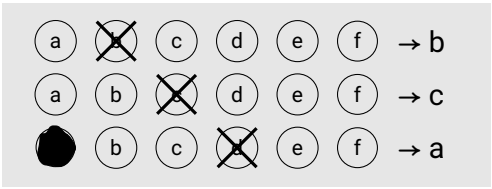
Exercises

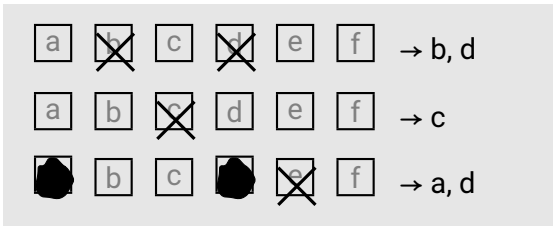
1	2	3	4	5	6	7	8	9	10
11	12								

1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
0	0	0	0	0	0	0

Surname, First name

KEN1110 Introduction to Data Science and AI
KEN1110 Introd. to DS & AI Final Exam





Fill in your answer(s) to the multiple-choice questions as shown above (circles = one correct answer, boxes = multiple correct answers possible).

Program: Bachelor Data Science and Artificial Intelligence
Course code: KEN1110
Examiners: Dr. Pietro Bonizzi, Dr. Rachel Cavill, and Prof. Anna Wilbik
Date/time: Wednesday 25 October 2023, 9:00-11:00
Format: closed book exam
Allowed aids: Pens, simple (non-programmable) calculator from the DACS-list of allowed calculators.
Instructions to students:

- The exam consists of 11 questions on 22 pages.
- Fill in your name and student ID number on the cover page and tick the corresponding numerals of your student number in the table (top right cover page).
- Answer every question in the reserved space below the question. **Do not write outside the reserved space or on the back of pages, this will not be scanned and will NOT be graded!** As a last resort if you run out of space, use the extra answer space at the end of the exam.
- *In no circumstance write on or near the QR code at the bottom of the page!*
- This exam consists of two parts. Both worth 50 points.
- Ensure that you properly motivate your answers.
- Only use black or dark blue pens, and write in a readable way. Do not use pencils.
- Answers that cannot be read easily cannot be graded and may therefore lower your grade.
- If you think a question is ambiguous, or even erroneous, and you cannot ask during the exam to clarify this, explain this in detail in the space reserved for the answer to the question.
- If you have not registered for the exam, your answers will not be graded, and thus handled as invalid.
- You are not allowed to have a communication device within your reach, nor to wear or use a watch.
- You have to return all pages of the exam. You are not allowed to take any sheets, even blank, home.
- **Success!**

Part 1 - Artificial Intelligence Question 1

A restaurant decides to use a robot server, like the one pictured above, to deliver dishes from the kitchen to the tables.

3p **1a** Suggest what actuators and sensors the robot would need to perform this task.

5p **1b** The robot's designer states that the environment is;

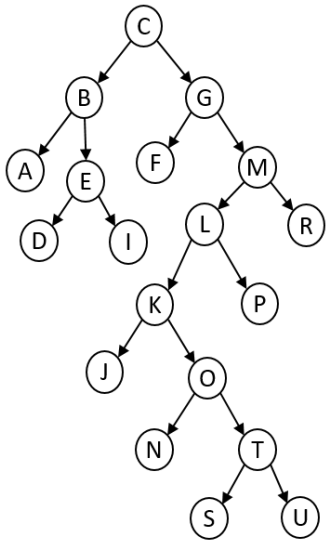
- Static because the tables do not move.
- Sequential because they need to plan a route through the restaurant, and history of decisions matters.
- Stochastic because there is randomness in what people order.
- Discrete because there is a limited menu.
- Partially observable because the robot can't see through walls/people/furniture.
- Multi agent as the robot must interact with people who are agents too.

Do you agree with this definition of the environment - explain all points you disagree with.

2p **1c** After a while the restaurant owner decides that it is taking the staff too long to instruct the robot on where to go for each delivery and that it would be better if the robot could understand verbal instructions. Would such an improved robot pass the Turing test? Explain.

