Maastricht University

Department of Data Science and Knowledge Engineering

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: Model answers part 1

Student ID: _____

Program: Bachelor Data Science and Knowledge Engineering

Course code: KEN1110

Examiners: Dr. Rachel Cavill & Dr. Pietro Bonizzi

Date/time: Friday October 25thth, 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 5 questions on 12 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.
- **Good luck!**

The following table will be filled by the examiner:

Question:	1	2	3	4	5	Total
Points:	13	10	5	12	8	48
Score:						

Part I

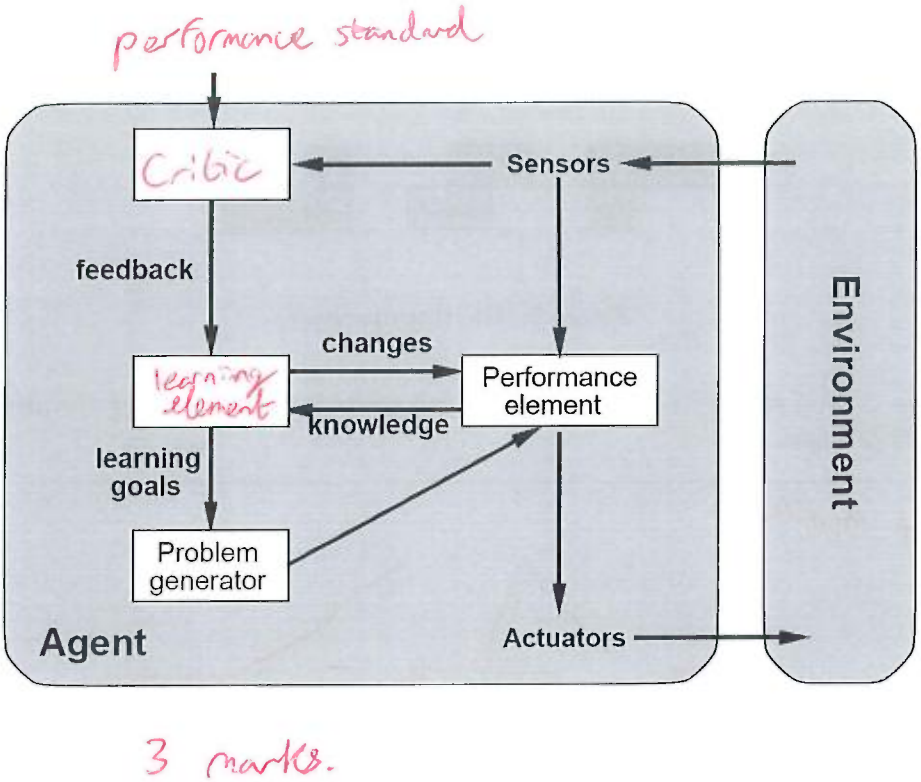
1. (13 points) a. An agent is used to talk to customers using a website and advise them about their holiday choices. Describe the environment for this agent using the standard terms explaining why each of the terms applies.

Main point here is that the definition of each term should be correctly applied.
1 point per item. (multiple correct answers were accepted if correctly justified)
model answer:
Fully observable - the agent can see the whole conversation it is responding to.
Multi-agent - agent + customer/s
Stochastic - actions do not have predictable consequences. we don't know if user will book a holiday.
Sequential - the history of the conversation is important when deciding what to say next.
Dynamic - the person may continue the conversation while the agent is deliberating.
Discrete - Only a limited number of conversational options are needed for the agent.

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.

2 marks for Turing test explanation, including how it works.
2 marks for Penelope/Richard's arguments.
Full marks only given to arguments about this agent.
eg. penelope says customers will be more comfortable if they think they are talking to a human.
Richard says developing a Turing test passing AI is too broad / expensive for this task.

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.



