CS540 Introduction to Artificial Intelligence
Lecture 3

Young Wu
Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

June 4, 2020

Test

Quiz

Socrative Room

CS540C

- A:
- → B: Choose this.
- C:
- D:
- E:

______ @wisc.edu

↑ login ID

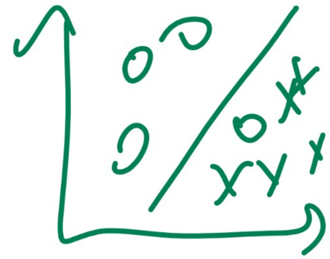
message "Questions" if you have questions.

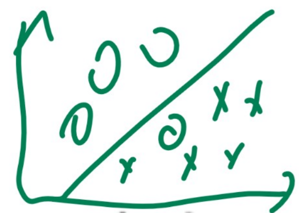
Single Layer Perceptron

Motivation

LTU Perceptron

Logistic Perceptron





- Perceptrons can only learn linear decision boundaries.
- Many problems have non-linear boundaries.
- One solution is to connect perceptrons to form a network.

Decision Boundary Diagram

Motivation

Multi Layer Perceptron

Motivation

- The output of a perceptron can be the input of another.

$$a = g\left(w^T x + b\right)$$
$$a' = g\left(w'^T a + b'\right)$$
$$a'' = g\left(w''^T a' + b''\right)$$
$$\hat{y} = \mathbb{1}_{\{a''>0\}}$$

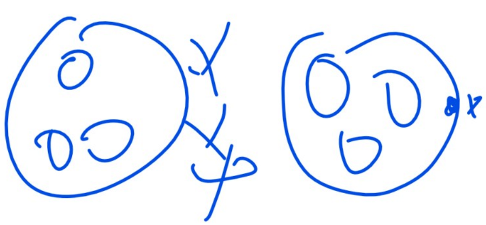
Neural Network Biology

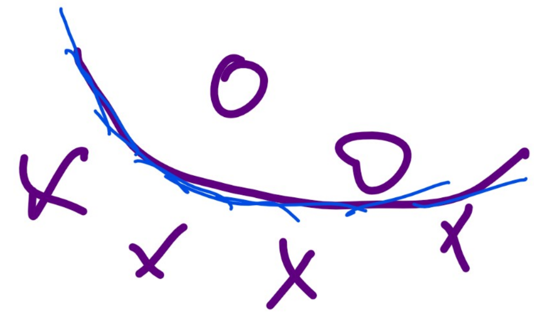
Motivation

- Human brain: $100,000,000,000$ neurons.
- Each neuron receives input from $1,000$ others.
- An impulse can either increase or decrease the possibility of nerve pulse firing.
- If sufficiently strong, a nerve pulse is generated.
- The pulse forms the input to other neurons.

Theory of Neural Network

Motivation





- In theory:

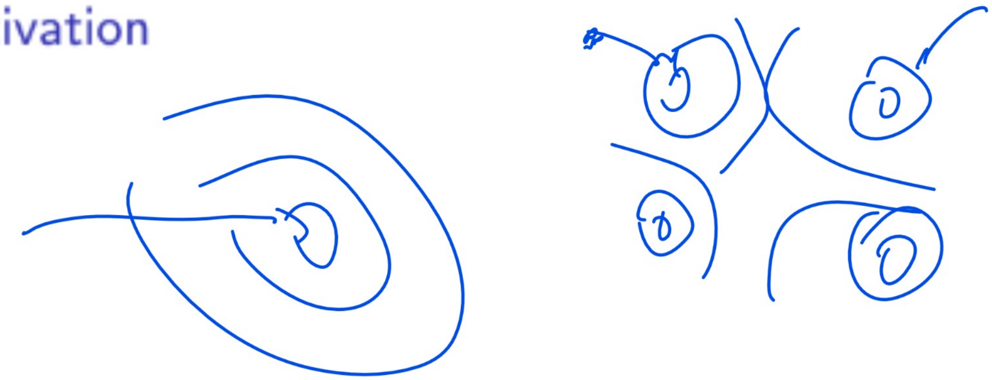
1. 1 Hidden-layer with enough hidden units can represent any continuous function of the inputs with arbitrary accuracy.
2. 2 Hidden-layer can represent discontinuous functions.