Question 2

Consider the following search tree.



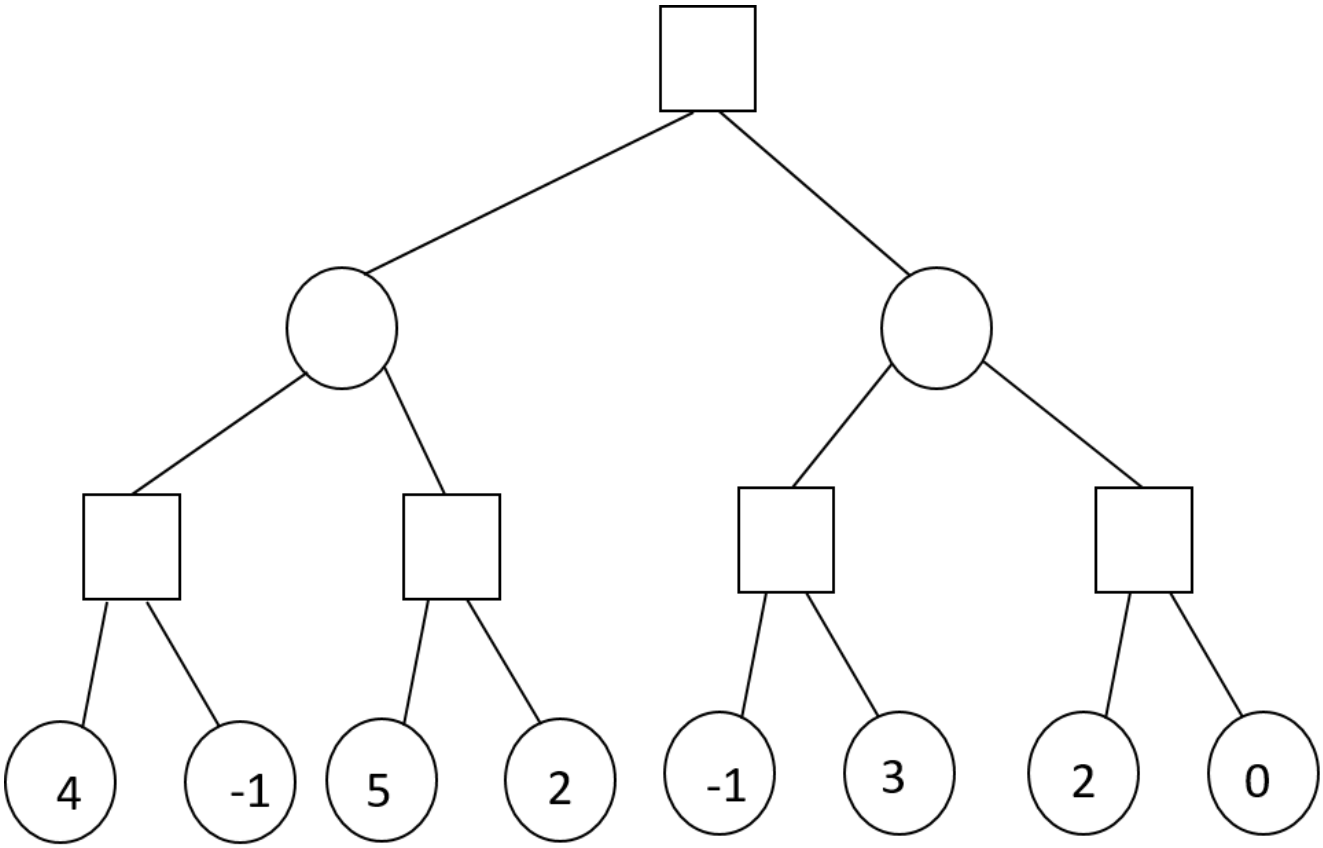
2p **2a** Which of the following statements is true? Select all that apply.

