

LOCATION: SWANSEA

LLEOLIAD: ABERTAWE

LEVEL 7

LEFEL 7

EXAMINATION PERIOD
JAN 2023

CYFNOD ARHOLIADAU
ION 2023

MODULE CODE: ACCA7026
ARTIFICIAL INTELLIGENCE

TIME: 3 HOURS INCLUDING UPLOAD TIME

ANSWER THREE QUESTIONS OUT OF FOUR

Equipment or materials required - Scientific Calculator

Adnoddau neu ddefnyddiau sydd eu hangen – cyfrifiannell wyddonol

QUESTION ONE

- 1a) Your team leader has put you in charge of developing a feature for an obstructive sleep disorder diagnosis application. This feature will classify someone as either having an elevated likelihood of obstructive sleep apnoea (OSA), or not. You have decided to use a Naive Bayes classifier for this problem and base the Naïve Bayes classifier decisions on age group, gender and BMI.

In order to demonstrate the method to your supervisor you plan to test the approach by manually classifying the likelihood of OSA for an individual with the following attributes:

- **BMI Group = 29.0;**
- **Age Group = Lower;**
- **Gender = Male**

Table 1, below, holds model training data that has been gathered previously from a variety of subjects.

Using the Naïve Bayes method, determine which classification, OSA=yes or OSA=no, is more likely for the individual, given the values above and the training data in table 1.

Sample No.	Age Group	Gender	BMI	OSA
1	Upper	Female	35	Yes
2	Lower	Female	30	Yes
3	Upper	Male	29	Yes
4	Upper	Male	28	Yes
5	Lower	Female	31	Yes
6	Upper	Male	26	Yes
7	Lower	Female	20	No
8	Lower	Female	24	No
9	Upper	Male	25	Yes
10	Lower	Female	20	No
11	Upper	Male	30	No
12	Upper	Male	36	Yes
13	Lower	Female	26	No
14	Lower	Male	22	Yes

Table 1. OSA Subject Data.

(Question continued overleaf).

QUESTION ONE (Continued)

Note, you will need to first calculate the probable BMI using the Gaussian distribution, $g(x)$ where,

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{equation 1,}$$

The sample standard deviation, $\sigma^2 = \sqrt{\frac{\sum(x-\mu)^2}{n}}$ equation 2,

μ is the mean, and n is the number of samples.

[20 Marks]

- 1b) The Naïve Bayes method typically results in very small probability values which can lead to numerical errors. Suggest a method that can overcome this problem and explain how it achieves this.

[5 Marks]

(P.T.O)

QUESTION TWO

- 2a) Describe the operation of Dijkstra's algorithm and provide an analysis of this algorithm highlighting its run time and space complexity. Discuss the purpose of any data structures used in its implementation.

