This question paper consists of 7 printed pages, each of which is identified by the Code Number COMP5400M.

This is an open book examination. Any written or printed material is permitted.

School of Computing

May 2018

COMP5400M

Bio-inspired Computing

Answer ALL THREE questions

Time allowed: 2 hours

Question 1

(a) In Figure 1 you see a data set that is linearly separable. A perceptron is in the process of being trained. The current decision line implemented by the perceptron is shown in the figure. The weight at the x input is known: $w_x = 1$. Find the weight of the y input and the threshold that implement the decision line. **[3 marks]**

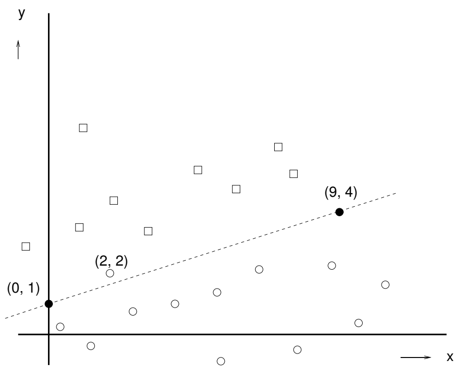

Figure 1: Linearly separable data with a misclassified point.

(b) In Figure 1, point $P = (2, 2)$ is being misclassified. Explain why the perceptron algorithm will change weights upon hitting this point and calculate the new weights. If you did not find weight and threshold values in the previous question, assume $w_x = w_y = \theta = 1$, where θ is the threshold. **[4 marks]**

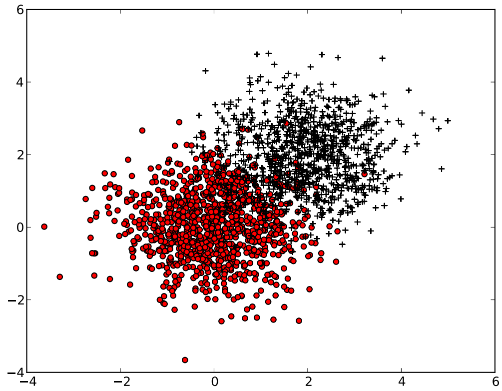

Figure 2: Two Gaussian distributions, one centred around (0,0), the other around (2,2).

In Figure 2 data points of two Gaussian distributions are shown. A Gaussian distribution is spherically symmetrical around its centre; the points are likely to cluster around the centre and less likely to be far away from the centre. One of the distributions has centre $(0, 0)$, the other $(2, 2)$.

(c) Given the data shown in Figure 2, the standard version of the perceptron algorithm would not converge. Explain why, how the perceptron algorithm can be modified to produce an optimal decision line, and describe in what sense it is optimal.

[4 marks]

(d) The data classification problem of Figure 2 is suitable for evaluation and debugging: you can estimate the optimal decision boundary. Find the boundary and present the result as perceptron weights and threshold.

[3 marks]

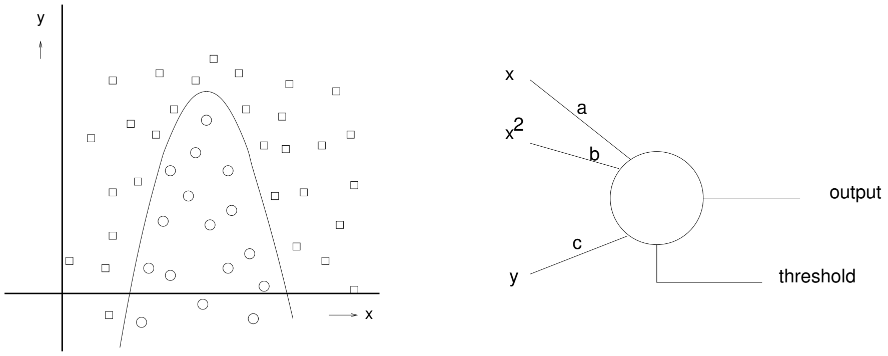

Figure 3: Data set for which the decision boundary is a parabola.

(e) In Figure 3 (left) a data set is shown which is obviously non linearly separable. If you know that the decision boundary always is a parabola, you can introduce a perceptron such as the one shown in Figure 3 right. This perceptron bases its output on:

$$o = f(ax + bx^2 + cy - \text{threshold})$$

Explain why such a perceptron would be able to learn to classify any data set with a parabolic decision boundary. **[2 marks]**

(f) Find a learning rule for such a perceptron based on steepest gradient descent. Assume that f is the sigmoidal function given by:

$$f(x) = \frac{1}{1 + e^{-x}}$$

[4 marks]

[question 1 total: 20 marks]

CONTINUED ON FACING PAGE

Question 2

Deneubourg *et al.* have proposed a model to account for corpse clustering and larval sorting in ant nests. They define a 'pick up' probability for an unladen agent as follows:

$$p_p = \left(\frac{k_1}{k_1 + f}\right)^2, \tag{1}$$

and also a probability for a randomly moving agent to deposit an item:

$$p_d = \left(\frac{f}{k_2 + f}\right)^2, \tag{2}$$

Here k_1 and k_2 are two dimensionless positive constants and f is the perceived fraction (so $0 \leq f \leq 1$) of items in the neighbourhood of the agent.

