Examination in Machine Intelligence

Thomas Dyhre Nielsen

January 8, 2020

On the next five pages you will find five questions covering different aspects of the course. The questions differ in their level of difficulty, and for each correctly answered question you will get a certain amount of points as indicated by each question. When solving the questions you are allowed to use all available material such as books, pocket calculator, etc., however, laptops/tablets and other networking devices are *not* allowed.

Before you answer a question make sure that you have read the question carefully. Moreover, make sure that you argue for your answers (e.g. include intermediate results) so that it is possible to follow your line of thought. Finally, it is important that your solutions are presented in a readable form. The answers to the questions should be written in English.

In addition to the five pages with questions, you are also provided with 10 pages that you can use when writing your answers to the questions.

- For each sheet of paper containing your response to the questions, please include your name, study number, current page number, and the total number of pages.
- If you need more paper, simply raise your hand to contact one of the guards in the examination room.

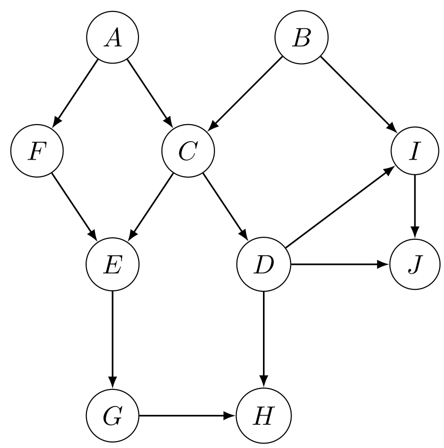
Lastly, please specify your name, study number, and study program here:

Full name	
Study number	
Study program	

Good luck with the questions
Thomas Dyhre Nielsen

Question 1 - 10 points

Consider the graph below:



1. What probability distributions (on the form $P(X|Y)$) should be specified in order to obtain a Bayesian network from the graph?
2. Assume that the variable J has three states labeled $\{a, b, c\}$ and that the remaining variables all have two states labeled t and f. Give an example of a table (containing probability values) representing a valid conditional probability distribution for variable J.
3. Which variables are d-separated from F?
4. Which variables are d-separated from F given hard evidence on C and E?
5. Which variables are d-separated from F given hard evidence on D?

Solution:

1. $P(A)$, $P(B)$, $P(F|C)$, $P(C|B, A)$, $P(I|B, D)$, $P(E|F, C)$, $P(E|A)$, $P(J|D, J)$, $P(G|E)$, $P(H|E, G)$
2. ...
3. B.
4. D, G, and H
5. None

Question 2 - 20 points

A contestant in a TV game show is placed in front of three doors (labeled D_1, D_2, and D_3). The game host explains that behind each of three doors there is either nothing, an angry bear, or a prize worth 10.000 EUR. In addition, the game host also explains that

- The contents behind D_2 is preferable to the contents behind D_1.
- The contents behind D_3 is less preferable than the contents behind D_2.
- The contents behind D_1 is different from the contents behind D_3.

The contestant in the show now needs to pick out the door holding the prize.

1. Model the problem as a constraint satisfaction problem, i.e., identify the variables, their domains as well as the constraints.
2. Represent the problem as a constraint network.
3. Make the network arc consistent.
4. Find a satisfying solution to the problem (if one exists) using variable elimination with elimination ordering D_1, D_2, and D_3. Show the operations involved when eliminating each of the variables.

Solution:

Answer 1 and 2

There is a variable for each bottle and each variable has domain $Dom(D_i) = \{nothing, bear, prize\}$. The constraints are as follows (numbered according to the ordering of the constraints above):

$C_1 = (D_1 < D_2)$:

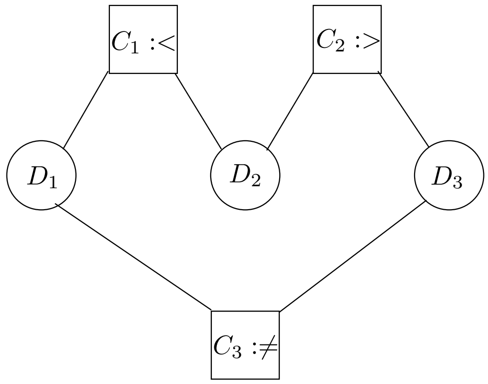
D_1	D_2
b	n
b	p
n	p

$C_2 = (D_3 < D_2)$:

