Last Name: First Name:

Matriculation Number:

Exam
Artificial Intelligence 2

April, 2025

	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	5.2	Sum	grade
total	8	10	7	10	11	11	8	8	6	7	4	90	
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 90 points if you fully solve all problems. You will only need 80 points for a perfect score, i.e. 10 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 (9 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, April, 2025 ....
(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.
- You can always ask for the translation or explanation of a non-technical word.
- 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 may or may not give you partial credit for a wrong answer.
- 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)

Below, 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 $J[x][y] = P(X = x, Y = y)$. All probabilities are represented as floating point numbers.

1. Complete the function below so that it returns the probability distribution of Y. 4 Points

```
def probY(J):
```

2. Consider the Python program below. 3 Points

```
def foo(J):
  numXs = len(J)
  numYs = len(J[0])
  for x in range(numXs):
    for y in range(numYs):
      px = sum([J[x][v] for v in range(numYs)])
      py = sum([J[v][y] for v in range(numXs)])
      if J[x][y] != px*py:
        return False
  return True
```

   Which probability-related operation does the function `foo` compute?

3. When using the function `foo` above, which effect may cause it to return an incorrect result? 1 Points

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

1. It is guaranteed that $a, b, c, d \in [0; 1]$. What other properties about the numbers a, b, c, d are guaranteed to hold? 2 Points
   For example, properties might be of the form $a + b = 1$ or $a < b$.
2. 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 Points
3. In terms of a, b, c, d, give $P(Y \neq 0)$ or argue why there is not enough information to compute the value. 2 Points
4. In terms of a, b, c, d, give $P(X + Y = 0 \mid X \cdot Y = 0)$ or argue why there is not enough information to compute the value. 2 Points
5. Now assume that X and Y are independent. Show that c can be computed from a and b. 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.
- When your business does well, the economy does 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. $P($ $) = 0.4$
   2. $P($ $) = 0.3$
   3. $P($ $) = 0.8$
2. Calculate the probability that your business does well when the economy does badly. 2 Points
3. How can you store the joint distribution of E and B in a minimally big table? 3 Points

Problem 2.2 (Bayesian Networks)

Consider the following situation about an alarm clock:

- It fails to ring if its batteries are empty.
- It fails to ring if it the volume is muted.
- You might oversleep if your alarm clock fails.
- You might oversleep if you stay up late.

You want to model this situation as a Bayesian network using Boolean random variables.

1. Give an appropriate set of random variables and their meaning and draw the Bayesian network (using an appropriate variable order). 3 Points
2. Give the probability of the clock failing in terms of the entries of the conditional probability tables of your network. 2 Points

Now you decide to model the failing of the clock as a deterministic node.

3. Explain (in about 3 sentences) whether that decision is justified by the description, and how it changes the conditional probability tables of your model. 3 Points
4. State the probability of the clock failing in terms of the entries of the conditional probability tables of your network (with the deterministic node for the failing clock). 2 Points

3 Markovian Reasoning

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

- You make daily observations about your business B. Each day business is either good (b_1) or bad (b_2).
- You know this is caused by the general economic situation G, which you cannot easily observe, and which can be getting worse (g_1), be stable (g_2), or getting better (g_3).
- You have previously obtained the following information:
  - when the economy gets worse, your business is good 36% of the time,
  - when the economy is stable, your business is good 84% of the time,
  - when the economy gets better, your business is good 90% of the time,
  - half the time, the economy is the same as on the previous day,
  - when the economy changes from one day to the next, each change is equally likely.

You want to model this situation as a hidden Markov model with two families of random variables indexed by day number d.

1. Give the state and evidence variables and their domains. 2 Points
2. How can you tell that the sensor model is stationary here? 1 Points
3. What order does the model have? 1 Points
4. Complete the following sentence: The transition model T is given by the matrix 2 Points

$$T = \begin{pmatrix} & & \\ & & \end{pmatrix} \qquad \text{where} \qquad T_{ij} = P(G_{d+1} = g_j \mid G_d = g_i).$$

5. Complete the following sentence: The sensor model S is given by the matrix (2 Points)

$$S = \begin{pmatrix} \quad & \quad & \quad \\ \\ \\ \end{pmatrix} \qquad \text{where} \qquad S_{ij} = P(B_d = b_j \mid G_d = g_i).$$

6. Let T be as above and let $\mathbf{v}$ be a 3-dimensional vector whose coefficients sum to 1. What is the intuitive meaning of the property $T \cdot \mathbf{v} = \mathbf{v}$? (2 Points)
7. Assume you want to apply filtering after observing good business at $t = 1$. Give the diagonal sensor matrix O_1 to use in this case. (1 Points)

Problem 3.2 (Decision Processes and Utility)
Consider a robot arranging 9 identical objects into a frame with 9 identical slots. Initially, all objects are stacked next to the frame.
Each object can be placed into every one of the 9 slots in two ways: correctly aligned or mis-aligned. The robot's task is to place one correctly aligned object into each slot.
In each move, the agent can do one of the following:

- Put the next unplaced object into a free slot. There is a 10% chance the placement is mis-aligned.
- Remove a mis-aligned object from the frame and add it back to the stack.
- Do nothing.

