Frank again) wakes up one morning feeling ill. Frank thinks that he may have caught the flu, and he now has to decide whether to go to the pharmacy and buy some medicine (at the cost of 150 kr). If Frank has the flu, then the medicine will relieve his discomfort during the sickness period; if he doesn't have the flu then the medicine will have no effect. Assuming that Frank doesn't suffer from the discomfort caused by a flu, then he can take some additional overtime work which will be worth 2000 kr.

Before Frank decides to go to the pharmacy, he can try to get more information by buying a thermometer (at a cost of 10 kr.) and test whether he has a fever; the thermometer is very precise and will indicate a fever if and only if Frank actually has a fever.

- Perform a myopic value of information analysis for the decision problem above and calculate the expected profit of performing the test (i.e., buying the thermometer at a cost of 10 kr and taking the temperature). Calculate the required probabilities from the joint probability table (over the variables Flu and Fever) specified below.

		Flu	
		yes	no
Fever	yes	0.00905	0.09702
	no	0.00095	0.89298

Answers

First we calculate $P(\text{Flu}) = (0.01; 0.99)$, $P(\text{Fever}) = (0.10607; 0.89393)$, and:

		Flu	
		yes	no
Fever	yes	0.08532	0.91468
	no	0.00106	0.99894

$P(\text{Flu}|\text{Fever})$

With $U(\text{Buy}, \text{Flu})$:

		Flu	
		yes	no
Buy	yes	1850	1850
	no	0	2000

we get:

$$EU(\text{Buy}) = (1850, 1980) \quad (1)$$

hence, $V(P(\text{Flu})) = 1980$.

Now $EU(\text{Buy}|\text{Fever}) =$

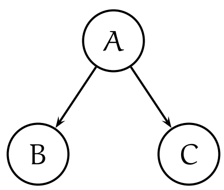
		Buy	
		yes	no
Fever	yes	1850	1829.36
	no	1850	2000

thus, $V(P(\text{Flu}|\text{Fever})) = (1850, 1997.88)$ and in particular:

$$\begin{aligned} EV(\text{Fever}) &= 1982.19; \\ EB &= 1982.19 - 1980 = 2.19; \\ EP &= 2.19 - 10 = -7.81. \end{aligned}$$

Exercise 8

Consider the Bayesian network as well as the conditional probability tables $P(B|A)$ and $P(C|A)$ specified below.



		B	
		b_1	b_2
A	a_1	0.355	0.645
	a_2	0.48	0.52

$P(B|A)$

		C	
		c_1	c_2
A	a_1	0.3	0.7
	a_2	0.3	0.7

$P(C|A)$

a) Calculate the size of the Bayesian network above.

b) Based on the joint probability table specified below, calculate the unspecified probability table for the Bayesian network above.

c) Calculate the acceptance measure for the Bayesian network above (w.r.t. the joint probability table specified below) using the Euclidean distance and with $k = 10$.

		B	
		b_1	b_2
A	a_1	$(0.171, 0.042)$	$(0.009, 0.378)$
	a_2	$(0.024, 0.168)$	$(0.096, 0.112)$

$P(A, B, C)$

Answers

a) 10.

b) $P(A) = (0.6; 0.4)$.

c) $P(A, B, C) =$

		B	
		b_1	b_2
A	a_1	$(0.0639; 0.1491)$	$(0.1161; 0.2709)$
	a_2	$(0.0576; 0.1344)$	$(0.0624; 0.1456)$

Hence: $\mathtt{dist}_q = 0.0504$ and $\mathtt{accept} = 10 + 10 \cdot \mathtt{dist}_q = 10.5$.

Exercise 9

A company has observed that one of their software systems is unstable, and they have identified a component which they suspect is the cause of the instability. The company estimates that the prior probability for the component being faulty is 0.01, and if the component is faulty then it causes the system to become unstable with probability 0.99; if the component is not faulty, then the system may still be unstable (due to some other unspecified element) with probability 0.001.

To try to solve the problem the company must first decide whether to "patch" the component at a cost 10000 Dkr: if the component is faulty, then the patch will solve the fault with probability 0.95 (there may be several things wrong and not all of which may be covered by the patch), but if the component is not faulty then the patch will have no effect. The company also knows that in the near future the vendor of the component will make another patch available at the cost of 20000 Dkr; the two patches focuses on different parts of the component. This new patch will solve the problem with probability 0.99, and (as for the first patch) if the component is not faulty then the patch will have no effect. Thus, after deciding on the first patch, the company observes whether or not the patch solved the problem (i.e., is the system still unstable?) and it then has to decide on the second patch. The company estimates that (after the final decision has been made) the value of having a fully functioning component is worth 100000 Dkr.

a) Construct an influence diagram for the company from the description above.

b) Specify the probability distributions and the utility functions for the influence diagram.

