Neural Networks, Part 1

Original Slides by:
Courtney Anne Ngo
Daniel Stanley Tan, PhD
Arren Antioquia

Updated (AY 2024 – 2025 T1s) by:
Thomas James Tiam-Lee, PhD

Recap: Binary Logistic Regression

X

	x_1	x_2	x_3	x_0
$x^{(1)}$	4	0	3	1
$x^{(2)}$	3	2	1	1
$x^{(3)}$	0	4	3	1
$x^{(4)}$	2	0	2	3
$x^{(5)}$	2	2	0	1

θ

0.4	θ_1
0.3	θ_2
0.2	θ_3
0.25	θ_0

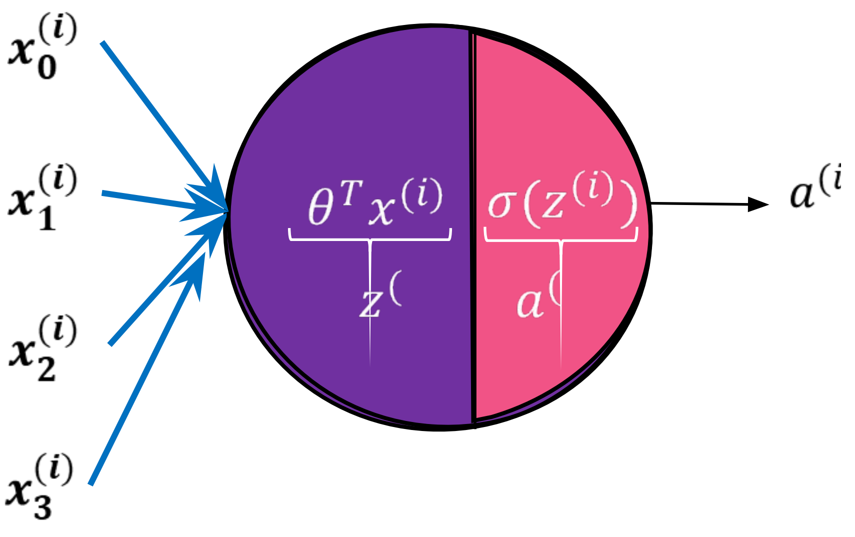
$z = X\theta$

2.45	$z^{(1)}$
2.25	$z^{(2)}$
2.05	$z^{(3)}$
1.95	$z^{(4)}$
1.65	$z^{(5)}$

$$\hat{y} = \sigma(z) = \sigma(X\theta)$$

0.92	$\hat{y}^{(1)}$
0.90	$\hat{y}^{(2)}$
0.89	$\hat{y}^{(3)}$
0.88	$\hat{y}^{(4)}$
0.84	$\hat{y}^{(5)}$

Alternative Representation:



Recap: Binary Logistic Regression

X

	x_1	x_2	x_3	x_0
$x^{(1)}$	4	0	3	1
$x^{(2)}$	3	2	1	1
$x^{(3)}$	0	4	3	1
$x^{(4)}$	2	0	2	3
$x^{(5)}$	2	2	0	1

