Introduction to Data Science and Knowledge Engineering 2019/2020 Exam Questions

— Do not turn this page before the official start of the exam! —

First Name, Surname: _____

Student ID: _____

Program: Bachelor Data Science and Knowledge Engineering

Course code: KEN1110

Examiners: Dr. Rachel Cavill & Dr. Pietro Bonizzi

Date/time: Friday October 25th, 2019, 9.00-12.00h

Format: Closed book exam

Allowed aides: Pens, simple (non-programmable) calculator from the DKE-list of allowed calculators.

Instructions to students:

- The exam consists of 11 questions on 18 pages (excluding the 1 cover page(s)).
- Fill in your name and student ID number on each page, including the cover page.
- This exam consists out of two parts. Both worth 50 points.
- Answer every question at the reserved space below the questions. If you run out of space, continue on the back side, and if needed, use the extra blank page.
- Ensure that you properly motivate your answers.
- Before answering the questions, please first read all the exam questions, and then make a plan to spend the three hours.
- Do not use red pens, and write in a readable way. Answers that cannot be read easily cannot be graded and may therefore lower your grade.
- 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.
- 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.
- **Success!**

The following table will be filled by the examiner:

Question:	1	2	3	4	5	6	7	8	9	10	11	Total
Points:	13	10	5	14	8	5	10	10	10	5	10	100
Score:												

Part I

1. (13 points) a. An agent is used to talk to customers using a website and advise them about their holiday choices. Describe the environment for this agent using the standard terms explaining why each of the terms applies.

b. Penelope says that it is important that the agent can pass the Turing Test. Richard disagrees. Explain what a Turing test is, and give one point backing each point of view.

c. Penelope says it will be best if the agent can learn. Complete the figure below with the elements needed to make the agent learn.

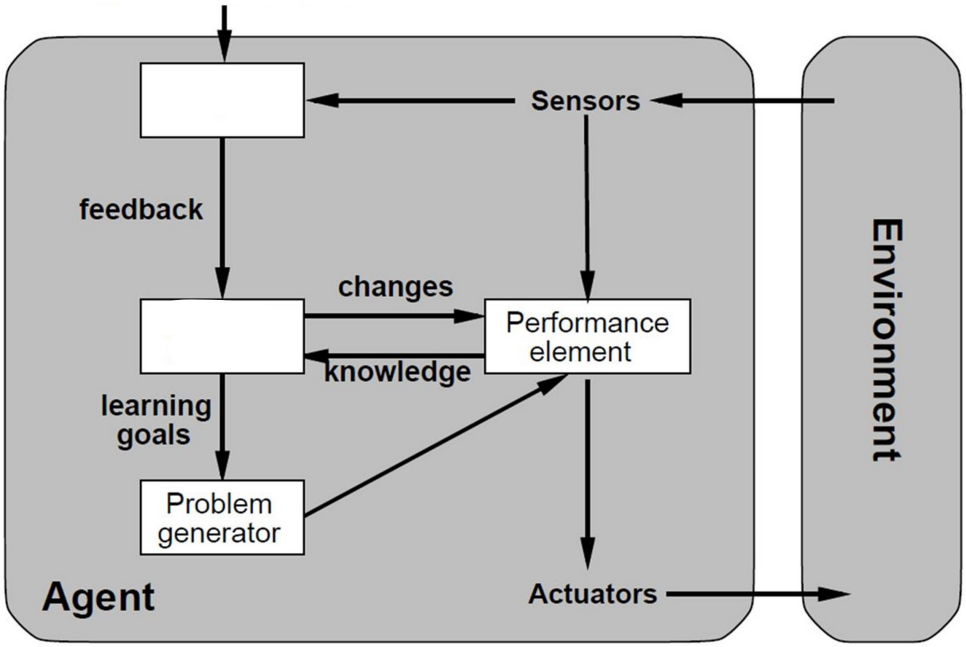

Figure 1: An incomplete learning agent

2. (10 points) A delivery company has 3 fragile packages to deliver. The packages must be kept upright (no rotations/reflections allowed). They want to fit all three packages in a standard packing box which is three units across, four units high. They have an agent to help with determining the best order for packing. The agent assumes gravity, so each package will be placed at the lowest point possible in the packing box.

