Last Name: First Name:

Matriculation Number:

Exam
Artificial Intelligence 2

Feb 16, 2023

	To be used for grading, do not write here										
prob.	1.1	1.2	2.1	2.2	3.1	3.2	4.1	4.2	5.1	Sum	grade
total	7	9	8	13	13	10	10	8	7	85	
reached											

Organizational Information

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

1. Please place your student ID card and a photo ID on the table for checking.
2. You can reach 85 points if you fully solve all problems. You will only need 80 points for a perfect score, i.e. 5 points are bonus points.
3. No resources or tools are allowed except for a pen.
4. You have 90 min (sharp) for the exam.
5. Write the solutions directly on the sheets, no other paper will be graded.
6. If you have to abort the exam for health reasons, your inability to sit the exam must be certified by an examination at the University Hospital. Please notify the exam proctors and have them give you the respective form.
7. Please make sure that your copy of the exam is complete (14 pages excluding cover sheet and organizational information pages) and has a clear print. **Do not forget to add your personal information on the cover sheet and to sign this declaration.**

Declaration: With my signature I certify having received the full exam document and having read the organizational information above.

Erlangen, Feb 16, 2023 . . . .
(signature)

Organisatorisches

Bitte lesen die folgenden Anweisungen genau und bestätigen Sie diese mit Ihrer Unterschrift.

1. Bitte legen Sie Ihren Studierendenausweis und einen Lichtbildausweis zur Personenkontrolle bereit!

2. Sie können 85 Punkte erreichen, wenn Sie alle Aufgaben vollständig lösen. Allerdings zählen 80 Punkte bereits als volle Punktzahl, d.h. 5 Punkte sind Bonuspunkte.
3. Es sind keine Hilfsmittel erlaubt außer einem Stift.
4. Die Bearbeitungszeit beträgt genau 90 min.
5. Schreiben Sie die Lösungen direkt auf die ausgeteilten Aufgabenblätter. Andere Blätter werden nicht bewertet.
6. 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.
7. Überprüfen Sie Ihr Exemplar der Klausur auf Vollständigkeit (14 Seiten exklusive Deckblatt und Hinweise) und einwandfreies Druckbild! **Vergessen Sie nicht, auf dem Deckblatt die Angaben zur Person einzutragen und diese Erklärung zu unterschreiben!**

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

Erlangen, Feb 16, 2023 ....
(Unterschrift)

Please consider the following guidelines to avoid losing points:

- If you continue an answer on another page, clearly give the problem number on the new page and a page reference on the old page.
- If you do not want something to be graded, clearly cross it out. Adding a wrong statement to a correct solution may lead to deductions.
- The instructions “Give X”, “List X” or similar mean that only X is needed. If you additionally justify your answer, we will try to give you partial credit for a wrong answer (but there is no guarantee that we will).

- The instruction "Assume X" means that X is information that you may use in your answer.
- The instruction "Model X as a Y" means that you have to describe X formally and exactly as an instance of Y using the definition of Y from the lecture.
- If you are uncertain how long or complex an answer should be, use the number of points as an indication: 1 point roughly corresponds to 1 minute.
- In all calculation questions, you have to simplify as much as reasonably possible without a calculator. For example, $\log 2$ or 3^7 should not be calculated, but $0.4 \cdot 0.3 \cdot 0.5 = 0.06$ should be.

1 Probabilities

Problem 1.1 (Python) 7 pt

Consider the Python *program* below.

1. Which operation does the function `foo` compute?

2. Assume random variables X with domain $\{0, \ldots, m-1\}$ and Y with domain $\{0, \ldots, n-1\}$.
   Assume the Python object C holds their joint probability distribution $P(X, Y)$, i.e., $C[i][j] = P(X = i, Y = j)$.
   Complete the definition of E in the program below in such a way that it holds the probability distribution $P(Y \mid X = 0)$, i.e., $E[j] = P(Y = j \mid X = 0)$.

```
# input: a list of numbers in the interval [0;1]
def foo(a):
    l = len(a)
    s = 0
    for i in range(l):
        s += a[i]
    res = []
    for i in range(l):
        res.append(a[i]/s)
    return res


E =
```

Hint: This can be done with relatively little code.