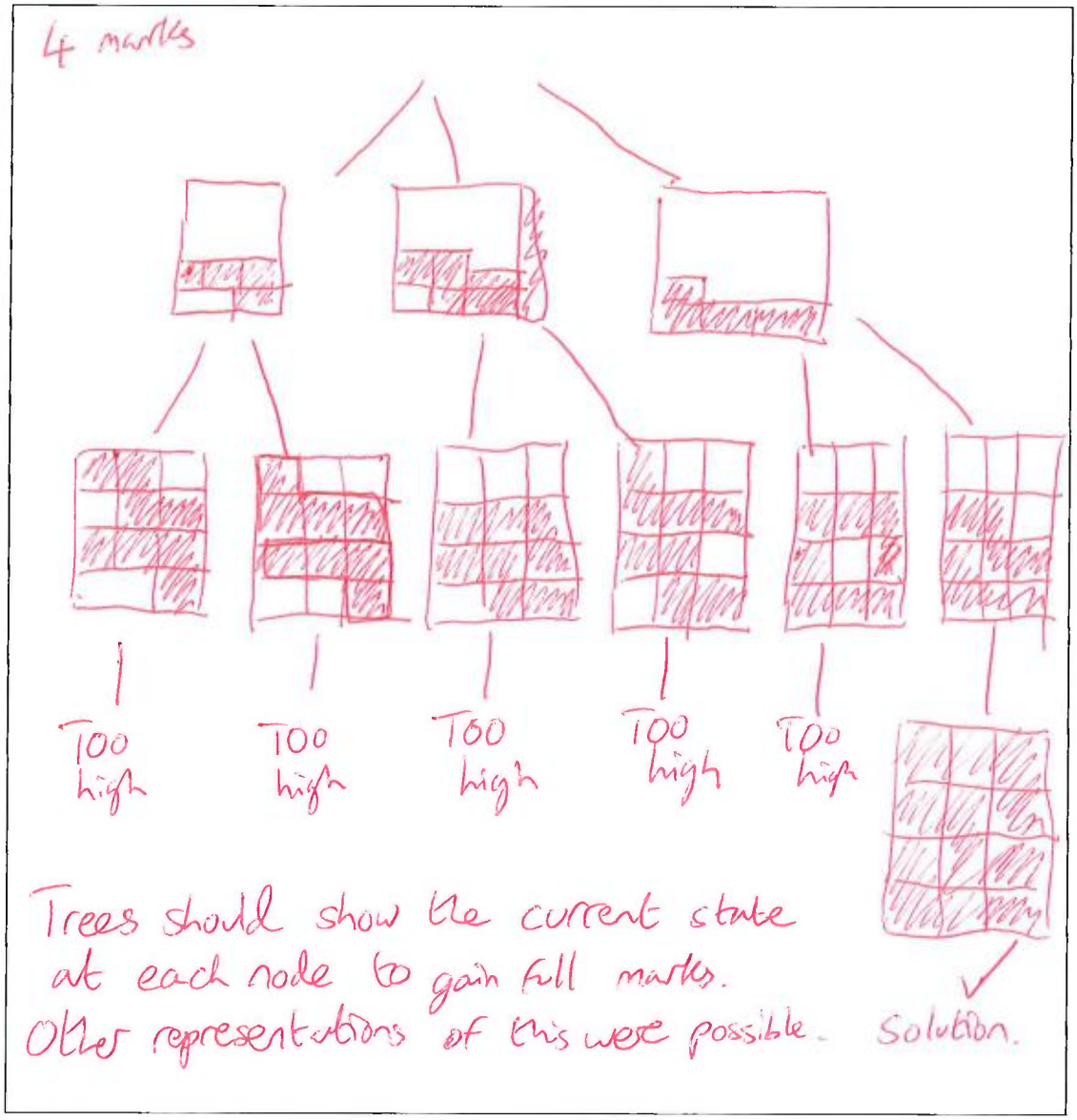
3 marks.

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.

i) If we are simulating it doesn't make a difference in this case, as solution is bottom right in tree.

ii) When placing the packages in the box depth first search is better as it will involve less moving of fragile packages.