θ

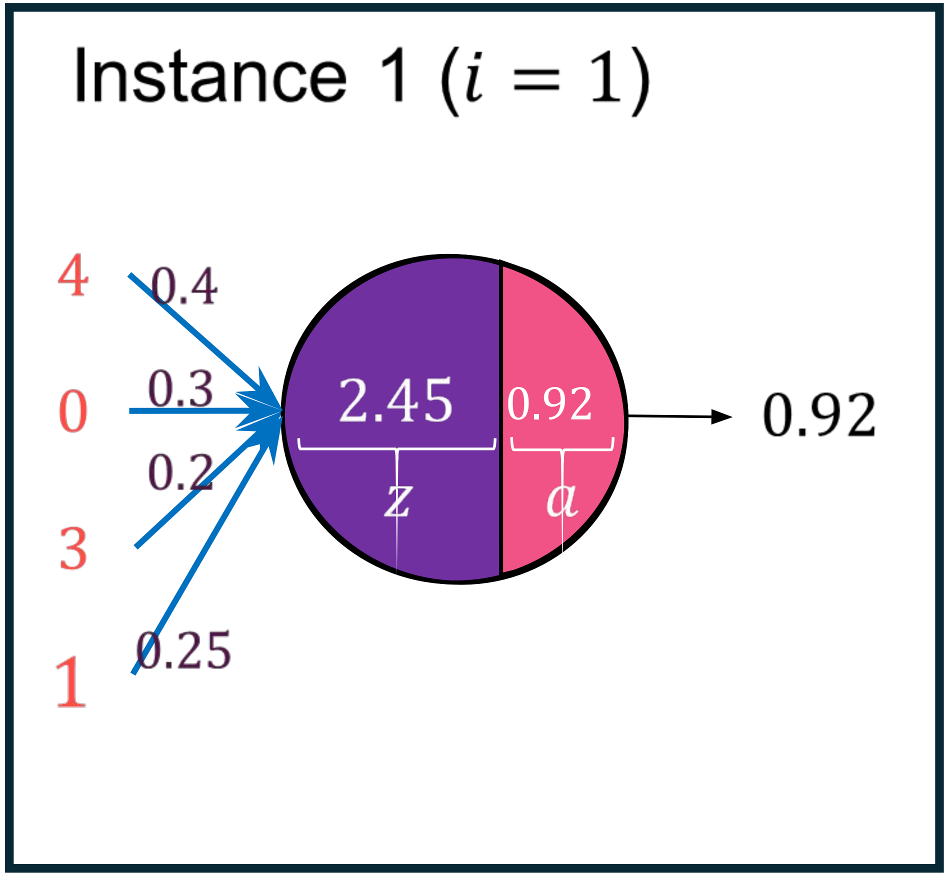
	θ
θ_1	0.4
θ_2	0.3
θ_3	0.2
θ_0	0.25

$z = X\theta$

