COMP9444-Neural Networks, Deep Learning - 2021 T2

Started on	Saturday, 14 August 2021, 6:40 PM
State	Finished
Completed on	Saturday, 14 August 2021, 8:50 PM
Time taken	2 hours 10 mins
Grade	**1.00** out of 29.00 (**3**%)

Information

COMP9444

Neural Networks and Deep Learning

Sample Examination

(1) Reading Time: 10 Minutes
(2) Time Allowed: 2 Hours
(3) Examination will count for 40% of your Final Mark
(4) Marks for each Question are As Indicated
(5) All Questions May Be Attempted
(6) Unless otherwise indicated, Numerical answers should be given correct to two decimal places (i.e. within 0.005 of the correct answer)
(7) For Multiple Choice Questions, you should select the One Best Answer
(8) Fractional Negative Marks for Incorrect Answers:

- -60% for Multiple Choice (Sub-)Question with Only Two Options
- -20% for Multiple Choice (Sub-)Question with Three or More Options
- No Negative Mark for Numerically Typed (Sub-)Question

For any queries during the exam, contact `a.blair@unsw.edu.au`
Any announcements during the exam will be sent to students via the course e-mail.

Declaration: By starting this exam as a student of The University of New South Wales, I do solemnly and sincerely declare that I have not seen any part of this specific examination paper for the above course prior to attempting this exam, nor have any details of the exam's contents been communicated to me. In addition, I will not disclose to any University student any information contained in this exam at any time. Violation of this agreement is considered Academic Misconduct and penalties may apply.

Fit to Sit Rule: By sitting this exam, you are declaring that you are fit to do so and cannot later apply for Special Consideration. If, during the exam, you feel unwell to the point that you cannot continue with the exam, you should take the following steps:

1. Stop working on the exam and take note of the time;
2. Contact the Course Convenor (`a.blair@unsw.edu.au`) immediately by e-mail or the course forum and advise them that you are unwell;
3. Immediately submit a Special Consideration application saying that you felt ill during the exam and were unable to continue;
4. Obtain a doctor's certificate within 24 hours and attach it to the Special Consideration application;
5. If you were able to advise the Course Convenor or Course Admin of the illness during the assessment, attach screenshots of this conversation to the Special Consideration application.

Technical Issues: If you experience a technical issue during the exam, take the following steps:

1. Take screenshots of as many of the following as possible (all screenshots must include the date and time the issue occurred):
   - error messages
   - screen(s) not loading
   - timestamped speed tests
   - power outage maps
   - messages or information from your internet provider regarding the issues experienced
2. Contact the Course Convenor (`a.blair@unsw.edu.au`) by e-mail or the course forum as soon as possible to advise them of the issue;
3. Submit a Special Consideration application immediately after the exam, including all appropriate screenshots.

Question **1**

Correct

Mark 1.00 out of 1.00

Note: Part A of this Sample Exam contains only One Multiple Choice Question, but Part A of the real Final Exam will contain 12 Multiple Choice Questions.

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

Select one:

- a. Long Short Term Memory ✔
- b. Simple Recurrent Network
- c. Feedforward network with sliding window
- d. Elman Network

Your answer is correct.

The correct answer is: Long Short Term Memory

Question **2**

Not answered

Marked out of 3.00

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.

Assume this Perceptron is being trained on the data in the following table, and that the current values of the weights are $w_0= -1.5$, $w_1= 0$ and $w_2= 2$.

Training Example	x_1	x_2	Class
(a)	0	1	−1
(b)	2	0	−1
(c)	1	1	+1

If the Perceptron Learning Rule is applied to the current weights, using training item (a) and a learning rate of η=1.0, the new values for w_0, w_1 and w_2 at the end of this training step will be:

* w_0: −1.5 ✖
* w_1: 0 ✖
* w_2: 1 ✖

η = 1.0

(a) $-1.5+0\times0+1\times2=0.5$ −1.5 0 2 −1 错.

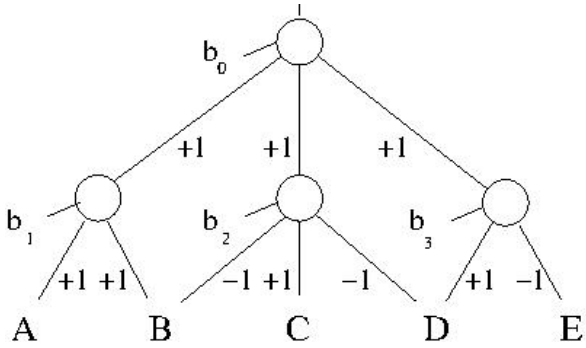
(b) $-2.5+0\times2+0\times1=-2.5$ −2.5 0 1 −1