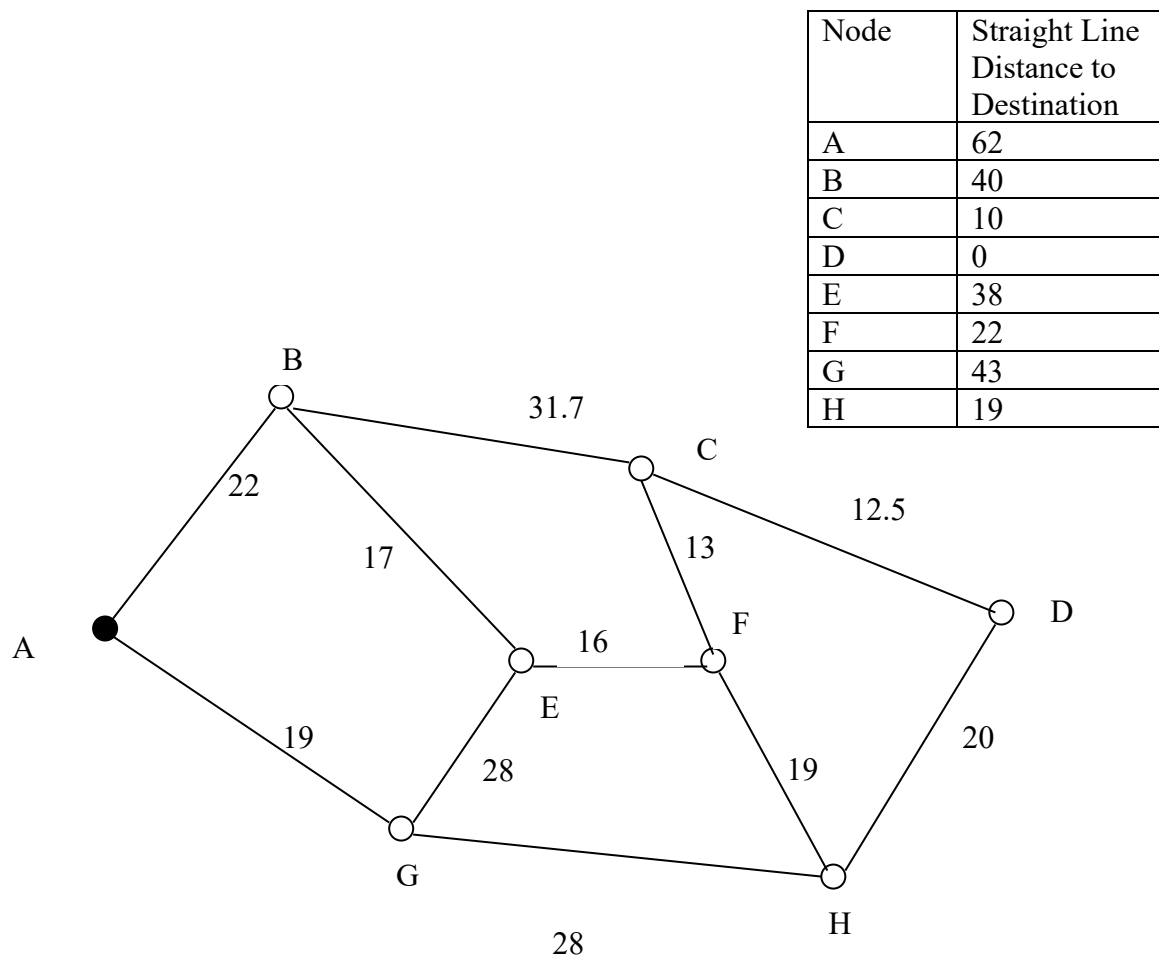
[8 Marks]

- 2b) Explain how Dijkstra's algorithm can be adapted to become the A* shortest path algorithm and identify any constraints required to ensure optimality of this algorithm.

[4 Marks]

- 2c) The following graph shows the travelled distance between several nodes and also includes a table showing the straight-line distance from the current node to the destination. Using this graph as an example and starting at the node labelled A with destination node D, illustrate the working of the A* algorithm, show the development of the path at each successive stage of the search, detailing the contents of the algorithm's open and closed lists.

[13 Marks]



P.T.O.

QUESTION THREE

- 3 (a) Explain what is meant by a “local search” heuristic and illustrate the key ideas involved by using as an example the hill climbing local search method. In your answer critically evaluate hill climbing and suggest approaches that could overcome any shortcomings that you have identified in your discussion.

[7 Marks]

- 3(b) Give a pseudo-code description of a standard Genetic Algorithm that uses the genetic operators of crossover and mutation, and identify key variables in the algorithm, explaining their function and purpose. Assume the genotype is an ordered fixed length string.

[4 Marks]

- 3(c) A fixed length chromosome (labelled, x) is composed of 8 genes each representing a variable a,b,c,d,e,f,g and h. The chromosome is represented by the string of variables, x=abcdefgh. Each variable can be a digit between 0 and 9. Show the result of performing a crossover operation on Parent 1 and Parent 2 (shown below) using the cut points indicated. Assume two children are produced.

