ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING EXAM JANUARY 11TH 2022

ACCA6011 – BSc Software Engineering

Siobhan Stephens - 1800709
University of Wales Trinity Saint David

Contents

Question 1 2
1A) [20 Marks] 2
1B) [5 Marks] 3
Question 3 4
3A) [5 Marks] 4
3B) [6 Marks] 5
3C) [6 Marks] 6
3D) [8 Marks] 7
Question 4 8
4A) [7 Marks] 8
4B) [6 Marks] 9
4C) [2 Marks] 10
4D) [10 Marks] 11

Question 1

1A) [20 Marks]

Question 1 a)

Siobhan Stephens 18007099

BMI Group = Medium
Age Group = Lower
Gender = Male

$P(OSA = yes) = 9/14$
$P(OSA = No) = 5/14$

$P(\text{BMI Group} = \text{medium} \mid OSA = \text{Yes}) = 4/9$

$P(\text{Age Group} = \text{Lower} \mid OSA = \text{Yes}) = 6/9$

$P(\text{Gender} = \text{Male} \mid OSA = \text{Yes} = 6/9$

$\overset{(BMI)}{(4/9)} \times \overset{(Age)}{(6/9)} \times \overset{(Gender)}{(6/9)} \times \overset{(OSA\ Yes)}{(9/14)} = 0.126984$

$\uparrow P(OSA\ Yes)$

$P(\text{BMI Group} = \text{Medium} \mid OSA = \text{No}) = 1/5$

$P(\text{Age Group} = \text{Lower} \mid OSA = \text{No}) = 1/5$

$P(\text{Gender} = \text{Male} \mid OSA = \text{No}) = 1/5$

$\overset{BMI}{(1/5)} \times \overset{Age}{(1/5)} \times \overset{Gender}{(1/5)} \times \overset{OSA\ No}{(5/14)} = 0.002857$

$\leftarrow OSA\ No$

$P(OSA\ Yes \mid x) > P(OSA\ No \mid x)$

$0.126984 > 0.002857$, so $OSA = Yes$

Probability is greater for $OSA = Yes$ given BMI Group = Medium, Age Group = Lower and Gender = Male as 0.126984 is greater than 0.002857.

Figure 1. Ans for Q 1A

1B) [5 Marks]

Q) 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.

Ans)

The formula for naïve bayes is as follows:

$$(A|B_1, B_2, \dots, B_n) = \frac{P(A)P(B_1|A)P(B_2|A) \dots P(B_n|A)}{\sum_{i=1}^{m} P(B_1, B_2, \dots, B_n|A_i)P(A_i)}$$

If any one of the numerator terms is zero, then the result will be zero and therefore we cannot rely on conditional independence.

If there are missing samples a way to overcome this issue is to set any zero values to be a finite but small constant such as 1. This process is known as additive or laplace smoothing.

Question 3

3A) [5 Marks]

Q) 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.

Ans)

```
determine objective function (OF)
assign number of generations to 0 (t = 0)
randomly create individuals in initial population P(t)
evaluate individuals in population P(t) using OF

while termination criterion is not satisfied do:
      t = t + 1
      select individuals to population P(t) from P(t - 1)
      crossover(P(t))
      mutation(P(t))

      evaluate new individuals in population P(t) using OF
end while
return the best individual found during the evolution
```

Key variables:

P – Population

t = Generations

P(t) – Individuals genotypes represented by fixed length ordered string

3B) [6 Marks]

Q) Critically evaluate the effectiveness of common terminating conditions used by genetic algorithms.