Question **3**

Not answered

Marked out of 2.00

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 and b_3 would this network compute the logical function

(A ∨ B) ∧ (¬B ∨ C ∨ ¬D) ∧ (D ∨ ¬E)

* b_0= -1.5 ✖
* b_1= -0.5 ✖
* b_2= 1.5 ✖
* b_3= 0.5 ✖

Question **4**

Not answered

Marked out of 2.00

One bag contains 2 red balls and 3 white balls. Another bag contains 3 red balls and 2 green balls. One of these bags is chosen at random, and two balls are drawn randomly from that bag, without replacement. Both of the balls turn out to be red. What is the probability that the first bag is the one that was chosen? (Give your answer correct to two decimal places)

Answer: 0.25 ✖

The correct answer is: 0.25

设第一个bag为F，两球都为红为R. 求P(F|R)

$P(F)=\frac{1}{2}$,

$P(R|F)=\frac{2}{5}\times\frac{1}{4}=\frac{1}{10}$, $P(R\wedge F)=\frac{1}{10}\times\frac{1}{2}=\frac{1}{20}$

$P(R|\neg F)=\frac{3}{5}\times\frac{2}{4}=\frac{3}{10}$, $P(R\wedge\neg F)=\frac{3}{10}\times\frac{1}{2}=\frac{3}{20}$

$P(R)=\frac{3}{20}+\frac{1}{20}=\frac{1}{5}$.

$P(F|R)=\frac{1}{20}/\frac{1}{5}=\frac{1}{4}$

Question **5**

Not answered

Marked out of 3.00

Consider these two probability distributions on the same space Ω = {A, B, C, D}

p = ⟨ ½, ¼, ⅛, ⅛ ⟩
q = ⟨ ¼, ⅛, ⅛, ½ ⟩

Compute, correct to two decimal places:

* The Entropy H(p):

1.75 ✖

$H(P)=-\frac{1}{2}(\log_2\frac{1}{2})-\frac{1}{4}(\log_2\frac{1}{4})-\frac{1}{8}(\log_2\frac{1}{8})-\frac{1}{8}(\log_2\frac{1}{8})$.

* The KL-Divergence $D_{KL}(p \| q)$:

0.5 ✖

$D_{KL}=\frac{1}{2}(\log_2\frac{1}{2}-\log_2\frac{1}{4})+\frac{1}{4}(\log_2\frac{1}{4}-\log_2\frac{1}{8})+\frac{1}{8}(\log_2\frac{1}{8}-\log_2\frac{1}{8})+\frac{1}{8}(\log_2\frac{1}{8}-\log_2\frac{1}{2})$.

$=\frac{1}{2}\times1+\frac{1}{4}\times1+\frac{1}{8}\times0+\frac{1}{8}\times(-2)=\frac{1}{2}$.

Question **6**

Not answered

Marked out of 2.00

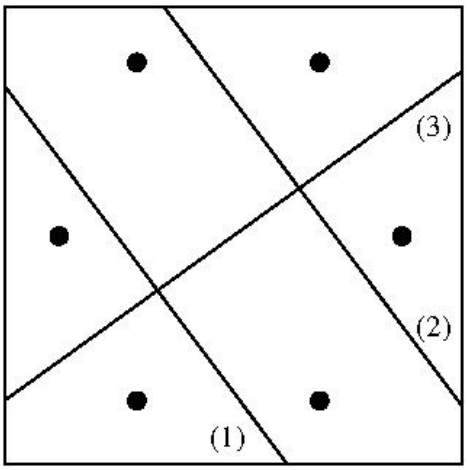
Consider a fully connected feedforward neural network with 6 inputs, 2 hidden units and 3 outputs, using tanh activation at the hidden units and sigmoid at the outputs. Suppose this network is trained on the following data, and that the training is successful.

Item	Inputs	Outputs
	123456	123
1.	100000	000
2.	010000	001
3.	001000	010
4.	000100	100
5.	000010	101
6.	000001	110

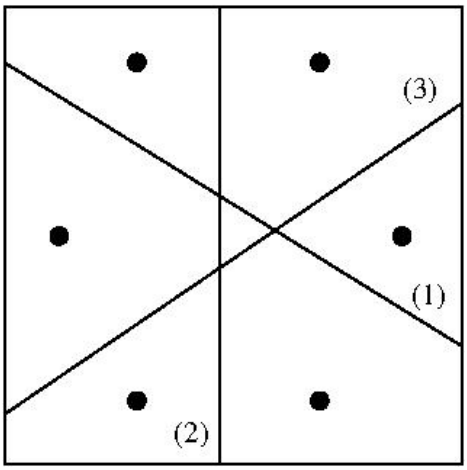
Which of these diagrams correctly shows a point in hidden unit space corresponding to each input, and, for each output, a (correctly labeled) line dividing the hidden unit space into regions for which the value of that output is greater/less than one half ?

