Last Name: First Name:

Matriculation Number:

Exam
Artificial Intelligence 2

October 10, 2023

Please ignore the QR codes; do not write on them, they are for grading support

	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	5.2	Sum	grade
total	7	8	8	10	11	11	9	9	6	6	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 (12 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, October 10, 2023 ....
(signature)

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.

1 Probabilities

Problem 1.1 (Python)

1. Consider the Python program below. 4 Points

```
# input: two lists of real numbers between 0 and 1
def bar(x, y):
    l = len(x)
    m = len(y)
    res = []
    for i in range(l):
        row = []
        for j in range(m):
          row.append(x[i]*y[j])
        res.append(row)
    return res
```

Assuming the inputs represent probability distributions, which probability-related operation does the function `bar` compute? State any assumptions that need to be made about the inputs.

2. Assume random variables X with domain $\{0, \dots, m-1\}$ and Y with domain $\{0, \dots, n-1\}$. Assume the Python object C holds their joint probability distribution $P(X, Y)$, i.e., $C[i][j] = P(X = i, Y = j)$. 3 Points

Complete the definition of E in the program below in such a way that it holds the probability distribution $P(X)$, i.e., $E[i] = P(X = i)$.

```
# m, n, C are defined as described above

E =
```

Problem 1.2 (Calculations)

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. Give all subsets of the probabilities $\{a, b, c, d, e, f, g, h, i\}$ that sum to 1.

   2 Points

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

   2 Points

3. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X + Y = 2)$.

   2 Points

4. In terms of $a, b, c, d, e, f, g, h, i$, give $P(X + Y = 2 | X > Y)$.

   2 Points

2 Bayesian Reasoning

Problem 2.1 (Bayesian Calculations)
Assume you are trying to relate economic development and your business results. You have collected the following data:

- The economy does well 40% of the time and badly otherwise.
- Your business does well 30% of the time and badly otherwise.
- If your business does well, the economy did well 80% of the time.

You model the problem using two Boolean random variables E (economy does well) and B (business does well). You also abbreviate the events $E = true$ and $B = true$ as e and b.

1. By filling in the gaps below, state for each number in the text above, which probability it describes. 2 Points
   1. $P(\underline{\hspace{10em}}) = 0.4$
   2. $P(\underline{\hspace{10em}}) = 0.3$
   3. $P(\underline{\hspace{10em}}) = 0.8$

2. Using Bayes' Rule, compute the probability that your business does well if the economy does. 2 Points

3. Explain how we can compute all values in the joint distribution of E and B. 4 Points
   (You can omit purely mathematical computations unrelated to probabilities if you mention what they do.)

Problem 2.2 (Bayesian Networks)
Consider the following situation about a chess game:

- The outcome O can be a win for white (w), a win for black (b), or a draw (d).
- The players have experience levels E_w and E_b, whose possible values are fresh (f), experienced (e), and professional (p), and that allow making predictions about the result of a game.
- You have placed a bet on the outcome (without knowing the players), and the outcome will determine if you gain (g) or lose (l) money (M).

You want to model this situation as a Bayesian network.

1. Give the set of random variables and their domains. 2 Points

2. Give a good variable order and draw the resulting Bayesian network. 2 Points

3. Assume your network is $E_w \to O \leftarrow E_b \to M$ (which may or may not be correct). How many entries does the conditional probability table for O have? 2 Points

4. Assume again your network is $E_w \to O \leftarrow E_b \to M$. Give the formula for $P(O|M, E_b)$ in terms of the entries of the probability tables of the network. 2 Points

5. You have already placed the bet. What does that mean for the relationship between O and M? How does that affect the memory needed for the conditional probability tables of the network? 2 Points

3 Markovian Reasoning

Problem 3.1 (Hidden Markov Models)

Consider the following situation:

- We 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).
- We know this causes a groundwater condition, which is either strong (g_1) or weak (g_2).

We have modeled this situation as a stationary and first-order hidden Markov model with two families of random variables R_a (rainfall) and G_a (groundwater), each indexed by year number a.