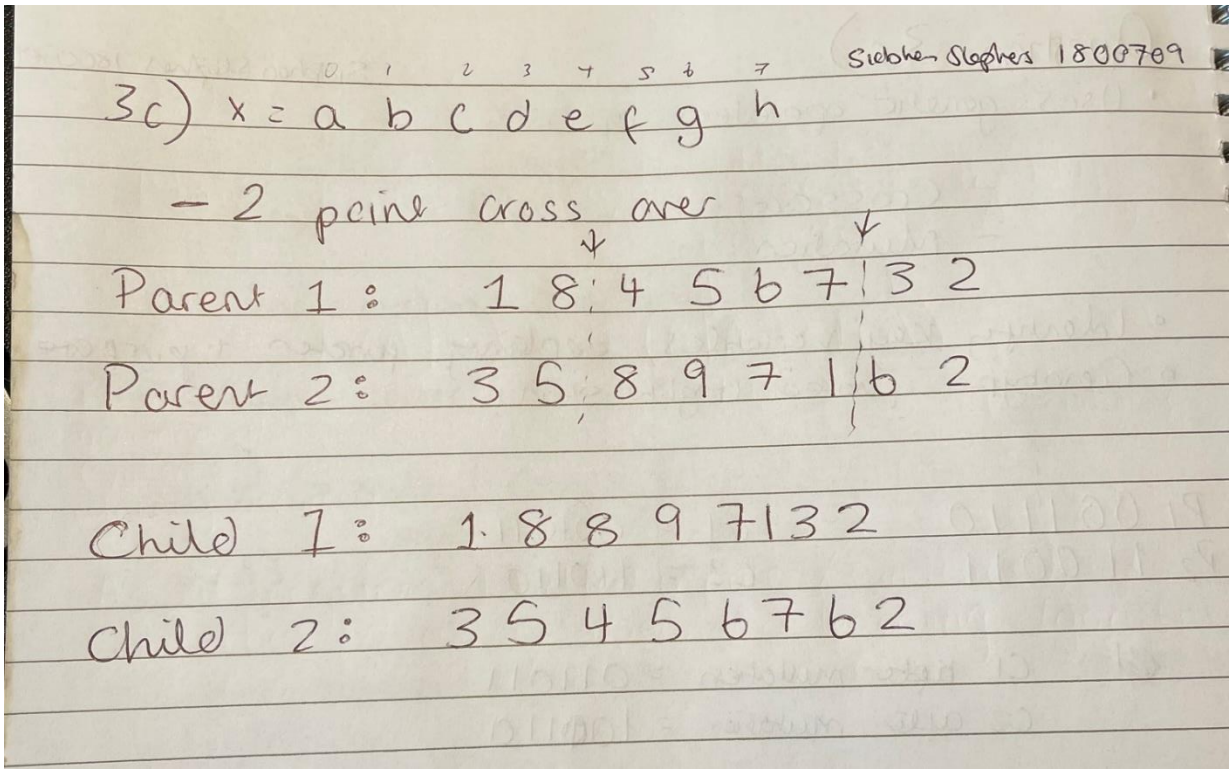
Ans)

Common terminating conditions used by Genetic Algorithms are:

- A solution is found that satisfies minimum criteria
- Fixed number of generations reached
- Allocated budget (computation time/money) reached - In this case, the algorithm runs for a predetermined execution time and returns the final solution.
- The highest-ranking solution's fitness is reaching or has reached a plateau such that successive iterations no longer produce better results
- Manual inspection
- Combinations of the above.

Direct termination criteria such as fixed number of generations or allocated budget reached are easy to implement but it can be difficult to determine a fixed value for these criteria whilst considering an optimal output. A major drawback for this type of termination is that it does not terminate the algorithm if it converges to a sub-optimal solution and also may not reach an optimal solution at all if not provided with enough iterations. In order to improve upon this a bound-condition such as highest ranking fitness has reached a plateau should be incorporated alongside a direct termination clause

3C) [6 Marks]