- In practice:

1. AlexNet: 8 layers.
2. GoogLeNet: 27 layers (or 22 + pooling).
3. ResNet: 152 layers.

Gradient Descent

Motivation



- The derivatives are more difficult to compute.
- The problem is no longer convex. A local minimum is longer guaranteed to be a global minimum.
- Need to use chain rule between layers called backpropagation.

Backpropagation

Description

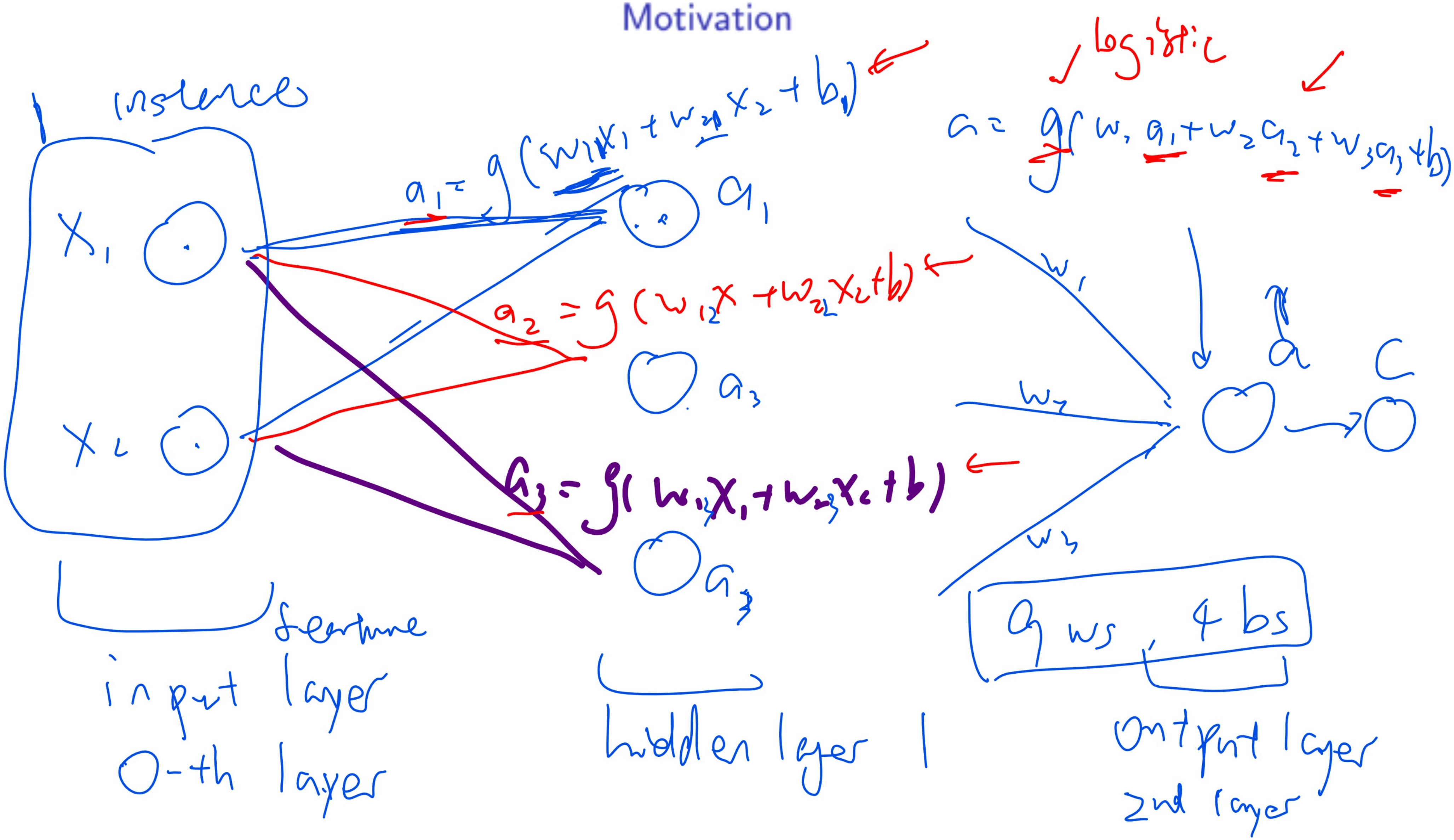
$$\frac{\partial C}{\partial w_j} = \sum^{h} a_i (1 - a_i) x_j \quad (a_i - y_i)$$

