Name: Alexander Mercer

Midterm Exam
Q355 Brain and Cognition
Spring 2022
February 28, 2022

(please put your name on all pages)

60.5/75

This exam is worth 75 pts. Please read the questions carefully before answering.

Multiple Choice (8 points; each worth 1 points. Choose the <u>best</u> answer and indicate your answer by CIRCLING THE LETTER of the answer you choose.)

1. An increase in synaptic weights is associated with _____
   a. Long term potentiation
   b. Long term depression
   c. increased eligibility traces
   d. weight transport

2. The fastest hardware for neural network computation currently is:
   a. CPU
   b. GPU
   c. TPU
   d. APU

3. The Kohonen map is a method of:
   a. supervised learning
   b. unsupervised learning
   c. applying the delta rule
   d. spike timing dependent plasticity

4. What happened in 1986 in terms of neural networks?
   a. The XOR problem was UNsolvable
   b. The XOR problem was SOLVABLE
   c. The AI winter began
   d. The vanishing gradient problem was solved

5. A simple perceptron can solve all of the following problems EXCEPT:
   a. The AND operation
   b. The OR operation
   c. The XOR operation
   d. The NOR operation

6. If a matrix is not invertible, then it must: it can still be square and not invertible
   a. not be square
   b. not be transposable
   c. have a row that is a linear combination of the other rows
   d. be a constant multiple of the identity matrix

-1

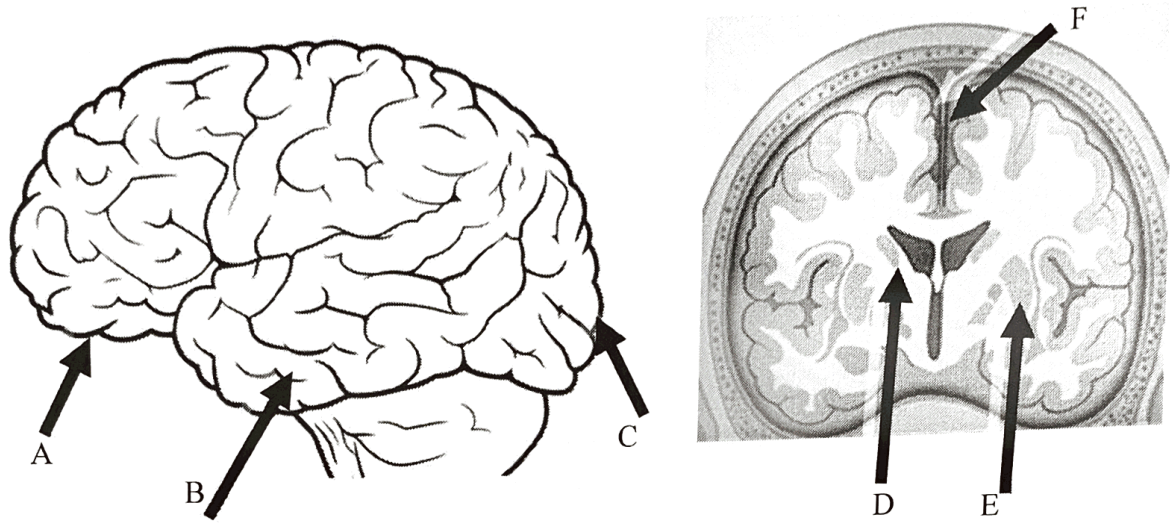
7. Neurotransmitters are chemicals usually released from:
   a. dendrites
   b. axons
   c. soma
   d. nodes of Ranvier

8. With spike timing dependent plasticity, synaptic weights are most increased when
   a. presynaptic spikes lead postsynaptic spikes in time
   b. postsynaptic spikes lead presynaptic spikes in time
   c. pre and postsynaptic spikes occur simultaneously
   d. only postsynaptic spikes occur

-1

Name Alexander Mercer

9. Fill in the blanks below to identify the brain regions or features indicated by the arrows **(6 points):**



A. Orbitofrontal cortex

B. Parietal (temporal) lobe

C. Cerebellum (occipital)

D. Amygdala (caudate nucleus)

E. Globus Pallidus (Putamen)

F. Central (longitudinal) fissure

Fill in the blank (2 points; each blank worth 1 points)

10. Outside the nerve cell there is a higher concentration of ~~calcium~~ (sodium) ions cations

11. Cortical gray matter consists of how many layers? 3 (6)

Short answers (1 point each):

12. The main current equation of the Hodgkin-Huxley equation is

$$I_{ion} = G_{Na}(V_m - E_{Na}) + G_K(V_m - E_K) + G_L(V_m - E_L).$$

What do the following terms represent (1 pt each):

a. GNa The Sodium channel weights

b. Vm Voltage Potential of a Cell

Name Alexander Merar

c. ENa Measured calcium within the cell Na Nernst potential -1

d. If GNa = 1 and GK=0 and GL=0, what will be the value of Vm at equilibrium?

