Exam: Artificial Intelligence – Algorithms and Application

Module Exam

Winter 2021 / 2022 Date: 31.03.2022

TECHNISCHE UNIVERSITÄT DARMSTADT

WIRTSCHAFTS INFORMATIK

Important Information

- Please check your exam copy for completeness.
  It covers **18 pages** (cover sheet included).
- Fill out the cover sheet immediately after receiving the exam.
- Use only the examination paper to solve the tasks. If you do not have enough space, you can receive additional paper during the examination. Additional papers must also be marked with your name and matriculation number.
- Please leave a **correction margin of 3 cm**.
- You have a total of **90 minutes** to complete the exam.
- Except for a **non-programmable calculator, no other aids** are allowed in the exam.

We wish you much success!

Please fill out clearly in block letters.

First Name .... Last Name .... Seat No. ....

Matr. No. .... Course of Study .... ❑ Master ❑ Diplom

Repeater:
❑ yes ❑ no

Section	Max. Points	Achieved Points
1	30	
2	34	
3	26	
Sum	90	

Exam Review („Klausureinsicht“):

(do not fill out before the review)

I have reviewed the corrected exam:

□ There are no complaints about the correction.

□ Complaints about the correction exist (see additional sheet).

Date: ………………………………………..

Signature: ……………………………….....

First Name................................ Last Name.. Matr. No...

1 Basic Concepts and Algorithms (30 Points)

1.1 Please **briefly compare** the Python Development in the **Spyder IDE** and **Jupyter notebooks**. Please **name one specific use case** for which **Spyder** is a better choice and **one** for which **Jupyter notebooks** are a better choice. (2 P)

1.2 Please **explain** the difference between ***feedback*** and ***feedforward*** **ANNs**. How do the

1.3 Please **briefly explain** the **concept** of an "***Intelligent Agent***" in Artificial Intelligence based on the definition of Russell and Norvig. (2 P)

1.4 Please **briefly explain** the **central statements** of the ***Moravec's Paradox***. Please **also provide an example** of your own regarding today's Artificial Intelligence. (4 P)

1.5 **Explain** the **difference** between a ***route-finding problem*** and the ***touring problem***. Which **type** of problem is the ***traveling salesman problem***? (3 P)

1.6 What **type** of **information system** is "***Mycin***"? (1 P)

1.7 What is the **conjunctive normal form** (**CNF**) in knowledge reasoning? Please also **provide one example** in **CNF**. (2 P)

1.8 **What** is a ***CAPTCHA***? **How** is it **related** to **Artificial Intelligence**? (2 P)

First Name................................ Last Name... Matr. No..

1.9 Please **briefly explain** the different **steps** of the ***CRISP-DM*** model that is used as a development framework in analytics and AI projects. (6 P)

1.10 Please **briefly explain** the **difference** between ***batch*** and ***online*** **learning** in machine learning. (2 P)

1.11 **What** is the **difference** between ***k-means*** and ***k-medians***? (2 P)

1.12 Please briefly **explain** the main idea of **PCA**. What is **PCA used for**? (2 P)

○

First Name................................ Last Name.. Matr. No...

2 Application of Markov Theory (34 Points)

Consider the following **AI problem**:

A genetic algorithm uses ***chromosomes*** in the form $c = G_1G_2G_3G_4G_5G_6G_7G_8$.

The chromosomes are of a ***fixed length*** of eight genes G_i.

Each gene can be any digit between 0 and 9.

The fitness of an individual *x* is calculated by the following ***fitness function***:

$$fitness = (G_1 + G_2) - (G_3 + G_4) + (G_5 + G_6) - (G_7 + G_8)$$

The ***initial population*** consists of four individuals with the following chromosomes:

$$c_1 = 56415333;\ c_2 = 78176601;\ c_3 = 32291258;\ c_4 = 14854321$$

2.1 Please briefly **describe** the **main idea** behind ***Genetic Algorithms***. Please also **explain each of the five components** of Genetic Algorithms. (1+ 5 = 6 P)

2.2 Please **describe** the **sequence** of the **process** of ***genetic algorithms*** using **Python** or **Pseudocode**. (2 P)

First Name.................................. Last Name... Matr. No...

2.3 Please **calculate** the ***fitness*** of **each** of the above given **individuals** and ***rank*** them from the **fittest** to the **least fit** one. (5 P)

2.4 Please **perform** the **following crossover operations**. (6 P)

a) Cross the two **fittest individuals** using ***one-point crossover*** in the **middle**.

b) Cross the two **least fit individuals** using a ***two-point crossover*** between $\boldsymbol{G_2}$ and $\boldsymbol{G_3}$ and **between** $\boldsymbol{G_6}$ and $\boldsymbol{G_7}$.

c) Make a ***uniform crossover*** of $\boldsymbol{c_1}$ and the **fittest** individual.

