Last Name: First Name:

Matriculation Number:

Exam
Artificial Intelligence 2

October, 2024

	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	4.3	5.1	Sum	grade
total	8	10	9	10	10	9	9	12	7	7	91	
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 91 points if you fully solve all problems. You will only need 80 points for a perfect score, i.e. 11 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, 2024 ....
(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.
- 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)

1. Consider the Python program below. 4 pt

```
def foo(J,x):
    px = sum(J[x])
    res = []
    for y in range(len(J[x])):
      res.append(J[x][y]/px)
    return res
```

The input J represents the joint probability distribution of random variables X with domain $\{0, \ldots, m-1\}$ and Y with domain $\{0, \ldots, n-1\}$ in such a way that $J[x][y] = P(X = x, Y = y)$. The input x represents an element in the domain of X.
Which probability-related operation does the function `foo` compute?

2. Assume a random variable X with domain $\{0, \ldots, m-1\}$ and distribution D, i.e., $D[x] = P(X = x)$. 4 pt
Assume that $D[x]$ is non-zero for some even value of x.
Complete the definition of p in the program below in such a way that $p(D, x)$ returns the conditional probability $P(X = x \mid X \text{ is even})$, i.e., the probability of $X = x$ under the condition that X takes an even value.

```
def even(x):
  return x % 2 == 0

def p(D,x):
```

Problem 1.2 (Calculations)
Assume random variables X, Y both with domain $\{0,1,2\}$. For some outcomes A, the probabilities are known as follows:

A	$P(A)$
$X = 0 \wedge Y = 0$	a
$X = 0 \wedge Y \neq 0$	b
$X \neq 0 \wedge Y = 0$	c
$X \neq 0 \wedge Y \neq 0$	d

1. Give all subsets of the probabilities $\{a, b, c, d\}$ that must sum to 1. 2 pt

2. In terms of a, b, c, d, give $P(X \neq 0)$ or argue why there is not enough information to compute the value. 2 pt

3. In terms of a, b, c, d, give $P(X > Y)$ or argue why there is not enough information to compute the value. 2 pt

4. In terms of a, b, c, d, give $P(X = 0 \mid Y = 0)$ or argue why there is not enough information to compute the value. 2 pt

5. Now assume we additionally know $e = P(X = Y)$. In terms of a, b, c, d, e, give $P(X + Y = 3)$. 2 pt

2 Bayesian Reasoning

Problem 2.1 (Basic Rules)
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.
- If your business does well, the economy does well 50% of the time.
- If the economy does well, your business does well 70% 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 pt
   1. $P(\qquad\qquad) = 0.4$
   2. $P(\qquad\qquad) = 0.5$
   3. $P(\qquad\qquad) = 0.7$
2. Calculate the probability that your business does well. 2 pt
3. If your business does well, you estimate your earnings to be \$100/month, otherwise \$60/month. Calculate your expected utility (in \$/month) if the economy does well. 2 pt
4. Calculate the probability that your business and the economy are doing badly at the same time. 3 pt

Problem 2.2 (Bayesian Networks)
Consider the following situation about a day out.

- If the sun shines, you are out long, and you did not use sunscreen, you may get sunburned.
- If the sun shines or you meet friends, you may stay out long.

You want to model this situation as a Bayesian network using Boolean random variables S (sunshine), O (staying out long), B (sunburn), U (sunscreen used), and F (meeting friends).

1. Using a good variable ordering, model this as a Bayesian network. 2 pt

2. Assume you meet your friends and use sunscreen, and you want to determine the probability of sunburn. What are the query/evidence/hidden variables? 2 pt

For the remaining questions, assume your network is $F \leftarrow S \rightarrow O \leftarrow B \rightarrow U$ (which may or may not be a correct solution to the previous question).

3. Which probabilities are stored in the conditional probability table of the node O? Which of those could be omitted and computed from the others? 3 pt

4. Give the probability distribution $P(S \mid O = o, U = u)$ (for fixed o, u) in terms of the entries of the probability tables of the network. 3 pt

3 Markovian Reasoning

Problem 3.1 (Hidden Markov Models)
Consider the following situation:

- Each year the groundwater level at your location is high or not.
- Each year your harvests are good or not.
- You want to study the groundwater level by observing the harvests.

You choose to model this situation as a stationary first-order hidden Markov model with a stationary sensor model with Markov property, using two families of year-indexed Boolean random variables.

