UNIVERSITY OF BRISTOL

SAMPLE PAPER Examination Period

FACULTY OF ENGINEERING

M Level Examination for the Degree of
Master of Engineering / Masters of Science

COMSM0094-A
Learning, Computation and the Brain

TIME ALLOWED:
2 hours

Answers to COMSM0094-A: Learning, Computation and the Brain

Intended Learning Outcomes:

Section A: short questions - answer all questions

Q1. Hebb's rule is often paraphrased as 'neurons that fire together wire together'; why is this no longer considered accurate?

Solution: It ignores the temporal structure; spike timing effects are now considered important.

Q2. The two principal forms of aphasia are expressive aphasia and fluent aphasia, one is distinguished by the inability to find words, the other by the inability to understand language. Expressive aphasia is associated with lesions in which brain areas.

Solution: Broca's area

Q3. What is the equilibrium value of h in the equation $dh/dt = 4 - 2h$? How do we know this value is stable?

Solution: The equilibrium value is $h = 2$. If $h > 2$ we have negative dh/dt and if $h < 2$ we have positive, so the value will also head towards two.

Q4. What are the advantages and disadvantages of Electroencephalography (EEG) as a tool to study neuroscience.

Solution: EEG is useful because it allows fine temporal resolution measurements on humans using an non-invasive technique; however the spatial resolution is very poor and the recordings are very noisy.

Q5. What caused the damage to patient HM's hippocampus and what did it cause?

Solution: An operation by William Scoville to treat his epilepsy. It made HM unable to form new memories. (Accept memory loss)

Q6. Give a typical value of the resting potential inside a neuron.

Solution: -70mV

Q7. What are dopaminergic cells believed to fire in response to.

Solution: unexpected rewards

Q8. Define Shannon's capacity $C(B, S)$.

Solution: $C = B\log_2(1 + S/N)$ where B=bandwidth, S=signal, N=noise

Q9. What was Dennards scaling law?

Solution: As the dimensions of a device go down, so does power consumption. or As transistors get smaller, their power stays constant hence power use stays in proportion with area. or something close that says the same thing.

Q10. Approximately how many neurons are in the human brain?

Solution: 10^{10} accept a single of order of magnitude up or down.

Q11. Consider the following fully connected, multilayer neural network, where all of the layers use the same activation function. A:(50 nodes) $->$ (40 nodes) $->$ (10 nodes)$->$ (40 nodes) $->$ (50 nodes)
What would you call this type of neural network architecture?

Solution: An auto-encoder.

Q12. In neural networks, what is an convolutional neural network and why does it need pooling layers?

Solution: A CNN learns features of data by filtering. (Accept any answer that is also domain specific such as sampling images) A CNN uses pooling to down sample and summarise the presence of features and to make the sample more robust to translational invariance. (Accept any single feature here)

Q13. Give two features of natural language that make it hard for neural networks to process?

Solution: Order matters, relationship between distant words, relative importance of words, conceptual or cultural frameworks. (Any two, also accept other forms of ambiguity and spelling and grammatical errors)

Q14. What, in the field of generative AI, is latent space?

Solution: An abstract multi-dimensional space that encodes a meaningful internal representation of externally observed events. (Accept lower dimensional, subspace of problem space, or other terms that demonstrate the concept well.)

Q15. Why do neural network activation functions need to be non-linear?

Solution: Because data is typically non-linear. Linear activation functions cause networks of any depth to act as single layer NN and cannot therefore process non-linear data. Non-linear neural networks can produce non-linear decision boundaries. (Accept solutions from any of these positions)

Section B: long questions - answer two questions

Q1. This question is about integrate-and-fire neurons.