Exercise 10

A guy (lets call him Frank) is working in a software company. Frank has decided to ask his boss for a yearly raise of 10000 Dkr. Before asking for the raise, Frank should decide whether to better himself by following a course (at his own expense) with the aim of improving his chances of getting the raise. Frank can decide either to follow an AttoSoft certification course (at a cost of 3000 Dkr), module 1 of the BSS course provided by AAU (at a cost of 5000 Dkr), or he can decide not to follow any of the courses. Frank only has time to follow one of the courses, but if he decides to follow module 1 of the BSS course (and if he passes the course), then he can decide to follow module 2 of the course (which is the final module and costs 2000 Dkr).

The probability that Frank will get the raise is 0.2 if he decides not to follow any of the courses, 0.4 if he passes the AttoSoft certification course, 0.6 if he passes module 1 of the BSS course, and 0.8 if he passes both module 1 and module 2 of the BSS course. The probability that Frank will get the raise if he fails a particular course is the same as if he had not followed the course at all. The probability that Frank will pass either the Attosoft course or module 1 of the BSS course is 0.9 and the probability that he will pass module 2 of the BSS course is 0.8.

a) Construct a decision tree for the Frank's decision problem with a two-year time horizon.

b) Calculate the maximum expected utility of the optimal strategy for the decision tree.

Exercise 11

Consider the Bayesian network shown below (all variables are binary), with the probability distributions $P(A) = (0.8, 0.2)$ and:

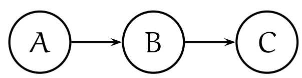
		A	
		a_1	a_2
B	b_1	0.1	0.6
	b_2	0.9	0.4

P(B|A)

and

		B	
		b_1	b_2
C	c_1	0.3	0.05
	c_2	0.7	0.95

P(C|B)



For $P(B|A = a_1)$ we have an initial sample size of 10, and for $P(B|A = a_2)$ we have an initial sample size of 20.

a) Perform fractional updating of $P(B|A)$ for the following sequence of cases:

$$< (a_1, b_1), (a_2, b_1), (a_1, b_1), (a_1, b_2), (a_2, b_2), (a_2, b_1), (a_1, b_1) > .$$

b) Use the initial tables and perform fractional updating of $P(B|A)$ for the following sequence of cases (the joint probability table $P(A, B, C)$ is shown below):

$$< (a_1, c_1), (c_2) > .$$

		A	
		a_1	a_2
B	b_1	(0.024; 0.056)	(0.036; 0.084)
	b_2	(0.036; 0.684)	(0.004; 0.076)

P(A, B, C)

Exercise 12

Consider a database of houses represented by the five training examples below. The target attribute *Acceptable*, which can have values `yes` or `no`, is to be predicted based on the other attributes of the house in question. These attributes indicate a) whether the house is situated in a good area (*GoodArea* having values `yes` and `no`), b) if it is close to a public school (*NearSchool* having values `yes` and `no`), c) how far there is to a shopping mall (*NearShop* having values `far`, `medium` and `close`), d) the number of rooms in the house (*#Rooms* having values `3` and `4`), and e) if a new kitchen is installed (`NewKitchen` having values `yes` and `no`).

	Attributes					Target
House	*GoodArea*	*NearSchool*	*NearShop*	*#Rooms*	*NewKitchen*	*Acceptable*
1	`yes`	`no`	`close`	3	`yes`	`yes`
2	`no`	`yes`	`close`	3	`no`	`no`
3	`yes`	`no`	`far`	4	`no`	`yes`
4	`yes`	`no`	`far`	3	`no`	`no`
5	`no`	`yes`	`medium`	4	`no`	`yes`

a) Calculate the entropy for the target attribute.[2]

b) Show the decision/classification tree that would be learned by the ID3 algorithm assuming that it is given the training examples above.

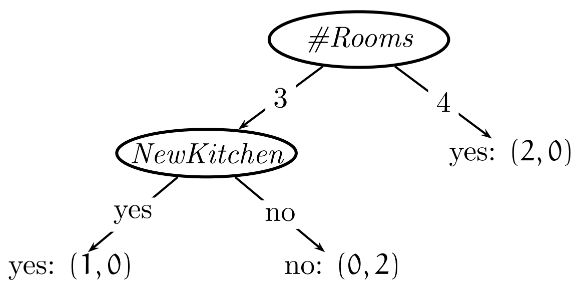
c) Show the value of the information gain for each candidate attribute at each step in the construction of the tree.

