COMP9444 Neural Networks and Deep Learning

Sample Exam

Instructions

1. Time allowed: 10 minutes reading + 2 hours working
2. This Exam will count for 45% of your Final Mark
3. Marks for each Question are as indicated
4. **All** Questions May be attempted
5. Unless otherwise indicated, Numerical answers should be given correct to two decimal places (i.e. within 0.005 of the correct answer)
6. For Multiple Choice Questions, you should select the One Best Answer
7. No material may be taken into the examination room
8. **Do not log out until you have saved and submitted your answers**

Part A: Multiple Choice (Questions 1-15)

Part A: Question 1 (1 mark)

Which of these archictectures would have the best chance of learning long range dependencies?

A. ◯ Elman Network

B. ◯ Feedforward network with sliding window

C. ◯ Long Short Term Memory

D. ◯ Simple Recurrent Network

Note: Part A of this Sample Exam contains only One Multiple Choice Question,

but Part A of the real Final Exam will contain 15 Multiple Choice Questions.

Part B: Numerical (Questions 16-21)

Part B: Question 16 (3 marks)

Consider a perceptron whose output is given by $h(w_0 + w_1x_1 + w_2x_2)$ where x_1, x_2 are inputs and $h()$ is the Heaviside (step) function. The current weights are $w_0 = -0.5, w_1 = -1, w_2 = -2$.

Training Example	x_1	x_2	Class
a.	2	1	+
b.	–2	2	+
c.	–1	–1	–

Suppose the Perceptron Learning Algorithm is applied to the current weights, using a learning rate of $\eta = 1.0$. We consider only the first step, where **item (a)** is learned. The new values for w_0, w_1 and w_2 after learning **item (a) only** will be:

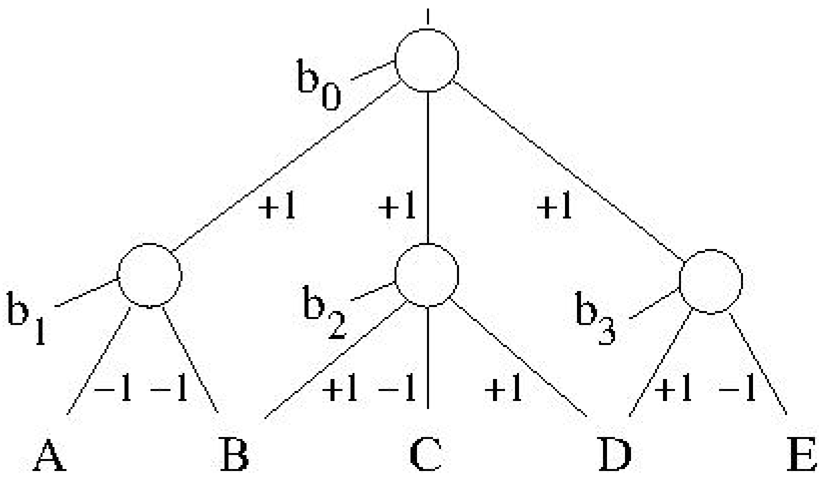
i. (1 mark) w_0 :

ii. (1 mark) w_1 :

iii. (1 mark) w_2 :

Part B: Question 17 (2 marks)

Consider the following multi-layer perceptron, using the threshold activation function, and assume that TRUE is represented by 1; FALSE by 0.



For which values of the biases b_0, b_1, b_2, b_3 would this network compute this logical function ?

$$(\neg A \vee \neg B) \wedge (B \vee \neg C \vee D) \wedge (D \vee \neg E)$$

i. ($\frac{1}{2}$ mark) b_0 :

ii. ($\frac{1}{2}$ mark) b_1 :

iii. ($\frac{1}{2}$ mark) b_2 :

iv. ($\frac{1}{2}$ mark) b_3 :

Part B: Question 18 (2 marks)

Only 30% of the population have been vaccinated against a certain disease. Among those who **are** vaccinated, only 1% of them have the disease. But, among those who are **not** vaccinated, 3% of them have the disease.

If a random person is found to have the disease, what is the probability that they have been vaccinated?