Problem 1.2 (Calculations) 9 pt

Assume random variables X, Y both with domain $\{0, 1, 2\}$, whose joint distribution $P(X, Y)$ is given by

x	y	$P(X = x, Y = y)$
0	0	a
0	1	b
0	2	c
1	0	d
1	1	e
1	2	f
2	0	g
2	1	h
2	2	i

1. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X = 0)$.

2. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X = 0 \mid Y = 1)$.

3. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X \neq 0 \mid Y \neq 1)$.

4. The table above is redundant. How can it be stored using less space?

5. Now assume X and Y are stochastically independent. How can the information in the table be stored using the least space?

2 Bayesian Reasoning

When working with an upper case Boolean random variable X, you may abbreviate the event $X = true$ by the corresponding lower-case letter x. If you do that, make sure the distinction between upper and lower case letters is clear in your writing.

Problem 2.1 (Bayes' Rule) 8 pt

Assume you are trying to predict whether a particular topic comes up in an exam. You have collected the following data:

- 30% of all topics come up in the exam.
- 40% of all topics come up in the tutorials.
- If a topic comes up in an exam, it was covered by an assignment 60% of the time.
- If a topic comes up in an exam, it came up in a tutorial 80% of the time.

You model this situation using 3 Boolean random variables X (comes up in exam), S (covered by assignments), and T (came up in a tutorial).

1. By filling in the gaps below, state for each number in the text above, which probability it describes.

   1. $P($ ☐ $) = 0.3$
   2. $P($ ☐ $) = 0.4$
   3. $P($ ☐ $|$ ☐ $) = 0.6$
   4. $P($ ☐ $|$ ☐ $) = 0.8$

2. Assume the topic you are interested in **did not** come up in the exam. Argue if and how we can obtain the probability that it was covered by a tutorial.

3. The topic you are interested in was covered by a tutorial. Using Bayes' rule, calculate the exact probability that it will come up in the exam.

Problem 2.2 (Bayesian Networks) 13 pt

Consider the following situation:

- Covid and influenza can cause fever.
- Fever causes stress.
- Tests can detect Covid. But a false-positive Covid-test causes stress as well.
- There are no other causal relationships.

You want to model this situation using Boolean random variables C (Covid infection), I (influenza infection), F (fever), S (stress), and T (positive Covid test).

1. Give a good variable ordering for forming a Bayesian network for this situation.

2. Give the resulting network.

3. You have a fever and have tested positive for Covid. Now you want to determine if you have influenza. What are the query, evidence, and hidden variables?

4. Assume your network is $I \leftarrow C \rightarrow F \leftarrow S \rightarrow T$ (which *may or may not* be a correct solution to the above question). Which probabilities are stored in the conditional probability table of node F?

5. Again using the network $I \leftarrow C \rightarrow F \leftarrow S \rightarrow T$, give the formula for

$$P(C = true, I = true, F = true, T = true)$$

in terms of the entries of the conditional probability table of that network. You may abbreviate the event $X = true$ by the lower-case name of the random variable X.

3 Markovian Reasoning

Problem 3.1 (Hidden Markov Models) 13 pt

Consider the following situation:

- You make annual observations about the rainfall at a certain location. Each year the rainfall is high (r_1), medium (r_2), or low (r_3).
- You know this is caused by an atmospheric condition, which is either strong (c_1) or weak (c_2).
- You have previously obtained the following information:
  - when the condition is strong, the rainfall is high 20% and medium 30% of the time,
  - when the condition is weak, the rainfall is high 35% and medium 15% of the time,
  - when the condition is strong, it stays strong next year 70% of the time,
  - when the condition is weak, it becomes strong next year 40% of the time,

You want to model this situation as a hidden Markov model with two families of random variables R_a (rainfall) and C_a (condition), each indexed by year number a.

1. Give the state and evidence variables and their domains.

2. How can you tell that the model is first-order here?

3. Complete the following sentences:
   4. The transition model T is given by the matrix

   $$T = \begin{pmatrix} & & \\ & & \end{pmatrix} \qquad \text{where} \qquad T_{ij} = P(C_{a+1} = c_j \mid C_a = c_i)$$

   5. The sensor model S is given by the matrix

   $$S = \begin{pmatrix} & & \\ & & \end{pmatrix} \qquad \text{where} \qquad S_{ij} = P(R_a = r_j \mid C_a = c_i)$$

6. The atmospheric condition has been strong last year, and the rainfall is low this year. You want to use filtering to obtain the probability distribution of this year's condition.
   We compute $f_{1:1}$ by applying the filtering equation once.
   7. Give the recursive filtering equation for $f_{1:a+1}$.
   8. Give the initial value $f_{1:0}$ to use in this case.