- ☐ M is leaf
- ☐ C is root
- ☐ C and L are internal nodes
- ☐ G is ancestor of E
- ☐ K has 3 grandchildren

2p **2b** What is the search order for this tree if you use breadth first search?

2p **2c** What is the search order for this tree if you use depth first search?

4p **2d** Consider the following tree of a 2 people game. Solve this tree.
Indicate the principle variation by circling it. Mark with the X, branch that can be pruned.



Question 3

4p **3a** Does $(A \Rightarrow B) \vee (B \Rightarrow A) \vDash A \wedge (A \Rightarrow \neg B) \wedge B$?
Show a truth table.

6p **3b** Which of the following statements is true. Select all that apply.

- ☐ The right hand side of the equation in part a is satisfiable.
- ☐ The left hand side of the equation in part a is unsatisfiable.
- ☐ The right hand side of the equation in part a is valid.
- ☐ The left hand side of the equation in part a is satisfiable.
- ☐ The left hand side of the equation in part a is valid.
- ☐ The right hand side of the equation in part a is unsatisfiable.

Question 4

NASA wants to predict when there will be thermal distress in the O-Ring part of their space shuttle flights. They have two attributes, firstly the launch temperature which is measured in degrees Fahrenheit to the nearest degree, secondly the pressure at which they performed the pre-flight leak-check, which is measured in psi and has 3 possible values.

Number of O-rings experiencing thermal distress	Launch temperature	Leak-check pressure used
0	66	50
0	67	50
0	68	50
0	69	50
1	70	50
0	72	50
0	70	100
0	73	100
2	53	200
1	57	200
1	58	200
1	63	200
0	67	200
0	67	200
1	70	200
0	70	200
0	75	200
2	75	200
0	76	200
0	76	200
0	78	200
0	79	200
0	81	200

2p **4a** Show that if we want to predict if **any** O-rings experience thermal distress the entropy of this dataset is 0.887 to 3 significant figures, where entropy is given by $E(S) = -p^{+}log_2p^{+} - p^{-}log_2p^{-}$

2p **4b** Complete the following table to calculate the gain for different temperature split points, where gain is given by $Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_V|}{|S|} E(S_V)$ and $S_V = \{s \in S | A(s) = V\}$

Show your working below the table.

Temp	+ve below	-ve below	Entropy below	+ve above	-ve above	Entropy above	Gain
64.5	4	0	0.000	3	16	0.629	0.367
69.5	4	6	0.971	3	10	0.779	0.024
71	6	8	0.985	1	8	0.503	0.090
74	6	10	0.954	1	6	0.592	0.043
75.5	7	11	0.964	0	5	0.000	

3p **4c** Calculate the gain for the Leak-Check pressure, using the values as three different options (like rainy, sunny, cloudy).

2p **4d** Given the above draw the first split point of the tree, and indicate how many samples of each class are placed at each branch.

1p **4e** What would change in our calculations if we wanted to predict (as a class) the number of O-rings that suffer distress, rather than whether any O-rings suffer thermal distress?

Question 5

A dictionary for a collection contains the following 10 terms, together with their document frequency:
horse -10
dog - 5
fox - 3
rabbit - 15
bear - 23
sheep - 12
cat - 7
cow - 4
giraffe - 18
zebra - 2

Consider a query: (zebra OR horse) AND (horse OR dog) AND bear AND (cow OR sheep) AND NOT giraffe AND fox.

1p **5a** Which of the following is processed first? Select one answer.

(a) zebra OR rabbit
(b) fox AND bear
(c) fox AND NOT giraffe
(d) horse OR dog

1p **5b** Which term is processed as last? Select one answer.

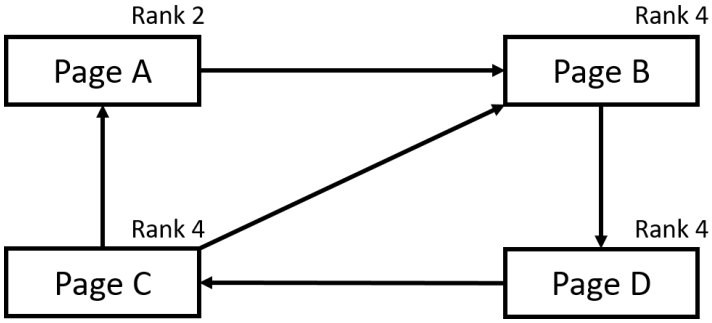
(a) fox (b) bear (c) zebra (d) cat

2p **5c** Which terms are processed before term "giraffe"? Select all that apply.