Answers

The entropy for the target attribute:

$$\mathsf{Ent}(S) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.971$$

The classification tree has the following form:



The information gain for the first step:

$$\begin{aligned}
\mathsf{Gain}(S, \mathit{GoodArea}) &= 0.971 - 0.9508 = 0.0202 \\
\mathsf{Gain}(S, \mathit{NearSchool}) &= 0.971 - 0.9508 = 0.0202 \\
\mathsf{Gain}(S, \mathit{NearShop}) &= 0.971 - 0.8 = 0.171 \\
\mathsf{Gain}(S, \mathit{\#Rooms}) &= 0.971 - 0.5508 = 0.4202 \\
\mathsf{Gain}(S, \mathit{NewKitchen}) &= 0.971 - 0.0.8 = 0.171
\end{aligned}$$

In the second step we continue for *#Rooms*=3, and for this partition the entropy for the target attribute is:

$$\mathsf{Ent}(S') = 0.918$$

[2]Note that $\log_2(x) = \frac{\log_{10}(x)}{\log_{10}(2)}$.

The information gain for the second step:

$$\begin{aligned}
\text{Gain}(S, \textit{GoodArea}) &= 0.918 - 2/30.2513 \\
\text{Gain}(S, \textit{NearSchool}) &= 0.918 - 2/3 = 0.2513 \\
\text{Gain}(S, \textit{NearShop}) &= 0.918 - 2/3 = 0.2513 \\
\text{Gain}(S, \textit{NewKitchen}) &= 0.918 - 0 = 0.918
\end{aligned}$$

Exercise 13

Consider a database of cars represented by the five training examples below. The target attribute *Acceptable*, which can have values `yes` and `no`, is to be predicted based on the other attributes of the car in question. These attributes indicate a) the age of the car (*Age* having values < 5 `years` and ≥ 5 `years`), b) the make of the car (*Make* having states `Toyota` and `Mazda`), c) the number of previous owners (*#Owners* having values 1, 2 and 3), d) the number of kilometers (*#Kilometers* having values > 150k and ≤ 150k) and e) the number of doors (*#Doors* having values 3 and 5).

	Attributes					Target
	Age	*Make*	*#Owners*	*#Kilometers*	*#Doors*	*Acceptable*
1	< 5	Mazda	1	> 150k	3	yes
2	≥ 5	Mazda	3	> 150k	3	no
3	≥ 5	Toyota	1	≤ 150k	3	no
4	≥ 5	Mazda	3	> 150k	5	yes
5	≥ 5	Toyota	2	≤ 150k	5	yes

a) Calculate the entropy for the attribute *#Owners*.[3]

b) Show the decision/classification tree that would be learned by the ID3 algorithm assuming that it is given the training examples in the database.

c) Show the value of the information gain for each candidate attribute at each step in the construction of the tree.

0.1 Answers

For question (a):

$$\text{ENT}(\#\text{Owners}) = 1.5219 \tag{2}$$

[3] Note that $\log_2(x) = \frac{\log_{10}(x)}{\log_{10}(2)}$.

For question (b):

$$\begin{aligned}
\mathtt{Ent}(\text{Accept}) &= 0.971 \\
\mathtt{Gain}(\text{Age}) &= 0.171 \\
\mathtt{Gain}(\text{Make}) &= 0.0202 \\
\mathtt{Gain}(\#\text{Owners}) &= 0.171 \\
\mathtt{Gain}(\#\text{Kilo}) &= 0.0202 \\
\mathtt{Gain}(\#\text{Doors}) &= 0.4202
\end{aligned}$$

I.e., we start with #Doors. For the second layer:

$$\begin{aligned}
\mathtt{Ent}(\text{Accept}|\#\text{Doors} = 3) &= 0.918 \\
\mathtt{Gain}(\text{Age}) &= 0.918 - 0 \\
\mathtt{Gain}(\text{Make}) &= 0.918 - 2/3 \\
\mathtt{Gain}(\#\text{Owners}) &= 0.918 - 2/3 \\
\mathtt{Gain}(\#\text{Kilo}) &= 0.918 - 2/3
\end{aligned}$$

Hence, for #Doors= 3 we pick Age as the next node. The classification tree can directly be constructed from these calculations.

Exercise 14

A travel agency has made a database (based on questionaires from previous customers) with information about the quality of their various hotels. The target attribute Satisfactory, which can have the values yes and no, is to be predicted based on three characteristics of the hotels: a) is there a swimming pool at the hotel (Pool with states yes and no)?, b) how close is it to the nearest town (Town with states < 1km and ≥ 1km)?, and c) how close is it to the beach (Beach with states < 500m, ≥ 500m;< 1km and ≥ 1km)?

	Attributes			Target
	Beach	Town	Pool	Satisfactory
1	< 500m	< 1km	no	yes
2	≥ 500m;< 1km	< 1km	yes	yes
3	< 500m	≥ 1km	yes	yes
4	≥ 1km	< 1km	yes	yes
5	≥ 500m;< 1km	≥ 1km	yes	no
6	≥ 1km	< 1km	no	no

a) Calculate the entropy of the attribute Beach.[4]

b) Show the decision/classification tree that would be learned by the ID3 algorithm assuming that it is given the training examples above.

c) Show the value of the information gain for each candidate attribute at each step in the construction of the tree.

[4]Note that $\log_2(x) = \frac{\log_{10}(x)}{\log_{10}(2)}$.