4 marks.

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

Some packages can now go in more than one place (L, R), so potentially 5 options instead of 3.
So branching factor increases.
However, at later stages in the tree there will be less options, as it goes too high more quickly.
This will also affect branching factor.
2 marks.

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

3 marks.

Smoke	Fire	Smoke ⇒ Fire	¬Smoke ⇒ ¬Fire	Whole
T	T	T	T	T
T	F	F	T	T
F	T	T	F	F
F	F	T	T	T

So it is satisfiable, as true in some worlds

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.

2 marks.

KB is true in worlds 1, 2, 4.
Smoke is true in words 1, 2

$\{1, 2, 4\} \not\subseteq \{1, 2\}$

So KB $\not\models$ smoke.

4. (12 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.

$-\frac{15}{24}\log_2\left(\frac{15}{24}\right) - \frac{9}{24}\log_2\left(\frac{9}{24}\right) = 0.954$

2 marks.

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 $S_V = \{s \in S | A(s) = V\}$

4 marks.

Gain (S, age) = 0.03

Gain (S, myopic) = 0.005

Gain (S, astigmatic) = 0.005 (Same numbers as myopic)

Gain (S, tear production rate) = 0.549

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.

The first split will be tear production rate from above.

if reduced tears => no lenses.

so just need to look at normal tears group.