- Initialize random weights.
- (Feedforward Step) Evaluate the activation functions.
- (Backpropagation Step) Compute the gradient of the cost function with respect to each weight and bias using the chain rule.
- Update the weights and biases using gradient descent.
- Repeat until convergent.

$$w = w - \alpha \frac{\partial C}{\partial w_j}$$

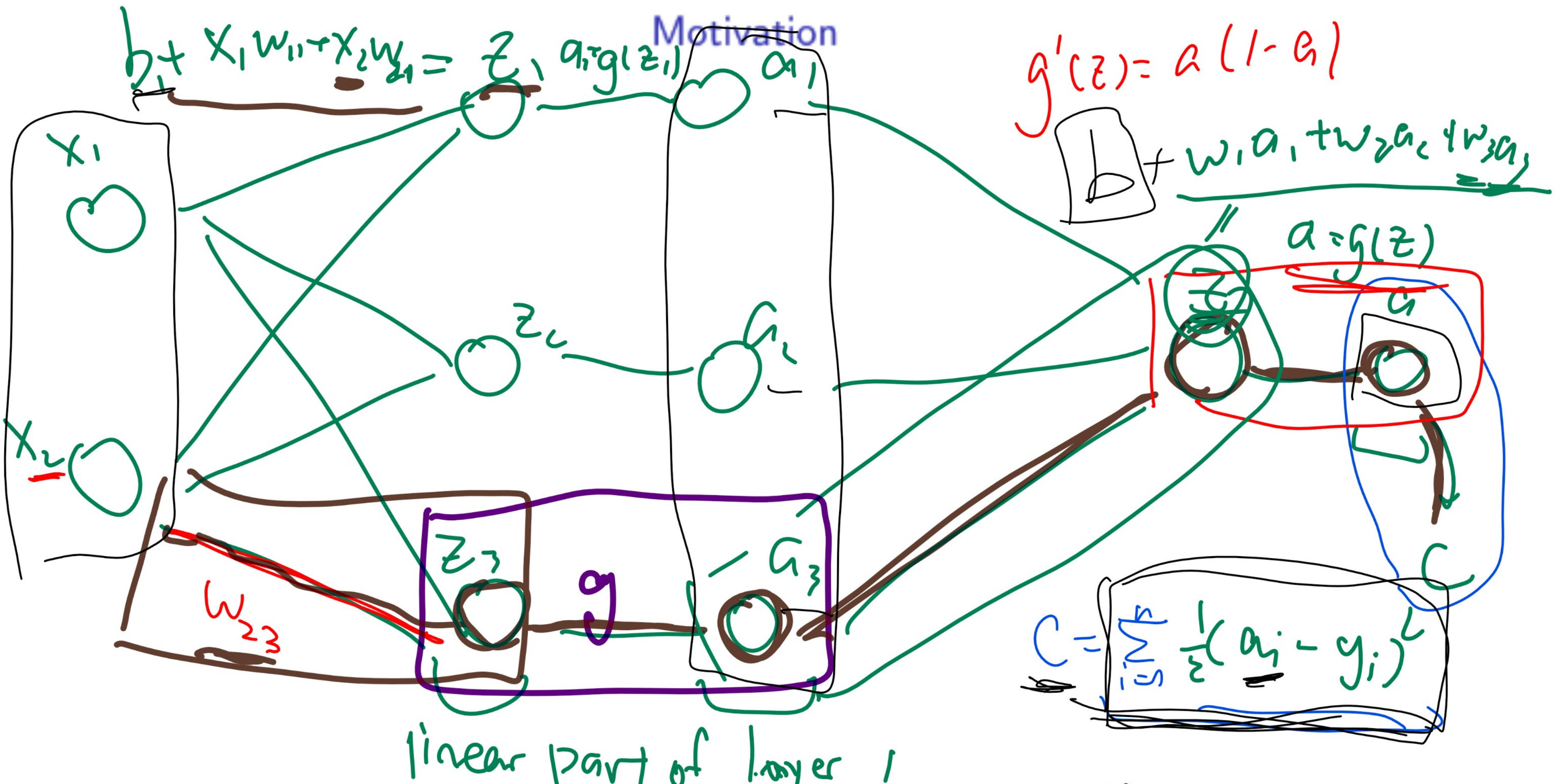
Two Layer Neural Network Weights Diagram 1

