Name:

Birth Date:

Matriculation Number:

Field of Study:

Final Exam
Künstliche Intelligenz 2

Aug 01., 2017

Sie können 14 Punkte erreichen, wenn Sie alle Probleme lösen. Die volle Punktzahl ist allerdings schon bei 125 erreicht; d.h. -111 sind Bonuspunkte.

To be used for grading, do not write here			
prob.	1.1	1.2	Sum
total	8	6	14
reached			

Exam Grade: Bonus Points: Final Grade:

Organizational Information

Please read the following directions carefully and acknowledge them with your signature.

- Bitte legen Sie Ihren Studentenausweis und einen Lichtbildausweis zur Personenkontrolle bereit!
- Die angegebene Punkteverteilung gilt unter Vorbehalt.
- Es sind keine Hilfsmittel erlaubt.
- Die Lösung einer Aufgabe muss auf den vorgesehenen freien Raum auf dem Aufgabenblatt geschrieben werden; die Rückseite des Blatts kann mitverwendet werden. Wenn der Platz nicht ausreicht, können bei der Aufsicht zusätzliche Blätter angefordert werden.
- Wenn Sie die Prüfung aus gesundheitlichen Gründen abbrechen müssen, so muss Ihre Prüfungsunfähigkeit durch eine Untersuchung in der Universitätsklinik nachgewiesen werden. Melden Sie sich in jedem Fall bei der Aufsicht und lassen Sie sich das entsprechende Formular aushändigen.
- Die Bearbeitungszeit beträgt 90 Minuten.
- Überprüfen Sie Ihr Exemplar der Klausur auf Vollständigkeit (18 Seiten inklusive Deckblatt und Hinweise) und einwandfreies Druckbild! Vergessen Sie nicht, auf dem Deckblatt die Angaben zur Person einzutragen!

Erklärung

Durch meine Unterschrift bestätige ich den Empfang der vollständigen Klausurunterlagen und die Kenntnisnahme der obigen Informationen.

Erlangen, Aug 01., 2017 ....

(Unterschrift)

1 Bayesian Reasoning

Problem 1.1 (Bayesian Rules)

Name the four basic rules in Bayesian inference and explain each with a short sentence and formula. 8pt 4min

Solution: Bayes rule, Normalization, Marginalization, Chain rule

Problem 1.2 (Conditional Independence)

Define *conditional independence*. 6pt

Solution: Two events A, B are conditionally independent given C, if $P(A \wedge B|C) = P(A|C)P(B|C)$. 3min

Problem 1.3 (Nuclear Test)

Assume it is your responsibility to monitor the Nuclear Test Ban treaty. You receive data from two different stations (seismometers), S_1 and S_2. Each S_i is modeled as a Boolean variable where "true" stands for "I detected a Nuclear test" and "false" stands for "I did not detect a Nuclear test". The seismometers are not fully reliable, however; they may not detect a Nuclear test even though there was one, and they may mistake an earthquake for a Nuclear test. We model this situation with two additional Boolean variables: N for Nuclear test, and E for Earthquake. 14pt 7min

Use the algorithm from the lecture to construct a Bayesian network for these 4 variables. More precisely:

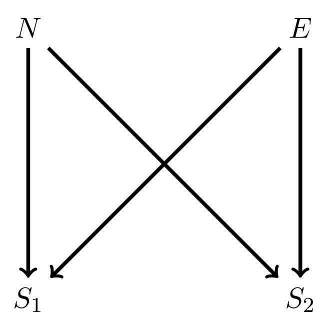
1. State the exact formal condition for when the algorithm inserts an edge between two nodes. 6 pt
2. Execute the algorithm for the variable order $X_1 = N$, $X_2 = E$, $X_3 = S_1$, $X_4 = S_2$. 8 pt

Justify your decisions.

Solution:

(a) $P(X_i|X_{i-1}, ..., X_1) = P(X_i|\texttt{Parents}(X_i))$

(b) With this variable order, we get the following network:



$X_2 = E$ does not need $X_1 = N$ as a parent because Earthquakes are independent from Nuclear tests. $x_3 = S_1$ needs both $X_1 = N$ and $X_2 = E$ as parents because each of these may influence the measurement; same for $X_4 = S_2$, i.e., here we also need the parents $X_1 = N$ and $X_2 = E$. However, given the values of N and E, the measurements of $X_3 = S_1$ and $x_4 = S_2$ are independent. So $X_4 = S_2$ does not require the parent $X_3 = S_1$.