9. Give the diagonal sensor matrix O_1 to use in this case.
10. Calculate the resulting distribution; Fully compute all values including the normalization. (This does not require approximations or a calculator.)

Problem 3.2 (Utility and Decision Processes) 10 pt

Consider an agent moving along a circular arrangement of 8 locations as indicated below.

$$\begin{array}{ccccccc} 0 & \rightarrow & 1 & \rightarrow & 2 & \rightarrow & 3 \\ \uparrow & & & & & & \downarrow \\ 7 & \leftarrow & 6 & \leftarrow & 5 & \leftarrow & 4 \end{array}$$

The agent's movement is as follows:

- It can move forward (in arrow direction) or backward (against arrow direction), and in each case it can move carefully or quickly.
- The careful moves result in moving 1 step in that direction with probability 60% and no move otherwise.
- The quick actions result in moving 1 step in that direction with probability 90% and moving 1 step in the opposite direction otherwise.

The agent's goal is to move to location 0.

1. Choose an appropriate reward function and model this situation as a Markov Decision Process.

2. State the Bellman equation relative to your model.

3. Give an optimal policy π^*.

Hint: This requires deciding whether careful or quick actions lead to the goal faster.

---

4. Now assume we use a POMDP because the agent is unable to tell what move an action resulted in. Assume we know the agent is initially in location 4. Calculate the belief state after moving backwards carefully twice.

4 Learning

Problem 4.1 (Decision Trees and Lists) 10 pt

Consider an unknown 3-letter word $W = L_1, L_2, L_3 \in \{BAT, BET, CAT, CAR\}$. Each letter L_i is currently covered (i.e., it cannot be seen) and can be uncovered individually.

1. Draw the decision tree for W that arises from uncovering the letters in the order L_1, L_2, L_3. (Do not uncover additional letters if the word can already be identified.)

2. Which choice would the information gain algorithm make first in this case? Justify your answer.

3. Give a decision list in 1-DL for W using only literals of the form $L_i = X$ for characters X.

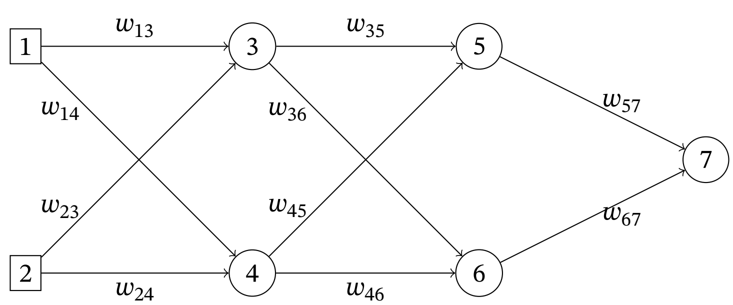
4. Give all minimal sets $A \subseteq \{L_1, L_2, L_3\}$ of letters such that $A \succ W$.

5. How would we have to change the set of possible words so that the determination $\{L_1, L_2\} \succ W$ holds?

Problem 4.2 (Neural Networks) 8 pt

Consider the neural network **without bias** given below where

- units 1, 2 are inputs,
- unit 7 is output,
- weights are given by the labels on the edges, and
- units 3, 4, 5, 6, 7 are perceptron units with activation function $T(x) = 1$ for $x > 0.5$ and $T(x) = 0$ otherwise.



1. There are ☐ hidden layers.
2. Assume $w_{ij} = 0.2$ for all weights and $a_1 = a_2 = 1$. Then the resulting output a_7 is ☐ ?
3. Assume we remove the edge from 5 to 7. Give two reasons why that would be a bad network to use.

4. Give the formula for the activation a_6 of unit 6 in terms of the inputs a_1 and a_2.

5. Assume the inputs a_1, a_2 are in $\{0, 1\}$. Choose weights such that the output a_7 is always 1 (no matter what the inputs are), or argue why that is impossible?

5 Natural Language Processing

Problem 5.1 (Language Models) 7 pt

1. How many different trigrams does a language with n words have?

2. What is a statistical language model?

3. Name two applications of statistical language models.

4. Why is it work-intensive in practice to build a good statistical language model for a natural language?