Motivation



Two Layer Neural Network Weights Diagram 2

Motivation



$$\frac{\partial C}{\partial w_{23}} = \sum_{i=1}^{n} \frac{\partial C}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial a_3} \cdot \frac{\partial a_3}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_{23}}$$

Two Layer Neural Network Weights Diagram 3

Motivation

$$\frac{\partial C}{\partial b_2} = \sum_{i=1}^{n} (a - y_i)\, a(1-a)\, w_3\, a_3(1-a_3) \cdot x_2$$

Gradient Step, Combined

Definition

- Put everything back into the chain rule formula. (Please check for typos!)

$$\frac{\partial C}{\partial w_{j'j}^{(1)}} = \sum_{i=1}^{n} (a_i - y_i)\, a_i\, (1 - a_i)\, w_j^{(2)} a_{ij}^{(1)} \left(1 - a_{ij}^{(1)}\right) x_{ij'}$$

$$\frac{\partial C}{\partial b_j^{(1)}} = \sum_{i=1}^{n} (a_i - y_i)\, a_i\, (1 - a_i)\, w_j^{(2)} a_{ij}^{(1)} \left(1 - a_{ij}^{(1)}\right)$$

$$\frac{\partial C}{\partial w_j^{(2)}} = \sum_{i=1}^{n} (a_i - y_i)\, a_i\, (1 - a_i)\, a_{ij}^{(1)}$$

$$\frac{\partial C}{\partial b^{(2)}} = \sum_{i=1}^{n} (a_i - y_i)\, a_i\, (1 - a_i)$$

Gradient Descent Step

Definition

- The gradient descent step is the same as the one for logistic regression.

$$w_j^{(2)} \leftarrow w_j^{(2)} - \alpha \frac{\partial C}{\partial w_j^{(2)}}, j = 1, 2, ..., m^{(1)}$$

$$b^{(2)} \leftarrow b^{(2)} - \alpha \frac{\partial C}{\partial b^{(2)}},$$

$$w_{j'j}^{(1)} \leftarrow w_{j'j}^{(1)} - \alpha \frac{\partial C}{\partial w_{j'j}^{(1)}}, j' = 1, 2, ..., m, j = 1, 2, ..., m^{(1)}$$

$$b_j^{(1)} \leftarrow b_j^{(1)} - \alpha \frac{\partial C}{\partial b_j^{(1)}}, j = 1, 2, ..., m^{(1)}$$

AND Operator Data

Quiz

- Sample data for AND

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

Learning AND Operator

Quiz

LTU perceptron

- Which one of the following is AND?
- A: $\hat{y} = \mathbb{1}_{\{1x_1+1x_2-1.5\geqslant 0\}}$
- B: $\hat{y} = \mathbb{1}_{\{1x_1+1x_2-0.5\geqslant 0\}}$
- C: $\hat{y} = \mathbb{1}_{\{-1x_1+0.5\geqslant 0\}}$
- D: $\hat{y} = \mathbb{1}_{\{-1x_1-1x_2+0.5\geqslant 0\}}$
- E: None of the above

x_1	x_2	y	$\hat{y}$
0	0	0	0
0	1	0	0
1	0	0	0
1	1	1	1

$0 + 0 - 1.5 < 0 \rightarrow \hat{y} = 0$

$1 + 1 - 1.5 = 0.5 \geq 0$

OR Operator Data

Quiz

- Sample data for OR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

Learning OR Operator

Quiz

- Which one of the following is OR?
- A: $\hat{y} = \mathbb{1}_{\{1x_1+1x_2-1.5\geqslant 0\}}$
- B: $\hat{y} = \mathbb{1}_{\{1x_1+1x_2-0.5\geqslant 0\}}$
- C: $\hat{y} = \mathbb{1}_{\{-1x_1+0.5\geqslant 0\}}$
- D: $\hat{y} = \mathbb{1}_{\{-1x_1-1x_2+0.5\geqslant 0\}}$
- E: None of the above

XOR Data

Quiz

- Sample data for XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Learning XOR Operator