(You can give your answer either as a percentage, or as a number between 0 and 1)

Part B: Question 19 (3 marks)

Consider these two probability distributions on the same space $\Omega = \{A, B, C, D, E\}$

$$p = \langle \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16} \rangle, \; q = \langle \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{2}, \frac{1}{16} \rangle$$

Compute, to two decimal places:

i. (1 mark) the Entropy $H(p)$:

ii. (2 marks) the KL-Divergence $D_{KL}(p \parallel q)$:

Part B: Question 20 (3 marks)

Consider a neural network trained using **softmax** for a classification task with three classes $1, 2, 3$. Suppose a particular input is presented, producing outputs

$$z_1 = 2.5, z_2 = 1.5, z_3 = 0$$

If we assume that the correct class for this input is Class 2, and that Prob(2) is the softmax probability of the network choosing Class 2, compute the following (correct to two decimal places):

i. (1 mark) $d(\log \text{Prob}(2))/\, dz_1$:

ii. (1 mark) $d(\log \text{Prob}(2))/\, dz_2$:

iii. (1 mark) $d(\log \text{Prob}(2))/\, dz_3$:

Part B: Question 21 (4 marks)

Consider a convolutional neural network which takes as input a 65×77 color image (i.e. with three channels R, G, B). The first convolutional layer has 18 filters that are 5-by-5, with stride 3 and no zero-padding.

Compute the number of:

i. (1 mark) weights per filter in this layer (including bias) :

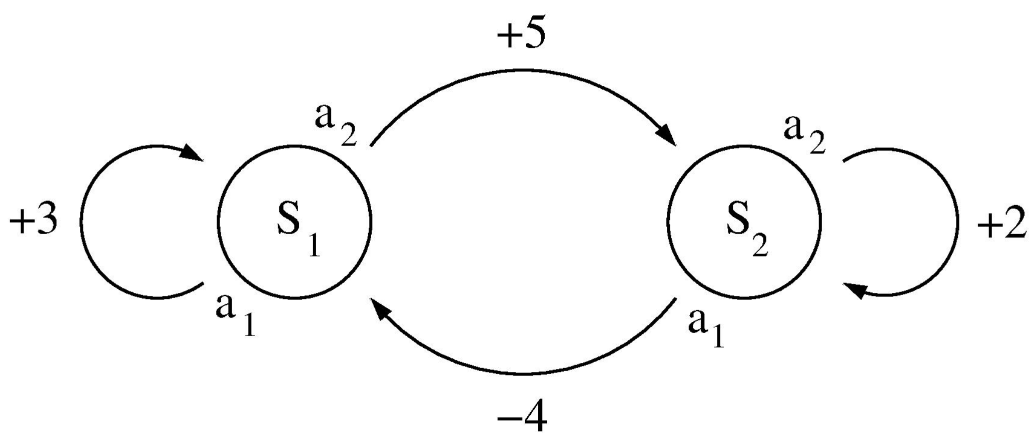
ii. (1 mark) neurons in this layer :

iii. (1 mark) connections into the neurons in this layer :

iv. (1 mark) independent parameters in this layer :

Part B: Question 22 (8 marks)

Consider an environment with two states $S = \{S_1, S_2\}$ and two actions $A = \{a_1, a_2\}$, where the (deterministic) transitions δ and reward R for each state and action are as follows:



Assuming a discount factor of $\gamma = 0.7$, compute:

i. (1 mark) $\pi^*(S_1)$

ii. (1 mark) $\pi^*(S_2)$

iii. (1 mark) $Q^*(S_1, a_1)$:

iv. (1 mark) $Q^*(S_1, a_2)$:

v. (1 mark) $Q^*(S_2, a_1)$:

vi. (1 mark) $Q^*(S_2, a_2)$:

If γ is allowed to vary between 0 and 1, for which range of values of γ is this policy optimal?

vii. (1 mark) **Minimum** value of γ :

viii. (1 mark) **Maximum** value of γ :

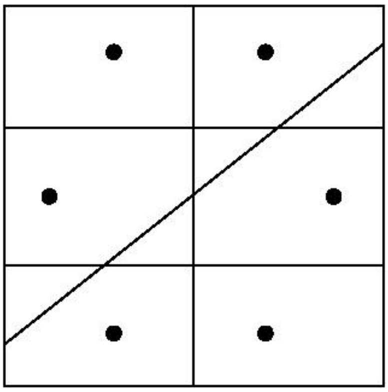
Part C: Graphical Questions (Questions 23-24)

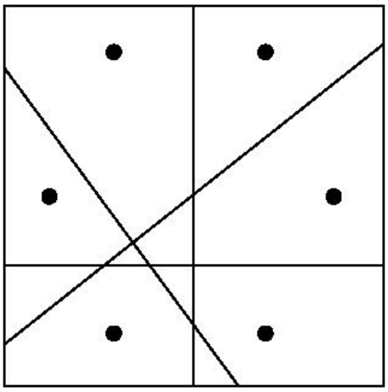
Part C: Question 23 (2 marks)

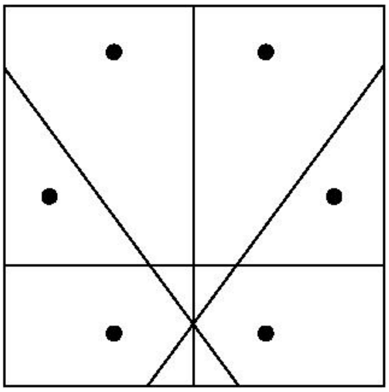
Consider a fully connected feedforward neural network with 6 inputs, 2 hidden units and 4 outputs, using tanh activation at the hidden units and sigmoid at the outputs. Suppose this network is trained on the following data, and that the training is successful.