Siobhan Stephens 1800709

3c) x = a b c d e f g h

– 2 point cross over

Parent 1: 1 8 4 5 6 7 3 2

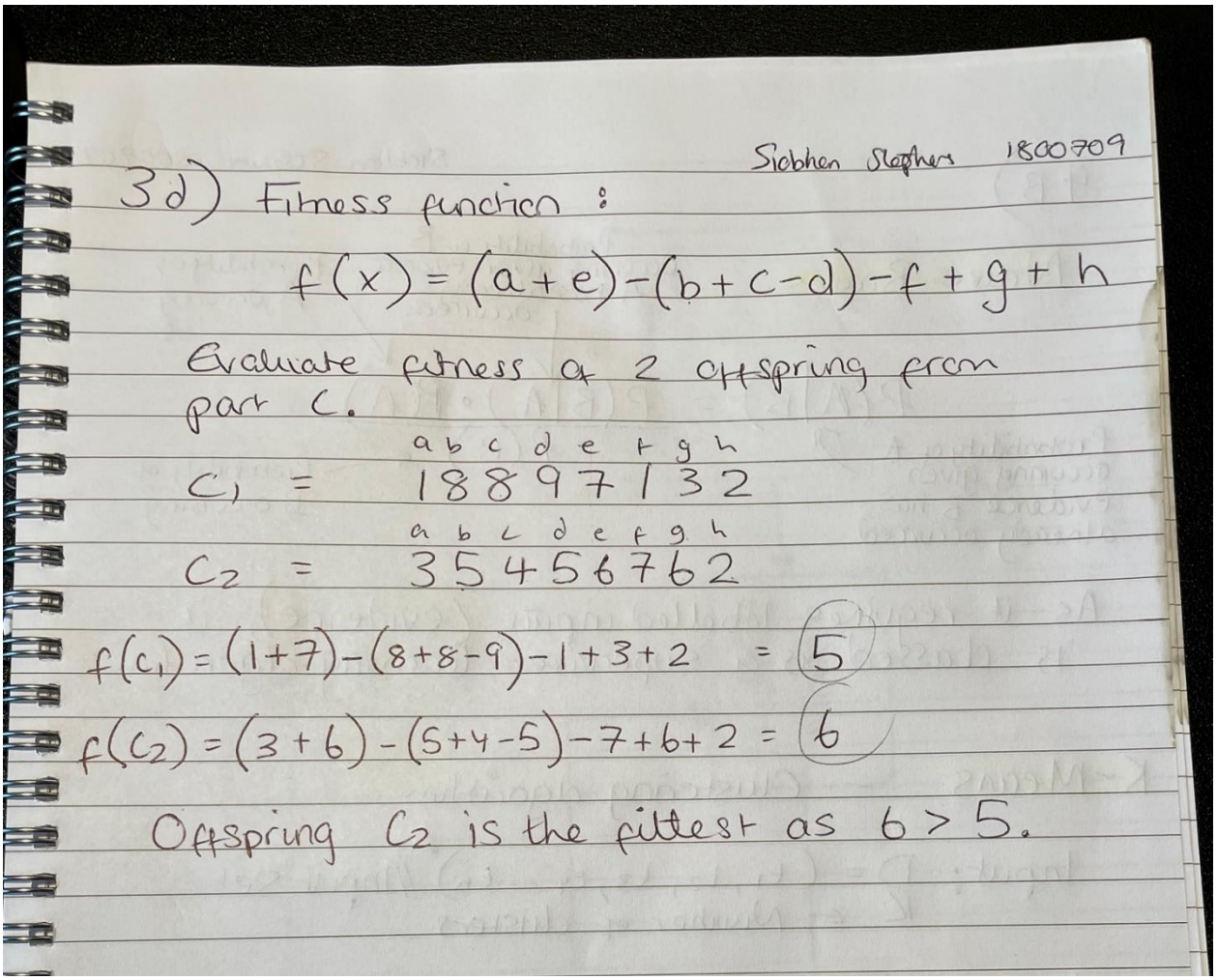
Parent 2: 3 5 8 9 7 1 6 2

Child 1: 1 8 8 9 7 1 3 2

Child 2: 3 5 4 5 6 7 6 2

Figure 2. Ans for Q 3C

3D) [8 Marks]



Siobhan Stephens 1800709

3d) Fitness function :

$$f(x) = (a+e)-(b+c-d)-f+g+h$$

Evaluate fitness of 2 offspring from part C.

a b c d e f g h

C_1 = 1 8 8 9 7 1 3 2

a b c d e f g h

C_2 = 3 5 4 5 6 7 6 2

$f(c_1) = (1+7)-(8+8-9)-1+3+2 = 5$

$f(c_2) = (3+6)-(5+4-5)-7+6+2 = 6$

Offspring C_2 is the fittest as $6 > 5$.