Select one:

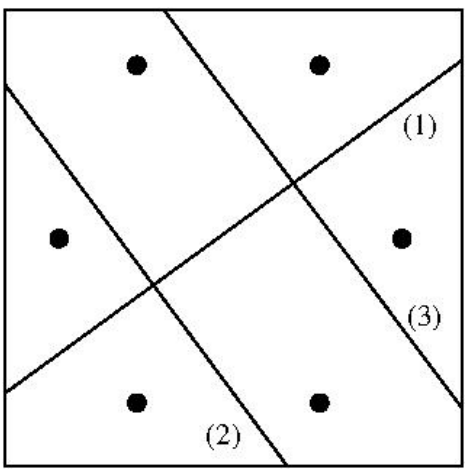
a.



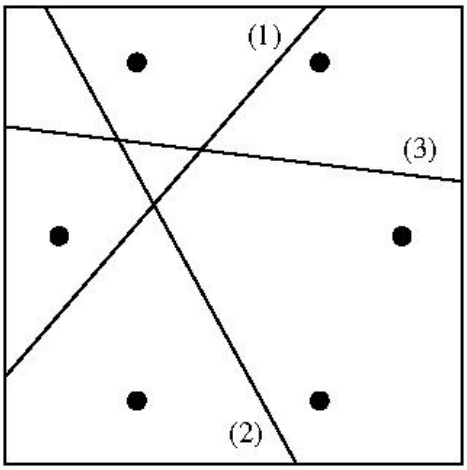
b.



c.

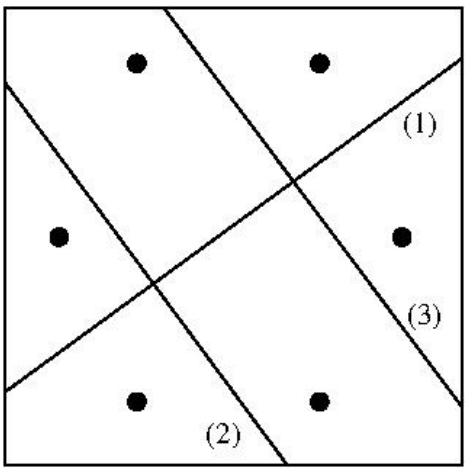


d.



Your answer is incorrect.

The correct answer is:



Question **7**

Not answered

Marked out of 3.00

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= 1.0, z_2= 2.0, z_3= 3.0

Assuming 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):

* d(log Prob(2))/dz_1=

-0.09

✖

* d(log Prob(2))/dz_2=

0.76

✖

* d(log Prob(2))/dz_3=

-0.17

✖

$$Prob(2) = \frac{e^2}{e^1 + e^2 + e^3}$$

Question **8**

Not answered

Marked out of 4.00

Consider a convolutional neural network which takes as input an 84×84×4 image (i.e. with four channels). The first convolutional layer has 16 filters that are 8-by-8, with stride 4 and no zero-padding.

Compute the number of:

* weights per neuron in this layer (including bias):

257

✖

* neurons in this layer:

6400

✖

* connections into the neurons in this layer:

1648800

✖

* independent parameters in this layer:

4112

✖

Question **9**

Not answered

Marked out of 3.00

Consider a Hopfield Network with the following weight matrix W:

| 0 -1 +1 -1 -1 |
|-1 0 -1 +1 +1|
|+1 -1 0 -1 -1|
|-1 +1 -1 0 +1|
|-1 +1 -1 +1 0 |

For each of the following vectors, state whether it is Stable or Not Stable for this network:

* [+1 +1 +1 +1 –1]: Not Stable ✖
* [+1 –1 +1 –1 –1]: Stable ✖
* [–1 +1 –1 +1 –1]: Not Stable ✖

Vector x (weight matrix) . T 后重复 vector .

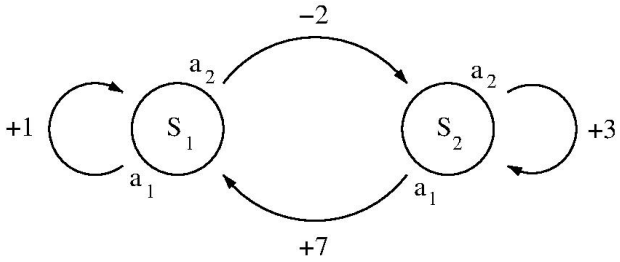
与原 vector 相等 则为 stable, 不同则为 Not stable.