D_3	D_2
b	n
b	p
n	p

$C_3 = (D_1 \neq D_3)$:

D_1	D_3
b	n
b	p
n	b
n	p
p	b
p	n



Answer 3

To make the network arc consistent we observe that, e.g., the constraint C_1 cannot be satisfied for $D_1 = p$ and, hence, p is removed from the domain of D_1. Continuing in this fashion we end up with the domains $Dom(D_1) = Dom(D_3) = \{b, n\}$ and $Dom(D_2) = \{n, p\}$.

Answer 4

Eliminating D_1:

$$C_4 = (C_1 \bowtie C_3)^{\downarrow D_2, D_3} = \left(\begin{array}{c|c|c} D_1 & D_2 & D_3 \\ \hline \text{b} & \text{n} & \text{n} \\ \text{b} & \text{p} & \text{n} \\ \text{n} & \text{p} & \text{b} \end{array} \right)^{\downarrow D_2, D_3} = \begin{array}{c|c} D_2 & D_3 \\ \hline \text{n} & \text{n} \\ \text{p} & \text{n} \\ \text{p} & \text{b} \end{array}$$

Eliminating D_2:

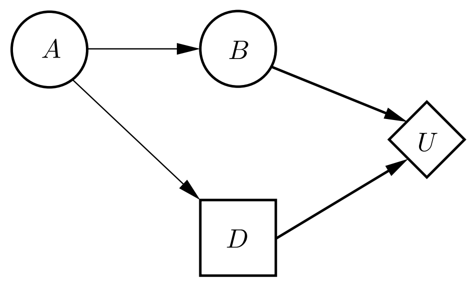
$$C_5 = (C_4 \bowtie C_2)^{\downarrow D_3} = \left(\begin{array}{c|c} D_2 & D_3 \\ \hline \text{p} & \text{n} \\ \text{p} & \text{b} \end{array} \right)^{\downarrow D_3} = \begin{array}{c} D_3 \\ \hline \text{n} \\ \text{b} \end{array}$$

A satisfying solution to the problem can now be found from C_5 and joining that with the intermediate tables/constraints produced during variable elimination:

$$\left(\begin{array}{c|c|c} D_1 & D_2 & D_3 \\ \hline \text{b} & \text{p} & \text{n} \\ \text{n} & \text{p} & \text{b} \end{array} \right).$$

Question 3 - 20 points

Consider the influence diagram defined by the graphical structure below together with the following conditional probability and utility tables:



A	
a_1	a_2
0.2	0.8

$P(A)$

		B			
		b_1	b_2	b_3	b_4
A	a_1	0.1	0.2	0.25	0.45
	a_2	0.5	0.3	0.1	0.1

$P(B|A)$

		B			
		b_1	b_2	b_3	b_4
D	d_1	100	90	20	0
	d_2	0	10	30	50

$U(B, D)$

1. Specify the temporal order in which a decision maker observes and decides upon the variables in the influence diagram.
2. Calculate the expected utility of each of the two decision options for D conditioned on its past. Specify the corresponding optimal decision function.
3. Assume that it is optional to observe A before deciding on D. What is the value of information about A for decision D.

Solution:

Sub-problem 1

The sequence in which the variables are observed and decided upon is: $\{A\} \prec D \prec \{B\}$.

Sub-problem 2

We start by calculating the expected utilities conditioned on the two possible outcomes of A:

$$EU(D|A = a_i) = \sum_B P(B|A = a_i)U(B, D)$$

Thus:

$$EU(D|A=a_1) = \sum_B \left(\begin{array}{|c|cccc|} \hline & & \text{B} & & \\ & b_1 & b_2 & b_3 & b_4 \\ \hline A=a_1 & 0.1 & 0.2 & 0.25 & 0.45 \\ \hline \end{array} \quad \begin{array}{|cc|cccc|} \hline & & & \text{B} & & \\ & & b_1 & b_2 & b_3 & b_4 \\ \hline \text{D} & d_1 & 100 & 90 & 20 & 0 \\ & d_2 & 0 & 10 & 30 & 50 \\ \hline \end{array} \right)$$

