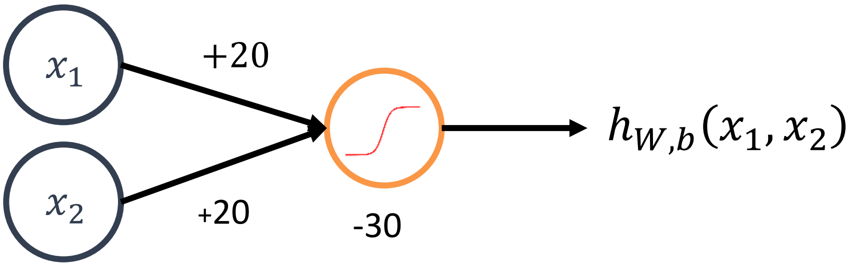
Exercise #1

Consider the following neural network which takes two binary-valued inputs $x_1, x_2 \in \{0,1\}$ and outputs $h_\theta(x)$. Which of the following logical function does it compute?

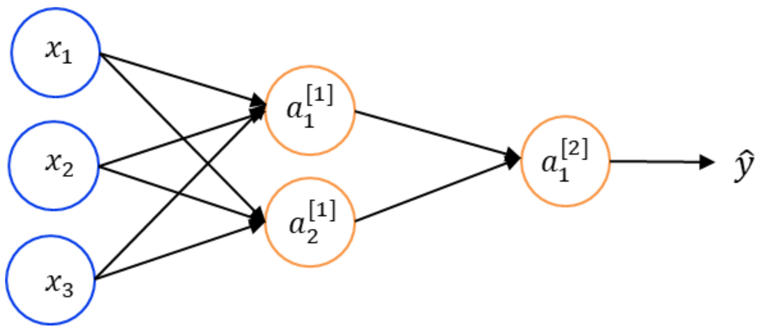


A. XOR
B. OR
C. NAND
D. AND

Solution: ___

Exercise #2

Consider the following neural network:



Let's suppose to have the following learned parameters and activation functions:

$W^{[1]} = \begin{bmatrix} 4 & -3 & 1 \\ -6 & 3 & 0 \end{bmatrix}; b^{[1]} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}; \quad W^{[2]} = [2 \quad 1]; b^{[2]} = [1].$

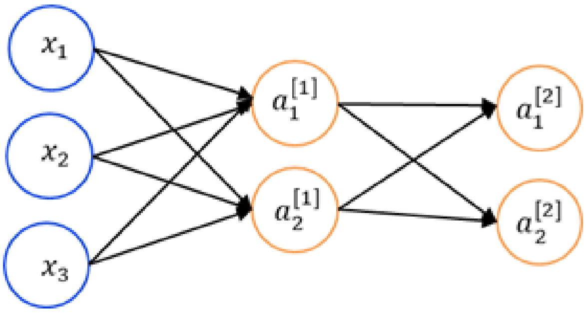
Activation Function: sigmoid

Compute the output of the neural network as $x = [2\ 5\ 3]^T$.

Solution: ____

Exercise #3

Consider the following neural network:



Let's suppose to have the following learned parameters and activation functions:

$$W^{[1]} = \begin{bmatrix} 4 & -3 & 2 \\ -6 & -2 & 0 \end{bmatrix}; b^{[1]} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}; \quad W^{[2]} = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}; b^{[2]} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

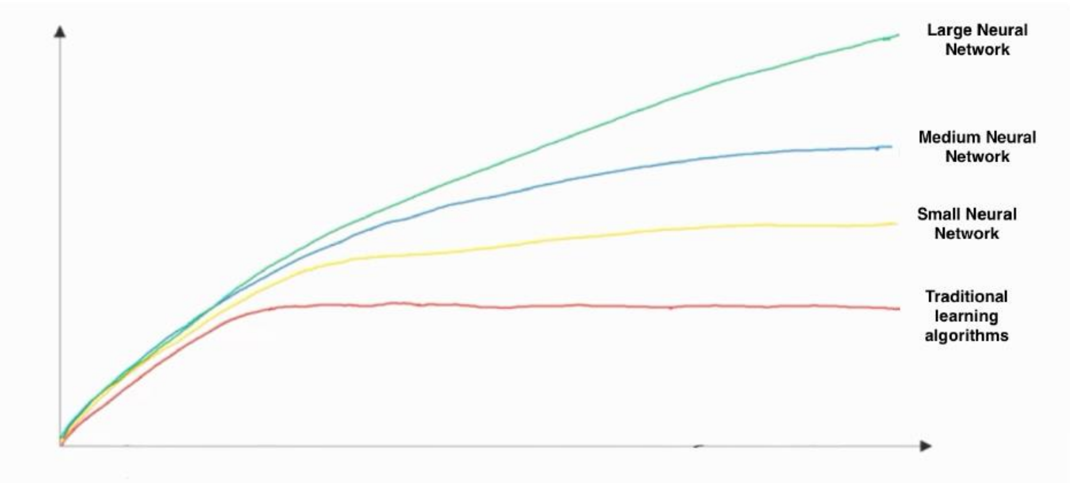
Activation Function $a^{[1]}$: sigmoid
Activation Function $a^{[2]}$: Sofmax

Compute the output $a_2^{[2]}$ of the neural network as $x = [2\ 3\ 1]^T$

Solution: ______

Exercise #4

In this diagram which we hand-drew in lecture, what do the horizontal axis (x-axis) and vertical axis (y-axis) represent?