Figure 3. Ans for Q 3D

Question 4

4A) [7 Marks]

Within the context of machine learning, explain the terms supervised and unsupervised learning and, in each case, outline the processes involved in creating a predictive model from a given data set.

Supervised learning is defined by its use of labelled data, whereby the datasets are used to train (supervise) algorithms such as classification or linear regression into classifying data or predicting outcomes most accurately. It is possible to measure the accuracy of supervised learning models using the labelled inputs and outputs.

Unsupervised learning however uses un-labelled data and machine learning algorithms to analyze and cluster data sets to discover hidden patterns, including clustering and association algorithms.

Therefore, the main difference between supervised and unsupervised learning is their use of labelled or un-labelled data, and the differing goals of each type.

The processes to create a predictive model for both an unsupervised and a supervised learning algorithm is shown in the diagram below, (note the key (highlighted) outlining the difference of labelled data):

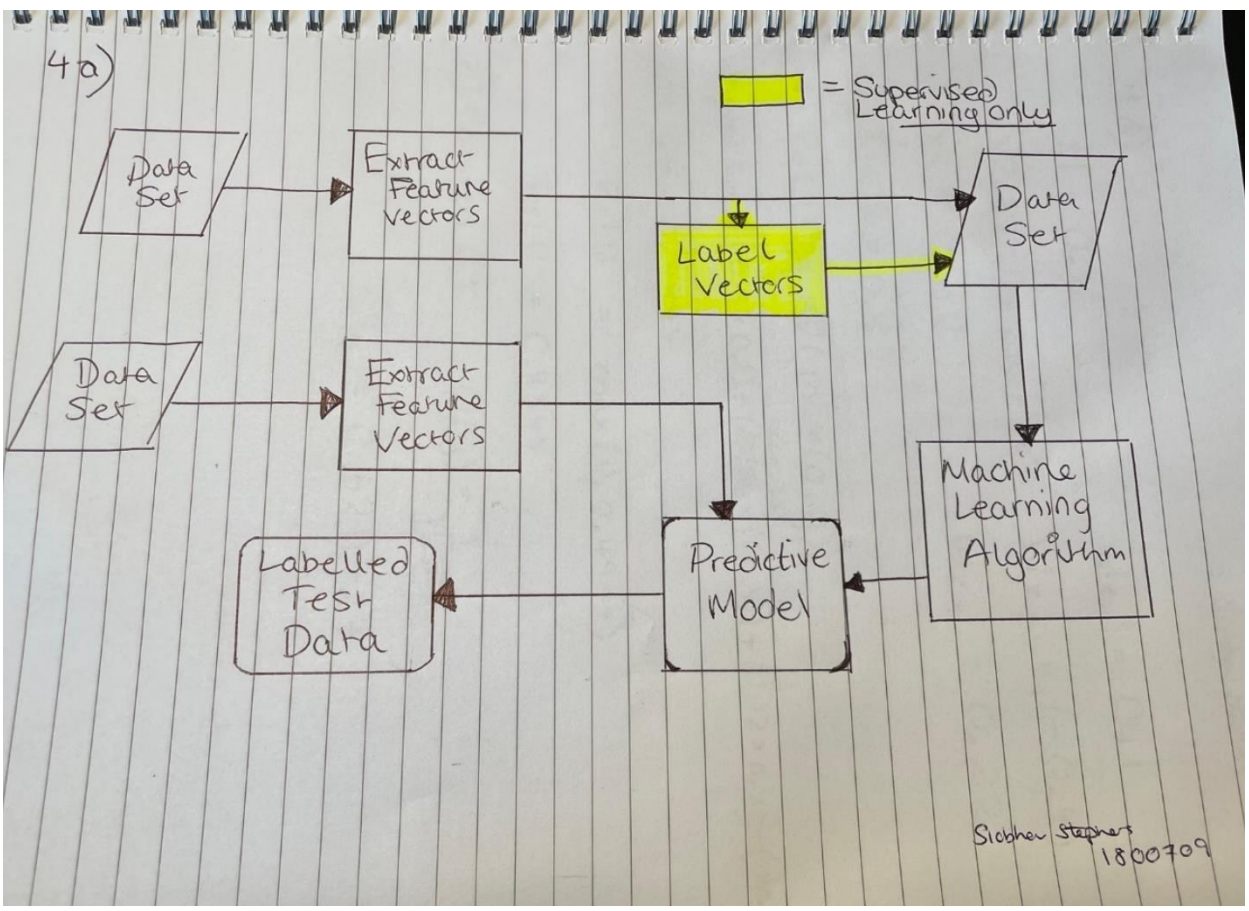


Figure 4. 4A Diagram of Supervised/unsupervised learning process

4B) [6 Marks]

Explain, with supporting arguments, whether the Naïve Bayes ML algorithm and the k-means algorithm are examples of supervised or unsupervised machine learning algorithms.