Chosen cut points
 ↓ ↓
Parent 1: **25795412**
Parent 2: **35485232**

[6 Marks]

- 3(d) If the fitness function is given by:-
$$f(x) = (a + c) - (b + e - d) - g + f + h,$$
then evaluate the fitness of each of the two offspring in part c) of this question stating any assumptions you make.

What would be the value of the chromosome’s gene alleles if the chromosome had maximum fitness? Explain your reasoning.

[8 Marks]

P.T.O.

QUESTION FOUR

4a) Identify and briefly describe the procedures involved in the operation of a fuzzy logic expert system.

[4 marks]

4b) State two defuzzification techniques and indicate which method you would choose for a real time application giving arguments for your choice.

[3 Marks]

4c) Your supervisor has asked you to develop an AI application that screens for Chronic Obstructive Pulmonary Disease (COPD) based upon the following symptoms. These are:-

- Incidence of breathlessness separated into three ranges (infrequent, moderate, frequent).
- Wheeziness (mild and severe).

The following fuzzy logic rules make up part of the AI system and are used to demonstrate the approach to your supervisor. The confidence of a positive diagnosis for each of the rules is classified as low, medium or high.

Rule 1

IF breathlessness is frequent
OR wheeziness is severe
THEN confidence is high

Rule 2

IF breathlessness is moderate
AND wheeziness is mild
THEN confidence is medium

Rule 3

IF breathlessness is infrequent
THEN confidence is low

A subject has an incidence of breathlessness score of 11 units and a wheeziness index of 6 units.

Using these rules and the membership functions defined overleaf, find the level of confidence score for a positive diagnosis. Use the Centroid Defuzzification method in your answer.

(Q4 continued overleaf).

QUESTION FOUR (continued)

The degree of membership (DOM) functions for the linguistic variables are as follows:

For Incidence of breathlessness:

$$\mu_{breathless}^{Moderate}(x) = \begin{cases} 0 & x \leq 10 \\ \frac{x-10}{5} & 10 < x \leq 15 \\ 1 & 15 < x \leq 25 \\ \frac{30-x}{5} & 25 < x \leq 30 \end{cases}$$

$$\mu_{breathless}^{Infrequent}(x) = \begin{cases} 1 & x \leq 10 \\ \frac{15-x}{5} & 10 < x \leq 15 \\ 0 & x > 15 \end{cases}$$

$$\mu_{breathless}^{frequent}(x) = \begin{cases} 0 & x \leq 25 \\ \frac{x-25}{25} & 25 < x \leq 50 \\ 1 & x > 50 \end{cases}$$

And for Degree of Wheeziness:

$$\mu_{Wheeziness}^{Mild}(x) = \begin{cases} 1 & x \leq 5 \\ \frac{7-x}{2} & 5 < x \leq 7 \\ 1 & x > 7 \end{cases} \quad \mu_{Wheeziness}^{Severe}(x) = \begin{cases} 1 & x > 10 \\ \frac{x-5}{5} & 5 < x \leq 10 \\ 0 & x \leq 5 \end{cases}$$

(Q4 continued overleaf).

QUESTION FOUR (continued)

The DOM of the output “positive diagnosis confidence” fuzzy sub-sets is given by the following membership functions:

$$\mu_{confidence}^{Low}(x) = \begin{cases} 1 & x \leq 0.2 \\ \frac{0.4 - x}{0.2} & 0.2 < x \leq 0.4 \\ 0 & x > 0.4 \end{cases} \quad \mu_{confidence}^{Medium}(x) = \begin{cases} 0 & x \leq 0.2; x > 0.8 \\ \frac{x - 0.2}{0.2} & 0.2 < x \leq 0.4 \\ 1 & 0.4 < x \leq 0.6 \\ \frac{0.8 - x}{0.2} & 0.6 < x \leq 0.8 \end{cases}$$
$$\mu_{confidence}^{High}(x) = \begin{cases} 0 & x \leq 0.6 \\ \frac{x - 0.6}{0.2} & 0.6 < x \leq 0.8 \\ 1 & x > 0.8 \end{cases}$$

[18 Marks]

Diwedd/End