$$= \begin{array}{|c|c|} \hline D=d_1 & \mathbf{33} \\ D=d_2 & 32 \\ \hline \end{array}$$

$$EU(D|A=a_1) = \sum_B \left(\begin{array}{|c|cccc|} \hline & & \text{B} & & \\ & b_1 & b_2 & b_3 & b_4 \\ \hline A=a_2 & 0.5 & 0.3 & 0.1 & 0.1 \\ \hline \end{array} \quad \begin{array}{|cc|cccc|} \hline & & & \text{B} & & \\ & & b_1 & b_2 & b_3 & b_4 \\ \hline \text{D} & d_1 & 100 & 90 & 20 & 0 \\ & d_2 & 0 & 10 & 30 & 50 \\ \hline \end{array} \right)$$

$$= \begin{array}{|c|c|} \hline D=d_1 & \mathbf{79} \\ D=d_2 & 11 \\ \hline \end{array}$$

The corresponding optimal decision function is

$$\delta(a_1) = d_1 \qquad \delta(a_2) = d_1$$

Sub-problem 3

Since the optimal decision for D is the same for both states of A, the VoI is 0. This is also reflected in the EU calculations.

We start by marginalizing out A:

$$P(B) = \sum_A P(B|A)P(A) = \sum_A \left(\begin{array}{|cc|} \hline \multicolumn{2}{|c|}{\text{A}} \\ a_1 & a_2 \\ \hline 0.2 & 0.8 \\ \hline \end{array} \quad \begin{array}{|cc|cccc|} \hline & & & \text{B} & & \\ & & b_1 & b_2 & b_3 & b_4 \\ \hline \text{A} & a_1 & 0.1 & 0.2 & 0.25 & 0.45 \\ & a_2 & 0.5 & 0.3 & 0.1 & 0.1 \\ \hline \end{array} \right)$$

$$= \sum_A \left(\begin{array}{|cc|cccc|} \hline & & & \text{B} & & \\ & & b_1 & b_2 & b_3 & b_4 \\ \hline \text{A} & a_1 & 0.02 & 0.04 & 0.05 & 0.09 \\ & a_2 & 0.4 & 0.24 & 0.08 & 0.08 \\ \hline \end{array} \right)$$

$$= (0.42, 0.28, 0.13, 0.17)$$

By combining $P(B)$ with the utility table we get:

$$
\begin{aligned}
EU(D) &= \sum_B P(B)U(B,D) \\
&= \sum_B \left((0.42, 0.28, 0.13, 0.17) \begin{array}{|cc|cccc|}
\hline
 & & & \mathrm{B} & & \\
 & & b_1 & b_2 & b_3 & b_4 \\
\hline
\mathrm{D} & d_1 & 100 & 90 & 20 & 0 \\
 & d_2 & 0 & 10 & 30 & 50 \\
\hline
\end{array} \right) \\
&= \sum_B \left(\begin{array}{|cc|cccc|}
\hline
 & & & \mathrm{B} & & \\
 & & b_1 & b_2 & b_3 & b_4 \\
\hline
\mathrm{D} & d_1 & 42 & 25.2 & 2.6 & 0 \\
 & d_2 & 0 & 2.8 & 3.9 & 8.5 \\
\hline
\end{array} \right) \\
&= (\mathbf{69.8}, 15.2)
\end{aligned}
$$

Question 4 - 25 points

An online newspaper company wants to make a targeted advertisement campaign to their online subscribers, trying to make them sign up for a full subscription. From a previous campaign the newspaper has collected data about the reading habits of their online subscribers and whether they reacted positively to the advertisement campaign (*Campaign* with states yes/no). The reading habits are characterized based on whether the subscribers read the sports section (*Sports* with states yes/no), the section on international affairs (*Int.* with states yes/no), and the entertainment section (*Ent.* with states yes/no). The data is listed in the table below covering 12 subscribers.