Siobhan Stephens 1800709

4 B)

Naïve Bayes

Probability of B occuring given evidence A has already occurred.

Probability of A occuring

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Probability of A occuring given evidence B has already occurred

Probability of B occuring

As it requires labelled inputs (evidence), it is classed as a supervised learning algorithm.

K-Means – Clustering algorithm

Input: $D = (t_1, t_2, t_3, t_4 \dots t_n)$ // Input set
K ← Number of clusters

Unlabelled data is inputted and is clustered according to its similarity or dissimilarity to centroids.

Therefore, it is an example of an un-supervised algorithm.

Figure 5. Ans for Q 4B

4C) [2 Marks]

Q) With reference to the perceptron neuron shown in figure 4a below, explain the purpose of the synaptic weights wk1, wk2 and wk3.

Ans)

The weights Wk1, Wk2, and Wk3 above represent the strength of the connection between units for each input. The greater the magnitude of the weight, the greater the influence it has on the output.

The use of weights lowers the importance of the input value, for example, weights near zero such as Wk_1 mean that changing the input x1 will not change the output.

Negative weights such as Wk_2 mean that increasing input x2 will decrease the output, representing an inhibitory input.

Positive weights such as Wk_3 will increase the influence of input x3 and therefore increase the output, representing an excitatory input.