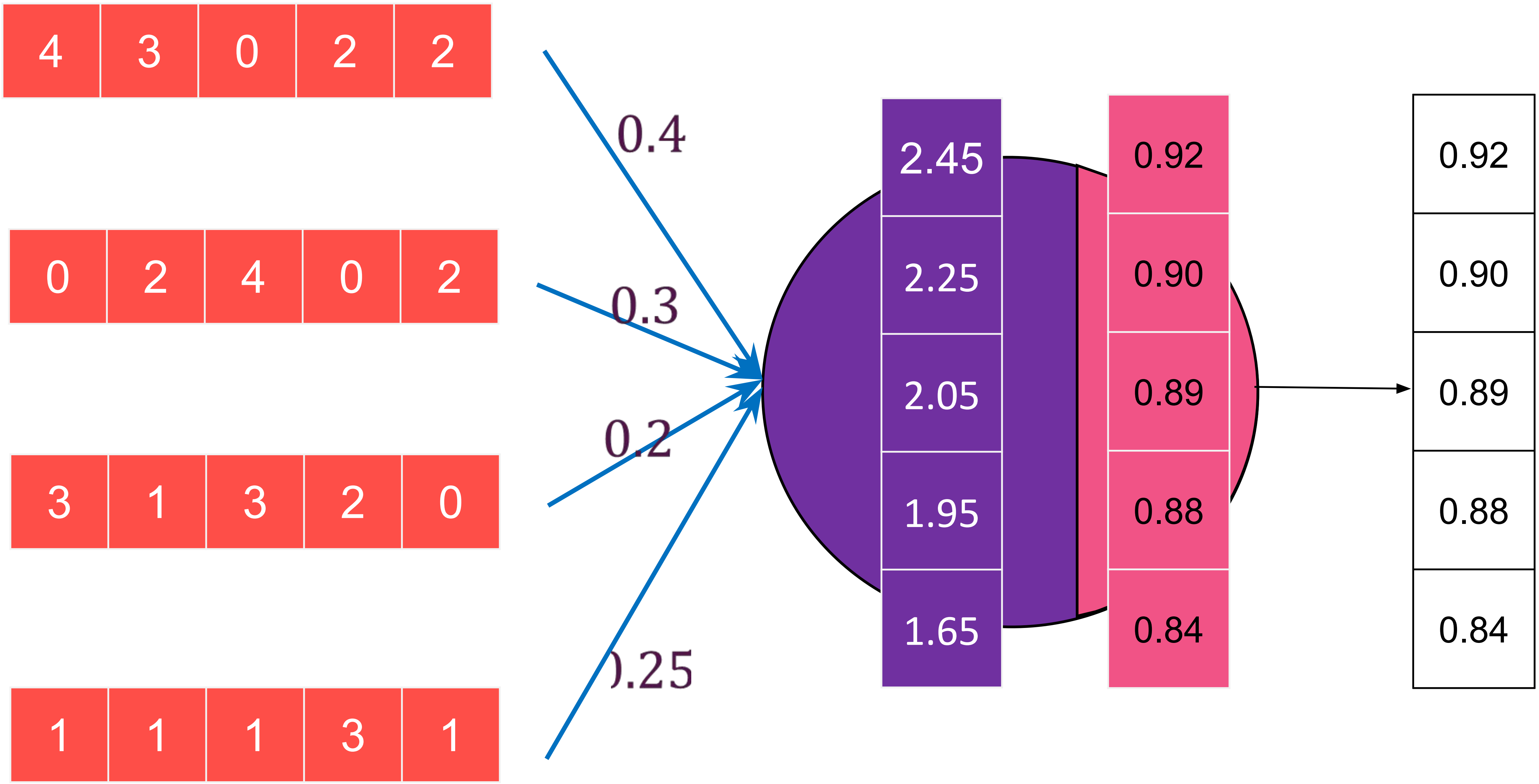
	z
$z^{(1)}$	2.45
$z^{(2)}$	2.25
$z^{(3)}$	2.05
$z^{(4)}$	1.95
$z^{(5)}$	1.65