$1(V_m - E_{NA}) + 0 + 0 = 0 \Rightarrow \boxed{V_m = E_{NA}}$

13. Solve the matrix multiplication problem (1 pt):

$\begin{bmatrix}5 & 2\\ 4 & 9\end{bmatrix} * \begin{bmatrix}7\\ 3\end{bmatrix} = \begin{bmatrix} \end{bmatrix}$ $\quad \begin{matrix}35+6\\ 28+27\end{matrix} = \begin{bmatrix}41\\ 55\end{bmatrix}$

14. Write the formula relating the dot product to the cosine of the angle between two vectors. What is the cosine of the angle between [3,4] and [4, 3]? (2 pts)

$\theta = \arccos(\text{dot Product} / [(V_1)^2 * (V_2)^2]$

-0.5 $\text{dotP} = 12 + 12 = 24$ $\quad \theta = \arccos(24/(25 * 25)) \quad \frac{24}{25}$

15. What is supervised learning (i.e. what is the general problem that supervised learning algorithms solve, in terms of vectors?) (2 pts)

Supervised learing is when you have a traing set of vectors with known outputs for your traing set. By using some sort of learning algorithm (like the delta rule), you are able to update the weights of your network to learn using that training set of inputs and outputs.

16. What is the advantage of using a cross-entropy loss function instead of a sum squared error loss function? (2 pts)

-1 Cross-entropy loss functions do not slow the learning rate when the output becomes very large or very negative. no signal func derivative, no vanishing gradient

17. What is rate coded model neuron? In qualitative terms, how does it work? (3 pts)

A rate coded mode measures the frequency of action potential within a cell. Without much detail and a few parameters, the complexity of an / action potential is simplified to measure the rate of a cell

18. Describe how the cortical sensory homunculus changes when just the middle and index fingers of a monkey are stimulated repeatedly. Explain how a Kohonen map simulates this effect. (3 pts)

When the middle and index fingers of a monkey are stimulated repeatedly, the cortical area responsible for the middle and index fingers of that monkey grows to fit the growth in use of that area and the lower use of the other fingers.

A Kohonen map mirrors this by being an unsupervised model capable of fitting the topography of a dataset that it is trying to learn.

-2.5

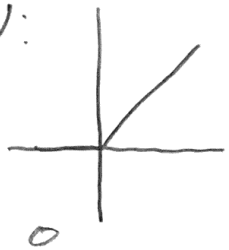
Name Alexander Mercer

19. Write the formula for the L2 norm of a vector x. What is the L2 norm of the vector [6, 8]? (2 pt) L2 is the euclidean distance of a vector X.

$L2 = \sqrt{a^2+b^2}$ L2 of $[6,8] = \sqrt{36+64} = \sqrt{100} = 100$

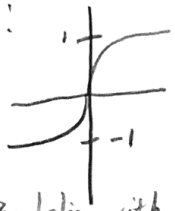
20. Describe the ReLU and sigmoidal signal functions. Draw a qualitative sketch of each. Why might we prefer a ReLU signal function for deep networks? (3 pts)

ReLU:



$x < 0: 0$
$x \geq 0: y = x$

Sigmoid:



Parabolic with $x \geq 0$ converging to 1 $x < 0$ converging to -1

The ReLU function is used most popularly to diminish the effects of the vanishing gradient problem. When a network becomes deeper, it is harder to incentivize learning in a model and a ReLU function helps preserve learning throughout training.

21. What is the danger of having too many hidden units in a multi-layer perceptron? (2 pts)

When there are too many hidden units, you need more train data in a multi-layer perceptron. This could cause overfitting, which would be detrimental when approaching novel data.

22. What is the difference between a McCullouch-Pitts neuron and a Perceptron? (1 pt)

A McCP neuron has all-or-none thinking with no weights affecting the inputs.

A perceptron's inputs are variable and can possess values other than 0 or 1. Additionally, a perceptron has weights, which means some inputs would affect the output more.

23. What is the k-means algorithm, and how does it work? (4 pts)

The k-means algorithm is an unsupervised learning algorithm that looks to cluster data into N possible clusters. The algorithm works by placing the k centroids of the algorithm and finding the euclidean distance of each point to the nearest centroid. Then, the centroids adjust to fit the average distance. By the end of this iterative cycle, the data is separated into N desired clusters.

24. A three-layer perceptron can theoretically learn any mapping from input to output, so what is the advantage of having more than 3 layers? (2 pts)

When there are more than three layers in a perceptron, the perceptron is able to accomplish more difficult tasks that may have emergent properties based on inputs found within previous layers. An example of this benefit can be found in image recognition of digits in the MNIST data set. More layers allows the perceptron to parse pieces of an image before joining all of that information together to guess at a digit 0-9. -0

Name Alexander Mercer

25. Explain how action potentials lead to neurotransmitter release at axon terminals (3 pts) calcium

-1 An action potential happens when the cell depolarizes → Potassium channels open → potassium enters the cell → transmitters bind to vesicles → Vesicles bind to the presynaptic membrane → the transmitter is released into the synaptic cleft.