1. Model this situation as a Markov Decision Process $\langle S, A, T, s_0, R\rangle$. Use a reward function that uses -0.1 for non-goal states. (5 Points)
2. Give an optimal policy π^*. (2 Points)
3. State the Bellman equation for $\gamma = 0.5$. Then using initial utilities $U(s) = 0$ for all states, compute the utility value of the initial state after two value iteration steps. (3 Points)
4. Following up on the previous question: What is the smallest number of iterations, after which the utility of the initial state can be positive? (1 Points)

4 Learning

Problem 4.1 (Decision Trees)
Consider an unknown natural number $N \in \{0, \ldots, 3\}$ that we want to determine from certain Boolean attributes. For example, knowing that N is positive and even determines that $N = 2$.

1. Give the entropies $I(positive)$ and $I(even)$. Which of the two has higher information gain? 3 Points
2. Define 3 Boolean-valued attributes of N such that the following hold: 3 Points
   - The smallest decision tree for N using these attributes has depth 2.
   - There is no decision tree for N that uses only 2 of the 3 attributes.
3. Now assume our goal is to learn the function that computes whether N is a square of a natural number, using the attributes *positive* and *even*. We use $N = 0$ and $N = 1$ as examples. Formally state this situation as an inductive learning problem $\langle \mathcal{H}, T \rangle$. 2 Points

Problem 4.2 (Statistical Learning)
You observe the values below for 50 games of a tennis player. You want to predict the result based on time of day and opponent.

		Number of	
Time	Opponent	wins	losses
Morning	Weaker	5	1
Afternoon	Weaker	6	2
Evening	Weaker	3	0
Morning	Similar	3	3
Afternoon	Similar	2	3
Evening	Similar	4	5
Morning	Stronger	2	2
Afternoon	Stronger	1	3
Evening	Stronger	1	4

1. What is the hypothesis space for this classification task, seen as a decision tree learning problem? 2 Points
2. Explain (in about 2 sentences) the key characteristic of this data that makes it difficult to use decision tree–based classification methods from the lecture and why. 2 Points
3. Now instead, consider this as a statistical learning problem. As hypotheses, we use the probability distributions $P(Result \mid Time, Opponent)$. 2 Points
   Relative to the observed data, give the likelihood of the following hypothesis: The player's probability to win is 80% if the game is against a weaker player, and it is 10% otherwise.
4. To learn a hypothesis via Bayesian learning, we model this situation as a Bayesian network $Time \rightarrow Result \leftarrow Opponent$. Give the resulting entries of the conditional probability table for 2 Points
   1. $P(Opponent = Weaker) =$
   2. $P(Result = win \mid Time = Afternoon, Opponent = Weaker) =$

Problem 4.3 (Inductive Learning)
Assume we already know the predicate $\mathtt{par}(x, y)$ for x being a parent of y. Our goal is to learn the predicate $\mathrm{gp}(x, y)$ for x being a grandparent of y, i.e., to find a first-order formula D such that $\forall x, y.\mathrm{gp}(x, y) \Leftrightarrow D(x, y)$.
We do not know D, but we have the following (counter-)examples for gp:

x	y	$\mathrm{gp}(x, y)$
A	H	yes
B	I	yes
A	E	no
A	F	no
A	B	no
A	C	no

1. Give the first-order formula D that represents $\mathrm{gp}(x, y)$. 2 Points
2. Explain (in about 2 sentences) the key advantages of using inductive learning to learn a formula for gp as opposed to using the formula that can be derived from a decision tree for the example set. 2 Points
3. Assume that a first-order inductive learning algorithm has partially learned the predicate as the clause 2 Points
   $$\mathrm{gp}(x, y) \Leftarrow \mathtt{par}(x, u) \wedge \ldots$$
   Give two options of non-recursive literals that the algorithm might reasonably try next to complete the clause.

5 Natural Language Processing

Problem 5.1 (Language Models)

1. How many different trigrams does a language with n words have? 1 Points
2. What is a statistical language model? 2 Points
3. Name two applications of statistical language models. 2 Points
4. Why is it work-intensive in practice to build a good statistical language model for a natural language? 2 Points

Problem 5.2 (Seq2Seq Translation)

1. Explain (in about 2 sentences) the key idea of seq2seq models for machine translation. 2 Points
2. Explain (in about 2 sentences) the role of beam search in the decoder. 2 Points