$$\hat{y} = \sigma(z) = \sigma(X\theta)$$

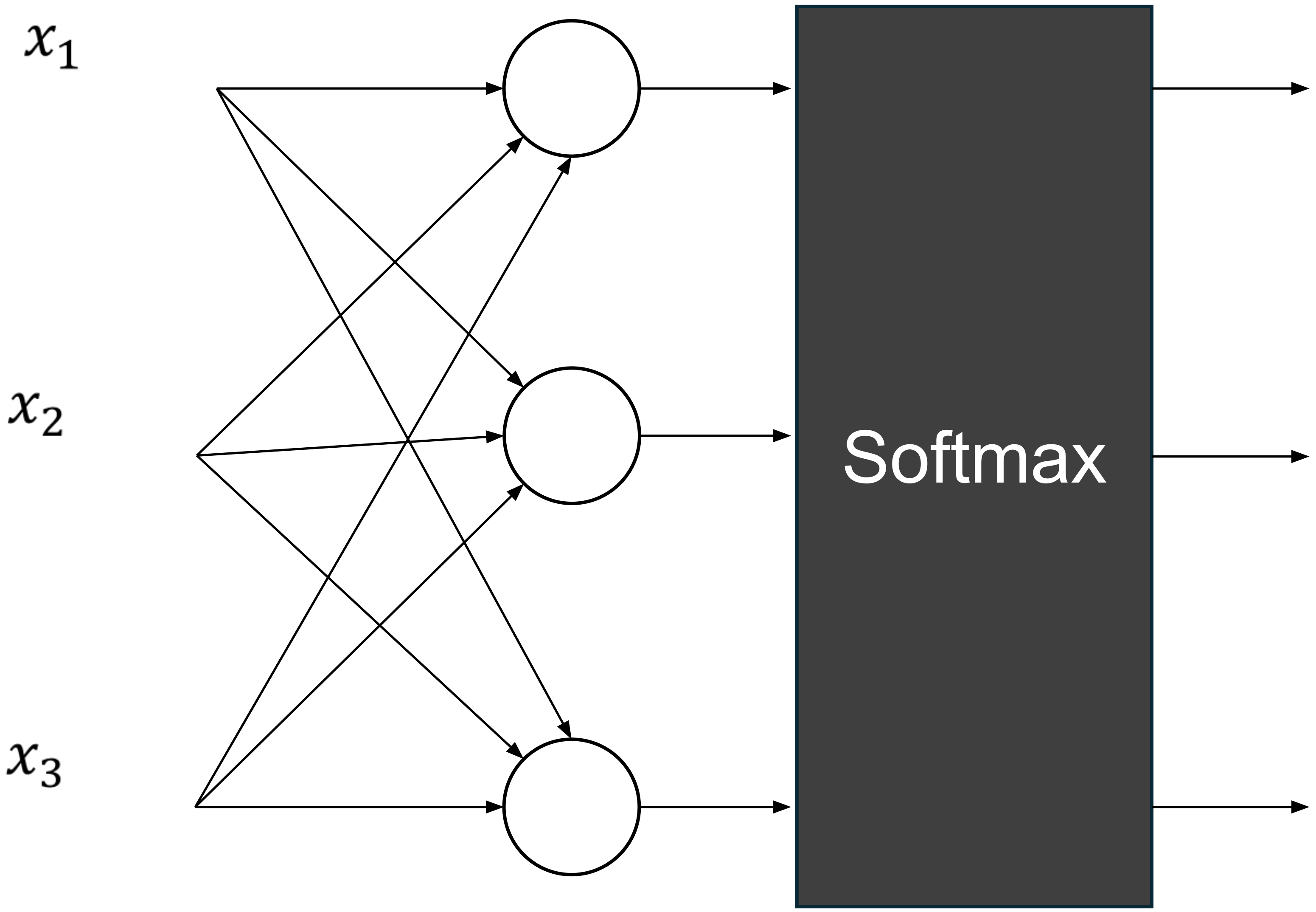
	$\hat{y}$
$\hat{y}^{(1)}$	0.92
$\hat{y}^{(2)}$	0.90
$\hat{y}^{(3)}$	0.89
$\hat{y}^{(4)}$	0.88
$\hat{y}^{(5)}$	0.84