Problem 1.4 (Causal and Diagnostic)

State the difference between *causal* and *diagnostic* edges in (e.g.) a Bayesian network. Use no more than four sentences. 8pt 4min

Solution:

2 Decision Theory

Problem 2.1 (Decision Preferences)

8pt

4min

1. Name and state three of the axioms for preferences (i.e. $\prec$).
2. How are preferences related to utility functions?

Solution:

1. Orderability $(A \prec B) \vee (B \prec A) \vee (A \sim B)$

   Transitivity $(A \prec B) \wedge (B \prec C) \Rightarrow (A \prec C)$

   Continuity $A \prec B \prec C \Rightarrow \exists p([p, A; (1-p), C] \sim B)$

   Substitutability $(A \sim B) \Rightarrow ([p, A; (1-p), C] \sim [p, B; (1-p), C])$

   Monotonicity $(A \prec B) \Rightarrow (p \geq q) \Leftrightarrow ([p, A; (1-p), B] \preceq [q, A; (1-q), B])$
2. Ramsey's theorem states that given a set of preferences that obey the constraints above, there is a utility function U with $(U(A) \geq U(B)) \Leftrightarrow (A \preceq B)$ and.

Problem 2.2 (Expected Utility)

What is the formal(!) definition of *expected utility*? Explain every variable in the defining equation. 8pt 4min

---

Solution: The expected utility EU is defined as $EU(a|e) = \sum_{s'} P(R(a) = s'|a, e) \cdot U(s')$, where

1. a is the action for which we want to find out the expected utility, given the evidene e.
2. $U(s')$ is the utility of a state s'.
3. $R(a)$ is the result of the action a.

Problem 2.3 (Textbook Decisions)

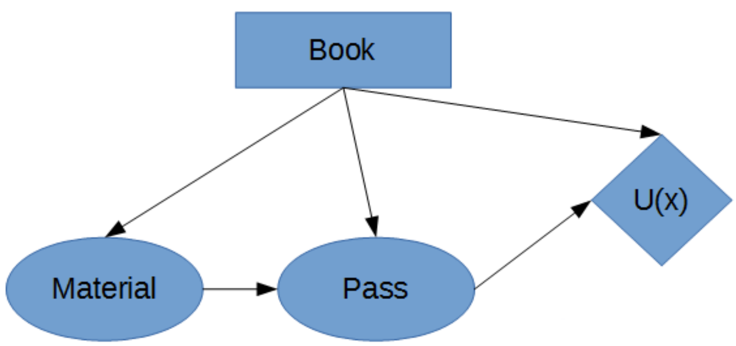
Abby has to decide whether to buy Russell&Norveig for 100\$. There are three boolean variables involved in this decision: B indicating whether Abby buys the book, M indicating whether Abby knows the material in the book perfectly anyway and P indicating that Abby passes the course. Additionally, we use a utility node U. 20pt 10min

Abbys utility function is additive, so $U(B) = -100$. Furthermore, she evaluates passing the course with a utility of $U(P) = 2000$. The course has an open book final exam, so B and P are not independent given M. Assume the conditional probabilities $P(P|B,M)$, $P(P|B,\neg M)$, $P(P|\neg B,M)$, $P(P|\neg B,\neg M)$, $P(M|B)$, $P(M|\neg B)$ are given.

1. Draw the decision network for this problem. 10 pt
2. Explain how to compute the utility of buying the book. 10 pt

---

Solution:



P	B	Utility
F	F	0
F	T	-100
T	F	2000
T	T	1900

B	P(B)
F	0.7
T	0.9

B	M	P(P)
F	F	0.3
F	T	0.8
T	F	0.5
T	T	0.9

P(B=T,P=T,M=T) = P(P=T | M=T,B=T) $\cdot$ P(B=T) $\cdot$ P(M=T | B=T) = $0.9 \cdot 1 \cdot 0.9$
P(B=T,P=T,M=F) = $0.5 \cdot 1 \cdot 0.1$
P(B=T,P=F,M=T) = $0.1 \cdot 1 \cdot 0.9$
P(B=T,P=F,M=F) = $0.5 \cdot 1 \cdot 0.1$