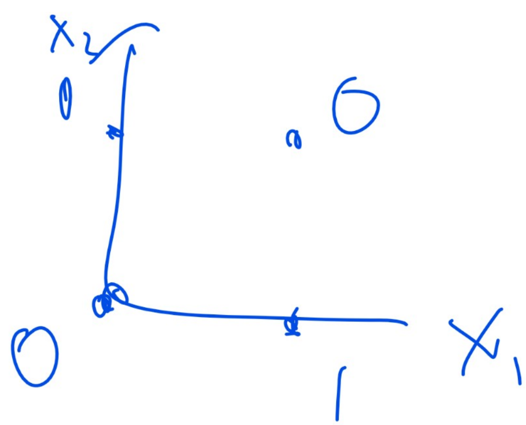
Quiz

0.3

Socrative Room CS540C

- Which one of the following is XOR?
- A: $\hat{y} = \mathbb{1}_{\{1x_1+1x_2-1.5 \geqslant 0\}}$
- B: $\hat{y} = \mathbb{1}_{\{1x_1+1x_2-0.5 \geqslant 0\}}$
- C: $\hat{y} = \mathbb{1}_{\{-1x_1+0.5 \geqslant 0\}}$
- D: $\hat{y} = \mathbb{1}_{\{-1x_1-1x_2+0.5 \geqslant 0\}}$
- E: None of the above

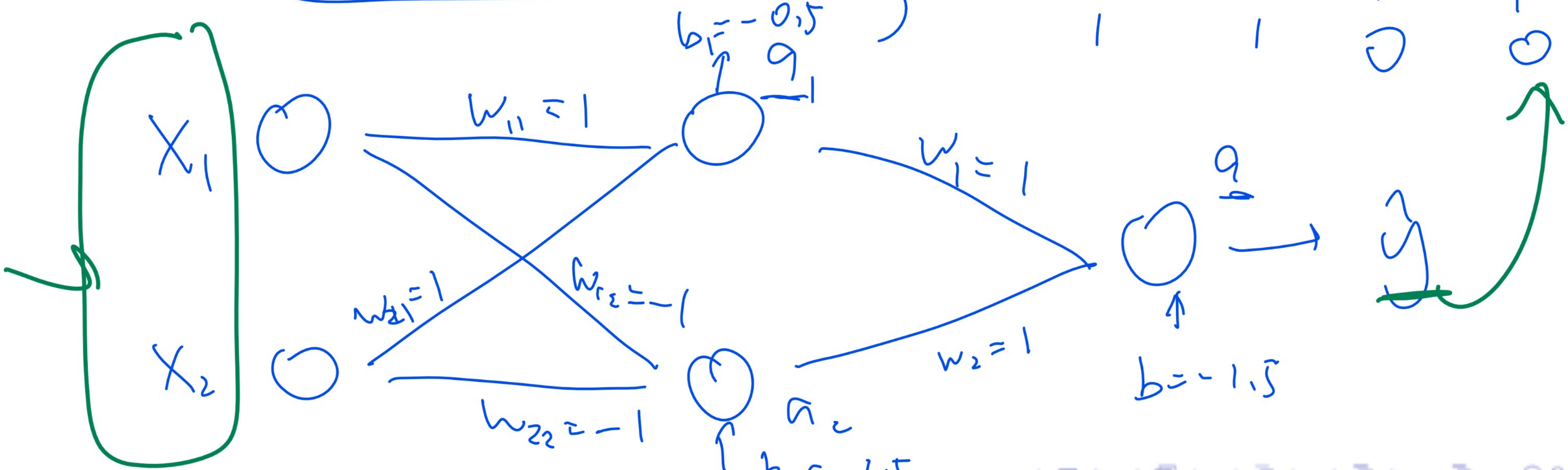
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0



Learning XOR Operator Network

Quiz

- $y = x_1 \text{ XOR } x_2$ is the same as
  $y = (x_1 \text{ OR } x_2) \text{ AND } (x_1 \text{ NAND } x_2)$



