Mock exam **Introduction to Data Science and Artificial Intelligence - Data Science Part**, Department of Data Science and Knowledge Engineering

Read each question thoroughly and motivate all of your answers with arguments and/or computations.

This is still a closed book exam. The use of a simple calculator is allowed. The use of graphical calculators, phones, watches and other electrical devices is prohibited. You can earn up to 50 points in this part.

1. **[5 points]** Given the following situation:

   Anti-spam techniques are used to prevent email spam and may be based, among all, on word checking.

   a. Explain how words in an email 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.

2. **[10 points]** Let $z = 1 + i$ and $w = 2 - i$.

   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: $\bar{x} = i(x - 1)$

θ	0	$\dfrac{\pi}{6}$	$\dfrac{\pi}{4}$	$\dfrac{\pi}{3}$	$\dfrac{\pi}{2}$	π	$\dfrac{3\pi}{2}$	2π
$\sin\theta$	0	$\dfrac{1}{2}$	$\dfrac{1}{\sqrt{2}}$	$\dfrac{\sqrt{3}}{2}$	1	0	-1	0
$\cos\theta$	1	$\dfrac{\sqrt{3}}{2}$	$\dfrac{1}{\sqrt{2}}$	$\dfrac{1}{2}$	0	-1	0	1
$\tan\theta$	0	$\dfrac{1}{\sqrt{3}}$	1	$\sqrt{3}$	∞	0	$-\infty$	0

3. **[10 points]** Given the three time series:

   $$x = \{1, 0, 1, 1\}, \quad y = \{0, 2, 1, -1\}, \quad z = \{1, 0, 0, 4\}$$

   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?

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

4. **[10 points]** In a lab experiment four different measurements y_k of a certain reaction are obtained, depending on a variable x_k:

x_k	-1	0	1	2
y_k	0	1	1	3

   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.

5. [**5 points**] Consider the experiment of picking a fruit at random from a bag, noting the type of fruit, and putting it back. The bag has 8 apples, 6 oranges, 3 bananas, and 5 pears.

    a. Complete the following table summarizing the probability distribution of this experiment:

Fruit type	Probability	Cumulative probability

    b. Write a rule that would allow you to simulate this experiment with a computer.

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

6. [**10 points**] Given the following game:
Grab the Dollar

A dollar is placed on a table between two players, A and B, who must each decide whether to grab it or not. If both grab, then they are both fined 1 dollar; if only one grabs, he keeps the dollar.

    a. Construct the payoff table.

    b. Is this a zero-sum game?

    c. Is there a Nash equilibrium for this game?

Success!