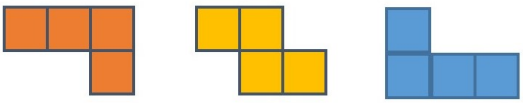

Figure 2: The three packages

a. Show the full search tree that this agent will generate when exploring the packing order for these packages.

b. Would you recommend that this agent implements breadth first or depth first search and why?
Consider two situations;
(i) Where the agent simulates packing the parcels.
(ii) Where the agent actually places the parcels in the packing box during the search process.

c. A new set of packages have arrived. Explain how the branching factor of the search tree for these packages will differ from your original tree and why.

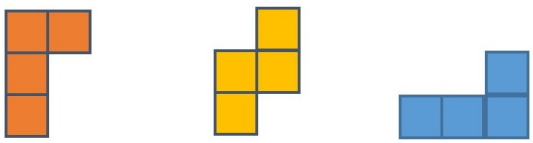

Figure 3: The three new packages

3. (5 points) a. Is the following statement satisfiable? Include the truth table in your answer. $(smoke \Rightarrow fire) \Rightarrow (\neg smoke \Rightarrow \neg fire)$.

b. If you had a knowledge base containing just the above statement would this knowledge base $\models smoke$. Explain your answer, using your work from the previous part.

4. (14 points) We wish to predict whether a patient can wear contact lenses or not, we have data on four variables which can affect this prediction.

Age	Myopic?	Astigmatic?	Tear production rate	Class
Young	Yes	No	Reduced	No
Young	Yes	No	Normal	Lenses
Young	Yes	Yes	Reduced	No
Young	Yes	Yes	Normal	Lenses
Young	No	No	Reduced	No
Young	No	No	Normal	Lenses
Young	No	Yes	Reduced	No
Young	No	Yes	Normal	Lenses
Middle Aged	Yes	No	Reduced	No
Middle Aged	Yes	No	Normal	Lenses
Middle Aged	Yes	Yes	Reduced	No
Middle Aged	Yes	Yes	Normal	Lenses
Middle Aged	No	No	Reduced	No
Middle Aged	No	No	Normal	Lenses
Middle Aged	No	Yes	Reduced	No
Middle Aged	No	Yes	Normal	No
Old	Yes	No	Reduced	No
Old	Yes	No	Normal	No
Old	Yes	Yes	Reduced	No
Old	Yes	Yes	Normal	Lenses
Old	No	No	Reduced	No
Old	No	No	Normal	Lenses
Old	No	Yes	Reduced	No
Old	No	Yes	Normal	No

a. Given the entropy formula $E(S) = -p^{+}log_2p^{+} - p^{-}log_2p^{-}$, calculate the entropy of the dataset.

b. Calculate the information gain of using each attribute as the first choice in the classification tree, where $Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_V|}{|S|} E(S_V)$ and $S_V = \{s \in S | A(s) = V\}$

c. Perform the additional calculations necessary and draw the first two decision splits of the classification tree (space on the following page). Show which samples are left at each leaf node where there is still a mixed class group.

Draw classification tree here.

5. (8 points) Five web pages have the following link structure and previous pagerank.