Learning XOR Operator Network Diagram

Quiz

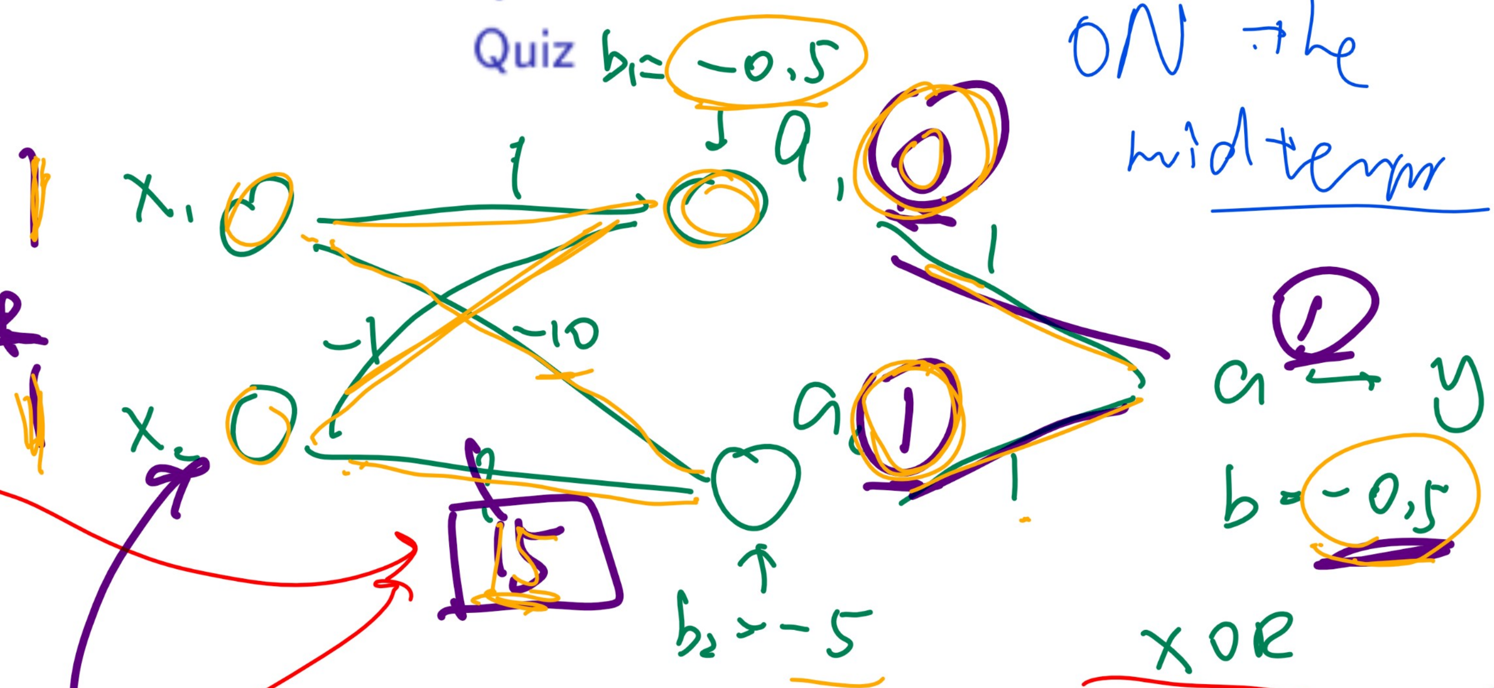
Another XOR Operator Network 1

Quiz

- Given $w_{11}^{(1)} = 1, w_{12}^{(1)} = -10, w_{21}^{(1)} = -1$ and $b_1^{(1)} = -0.5, b_2^{(1)} = -5$ and $w_1^{(2)} = 1, w_1^{(2)} = 1, b^{(2)} = -0.5$. Which of the following value for $w_{22}^{(1)}$ does NOT make the network compute XOR?
- A: 5
- B: 7.5
- C: 10
- D: 12.5
- E: 15