Ent.	*Sports*	*Int.*	*Campaign*
no	yes	no	yes
yes	no	yes	no
yes	no	no	no
no	no	yes	yes
yes	yes	yes	yes
no	yes	no	yes
no	no	no	no
no	no	yes	no
no	no	no	no
no	yes	no	yes
yes	no	no	yes
yes	yes	no	no

The newspaper company now wants to build a classifier for predicting customers' reactions to an advertisement campaign.

1. Show the structure of a naive Bayes classifier covering the variables in the domain/table above.
2. Specify the conditional probability distributions (on the form $P(X|Y)$) needed to obtain a full Bayesian network.
3. Based on the data in the table above, estimate the conditional probability tables for your classifier.
4. Calculate the probability $P(Campaign|Ent. = y, Sports = y, Int. = n)$. How would you use this probability to decide on whether to advertise?
5. Calculate the probability $P(Campaign|Sports)$
6. Describe at least one other machine learning task that one might perform on the data set above besides doing classification wrt. *Campaign*.

Solution:

The probability tables are as follows: $P(C) = (6/12, 6/12)$

$$P(Ent|Campaign) = \begin{array}{c|cc} & \text{C=y} & \text{C=n} \\ \hline \text{T=y} & 2/6 & 3/6 \\ \text{T= n} & 4/6 & 3/6 \end{array}$$

$$P(Sports|Campaign) = \begin{array}{c|cc} & \text{C=y} & \text{C=n} \\ \hline \text{S=y} & 4/6 & 1/6 \\ \text{S= n} & 2/6 & 5/6 \end{array}$$

$$P(Int|Campaign) = \begin{array}{c|cc} & \text{C=y} & \text{C=n} \\ \hline \text{P=y} & 2/6 & 2/6 \\ \text{P= n} & 4/6 & 4/6 \end{array}$$

The conditional probability $P(Campaign|Tech = y, sports = y, politics = n)$ can be calculated as:

$$P(Campaign|Tech = y, sports = n, politics = n)$$
$$= \frac{6/12 \cdot 2/6 \cdot 4/6 \cdot 4/6}{6/12 \cdot 2/6 \cdot 2/6 \cdot 4/6 + 3/6 \cdot 1/6 \cdot 4/6 \cdot 6/12} = \approx (0.73, 0.27).$$

Calculating the probability $P(Campaign|Sports)$

$$\begin{aligned} P(Campaign|Sports) &= \frac{P(Campaign, Sports)}{P(Sports)} \\ &= \frac{P(Sports|Campaign)P(Campaign)}{P(Sports)} \\ &= \begin{array}{c|cc} & \text{C=y} & \text{C=n} \\ \hline \text{S=y} & 4/5 & 1/5 \\ \text{S= n} & 2/7 & 5/7 \end{array} \end{aligned}$$

Question 5 - 25 points

A travel agency has made a database (based on questionnaires 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 the hotel is to the nearest town (*Town* with states $< 1km$ and $\geq 1km$)?, and c) how close the hotel is to the beach (*Beach* with states $< 500m$, $\geq 500m;< 1km$, and $\geq 1km$)?

	Attributes			Target
	Beach	*Town*	*Pool*	*Satisfactory*
1	$\geq 1km$	$< 1km$	*yes*	*yes*
2	$\geq 500m;< 1km$	$< 1km$	*yes*	*yes*
3	$< 500m$	$\geq 1km$	*yes*	*yes*
4	$\geq 500m;< 1km$	$\geq 1km$	*yes*	*no*
5	$\geq 1km$	$< 1km$	*no*	*no*
6	$\geq 1km$	$\geq 1km$	*no*	*yes*
7	$< 500m$	$< 1km$	*no*	*yes*

a) Calculate the entropy of the attribute *Beach*.[1]

b) Show the decision/classification tree that would be learned by the decision tree algorithm assuming that it is given the training examples above and uses information gain for selecting the attributes.

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

d) Using the constructed decision tree, what class value will you assign to a customer described by [*Beach* = ($\geq 500m;< 1km$), *Town* = $<$ 1km, *Pool* = *no*]?