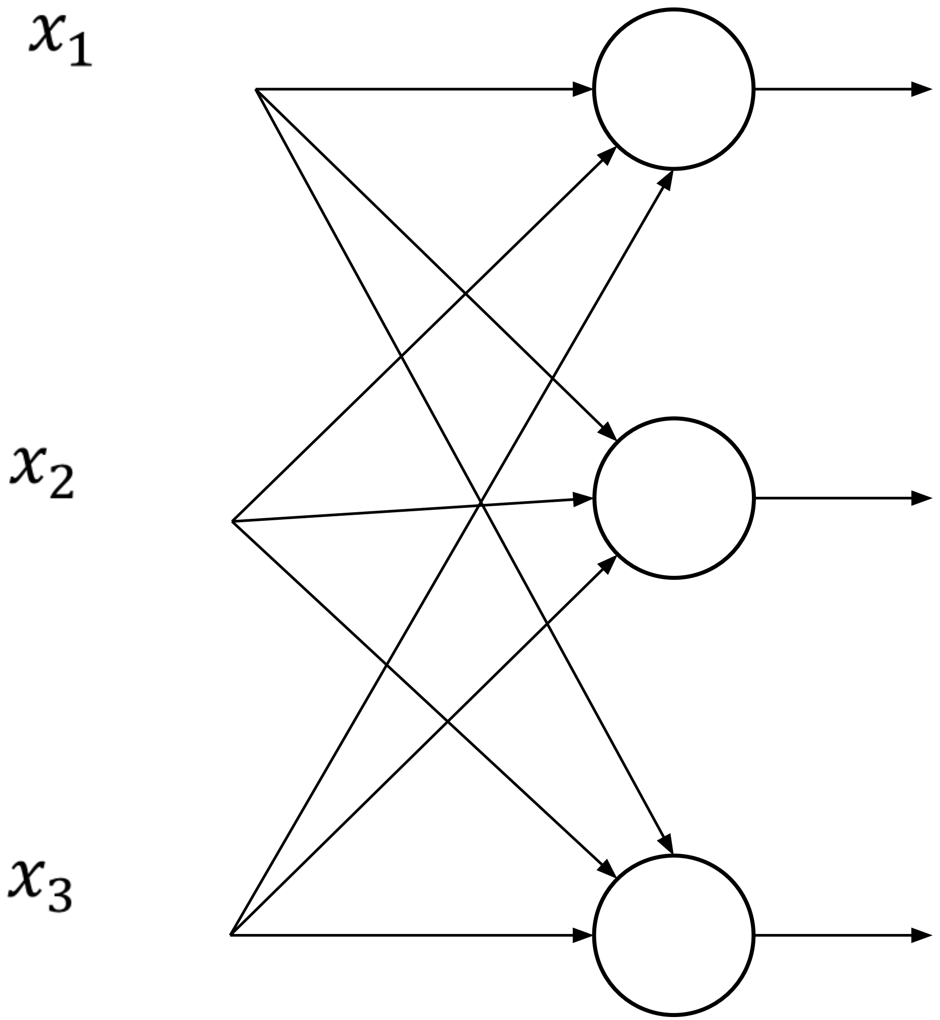
Recap: Binary Logistic Regression



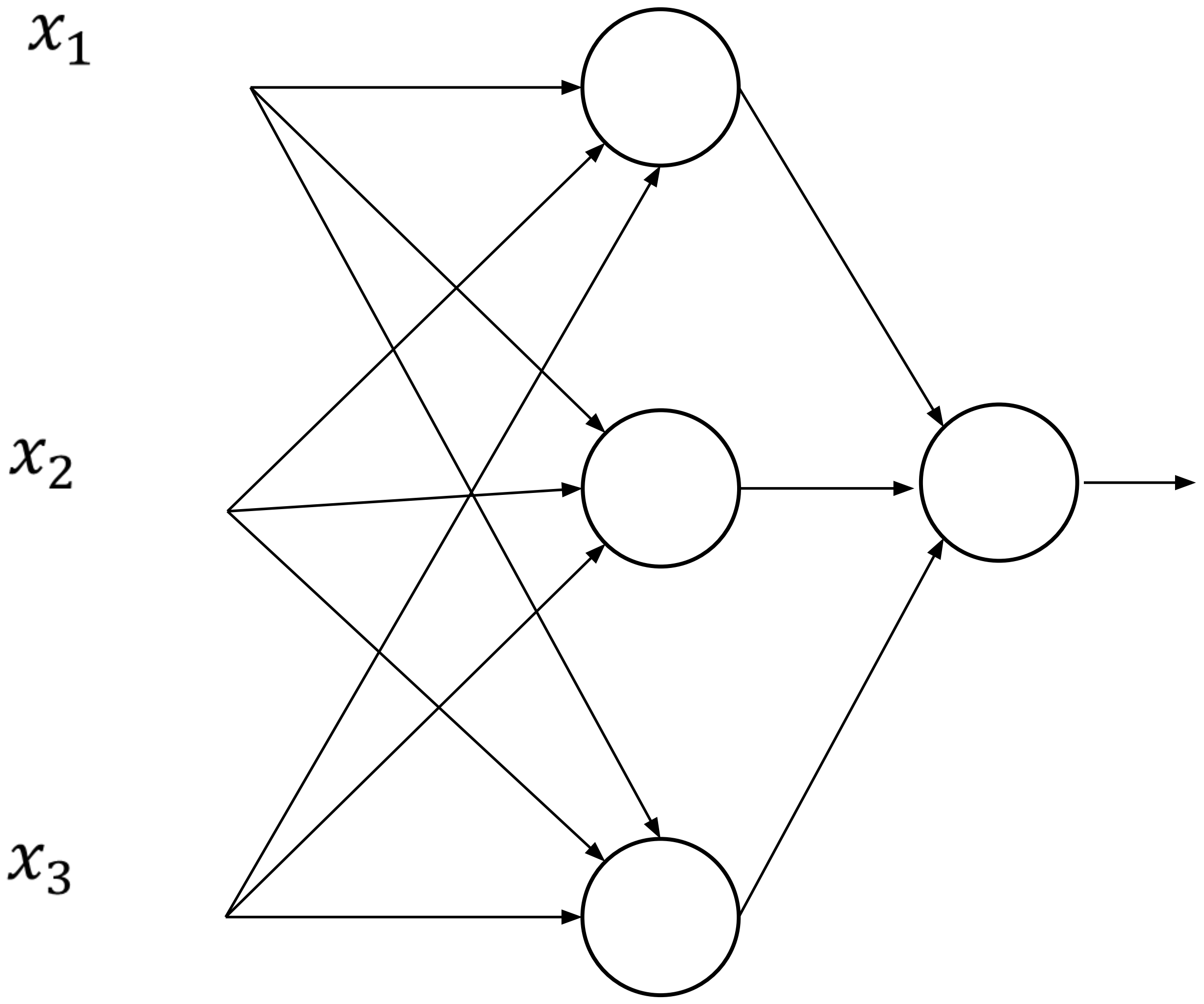
Multinomial Logistic Regression



Neural Network

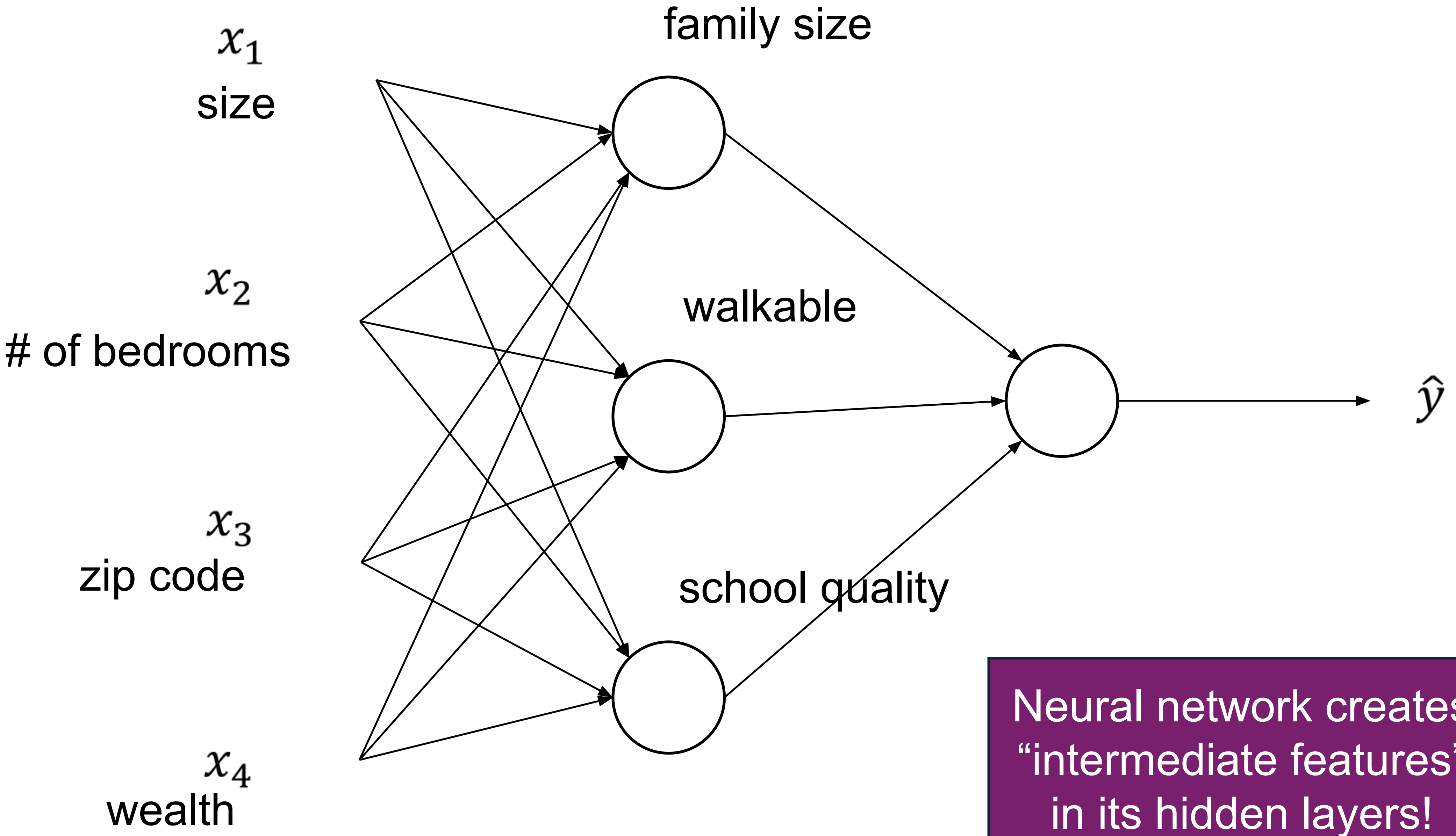


Neural Network



Neural Networks

- Intuitively, you can think of neural networks as just a **bunch