☐ cat ☐ bear ☐ horse ☐ cow

1p **5d** What is the maximal number of documents this query will return?

3p **5e** Four web pages have the following link structure and ranks of the pages.



Has the PageRank algorithm converged? Show your calculations.

Part 2 - Data Science: Question 6

3p **6a** Assume having a book about Data Science & AI. Reading and understanding a few concepts in the book can be considered:

(a) Information
(b) Data
(c) Knowledge

2p **6b** A car collector would like to make an inventory of their cars to make an estimate of the amount of vintage cars in their collection (cars from the period of 1919 to 1930). Which one of the following represents a sensible rule that they could implement in an algorithm, to automate this task? (assume numbers mean years - tick the right answer):

(a) If car's age > 92, then vintage.
(b) If $92 <$ car's age < 103, then vintage.
(c) If car old enough, then vintage.
(d) If car is vintage, then $92 <$ car's age < 103.

2p **6c** Which one of the following can be classified as "knowledge"? (Tick the right answer):

(a) Information stored in case of need.
(b) Data after they have been understood and provided with meaning.
(c) Information after it has been understood and provided with meaning.
(d) Information used to make a decision.

Question 7

4p **7a** Solve this expression $(1+i)(-2-i)$, and plot the result in the complex plane.

3p **7b** Write the complex number e^{π} in the form $a+ib$.

3p **7c** Find $x \in \mathbb{C}$ such that: $x\bar{x} - x = 0$.

Question 8

Observant wants to better distribute their paper editions to the newspaper boxes. They are interested in knowing how many newspapers should be allocated to each box.

1p **8a** Is this a good problem for data mining?

(a) no (b) yes

3p **8b** Motivate your answer with at least 3 arguments.

4p **8c** List 4 conditions of a distance.

2p **8d** Consider two vectors A={1, -1, 2, -6} and B={-2, -2, 2, 5}. Calculate the Euclidean and absolute (Manhattan) distances for those vectors.

