Exercises for Partial Exam AI: Reasoning with Uncertainty Theory: from Ertel's book: chapter 7, and from chapter 8: sections 8.4,8.5 and 8.6.

13 mei 2013

1. A witness of a nighttime accident involving a taxi declares that the taxi was blue.

   All taxis in town are blue or green. It is known that under dim lighting conditions discrimination between blue and green is 75% reliable; which means that $P(LB|B)$ as well as $P(\neg LB|\neg B)$ are 0.75, where B is the boolean variable for the predicate *"the taxi is blue"* and LB is the boolean variable for the predicate *"the taxi looked blue"*.

   Suppose that 9 out of 10 taxis are actually green. Given the declaration of our witness what is the probability that the taxi is indeed blue?

   Hint: pas Bayes' rule toe.

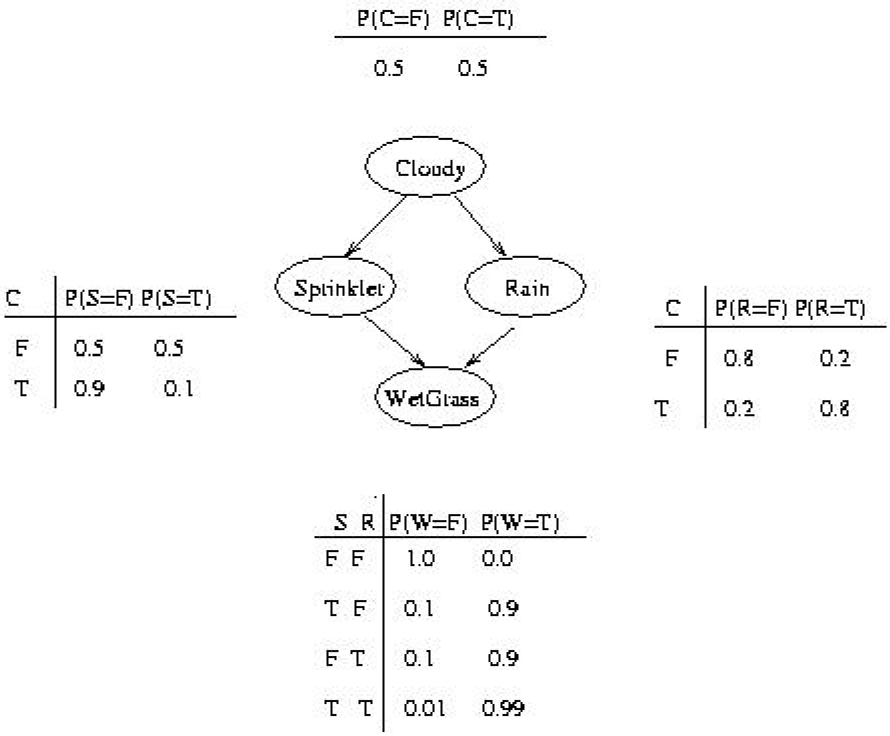
   a) 0.25
   b) 0.1
   c) 0.9
   d) 0.75

2. Given the Sprinkler network shown in Figure 1. Which of the following answers is closest to the correct value of $P(S = True|W = True)$ (the probability that the **S**prinkler was on given that the grass is **W**et)?

   a) 0.2781
   b) 0.6471
   c) 0.1945
   d) 0.4298

3. Given the Sprinkler network shown in Figure 1. Which value comes closest to the correct value of $P(S = True|W = True, R = True)$ (the probability that the **S**prinkler was on given that the grass is **W**et and that it was **R**aining)?

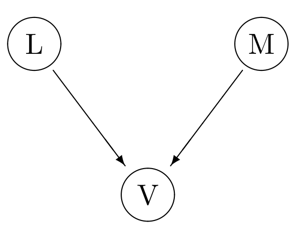
   a) 0.2781
   b) 0.6471



Figuur 1: The Sprinkler Bayesian network

c) 0.1945
d) 0.4298

4. In the Bayesian Network below with three boolean variables the probabilities for P and M are: $P(M = true) = 0,1$ and $P(L = true) = 0.7$ and the conditional probabilities for variable V are as shown in the table.