(a) Explain how clustering of items can emerge when agents capable of picking up, carrying and dropping off these items are in a confined space and use the probabilities defined above. Your answer should include a justification for the functional forms of p_d and p_p. You may use numerical examples to illustrate the principles. Explain why the values of k_1 and k_2 should not be chosen close to or larger than 1.

[6 marks]

(b) Explain the concept of stigmergy. Discuss the role of stigmergy in the clustering algorithm.

[4 marks]

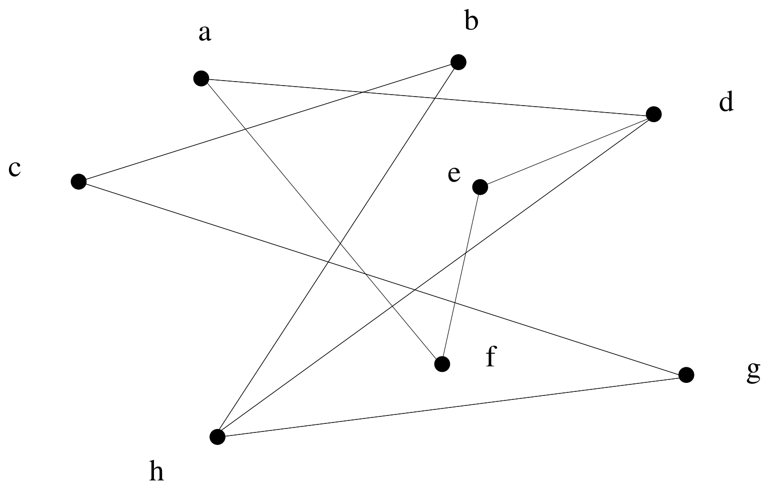

Figure 4: Example of a graph that could be partitioned by the KLS algorithm.

In a graph G, the set of vertices is denoted by $\{v_i\}, i = 1, \cdots, N$, where N is the number of vertices in G. Let $\rho(v_i)$ be the set of vertices neighbouring v_i, including v_i itself. Let $\mid A \mid$ be the number of elements in set A.

The distance $d(v_i, v_j)$ between vertices is defined in terms of their sets of neighbouring vertices as follows:

$$d(v_i, v_j) = \frac{\mid D(\rho(v_i), \rho(v_j)) \mid}{\mid \rho(v_i) \mid + \mid \rho(v_j) \mid} \tag{3}$$

The function $D(A, B)$ is defined by:

$$D(A, B) = A \cup B - A \cap B$$

(c) Explain, in words, how the ant-inspired graph partitioning algorithm of Kuntz, Layzell and Snyers (KLS) works, in particular explain the role that the distance defined in Equation 3 plays. **[4 marks]**

(d) Assume the KLS algorithm works correctly in the case of Figure 4. Draw a possibly resulting graph. Show by calculating the vertex distance between two vertices that are far apart and the distance between two vertices that are close together that your graph is consistent with those distances.

[6 marks]

[question 2 total: 20 marks]

neuron	1	2	3	4	5
1	0	0	$-\frac{2}{5}$	$\frac{2}{5}$	0
2	0	0	0	0	$\frac{2}{5}$
3	$-\frac{2}{5}$	0	0	$-\frac{2}{5}$	0
4	$\frac{2}{5}$	0	$-\frac{2}{5}$	0	0
5	0	$\frac{2}{5}$	0	0	0

Table 1: Weight matrix of a Hopfield network.

Question 3

A 5-node Hopfield network has been trained with two patterns. The resulting weight matrix is given in Table 1.

(a) Assume the values of the network's neurons are given by $\vec{x} = (x_1, x_2, x_3, x_4, x_5)$. Give Hebb's law.

[2 marks]

(b) Consider the pattern $(1, 1, -1, 1, 1)$. Show by explicit calculation that this pattern is stable.

[3 marks]

(c) The pattern in the previous question was actually one of two training patterns. Which other pattern has contributed to the weight matrix of Table 1? Explain how you identified this pattern.

[4 marks]

(d) Consider an N-node Hopfield network. Give the definition of the energy of a training pattern, given a weight matrix w_{ij}. Show that when a Hopfield network is trained with a single training pattern, that pattern is an energy minimum, and that the pattern is a steady state.

[5 marks]

(e) Explain Hopfield and Tank's approach to solve the Travelling Salesman Problem. Your explanation should contain the following: the state space of their Hopfield network; the representation of a valid tour; and a motivation for their choice of the energy function.

[6 marks]

[question 3 total: 20 marks]

[grand total: 60 marks]