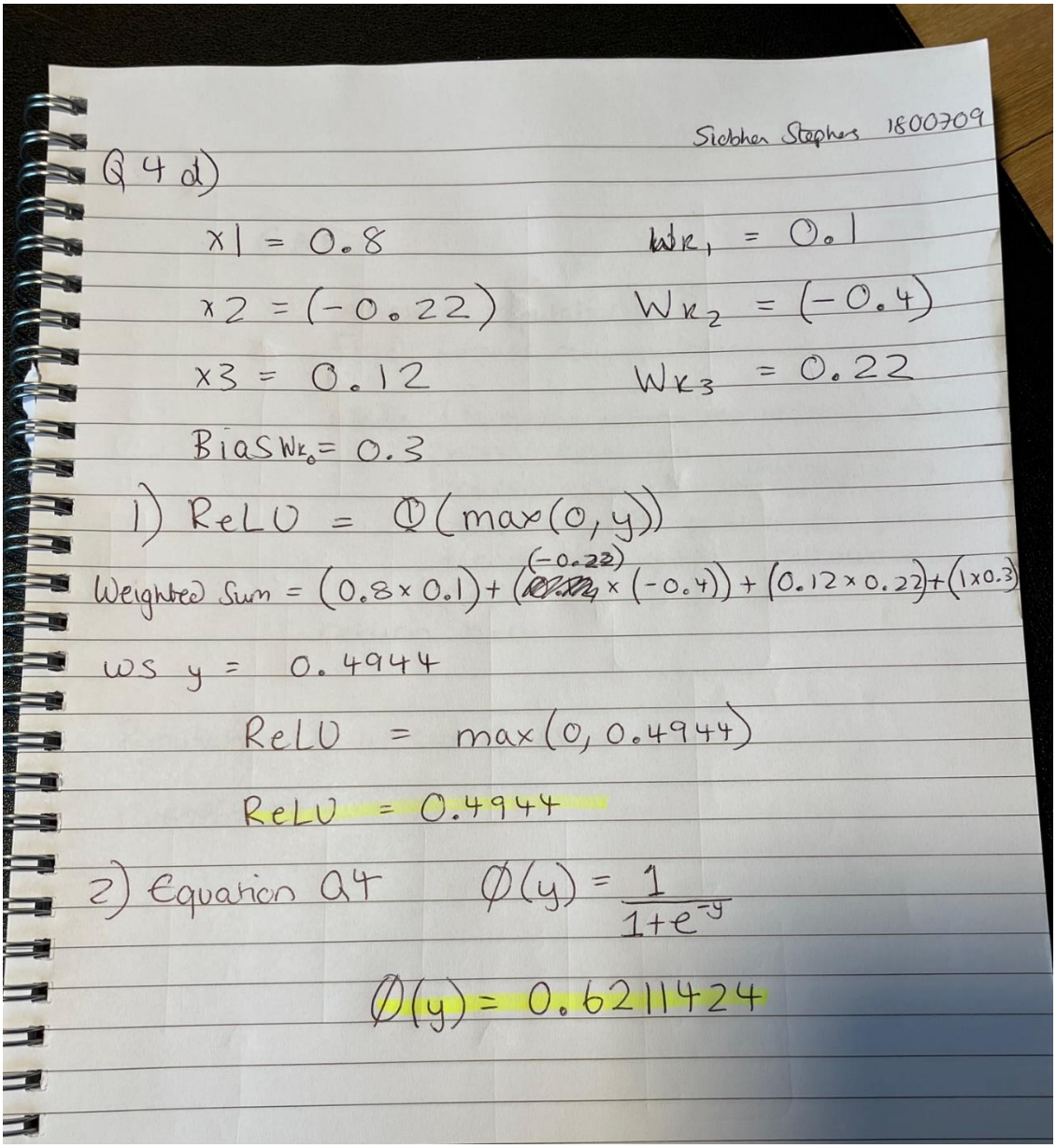
In short, each weight decides how much influence each input will have on the output.

4D) [10 Marks]

Q) Calculate the output of the activation function of the perceptron in figure 4a when x1=0.8 , x2=-0.22, x3=0.12 and the activation function is a Rectified Linear Unit (ReLU) given by Φ(max(0,y)). Repeat this calculation using the logistic activation function, given by equation Q4. Explain why the use of a ReLu is sometimes preferred to the logistic function.

Ans)

ReLU is sometimes preferred to the logistic function as it does not suffer from the vanishing gradients problem that arises when a>0. It is also quick to compute, and does not saturate for positive values as sparsity arises when a≤0.



Siobhan Stephens 1800709

Q 4 d)

$x1 = 0.8$ $\quad$ $Wk_1 = 0.1$

$x2 = (-0.22)$ $\quad$ $Wk_2 = (-0.4)$

$x3 = 0.12$ $\quad$ $Wk_3 = 0.22$

$Bias\ Wk_0 = 0.3$

1) $ReLU = \Phi(\max(0, y))$

Weighted Sum $= (0.8 \times 0.1) + ((-0.22) \times (-0.4)) + (0.12 \times 0.22) + (1 \times 0.3)$

$ws\ y = 0.4944$

$ReLU = \max(0, 0.4944)$

$ReLU = 0.4944$

2) Equation Q4 $\quad \phi(y) = \frac{1}{1+e^{-y}}$

$\phi(y) = 0.6211424$

Figure 6. Ans for Q 4D