(a) In the leaky integrate-and-fire neuron the voltage, v, satisfies

$$\tau_m \frac{dv}{dt} = E_l - v + R_m I_e$$

with the rule that if $v > V_t$ the voltage is reset to V_r. What is the term E_l and where does it come from? *[5 marks]*

(b) In an experiment a constant current input I_e is applied with successively larger values. What value of I_e will make the neuron spike? *[5 marks]*

(c) Draw the f-I curve for the integrate and fire neuron. *[3 marks]*

(d) Derive a formula for the interspike interval for this neuron when there is a constant current large enough to cause spiking. *[7 marks]*

Solution: a) E_l is the reversal potential [2 mark] and is the result of chemical gradients, largely in potassium, across the cell membrane [3 marks]
b) The neuron will spike if $I_e > (V_t - E_l)/R_m$ since then the equilibrium point is higher than threshold [5 marks, 2 for some attempt].
c) This is a curve that is zero until I_e causes the equilibium value to reach threshold, then it rises sharply. [3 for nice graph, 1 for graph missing labels or with the wrong I_e value]
d) In the model

$$\tau_m \frac{dV}{dt} = E_L - V + R_m I_e$$

which we can solve from our study of odes, it gives

$$V(t) = E_L + R_m I_e + [V(0) - E_L - R_m I_e]e^{-t/\tau_m}$$

[2 marks] so if the neuron has spiked and is reset at time $t = 0$ and reaches threshold at time $t = T$, assume $V_R = E_L$ we have [1 marks]

$$V_T = E_L + R_m I_e - R_m I_e e^{-T/\tau_m}$$

[1 mark] so

$$e^{-T/\tau_m} = \frac{E_L + R_m I_e - V_T}{R_m I_e}$$

[1 mark] Taking the log of both sides we get

$$T = \tau_m \log\left[\frac{R_m I_e}{E_L + R_m I_e - V_T}\right]$$

[2 marks]

Q2. This question is about Shannon's capacity theorem and its application to the energy efficiency of communication.

(a) Define Shannon's capacity $C(B, S)$. *[4 marks]*

(b) Calculate the channel capacity for two cases $S = 3N$ and $S = 15N$. *[5 marks]*

(c) Given the energy efficiency is $F = C/(S + N)$, where S is defined in terms of a constant times N, Compare the energy efficiency of the two cases from the previous part of the question, given $B = 1/2$. What can you say about the difference between the two cases? *[3 marks]*

(d) Explain, with reference to Shannon's capacity theorem, why the channel capacity cannot become unboundedly large even as we increase the bandwidth without bound? *[3 marks]*

(e) How can we show that the channel capacity has a bound? You do not have to calculate the limit. *[5 marks]*

Solution: a) C = B log2 (1+S/N) where B=bandwidth, S=signal, N=noise
b)
C = B log2 (1+3N/N) = 2B and C = B log2 (1+255N/N) = 4B c)
For S=3N, F=1/4N and for S=15N, F=2/16N=1/8N hence for two times the capacity the energy has increased 4 times so ignoring external influences it would seem that two cases of S=3N would be more efficient than one case of S=15N
d)
The channel capacity does not become infinite since, with an increase in bandwidth, the noise power also increases.
e)
Say N=nB then C = S/n(nB/S)log2 (1 + S/nB) and in the limit pow((1 + S/nB),(nB/S)) becomes a constant hence in the limit C = someconstant (S/n).

Q3. This question is about the neural network model in Figure [1]. The network has two input nodes with fixed values, one hidden layer and an output layer.

(a) Calculate the outputs of the neural network as described with activation function $f(x) = \frac{1}{1+e^{-x}}$. *[6 marks]*

(b) Prove that this activation function is non-linear? *[3 marks]*

(c) Given the following back formula:

$$\partial C/\partial w_{j,k}^{(L)} = \partial C/\partial a_j^{(L)} \times \partial a_j^{(L)}/\partial z_j^{(L)} \times \partial z_j^{(L)}/\partial w_{j,k}^{(L)} \qquad \text{Partial Derivative}$$

$$C = \sum (output - y)^2 \qquad \text{Cost function}$$

$$z_j^{(L)} = w_{j,k}^{(L)} a_k^{(L-1)} + b^{(L)} \qquad \text{Summation of weights plus bias}$$

$$a_j^{(L)} = f\left(z_j^{(L)}\right) \qquad \text{Activation function}$$

(cont.)