L	M	$P(V = true \mid L, M)$
true	true	0,9
true	false	0,5
false	true	0,3
false	false	0,05

What is the value of $P(V = true \mid L = true)$?

a) 0.72
b) 0.54
c) 0.46
d) 0.28

5. A full joint distribution for the "Toothache, Cavity, Catch Worldïs given by the table below, copied from Figure 13.3 in the book Artificial Intelligence (2nd edition) of Russel and Norvig.

	toothache		∼ toothache	
	catch	∼catch	catch	∼catch
cavity	0.108	0.012	0.072	0.008
∼cavity	0.016	0.064	0.144	0.567

A full joint distribution for the "Toothache, Cavity, Catch World"

How will a Bayesian network for this world look like, if we look at the conditional independencies that follow from the probabilities?

a) Catch and Toothache both have one parent node: Cavity.

b) The BN is a linear model with Toothache as child node of Catch and Catch as child node of Cavity

c) like a) but Catch is also parent of Toothache.

d) Toothache has two parents Cavity and Catch, and there are no other arrows in the network.

6. Let D be a data set. A data element d in this set is a vector of k feature values $d =< v_0, ...v_{k-1} >$, where v_j is the value of the $j-th$ feature A_j of d. With $v_j(d)$ we denote the value of this $j-th$ component of the vector d. Let A_{k-1} be the class feature. The value of this class feature determines the class the data element belongs to.
Let A be a feature with n distinct possible values $a_i(i = 1..n)$. Let D_i be the following subset of D: $D_i = \{d\epsilon D|v_j(d) = a_i\}$. Thus D_i is the subset of D that contains those elements of D that have the feature value of feature A equal to the $i-th$ value of feature A.

For a data set D the InfoGain for feature A is defined as:

$$InfoGain(D, A) = H(D) - \sum_{i=1}^{i=n} \frac{|D_i|}{|D|}.H(D_i)$$

where $H(D)$ is the Shannon entropy of the probability distribution P that is the most likely estimator (i.e. determined by relative frequencies) of the probability distribution of the class values in the data set D. Similarly, $H(D_i)$ is the Shannon entropy of the subset D_i of D.
The Shannon Entropy of the probability distribution $p =< p_0, ...p_{n-1} >$ equals:

$$H(p) = -\sum_{i-1}^{i=n} p_i.log_2(p_i)$$

A retailer wants for marketing purposes distinguish between costumers younger then 35 (class Y) and customers older then 35 (class O). The following table summarizes the data set in the data base of the retailer in an abstract form. The relevant attributes, determined by domain knowledge, are for convenience denoted by A with values *a1*, *a2* and *a3*, B with values *b1* and *b2*, C with values *c1* and *c2* and D with values *d1* and *d2*

A	B	C	D	Number of Instances	
				Y	O
a1	b1	c1	d1	12	4
a2	b1	c1	d2	4	6
a3	b1	c1	d1	6	0
a1	b2	c1	d2	0	12
a2	b2	c1	d1	4	2
a3	b2	c1	d2	0	4
a1	b1	c2	d1	0	8
a2	b1	c2	d2	8	0
a3	b1	c2	d1	4	0
a1	b2	c2	d2	0	4
a2	b2	c2	d1	7	0
a3	b2	c2	d2	5	0

The analyst wants to learn the above classification problem using decision trees. If he uses "information gain" as selection criteria what will be the first attribute for splitting the examples?

(a) A
(b) B
(c) C
(d) D

7. Given three random variables A, B, and C. According to probability theory one of the following does **not** hold. Which one?

   a $P(A, B, C) = P(A)P(B|A)P(C|A, B)$
   b $P(A, B, C) = P(B)P(A|B)P(C|A, B)$
   c $P(A, B, C) = P(C)P(A|C)P(B|C, A)$
   d $P(A, B, C) = P(A)P(B)P(C)$

8. Construct a probability distribution $P(A, B, C)$ that disproves

$$P(A, B) = P(A)P(B) \rightarrow P(A, B|C) = P(A|C)P(B|C)$$

   Draw a Bayesian network that illustrates a counter model.