Transition Model Sensor Model

$$T = \begin{pmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.3 & 0.6 \\ 0 & 0.1 & 0.9 \end{pmatrix} \quad S = \begin{pmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \\ 0.25 & 0.75 \end{pmatrix}$$

1. Give the state and evidence variables and their domains. 2 Points

2. Which probabilities are captured by the entries T_{ij} and S_{ij}? 2 Points

3. The rainfall was high last year and is low this year. Give the probability distribution of this year's groundwater condition. 2 Points

4. What is the purpose of the smoothing algorithm? 2 Points

5. Given evidence $G_1 = e_1, \ldots, G_a = e_a$, the smoothing algorithm can be written in matrix form as $P(R_k|e_{1:a}) = \alpha f_{1:k} b_{k+1:a}$. Give the recursive equations for f and b and explain the values of the matrices O. 3 Points

Problem 3.2 (Decision Processes and Utility)
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 -2, -1, 0, 1, or 2 steps (negative numbers represent backwards movement).
- The double steps result in moving 2 locations with probability 60%, and 1 location in the opposite direction otherwise.
- The single steps result in moving 1 location with probability 90%, and no move otherwise.
- The zero step results in no move.

The agent's goal is to move to location 7.

1. Model this situation as a Markov Decision Process $\langle S, A, P, R\rangle$. Use a reward function that uses a constant reward for non-goal states. 4 Points

2. State the Bellman equation for $\gamma = 0.5$. Then using initial utilities $U(s) = 0$ for all states, compute the value of $U(4)$ after two value iteration steps. 3 Points

3. Give an optimal policy π^*. 2 Points

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. Give the belief state after a double step forward. 2 Points

4 Learning

Problem 4.1 (Decision Trees and Lists)
Consider an unknown word $W \in \{I, you, he, she, it, we, they\}$. You are allowed to ask the following questions about W:

A length of the word, returning from $\{1, 2, 3, 4\}$

B occurrence of the letter h, returning yes/no

C last letter of the word, returning from $\{I, u, e, t, y\}$

D first letter of the word, returning from $\{I, y, h, s, i, w, t\}$

1. Draw the decision tree for W that arises from asking the questions in the order given above. (Do not ask additional questions if the word can already be identified.) 3 Points

2. Which choice would the information gain algorithm make first? Justify your answer. 2 Points

3. Give the size of the smallest decision list (measured as the sum of the numbers of literals in all tests) if tests may use arbitrarily many literals of the form $Question = Answer$. 2 Points

4. Give all minimal sets $Q \subseteq \{A, B, C, D\}$ of questions for which the determination $Q \succ W$ holds? 2 Points

Problem 4.2 (Support Vector Machines)

Consider the following dataset of points $\mathbf{x} = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle$ in $\mathbb{R}^2$ that are classified as either $y = +1$ or $y = -1$:

$\mathbf{x}_1$	$\mathbf{x}_2$	y
2	3	1
−2	−2	−1
4	−3	−1

1. Give the hypothesis space for finding a linear separator. 2 Points

2. Give a linear separator $h(\mathbf{x})$ for the dataset. 3 Points

3. Transform the dataset into a 1-dimensional dataset using the transformation $T(\mathbf{x}) = \mathbf{x}_1^2 + \mathbf{x}_2^2$. 2 Points

4. What does it mean, intuitively, if a linear separator exists for a dataset after this transformation? 2 Points

5 Natural Language Processing

Problem 5.1 (Part-of-Speech Tagging)

1. Briefly explain what part-of-speech tagging means. 2 Points

2. What is the role of the window width when machine-learning part-of-speech tags? 2 Points

3. Explain (in about 2 sentences) the role of word embeddings when learning part-of-speech tags, and the idea behind $tfidf$. 2 Points

Problem 5.2 (Grammars)
Consider the following probabilistic grammar:

$$\begin{array}{lll} S & \rightarrow & NP\ VP[1] \\ NP & \rightarrow & Article\ Noun[0.6] \mid Name[0.4] \\ VP & \rightarrow & Verb[0.5] \mid TransVerb\ NP[0.5] \\ Article & \rightarrow & \text{the}[0.7] \mid \text{a}[0.3] \\ Noun & \rightarrow & \text{stench}[0.2] \mid \text{breeze}[0.3] \mid \text{wumpus}[0.5] \\ Name & \rightarrow & \text{John}[0.3] \mid \text{Mary}[0.7] \\ Verb & \rightarrow & \text{smells}[1] \\ TransVerb & \rightarrow & \text{sees}[0.6] \mid \text{shoots}[0.4] \end{array}$$

1. Using this grammar as an example, explain the difference between grammar rules and lexicon. 2 Points

2. Give the probability of the sentence 2 Points

   John sees the wumpus

   (You have to give the expression with concrete values plugged in, but you do not have to compute the result.)

3. Now assume we do not know the probabilities of the productions, and our corpus is 2 Points

   John sees the wumpus. The wumpus smells. John shoots the wumpus.

   Give the probability that we can learn for $NP \rightarrow Name$ from this corpus.

This page was intentionally left blank for extra space

This page was intentionally left blank for extra space