Question **10**

Not answered

Marked out of 6.00

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 γ = 0.7, determine:

* $\pi^*(S_1)$= a_2 ✖

* $\pi^*(S_2)$= a_1 ✖

Again assuming γ = 0.7, compute these values (correct to two decimal places):

* $Q^*(S_1,a_1)$

4.98 ✖

* $Q^*(S_1,a_2)$

5.69 ✖

* $Q^*(S_2,a_1)$

10.98 ✖

* $Q^*(S_2,a_2)$

10.69 ✖

$V^*(S_1) = -2+0.7\times V^*(S_2)$ $V(S_1)=5.69$

$V^*(S_2) = +7+0.7\times V^*(S_1)$ $V(S_2)=10.98$

$Q^*(S_1,a_1) = 1+0.7\times V^*(S_1) = 1+0.7\times 5.69 = 4.98$

$Q^*(S_2,a_2) = 3+0.7\times V^*(S_2) = 3+0.7\times 10.98 = 10.69$.

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

* Minimum value of γ:

0.5 ✖

* Maximum value of γ:

0.8 ✖

(a_1,a_1):

$V(S_1) = 1+\gamma V(S_1)$ $S_1 = \frac{1}{1-\gamma}$

$V(S_2) = 7+\gamma V(S_1)$ $S_2 = \frac{\gamma+7-7\gamma}{1-\gamma} = \frac{-6\gamma+7}{1-\gamma}$

(a_2,a_1):

$V(S_1) = -2+\gamma V(S_2)$ $V(S_1) = \frac{7\gamma-2}{1-\gamma^2}$

$V(S_2) = 7+\gamma V(S_1)$ $V(S_2) = \frac{7-2\gamma}{1-\gamma^2}$

(a_1,a_2):

$V(S_1) = 1 + rV(S_1)$ $\quad V(S_1) = \frac{1}{1-r}$

$V(S_2) = 3 + rV(S_2)$ $\quad V(S_2) = \frac{3}{1-r}$

(a_2, a_1):

$V(S_1) = -2 + rV(S_2)$ $\quad V(S_1) = \frac{5r-2}{1-r}$

$V(S_2) = 3 + rV(S_1)$ $\quad V(S_2) = \frac{3}{1-r}$

S_1:

$$\frac{7r-2}{1-r^2} \geq \frac{1}{1-r} \quad , \quad \frac{7r-2}{1-r^2} > \frac{5r-2}{1-r}$$

$$\frac{7r-2}{(1-r)(1+r)} \geq \frac{1}{1-r}$$

$$7r-2 \geq 1+r$$

$$6r \geq 3$$

$$r \geq \frac{1}{2}$$

$$\frac{7r-2}{1+r} \geq 5r-2$$

$$7r-2 \geq (5r-2)(1+r)$$

$$7r-2 \geq 5r+5r^2-2-2r$$

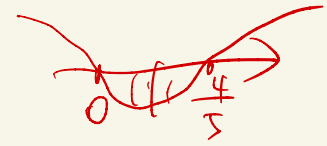
$$4r-5r^2 \geq 0$$

$$r(4-5r) \geq 0$$

$$0 \leq r \leq \frac{4}{5}$$

$$r(5r-4) \leq 0$$

$0 \quad \frac{4}{5}$



即 $\frac{1}{2} \leq r \leq \frac{4}{5}$

S2:

$\frac{7r-2}{1-r^2} \geq \frac{3}{1-r}$, $\frac{2r-2}{1-r^2} > \frac{6r+7}{1-r}$

$\frac{7r-2}{1+r} \geq 3$

$7r-2 \geq 3+3r$

$4r \geq 1$

$r \geq \frac{1}{4}$

$\frac{7r-2}{1+r} > -6r+7$

$7r-2 \geq (6r+7)(1+r)$

$7r-2 \geq -6r+7-6r^2+7r$

$-6r^2+r-1 \leq 0$

$-6(r^2+r-\frac{1}{6}) \leq 0$

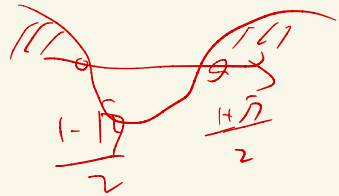
$r^2+r-\frac{1}{6} \geq 0$

综上 因为 $0 \le r \le 1$

$\therefore \frac{1}{2} \le r \le \frac{4}{5}$

$\frac{1-\sqrt{5}}{2} \le r \le \frac{1+\sqrt{5}}{2}$

$\frac{-1 \pm \sqrt{1^2 - 4 \times 1 \times (-1)}}{2}$



$\Delta = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

$\frac{-1 \pm \sqrt{7}}{2}$