Solution:

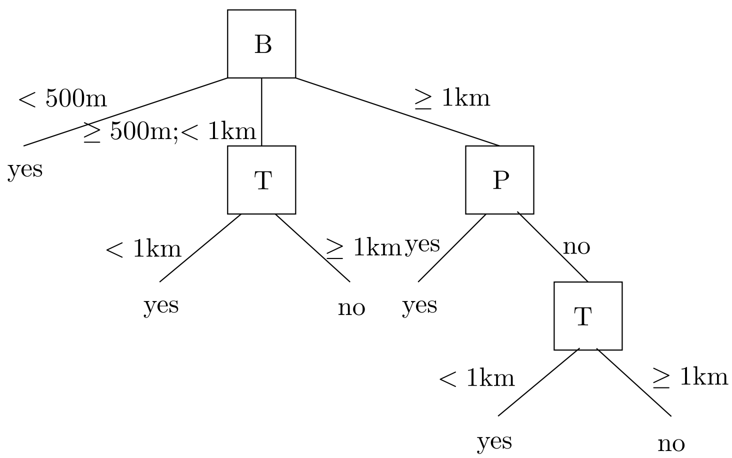
Answer (a)

$$Ent(Beach) = -\frac{2}{7} \cdot \log\left(\frac{2}{7}\right) - \frac{2}{7} \cdot \log\left(\frac{2}{7}\right) - \frac{3}{7} \cdot \log\left(\frac{3}{7}\right) \approx 1.56$$

Answer (b)

There are two possible solutions. The one given below and an equivalent tree where the two nodes in the right-most branch are swapped.

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



Answer (c)

First iteration:

$$
\begin{aligned}
Ent(Target) &= 0.863 \\
Ent(1,2) &= 0.918 \\
Ent(3,1) &= 0.811 \\
E\text{-}ent(Beach) &= \frac{2}{7} \cdot 0 + \frac{2}{7} \cdot 1 + \frac{3}{7} \cdot Ent(1,2) \approx 0.68 \\
E\text{-}ent(Town) &= \frac{4}{7} \cdot Ent(3,1) + \frac{3}{7} \cdot \mathit{2,1} \approx 0.857 \\
E\text{-}ent(Pool) &= \frac{4}{7} \cdot Ent(3,1) + \frac{3}{7} \cdot \mathit{2,1} \approx 0.857 \\
VOI(Beach) &= 0.863 - 0.68 = 0.183 \\
VOI(Town) &= 0.863 - 0.857 = 0.006 \\
VOI(Pool) &= 0.863 - 0.857 = 0.006
\end{aligned}
$$

Second iteration:

For B=($< 500m$) we are finished.

For B=($\geq$ 500m;$<$ 1km) we have

$$
\begin{aligned}
Ent(\mathit{Target} \,|\, B = (\geq 500m; < 1km)) &= 1 \\
E\text{-}ent(\mathit{Town} \,|\, B = (\geq 500m; < 1km)) &= \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = 0 \\
E\text{-}ent(\mathit{Pool} \,|\, B = (\geq 500m; < 1km)) &= \frac{2}{2} \cdot 1 = 1 \\
VOI(\mathit{Town} \,|\, B = (\geq 500m; < 1km)) &= 1 \\
VOI(\mathit{Pool} \,|\, B = (\geq 500m; < 1km)) &= 0
\end{aligned}
$$

and so we select *Town* and finishes.

For B=($\geq 1km$) we have.

$$
\begin{aligned}
Ent(\mathit{Target} \,|\, B = (\geq 1km)) &= 0.918 \\
E\text{-}ent(\mathit{Town} \,|\, B = (\geq 1km)) &= \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{2}{3} \\
E\text{-}ent(\mathit{Pool} \,|\, B = (\geq 1km)) &= \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3} \\
VOI(\mathit{Town} \,|\, B = (\geq 1km)) &= 0.918 - 2/3 = 0.2513 \\
VOI(\mathit{Pool} \,|\, B = (\geq 1km)) &= 0.918 - 2/3 = 0.2513
\end{aligned}
$$

Here we arbitrarily pick *Pool* as the next feature in the right branch of the tree.

Third iteration

This iteration only involves the branch defined by [B=($\geq 1km$), P=*no*], where the only variables left, *Town*, has VOI=1.