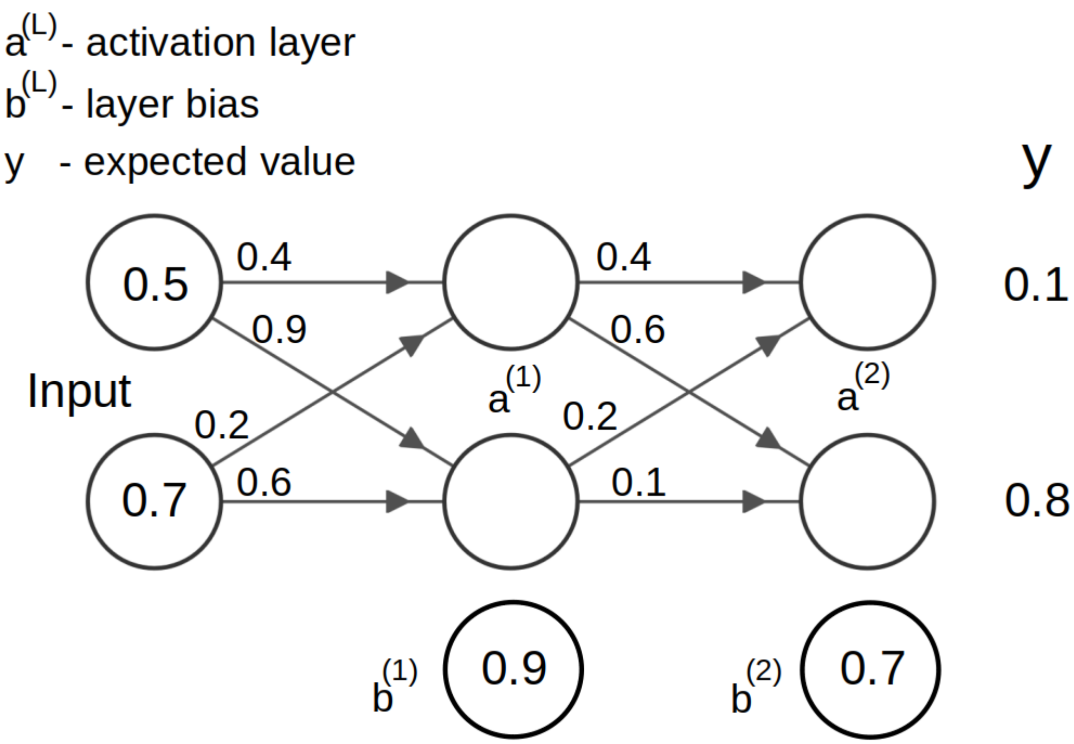

Figure 1: **Neural network for question 3.**

where L is the layer, $_j$ is the node index, $_k$ is the weight index and $f(x) = \frac{1}{1+e^{-x}}$, the logistic function. Calculate the formula for adjusting the weights from layer $L = 1$ to $L = 2$, by back propagation, using the above formula. *[2 marks]*

(d) Calculate the change to the weights $1, 0$ and $1, 1$, the bottom two weights, between layer 1 and layer 2, which have values 0.2 and 0.1, using standard back propagation. *[4 marks]*

(e) What is the learning rate of a neural network and how can it affect learning? *[5 marks]*

Solution: a) 0.7651..., 0.7774... anything over 1sf is fine
b) Use f(x+y)=f(x)+f(y) and f(ax)=af(x) to prove
c) Giving: $\partial C/\partial a_0^{(2)} = 2\left(a_0^{(2)} - y_0\right)$
$\partial a_0^{(2)}/\partial z_0^{(2)} = f\left(z_0^{(2)}\right)\left(1 - f\left(z_0^{(2)}\right)\right)$
$\partial z_0^{(2)}/\partial w_{0,0}^{(2)} = a_0^{(1)}$ [2 marks]
d) 0.2042... and -0.0066... (accept plus or minus, must have different signs) (credit for good working if wrong answer) [4 marks]
e) The learning rate is a variable that adjusts the results of back-propagation before they are applied to the weights. Too small may result in a long training process that could get stuck. Too large may result in learning a sub-optimal set of weights too fast or an unstable training process. (Accept discussion on sub-optimal gradient

descent) [5 marks]

END OF PAPER