9. Construct a probability distribution $P(A, B, C)$ that disproves

$$P(A, B|C) = P(A|C)P(B|C) \rightarrow P(A, B) = P(A)P(B)$$

   Draw a Bayesian network that illustrates a counter model.

10. Someone throws with two fair dice. He reports you that at least one of the outcomes is a six. What is the probability that the outcome is two sixes?

11. Draw a Bayesian Network for the previous problem. The observed node is called AtLeastOneSix (with observed value True), the node whose probability distribution is to be derived is called BothSix.

12. (From R&N, 3rd edition, page 558). We have a bag of three biased coins a,b, and c with probabilities of coming up heads of 30%,60% and 70%, respectively. One coin is drawn randomly from the bag (with equal likelihood of drawing each of the three coins) and then the coin is flipped three time to generate outcomes X_1, X_2 and X_3.

    a) Draw the Bayesian network correspond to this setup and define the condtional probability table (CPTs) of each of the nodes.

    b) Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

    In the last exercise you were asked to find the most likely model (urn) given some data. The models were assigned (prior) probabilities. What you were asked is compute the a-posterior probabilities of the models given the outcome of the experiment. And then tell what the most likely model is. You have learned that "probability" is a problematic notion. Indeed there are different schools. Some schools learn that you can only assign probabilities to outcomes of experiments that you can repeat over and over again: probability is seen as the limit of a sequence of relative frequencies (this is **not** a mathematical limit!). So models or hypotheses don't have probabilities. They are parameter values, not probability variables. Our AI robot belongs to the Bayesian school. He has a more liberal view. He may assign probabilities to any belief he has. The only constraint is that they should obey the axioms of probability theory. One of the questions answered differently by these schools is: "what has probability theory to do with statistics?".

    Remember that Bayes' rule is the following:

    $$P(H|D) = P(D|H)P(H)/P(D)$$

    If H stands for Hypotheses and D stands for Data, then the prob of H given D can be computed from the Likelihood of H, the prior probability of H and the probability of the data D. If you have followed a statistics course at the University you are familiar with the procedure that you follow according to the Standard Theory of Statistics (sometimes called frequentist theory). If you forgot, read for example "Statistical Analysis and the Illusion of Objectivity" (American Scientist, 1988, Vol. 76, pp.159-166) by J.O. Berger and D.A. Berry. If you belief that the outcome of this procedure only depends on the available Data, as it should be. You should make the following two related exercises and compare the answers.

13. You throw a coin 17 times. The outcome is 13 times head. Your H_0 hypotheses is: the coin is fair, i.e. $P(H) = 0.5$. Does your data mean you should reject the H_0 hypotheses? What is the P-value?

14. You perform the following experiment. You throw a coin until you have thrown 4 times head as well as 4 times tails. It turns out that this happens on the 17 throw. Does your data mean you should reject the H_0 hypotheses? What is the P-value?

    An answer can be found in Berger and Berry's paper (see above). The P-values (the critical values for deciding if you reject the H_0 hypothesis depends on the set up of the experiment, not only on the outcome, not only on the data D. Isn't that strange? For Bayesians, statistics is just applied probability theory. Bayes' rule plays a key role (the term "inverse probability" reverse to Bayes' inversion rule).

dinsdag
28
februari

Erik van Muiswinkel

4-8-'61

cabaret

Op 4 augustus 1961 werden ter weerszijden van de aardbol Barack Obama en Erik van Muiswinkel geboren. Als dat toeval is, zou het wel heel toevallig zijn! Daarom gooit Van Muiswinkel 25 jaar theaterervaring in de strijd voor zijn eerste grote soloshow: '4-8-'61'. Het verhaal van twee levens verteld in songs, sketches, imitaties, poëzie, conferences en een toespraak die de wereld zal veranderen. Begeleid door Omnibuzz, zijn vijfkoppige, twaalf instrumenten tellende band, haalt Van Muiswinkel twee uur lang

Figuur 2: Dat is wel heeeel toevallig... Om te berekenen hoe toevallig, moet je weten wat het veld van mogelijkheden is. Er zijn heel veel mensen op dezelfde dag geboren en bijna iedereen is geboren op een dag dat er ook iemand geboren is die bekend is geworden; en anders wel op een dag dat er iets anders opmerkelijks gebeurde.