A) x-axis is the amount of data; y-axis (vertical axis) is the performance of the algorithm.

B) x-axis is the amount of data; y-axis is the size of the model you train.

C) x-axis is the input to the algorithm; y-axis is outputs.

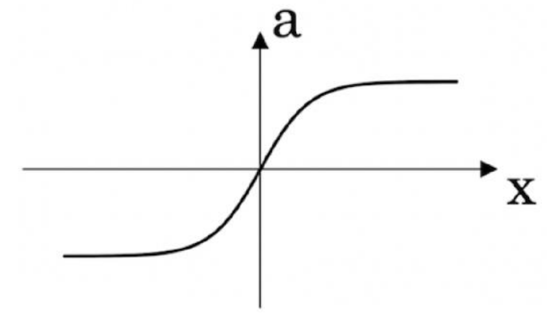
D) x-axis is the performance of the algorithm; y-axis (vertical axis) is the amount of data.

Solution: ____

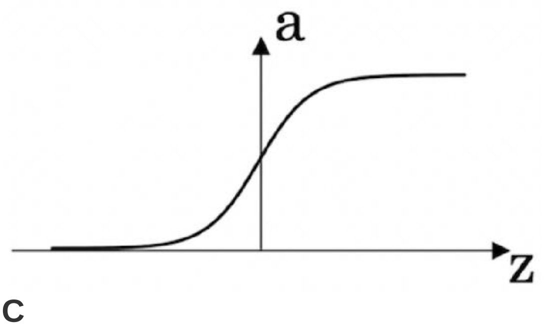
Exercise #5

Which one of these plots represents a ReLU activation function?

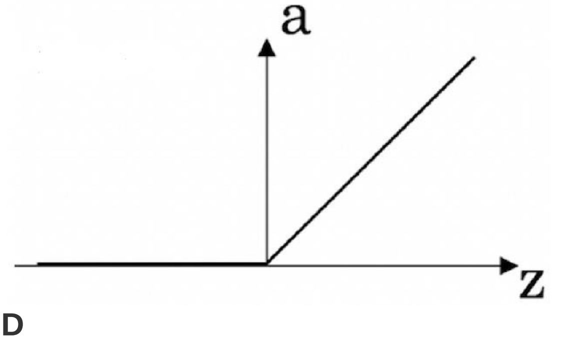
A



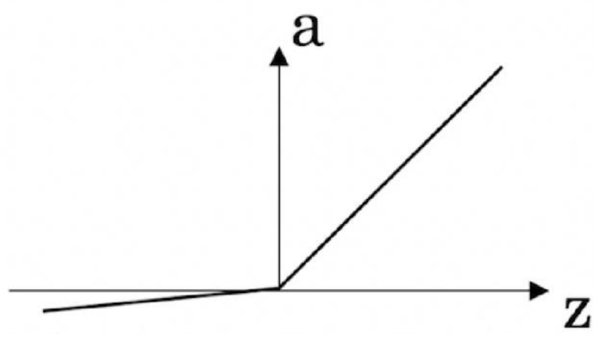
B



C



D



Solution: ______

Exercise #6

Which of these is the "Logistic Loss"?

A

$$\mathcal{L}^{(i)}(\hat{y}^{(i)}, y^{(i)}) = -(y^{(i)}\log(\hat{y}^{(i)}) + (1 - y^{(i)})\log(1 - \hat{y}^{(i)}))$$

B

$$\mathcal{L}^{(i)}(\hat{y}^{(i)}, y^{(i)}) = \mid y^{(i)} - \hat{y}^{(i)} \mid$$