P(B=T,P=T) = $0.9 \cdot 1 \cdot 0.9 + 0.5 \cdot 1 \cdot 0.1 = 0.86$
P(B=T,P=F) = $0.1 \cdot 1 \cdot 0.9 + 0.5 \cdot 1 \cdot 0.1 = 0.14$
$\rightarrow$ U(B=T) = $0.86 \cdot 1900 + 0.14 \cdot -100 = 1620$

P(B=F,P=T,M=T) = P(P=T | M=T,B=F) $\cdot$ P(B=F)
$\cdot$ P(M=T | B=F) = $0.8 \cdot 1 \cdot 0.7$
P(B=F,P=T,M=F) = $0.3 \cdot 1 \cdot 0.3$

$P(B{=}F,P{=}F,M{=}T) = 0.2 \cdot 1 \cdot 0.7$
$P(B{=}F,P{=}F,M{=}F) = 0.7 \cdot 1 \cdot 0.3$

$\rightarrow P(B{=}F,P{=}T) = 0.8 \cdot 1 \cdot 0.7 + 0.3 \cdot 1 \cdot 0.3 = 0.65$
$P(B{=}F,P{=}F) = 0.2 \cdot 1 \cdot 0.7 + 0.7 \cdot 1 \cdot 0.3 = 0.35$
$\rightarrow U(B{=}F) = 0.65 \cdot 2000 + 0.35 \cdot 0 = 1300$

---

3 Markov Models

Problem 3.1 (Prediction, Filtering, Smoothing)

Explain the goals of *prediction, filtering* and *smoothing* in markov models and briefly explain algorithms to compute them. 8pt 4min

Solution:

Prediction $P(X_{t+k}|e_{1:t})$

Filtering $P(X_t|e_{1:t})$

Smoothing $P(X_k|e_{1:t})$ for $0 \leq k < t$

Problem 3.2 (Markov Mood Detection)
On any given day d, your roommate Moody is in one of two states – either he is happy (H_d) or he is in a bad mood (B_d). Usually when he's in a bad mood, it's because he had a fight with his boyfriend and those tend to go on for a couple of days, so $P(B_{d+1}|B_d) = 0.7$, but aside from that he's a cheery guy, so ($P(H_{d+1}|H_d) = 0.85$). 16pt 8min

Of course you try to avoid talking to people, but you can hear his music blasting all day which tends to shift depending on his mood. On a good day he usually listens to Jazz (i.e. $P(J_d|H_d) = 0.7$), on a bad day he slightly prefers Death Metal ($P(DM_d|B_d) = 0.6$). He has a limited taste in music, so it's always one of the two.

You accidentally talked to Moody yesterday ($d = 0$) and know that he was in a good mood then, but today he listens to death metal.

1. How likely is it that he's in a bad mood today? 8 pt
2. Assume he's listening to death metal for n days straight. Explain how to compute the probability that he is in a bad mood on day $n + 1$. 8 pt

---

Solution: We have $P(H_0) = 1$ and

$$\langle P(H_d), P(B_d)\rangle = \langle P(H_d|H_{d-1}) + P(H_d|B_{d-1}), P(B_d|H_{d-1}) + P(B_d|B_{d-1})\rangle$$

which allows us to update using the information DM_d:

$$\langle P(H_d|DM_d), P(B_d|DM_d)\rangle = \alpha\langle P(DM_d|H_d)P(H_d), P(DM_d|B_d)P(B_d)\rangle$$

Problem 3.3 (Bellman Equation)

State the Bellman Equation and explain 1) every symbol in the equation and 2) what the equation is used for and how. 8pt

4min

Solution:

$$U(s) = R(s) + \gamma \cdot \max_a \left(\sum_{s'} U(s') \cdot T(s,a,s') \right)$$

4 Learning

Problem 4.1 (Gradient Descent)

Explain a gradient descent algorithm. 8pt

Solution: 4min

Problem 4.2 (Information Entropy)