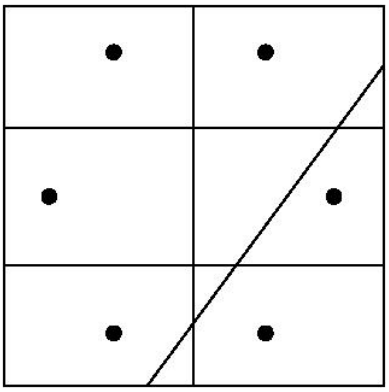
Item	Inputs	Outputs
	123456	1234
1.	100000	0001
2.	010000	0011
3.	001000	0100
4.	000100	1010
5.	000010	1011
6.	000001	1110

Which of these diagrams correctly shows a point in hidden unit space corresponding to each input, and, for each output, a line dividing the hidden unit space into regions for which the value of that output is greater/less than one half ?

A. 

B. 

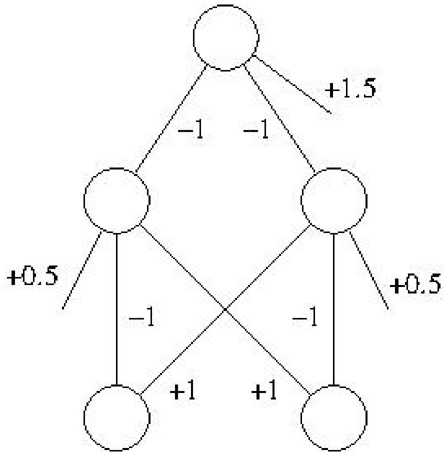
C. 

D. 

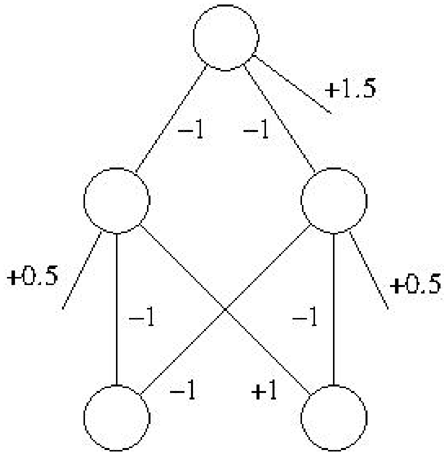
Part C: Question 24 (2 marks)

If 0=FALSE and 1=TRUE, which of these networks (with threshold activations at both the hidden and output layer) correctly computes the XOR function of two inputs?

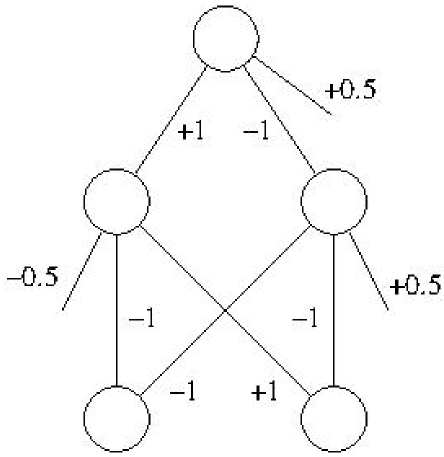
A. ◯



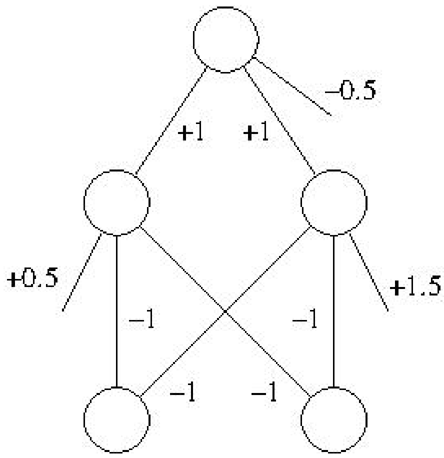
B. ◯



C. ◯



D. ◯



End of Exam