First Name................................ Last Name... Matr. No...

2.5 After multiple evaluations you get the following **population** of **offsprings *O***.

Please **evaluate** the ***fitness*** of the **new population** and **explain** if the **fitness** has **improved** or **not**. (5 P)

o_1 = 44444444; o_2 = 33333333; o_3 = 12341234; o_4 = 43214321; o_5 = 48163248

2.6 Please **consider only** the ***fitness function*** and that the ***genes of the distinct chromosomes*** can **only** be ***digits***. Please **find** the ***optimal solution*** with the **maximum fitness**. **Explain** your **solution** and your **calculations**. (6 P)

2.7 If you **consider** the ***initial*** **chromosomes**, is a genetic algorithm able to **reach** the **optimal solution from above without** the **mutation** operator? Please explain why or why not. (4 P)

First Name................................ Last Name... Matr. No...

3 Data Preprocessing with Python (26 Points)

In your current project for a big German sport car company, you plan to investigate unknown car failures. You aim to find correlations between different parts to identify the root cause. During your first analysis, you face the following dataset that you received from the R&D department:

PartID	Height (cm)	Height (inches)	Weight
1	68	26,77	120
2	74	29,13	130
3	76	29,92	141
…	…	…	…

3.1 Firstly, you try to **load the dataset** in your IDE. However, the upload **does not work** and **returns** the following **errors**.

Python Code

```
connection = sqlite3.connect("failures.db")
interface = connection.cursor()
sql = "SELECT * FROM failure_report"
cursor.execute(sql)

Traceback (most recent call last):
  File "<ipython-input-40-ce8946a7ae34>", line 9, in <module>
    cursor.execute(sql)

ProgrammingError: Cannot operate on a closed database.
```

Please **explain** the (a) ***reason for the errors*** and (b) ***how to solve them***. By doing so, please **provide code** in **Python** that solves this issue. (1 P)

3.2 In a first step, you plan to **cluster similar parts**. For that purpose you define the following **`distance()` function** based on the **Euclidian distance**:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

Please **implement** this function in **Python**. Write down the function in **code** and **comment** your code to explain it. (4 P)

(**Note**: One should be able to use your implemented distance function with input as in the task 3.3)

3.3 Please **compute** the **results** of the following **code** (rounded to three decimal places):

```
sim_12_cm = distance([68, 120], [74, 130])
```

Please **do the same** for the **following similarities.** (6P)

(**Note** that, e.g., `sim_23_inches` means that you should compute the distance from the point with *PartID=2* to the point with *PartID=3* based on their respective *Height (inches)* and *Weight*.)

Similarity	Result
sim_23_inches	
sim_13_inches	
sim_12_inches	
sim_23_cm	
sim_13_cm	

3.4 Compare the results of **task 3.3** regarding inches and centimeters. Please **discuss** and **explain** the results. **How** do you **handle such behavior** in data preprocessing ***<u>in general</u>***? (2P)

First Name.................................. Last Name... Matr. No...

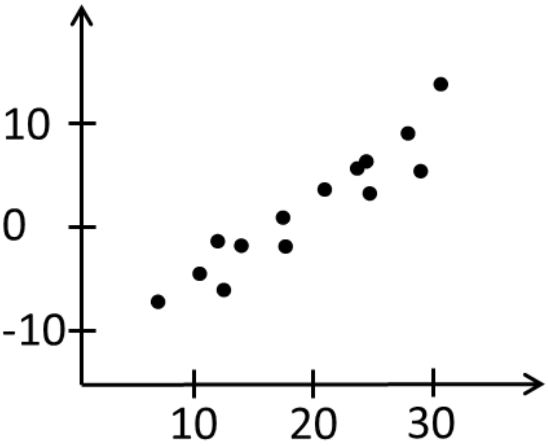
3.5 To handle this problem, you decide to write the following **rescaling function**:

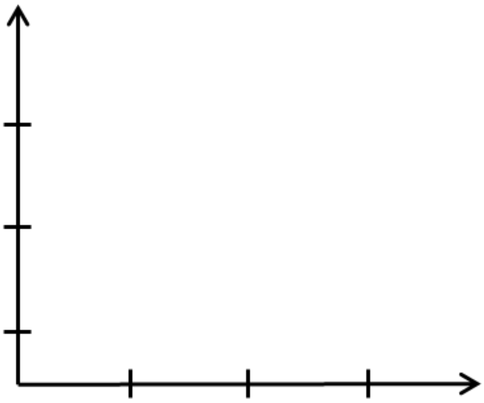
Python Code

```
def scale(data):
    num_rows, num_cols = shape(data)
    means = [mean(get_column(data, j))
                    for j in range(num_cols)]
    stdevs = [standard_deviation(get_column(data, j))
                    for j in range(num_cols)]
    return means, stdevs

def rescale(data):
    means, stdevs = scale(data)
    def rescaled(i,j):
        if stdevs[j] > 0 :
            return(data[i][j] - means[j]) / stdevs[j]
        else:
            return data[i][j]
    num_rows, num_cols = shape(data)
    return make_matrix(num_rows, num_cols, rescaled)
```