C

$$\mathcal{L}^{(i)}(\hat{y}^{(i)}, y^{(i)}) = max(0, y^{(i)} - \hat{y}^{(i)})$$

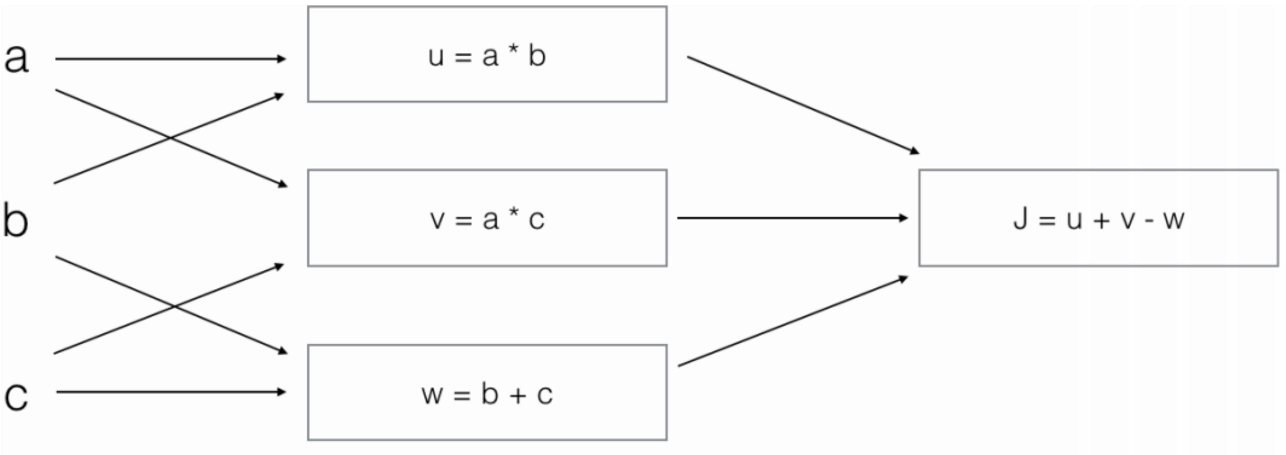
D

$$\mathcal{L}^{(i)}(\hat{y}^{(i)}, y^{(i)}) = \mid y^{(i)} - \hat{y}^{(i)} \mid^2$$

Solution: _____

Exercise #7

Consider the following computation graph.



What is the output J?

A. J = (c - 1)*(b + a)
B. J = (a - 1) * (b + c)
C. J = a*b + b*c + a*c

D. J = (b - 1) * (c + a)

Solution: ______

Exercise #8

You are building a binary visual classifier for recognizing apples (y=1) vs. tomatoes (y=0). Which one of these activation functions would you recommend using for the output layer?

A. ReLU
B. Leaky ReLU
C. Sigmoid
D. Tanh

Solution: ________

Exercise #9

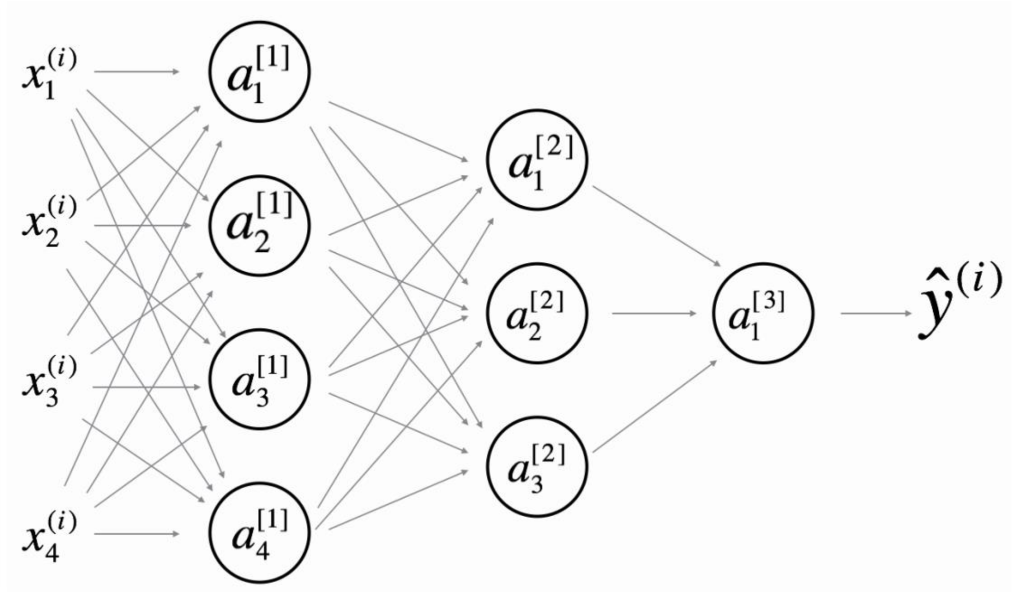
Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements is true?

A. Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.
B. Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have “broken symmetry”.
C. Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished “symmetry breaking” as described in lecture.
D. The first hidden layer’s neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

Solution: _________

Exercise #10

Consider the following 2 hidden layer neural network:



Which of the following statements are True? (Check all that apply).

A

$W^{[1]}$ will have shape (4, 4)

B

$b^{[1]}$ will have shape (4, 1)

C

$W^{[1]}$ will have shape (3, 4)

D

$b^{[1]}$ will have shape (3, 1)

E

$W^{[3]}$ will have shape (3, 1)

F

$b^{[3]}$ will have shape (1, 1)

G

$b^{[3]}$ will have shape (3, 1)

Solution: _____

Exercise #11

The dev and test set should:

A. Come from the same distribution
B. Come from different distributions
C. Be identical to each other (same (x,y) pairs)
D. Have the same number of examples

Solution: ____

Exercise #12

If your Neural Network model seems to underfit your data, what of the following would be promising things to try? (Check all that apply.)

A. Add regularization
B. Make the Neural Network deeper
C. Increase the number of units in each hidden layer
D. Get more training data
E. Get more test data

Solution: _____