Explain and define *information entropy*. 6pt

Solution: 3min

$$I(\langle P_1, ..., P_n\rangle) = \sum_{i=1}^{n} -P_i \log_2(P_i)$$

Problem 4.3 (Home Decisions)

Eight people go sunbathing. Some of them got a sunburn, others didn't: 20pt

10min

Name	Hair	Height	Weight	Lotion	Result
Sarah	Blonde	Average	Light	No	Sunburned
Dana	Blonde	Tall	Average	Yes	None
Alex	Brown	Short	Average	Yes	None
Annie	Blonde	Short	Average	No	Sunburned
Julie	Blonde	Average	Light	No	None
Pete	Brown	Tall	Heavy	No	None
John	Brown	Average	Heavy	No	None
Ruth	Blonde	Average	Light	No	None

Apply the decision tree learning algorithm on this table to predict whether people will get sunburned based on the attributes provided.

Solution:

$$E_0 := I(\langle\frac{2}{8},\frac{6}{8}\rangle) = -\frac{2}{8}\log_2(\frac{2}{8}) - \frac{6}{8}\log_2(\frac{6}{8}) \approx 0.81$$

$$\texttt{Gain(Hair)} = E_0 - \underbrace{\frac{5}{8}I(\langle\frac{2}{5},\frac{3}{5}\rangle)}_{\texttt{Blonde}} - \underbrace{\frac{3}{8}I(\langle 0,1\rangle)}_{\texttt{Brown}} \approx 0.20$$

$$\texttt{Gain(Height)} = E_0 - \underbrace{\frac{4}{8}I(\langle\frac{1}{4},\frac{3}{4}\rangle)}_{\texttt{Average}} - \underbrace{\frac{2}{8}I(\langle 0,1\rangle)}_{\texttt{Tall}} - \underbrace{\frac{2}{8}I(\langle\frac{1}{2},\frac{1}{2}\rangle)}_{\texttt{Short}} \approx 0.16$$

$$\texttt{Gain(Weight)} = E_0 - \underbrace{\frac{3}{8}I(\langle\frac{1}{3},\frac{2}{3}\rangle)}_{\texttt{Average}} - \underbrace{\frac{3}{8}I(\langle\frac{1}{3},\frac{2}{3}\rangle)}_{\texttt{Light}} - \underbrace{\frac{2}{8}I(\langle 0,1\rangle)}_{\texttt{Heavy}} \approx 0.12$$

$$\texttt{Gain(Lotion)} = E_0 - \underbrace{\frac{2}{8}I(\langle 0,1\rangle)}_{\texttt{Yes}} - \underbrace{\frac{6}{8}I(\langle\frac{2}{6},\frac{4}{6}\rangle)}_{\texttt{No}} \approx 0.12$$

`Hair` has the highest information gain, so we split here. All table entries with `Brown` have result `None`, so we continue with `Hair` = `Blonde`:

$$E_1 := I(\langle\frac{2}{5},\frac{3}{5}\rangle) \approx 0.97$$

$$\texttt{Gain}(\texttt{Height}) = E_1 - \underbrace{\frac{3}{5}I(\langle\frac{1}{3},\frac{2}{3}\rangle)}_{\texttt{Average}} - \underbrace{\frac{1}{5}I(\langle 0,1\rangle)}_{\texttt{Tall}} - \underbrace{\frac{1}{5}I(\langle 1,0\rangle)}_{\texttt{Short}} \approx 0.42$$

$$\texttt{Gain}(\texttt{Weight}) = E_1 - \underbrace{\frac{2}{5}I(\langle\frac{1}{2},\frac{1}{2}\rangle)}_{\texttt{Average}} - \underbrace{\frac{3}{5}I(\langle\frac{1}{3},\frac{2}{3}\rangle)}_{\texttt{Light}} - \underbrace{0}_{\texttt{Heavy}} \approx 0.02$$

$$\texttt{Gain}(\texttt{Lotion}) = E_1 - \underbrace{\frac{1}{5}I(\langle 0,1\rangle)}_{\texttt{Yes}} - \underbrace{\frac{4}{5}I(\langle\frac{2}{4},\frac{2}{4}\rangle)}_{\texttt{No}} \approx 0.17$$

`Height` has the highest information gain, so we proceed here. All short blondes are sunburned, all tall blondes are not, hence we only need consider `Average`...