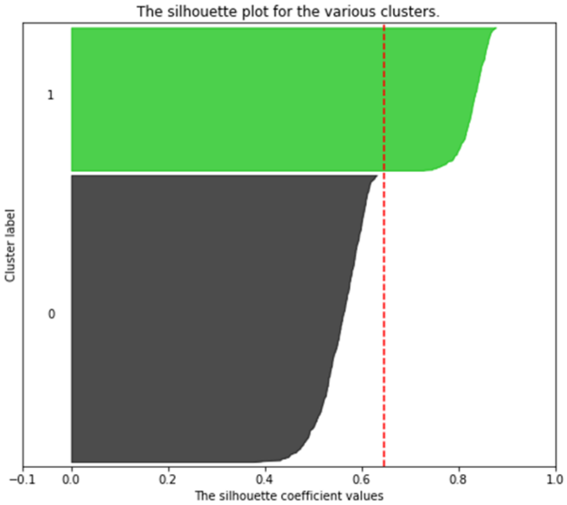
Then, you apply the `rescale()` function on the following dataset.
Please **illustrate how** the `rescale()` function **changes** the **data** by **painting a new plot**. For this purpose, please **explain** what the **two functions do** and **how they work** and **how** they **influence the data**. (8 P)

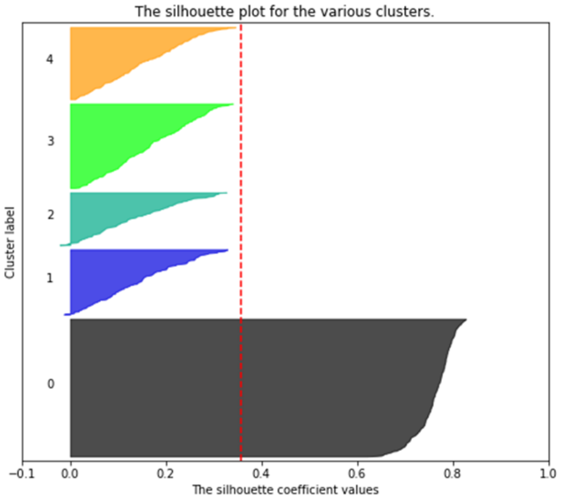


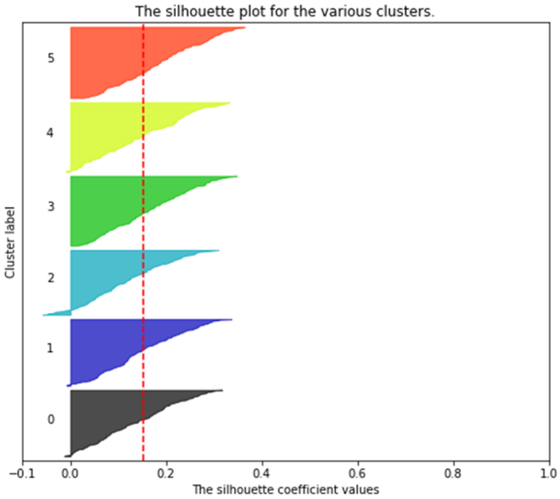


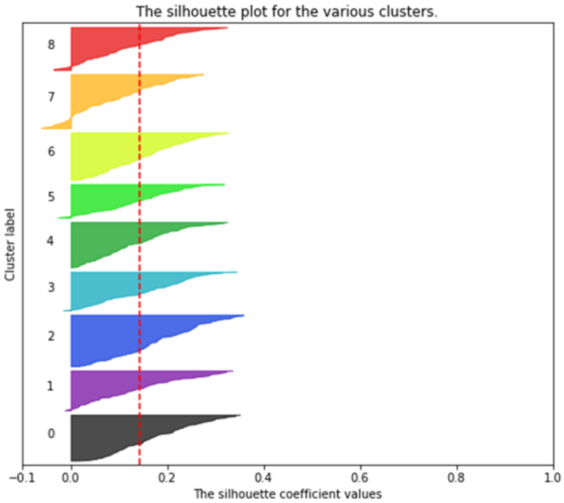
First Name................................ Last Name... Matr. No...

3.6 After pre-processing your data, you compute four different cluster models. Because you cannot decide which one fits best, you perform a silhouette analysis which results in the following four plots:









a) Please briefly **explain:**

i. the **idea behind this method** (3P)

ii. **how many clusters** the **best clustering model** has (1P)

iii. **why** it is the **best model** (1P)

b) Please also **provide** a **rough estimation** of the **silhouette coefficient** for each of the **four models**. (1P)