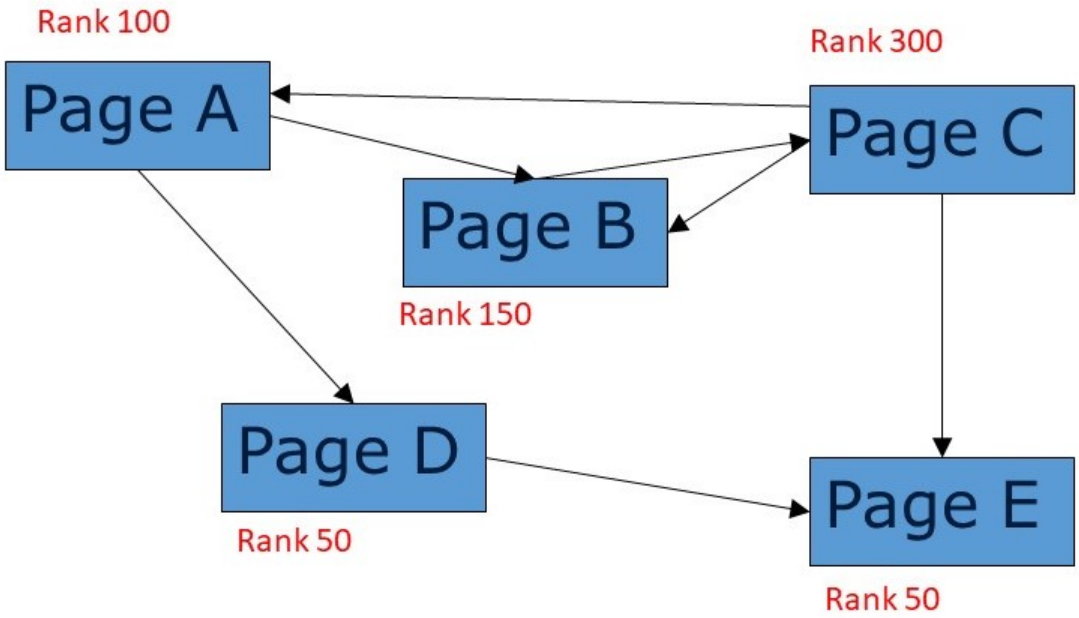

Figure 4: Link structure and previous pagerank of 5 web pages

a. Which pages have a new rank after one iteration of the Page rank algorithm?

b. After many iterations of the algorithm the new ranks of the pages are; Page A = 72, Page B = 216, Page C = 216, Page D = 36, Page E = 108. Has the algorithm converged? Explain your answer.

Part II

6. (5 points) Given the following situation:

   A pressure sensor is a device for pressure measurement of gases or liquids, which can be used to monitor pressure in a pipeline and keep it below dangerous levels.

   a. Explain how measurements of pressure in a pipeline could be transformed from data into information and then into knowledge.

   b. Express the knowledge generated at the previous point into a rule which could be implemented into a knowledge based system.

7. (10 points) Let $z = -1 + i$ and $w = i$.

   a. Compute $\bar{z}$

   b. Compute $|z|$

c. Express z in trigonometric form and plot it in the complex plane.

d. Compute $\frac{\bar{z}}{w}$

e. Find $x \in \mathbb{C}$ such that: $x\bar{x} = x(1-i)$

8. (10 points) Given the three sequences:

$$x = \{1, 0, -2, 1\}, \quad y = \{0, 1, 1, -1\}, \quad z = \{1, 0, 1, 1\}$$

a. Compute the absolute distances $d(x, y)$, $d(x, z)$, and $d(y, z)$. What are the most similar time series?

b. Compute the infinity distances $d(x, y)$, $d(x, z)$, and $d(y, z)$. What are the most similar time series? What do you notice compared to the result obtain at the previous point?

9. (10 points) In a physical experiment four different measurements y_k are obtained over four different x_k

x_k	0	1	0	2
y_k	1	1	-1	1

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

Compute the optimal values for a and b according to the chosen model and the data collected during the experiment.

10. (5 points) a. Consider the following data set of five values $\{1, -2, 1, 0, 3\}$. Generate four bootstrap samples with replacement of the original data set, each containing five values (same amount of values as in the original dataset).

b. Consider the experiment of tossing a fair coin. Write a rule that would allow you to simulate this experiment with a computer.

11. (10 points) Given the following game:

		Player 2		
Player 1	Strategy	1	2	3
	1	6,3	4,4	4,1
	2	5,4	6,5	0,2
	3	5,0	3,2	6,1

a. Can the table be simplified by looking for strictly dominated strategies and removing those?

b. Does the game have a Nash equilibrium?