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

2. Which probabilities do you need for this model? 3 pt

3. What would change about the answer to the previous question if the HMM were not stationary? 2 pt

4. Assume a sequence $e_1, e_2, \ldots$ of Booleans such that e_a gives the quality of the harvest in year $a = 1, 2, \ldots$. Then the filtering algorithm can be written in recursive matrix form as 3 pt

$$\mathbf{f}_{1:a} = \alpha(O_a T^t \mathbf{f}_{1:a-1})$$

(where T^t is the transposed transition matrix).
Which probability distributions are represented by the $\mathbf{f}_{1:a}$, and how do we obtain the values of the O_a?

Problem 3.2 (Decision Processes and Utility)
Consider an agent moving along a rectangular grid of 3×3 locations.
The agent can stand still, or move 1 location up, down, left, or right except when already at the edge.
A movement step succeeds with probability 90%. In the remaining cases, the agent does not move.
The agent is initially in the bottom-left corner location. Its goal is to move to the top-right location.

1. Model this situation as a Markov Decision Process $\langle S, A, T, s_0, R\rangle$. Use a reward function that uses a constant reward for non-goal states. 5 pt

2. Give an optimal policy π^*. 2 pt

3. Now assume the agent is unable to tell whether an action resulted in movement. Explain (in about 2 sentences) how we can still represent this situation as an MDP. 2 pt

4 Learning

Problem 4.1 (Decision Trees and Lists)
Consider an unknown natural number $N \in \{1, \ldots, 10\}$. You are allowed to ask the following attributes/questions about N:

A Is N prime? (possibly answers: yes, no)

B What is N modulo 3? (possible answers: $0, 1, 2$)

C Is $N > 5$? (possibly answers: yes, no)

D Is N a root of $X^2 + X - 7$? (possibly answers: yes, no)

E What is the result of $\sin((N + 0.5)\pi)$? (possibly answers: $-1, 1$)

F What is N modulo 4? (possible answers: $0, 1, 2, 3$)

1. Give the shortest (in terms of tree depth) decision tree for identifying N that uses at most the above questions. 3 pt

2. We can see each question as a random variable. Give the entropies $I(A)$, $I(C)$, and $I(D)$, i.e., the number of bits obtained by asking the question. (Simplify as much as possible without introducing approximate values.) 3 pt

3. Now assume our goal is to learn the function that computes whether any natural number is a square number, given the answers to the questions A-F, using $N = 1$ and $N = 2$ for training data. Formally state this situation as an inductive learning problem $\langle \mathcal{H}, T \rangle$. 3 pt

Problem 4.2 (Classifiers)
Consider a set E of examples of the form (x, y) where $x \in \mathbb{R}^2$ and $y \in \{0, 1\}$.
We want to learn a linear classifier h_w with a hard threshold. For the hard threshold, we use $\mathcal{T}(u)$ that returns 1 if $u > 0$ and 0 otherwise.

1. Give the general form of such a classifier. 2 pt

2. Give a neural network that can be used to represent such a classifier. 3 pt

3. Give the formula for the squared error loss of such a classifier. 2 pt

4. Why is gradient descent not applicable to minimize the loss in this case? 1 pt

5. Instead, we can use the perceptron learning rule to update the weights. What is that rule, and how do we apply it using the examples from E? 2 pt

6. Now assume our classifiers are multi-layer neural networks and that gradient descent is applicable. Explain informally (in about 2 sentences) the basic idea of back-propagation. 2 pt

Problem 4.3 (Inductive Learning)
Consider the family tree given by the following relations:

couple	children
A, B	E, F
C, D	G
F, G	H,I

Assume we already know the predicate $\mathtt{par}(x, y)$ for x being a parent of y. Our goal is to learn the predicate $\text{gp}(x, y)$ for x being a grandparent of y, i.e., to find a formula D such that $\forall x, y.\text{gp}(x, y) \Leftrightarrow D(x, y)$.
We do not know D, but we have the following (counter-)examples for gp:

person-pair	grandparent
A, H	yes
B, I	yes
A, E	no
A, F	no
A, B	no
A, C	no

1. Give the intended formula D_1, i.e., the correct definition of grandparent. 2 pt

2. Give a formula D_2 that is true exactly for the positive examples. 2 pt

3. Explain (in about 3 sentences) the commonalities and pros and cons of learning the formula D as D_1 vs. D_2. 3 pt

5 Natural Language Processing

Problem 5.1 (Information Retrieval)

1. Explain (in about 2 sentences) the connection between tfidf and cosine similarity to relate texts. 2 pt

2. Explain (in about 2 sentences) the difference between information retrieval and information extraction. 2 pt

3. Explain (in about 3 sentences) what kind of function is learned in the continuous bag of words algorithm, as used e.g. in Word2Vec, and how the training proceeds. 3 pt

This page was intentionally left blank for extra space