26. The error backpropagation algorithm has been criticized with respect to its biological plausibility. What are those critiques? How does the feedback alignment algorithm address those critiques? (5 pts)

-1 Backprop is critiqued because of the fact that cells are unable to send information back from the post-synaptic cells to the pre-synaptic cells. In backprop, this is done with no problem but, cells have no infrastructure to model this kind of learning within biological plasticity. Feedback alignment protects the error at backprop into a network of backwards weights, which better simulates biological plasticity. In this way, information and learning takes place without going back through the same cell.

fixed random backward wts

27. The modified learning rule "Adam" incorporates two modifications to the usual gradient descent rule. Describe them and why they are useful. (4 pts)

-1 -Momentum: Keeps a running average of the gradient descent to avoid zig-zag patterns.
-RMSprop: Keeps track of the changes in the gradient to avoid understeering or overshooting a particular local minimum root mean square of gradient

28. Consider a simple 2-layer neural network, with input node activation x, and output node activation y. Consider the error function $E=\sum_{i=1}^{N}(d_i - y_i)^2$, where E is the sum squared error, $e=d_i - y_i$, d is the desired output, and y is the actual output. Assume that y=Wx, where y and x are vectors, and W is a weight matrix. Assume that learning proceeds by gradient descent as $\Delta W = -\alpha\frac{\partial E}{\partial W}$, where α is the learning rate. Assuming a single (scalar) input x and a single output y, derive the delta rule using the chain rule. Your answer should take the form $\Delta W =$ ____, showing the steps along the way. (8 pts)

-6

$$\Delta W = -\alpha\frac{\partial E}{\partial W}$$

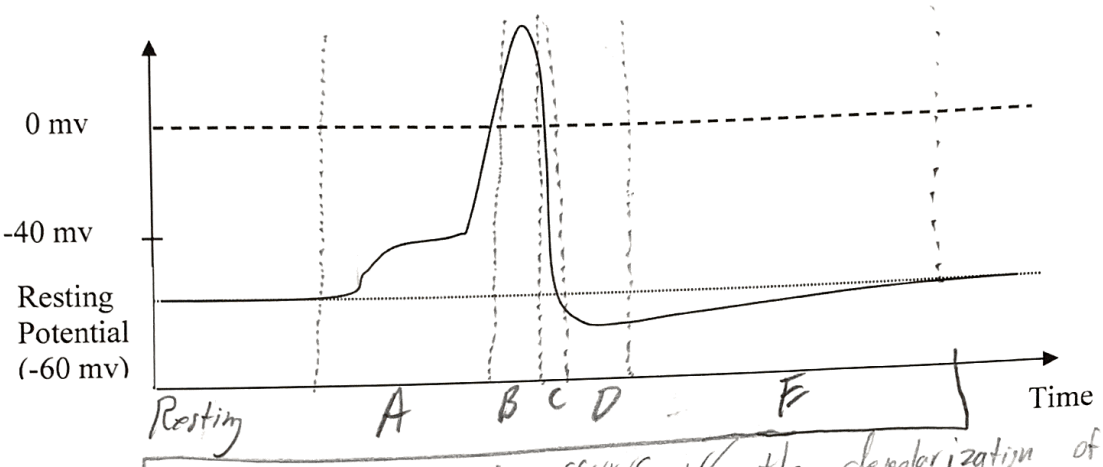
$$= -\alpha \cdot \frac{\partial E}{\partial e} \cdot \frac{\partial e}{\partial y} \cdot \frac{\partial y}{\partial W}$$

$$\therefore \Delta W = -\alpha \frac{\partial E}{\partial e} \cdot \frac{\partial e}{\partial y} \cdot \frac{\partial y}{\partial W}$$

-9

Name Alexander Mervar

29. The graph below depicts an action potential, which is a fundamental unit of neural signaling. Partition the action potential into its five phases and label them as A, B, C, D, E. Below the graph, *clearly and succinctly* explain, in order of events as they occur, how the action potential is generated. In your answer, name each labeled phase and describe BOTH the function and state of specific ion channels involved in each phase. Use the following abbreviations for ion channels: PK = passive potassium channels, VGNA = voltage-gated sodium channels, VGK = voltage-gated potassium channels. For this question, you may continue your answer on the back of this page if necessary. (8 pts)



A = Rising Phase
B = Overshoot Phase
C = Falling Phase
D = Undershoot Phase
E = Recovery Phase

An action potential occurs w/ the depolarization of the cell. With this depolarization, the rising phase (A) begins with the opening of ~~VGK~~ ~~VGNA~~ channels. This increases the voltage of the cell. Because the ~~VGNA~~ VGK channels are slower to respond, the voltage of the cell overshoots (B) before decreasing the voltage w/ the addition of sodium ions. As the sodium is taken on, the VGK channels begin to close w/ the falling voltage (C). Then, because the sodium channels are slow, they allow the cell to undershoot the desired resting voltage (D). Then, once all the voltage-gated VGNA & VGK channels have completed the stages of the action potential, the passive channels PK allow for the cell to recover and return to resting voltage (E). −1 VGNA & VGK flipped

−1