Another XOR Operator Network 2

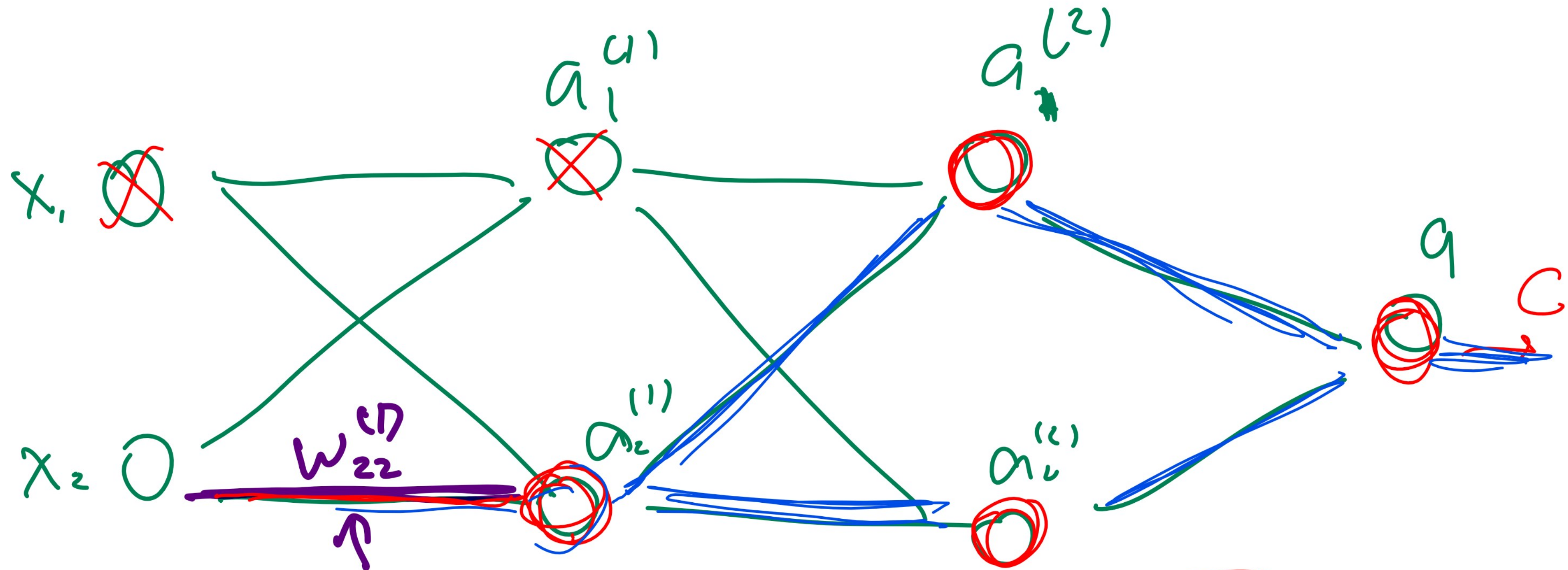
Quiz



- A: 5
- B: 7.5
- C: 10
- D: 12.5
- E: 15

Three Layer Neural Network Weights Diagram 1

Motivation



$$\frac{\partial C}{\partial w_{22}^{(1)}} = \sum_{i=1}^{n} \frac{\partial C}{\partial a} \cdot \frac{\partial a}{\partial z} \left[\frac{\partial z}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial a_2^{(1)}} + \frac{\partial z}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial z_2^{(2)}} \cdot \frac{\partial z_2^{(2)}}{\partial a_2^{(1)}} \right] \cdot \frac{\partial a_2^{(1)}}{\partial z_2^{(1)}} \cdot \frac{\partial z_2^{(1)}}{\partial w_{22}^{(1)}}$$

Three Layer Neural Network Backpropogation

Quiz

- Which of the following is correct for a three layer network? Assume there are 10 units in the first layer and 5 units in the second layer.
- Choices on the next page.

Three Layer Neural Network Backpropogation

Quiz

Q5

- A: $\frac{\partial C}{\partial w_{12}^{(1)}} = \sum_{j'=1}^{10} \sum_{j=1}^{5} \frac{\partial C}{\partial a_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial a_{j'}^{(1)}} \frac{\partial a_{j'}^{(1)}}{\partial w_{1j'}^{(1)}}$
- B: $\frac{\partial C}{\partial w_{12}^{(1)}} = \sum_{j=1}^{5} \frac{\partial C}{\partial a_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial a_1^{(1)}} \frac{\partial a_1^{(1)}}{\partial w_{12}^{(1)}}$
- C: $\frac{\partial C}{\partial w_{12}^{(1)}} = \sum_{j=1}^{5} \frac{\partial C}{\partial a_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial a_2^{(1)}} \frac{\partial a_2^{(1)}}{\partial w_{12}^{(1)}}$
- D: $\frac{\partial C}{\partial w_{12}^{(1)}} = \frac{\partial C}{\partial a_1^{(2)}} \frac{\partial a_1^{(2)}}{\partial a_1^{(1)}} \frac{\partial a_1^{(1)}}{\partial w_{12}^{(1)}}$
- E: $\frac{\partial C}{\partial w_{12}^{(1)}} = \frac{\partial C}{\partial a_2^{(2)}} \frac{\partial a_2^{(2)}}{\partial a_2^{(1)}} \frac{\partial a_2^{(1)}}{\partial w_{12}^{(1)}}$

10 units in layer 1
5 units in layer 2

1 instance.

$\frac{\partial a}{\partial b} = \frac{\partial a}{\partial z} \frac{\partial z}{\partial b}$

layer 1 unit 2