Question 9

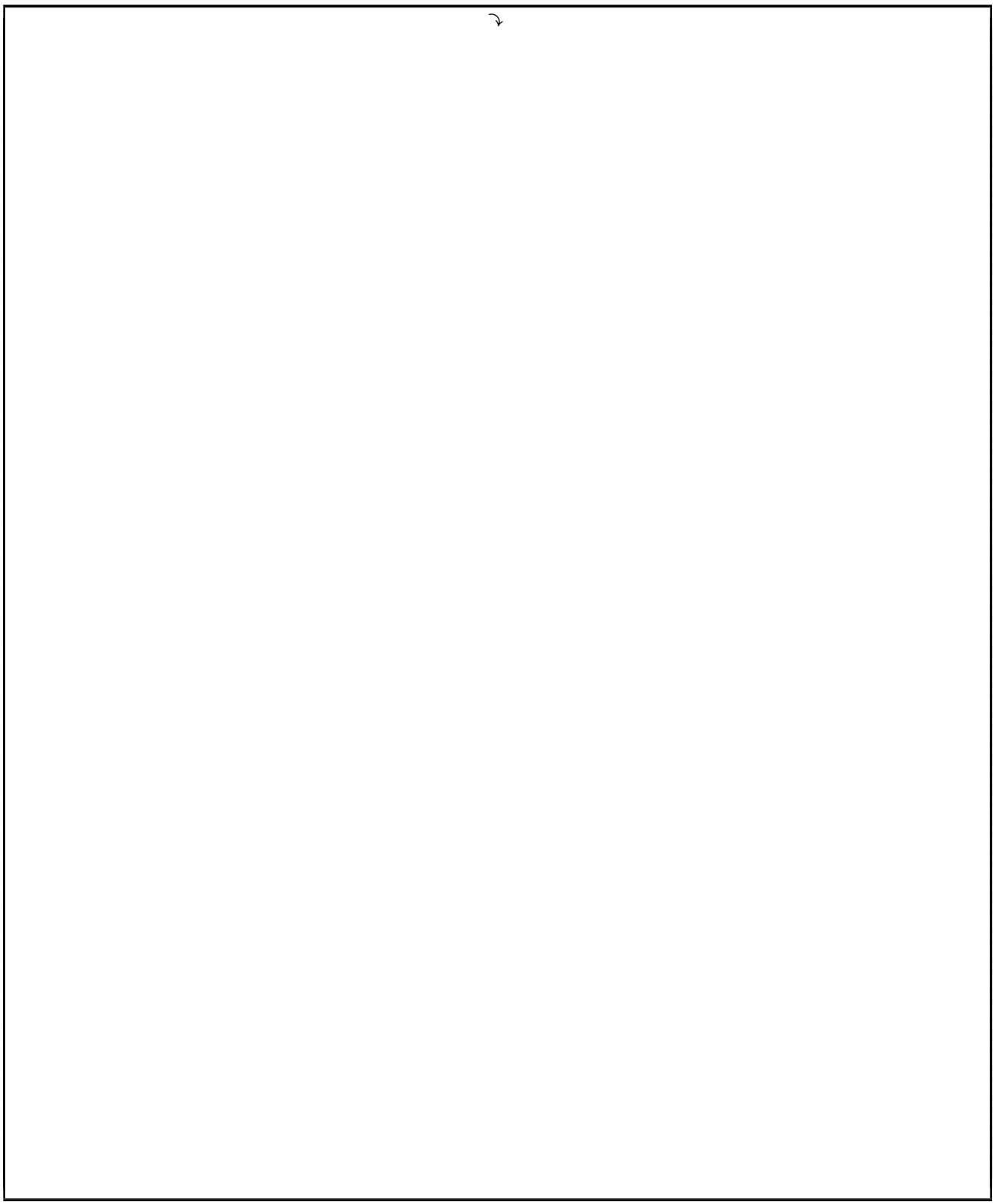
In a physical experiment four different measurements y_k are obtained over four different x_k.

x_k	-1	0	2	1
y_k	1	1	0	2

The physicist who collected those is interested in fitting these measurements to a model of the form: $y_k = ax_k + b$.

3p **9a** Plot the data and draw a qualitative estimate of what you expect to be the best fitting line through the data.

7p **9b** Compute the optimal values for a and b according to the chosen model and the data collected during the experiment. Does the best fitting line you obtained agree with your estimate at point a.?



Question 10

2p **10a** Which of the following statements are true? Select all that apply.

- ☐ In a union distribution every point has equal chances of being selected
- ☐ Normal distribution is symmetric around the mean
- ☐ Union distribution is defined on [0,1] interval
- ☐ Uniform distribution can be used to simulate a result of tossing a fair dice.

For questions b-d indicate which of the following are correct bootstrapped samples of X={1,1,2,5,6,6,7,10}, assuming equal length.

1p **10b** A={1,1,1,1,2,2,6,6}.

(a) yes (b) no

1p **10c** B={1,2,3,5,6,6,7,10}

(a) yes (b) no

1p **10d** C={1,2,2,5,6,6}

(a) no (b) yes

Question 11

Two friends want to meet and they need to agree on the location. If they choose a location that they both like, they both receive a reward of 2. If they choose a location that only one of them likes, the person who likes it receives a reward of 1, and the other person receives a reward of -1. If they choose a location that none of them likes, they both receive a reward of -2.

5p **11a** Build the payoff table of this game.

5p **11b** How many Nash equilibria exist for this game?

Extra space for answers

If you use these extra answer boxes, please mention clearly in your main answer that part of your answer can be found here!

12a Extra space 1

12b Extra space 2