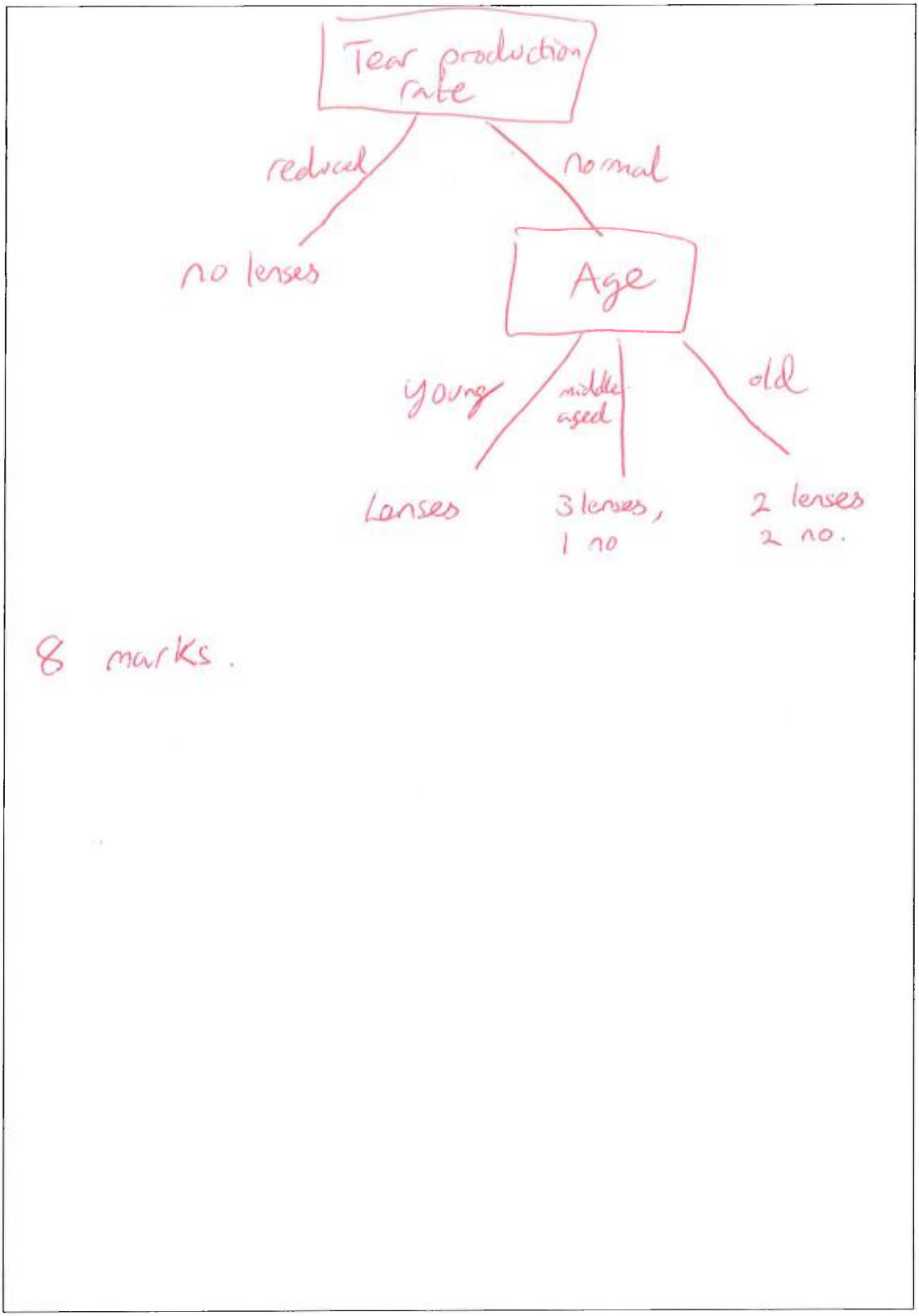
The entropy of this group is 0.811

Gain (Normal, Age) = 0.208

Gain (Normal, myopic) = 0.027

Gain (Normal, astigmatic) = 0.027

Draw classification tree here.



8 marks.

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

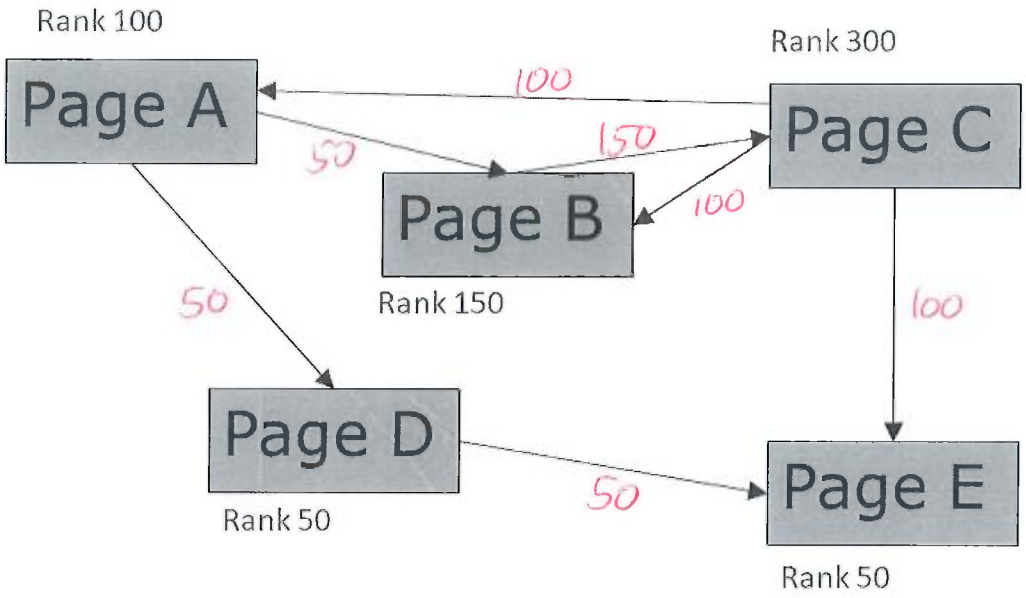


Figure 4: The three new packages

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

	old		new	
Page A	100	→	100	same
B	150	→	100+50	same
C	300	→	150	changed
D	50	→	50	Same
E	50	→	150	Changed.

5 marks.

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.

If it has converged then when we calculate another iteration from this point we will get the same answer as above.

In this case, all pages stay the same except B which changes from 216 to 108.

So it has not converged.

Part II