Family Name: .....

Other Names: .....

Signature: .....

Student Number: .....

THE UNIVERSITY OF NEW SOUTH WALES
SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

Sample Examination

COMP 9444 Neural Networks and Deep Learning

Exam Duration: 2 hours
Reading Time: 10 minutes
This paper has 12 pages including this cover page.
Students should answer ALL Questions (1 to 11)

Authorised materials:

- Approved Calculator.

Instructions to Students:

- This paper counts for 60% of your final grade.
- Write answers in ink, in the script books provided.
- Hand in question paper and script books when you are finished.
- All questions may be attempted.
- Start your answer to each question on a new page.

Question 1.

X	Y	OUTPUT
0	1	1
2	1	0
2	2	1

(a) (2 marks) Construct by hand a Perceptron network which correctly classifies the above three data items. Your network should have two inputs (X,Y) and one output. Make sure the connections and weights of your network are clearly specified.

(b) (1 mark) If a perceptron were trained on the above data using the perceptron learning algorithm, would it be guaranteed to find a solution in finite time?

Question 2.

(a) (2 marks) Briefly explain the difference between Perceptron learning and backpropagation.

(b) (2 marks) Briefly explain the difference between batch learning and online learning.

(c) (3 marks) What would happen if the transfer functions at the hidden layers in a multi-layer perceptron would be omitted; i.e. if the activation would simply be the weighted sum of the activations at the previous layer? Explain why this (simpler) activation scheme is not normally used in MLPs although it would simplify and accelerate the calculations for the backpropagation algorithm?

Question 3.

(4 marks) Consider a simple recurrent (Elman) network with 2 inputs, 2 hidden units and 2 outputs, using tanh activation at the hidden units and sigmoid at the outputs. Suppose this network is trained to predict the formal language $a^n b^n$ and that the training is successful.

Draw a diagram showing:

(a) the trajectory that might be traversed in hidden unit space as the network processes the string $a^5 b^5$

(b) the line that divides the hidden unit space into the region for which a is predicted, and the region for which b is predicted.

Question 4.

(a) (3 marks) Sketch the following activation functions, and write their formulas:

(i) sigmoid
(ii) tanh
(iii) ReLU

(b) (2 marks) Write the formula for Bayes' Rule, in terms of a cause A and an effect B.

(c) (3 marks) In the context of supervised learning, explain the difference between maximum likelihood estimation and Bayesian inference.

(d) (3 marks) Briefly explain the concept of Data Augmentation, and how it has been used in a neural network application of your choosing.

Question 5.

(4 marks) Consider a convolutional neural network which takes as input a 26-by-32 color image (i.e. with three channels R, G, B). The first convolutional layer has 42 filters that are 4-by-4, with stride 3 and zero-padding of width 1.

Compute the number of:

(a) weights for each neuron in this layer (including bias)

(b) neurons in this layer

(c) independent parameters in this layer

Note: if you do not have a calculator, it is sufficient to write an arithmetic expression in each case.

Question 6.

(a) (3 marks) Assume you are given a set of pre-trained word embeddings from word2vec or GloVe. Briefly explain how you would use these word embeddings to answer questions of the form:

A is to B as C is to X?

(where A, B, C are given words and you have to choose between four other words X_1, X_2, X_3, X_4 to replace X).

(b) (5 marks) The equation for Hierarchical Softmax is this:

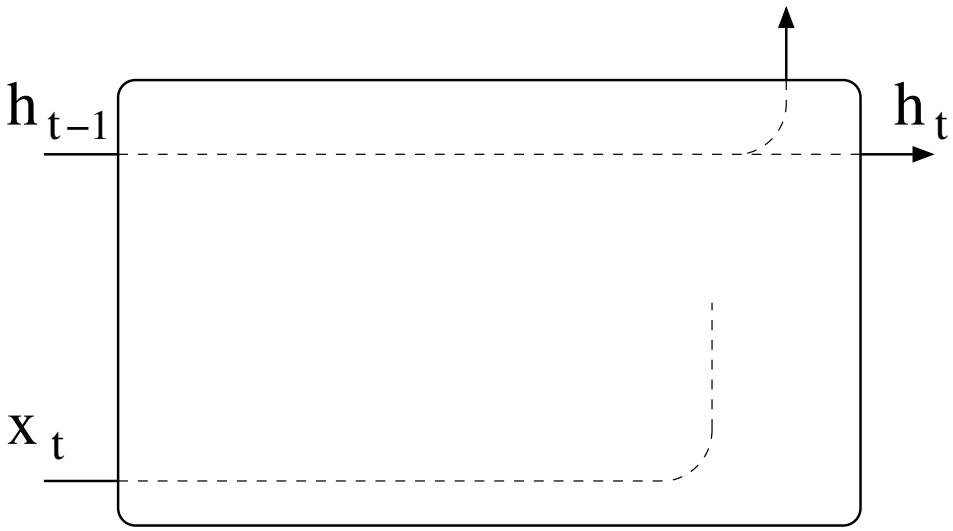
$$\text{prob}(w = w_t) = \prod_{j=1}^{L(w)-1} \sigma([\![n(w, j+1) = \text{child}(n(w, j))]\!]\mathbf{v}'^{\mathrm{T}}_{n(w,j)} \mathbf{h})$$

Identify the following symbols in the above equation:

$L(w)$
$n(w, j)$
$\mathbf{v}_{n(w,j)}$
$\mathbf{h}$
$\sigma()$
$[\![n(w, j+1) = \text{child}(n(w, j))]\!]$

Question 7.

(5 marks) Copy this outline into your script book, and use it to complete a diagram of the inner workings of a GRU (Gated Recurrent Unit).



Your diagram should include:

(a) the paths along which data travels through the system,

(b) the places where matrix multiplication occurs,

(c) the places where sigmoid and tanh are applied, and

(d) the junctions where values are combined by addition or multiplication.

Note: You do not have to write any equations.

Question 8.

(a) (4 marks) In the context of Deep Q Learning, explain the concept of Experience Replay, including why and how it is applied, and the relevant equation(s).

(b) (4 marks) Briefly describe how an Evolutionary Strategy or HillClimbing algorithm could be applied to one of the following domains:

(a) Backgammon

(b) Simulated Hockey

(c) Atari Pong

(d) MuJoCo humanoid walking

(Note: you do NOT have to describe the domain; you only need to describe the algorithm for updating the weights)

Question 9.

(5 marks) Consider a world with two states $S = \{S_1, S_2\}$ and two actions $A = \{a_1, a_2\}$, where the transition δ and reward r for each state and action are as follows:

$$\begin{array}{ll} \delta(S_1, a_1) = S_1, & r(S_1, a_1) = +1 \\ \delta(S_1, a_2) = S_2, & r(S_1, a_2) = -2 \\ \delta(S_2, a_1) = S_1, & r(S_2, a_1) = +4 \\ \delta(S_2, a_2) = S_2, & r(S_2, a_2) = +3 \end{array}$$

Assuming the discount factor is $\gamma = \frac{1}{2}$, compute:

(a) the optimal policy $\pi^* : S \mapsto A$

(b) the value function $V^* : S \mapsto R$

(c) the Q function $Q^* : S \times A \mapsto R$

Note: it may by helpful to know that, for $0 \leq x < 1$,

$$1 + x + x^2 + x^3 + ... \quad = \quad \frac{1}{1 - x}$$

Question 10.

(5 marks) Consider a Hopfield Network with the following weight matrix W:

$$\begin{pmatrix} 0 & +1 & -1 & +1 \\ +1 & 0 & +1 & -1 \\ -1 & +1 & 0 & -1 \\ +1 & -1 & -1 & 0 \end{pmatrix}$$

(a) If the network begins in state $[+1, -1, +1, +1]$, and assuming synchronous updates, compute the state at the next three timesteps.

(b) Assuming asynchronous updates, give one possible sequence leading to a stable state.

Question 11.

(a) (2 marks) Write the Energy function for a Boltzmann Machine, with units $\{x_i\}_{1 \leq i \leq n}$

(b) (3 marks) The Variational Auto-Encoder is trained to maximize

$$\mathbf{E}_{z \sim q_\phi(z|x^{(i)})}[\log\, p_\theta(x^{(i)}|z)\,] \;\; - \;\; D_{\mathrm{KL}}\big(q_\phi(z|x^{(i)}) \| p(z)\big)$$

Briefly state what each of these two terms aims to achieve.

(c) (2 marks) Briefly describe two ways in which a GAN may fail to converge.

END OF EXAM