of logistic regression units stacked together**!
- Why does this make sense?



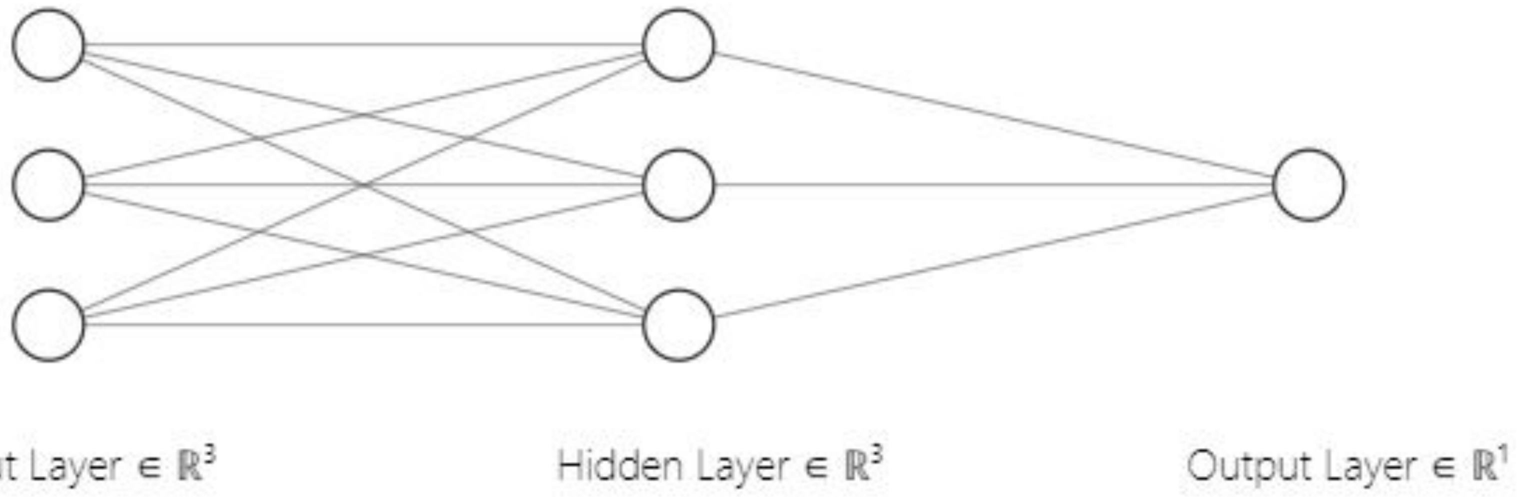
Neural network creates "intermediate features" in its hidden layers!

“Intermediate” Features

- “Intermediate” features are automatically learned by the network.
- The “intermediate” features the neural network creates are **not the same** as the features we humans would create.
- Neural networks have no concept of “walkable” or “school quality” or “family size”
- They will simply create whatever features that would **transform the raw the data into something more linearly separable**!

Neural Networks

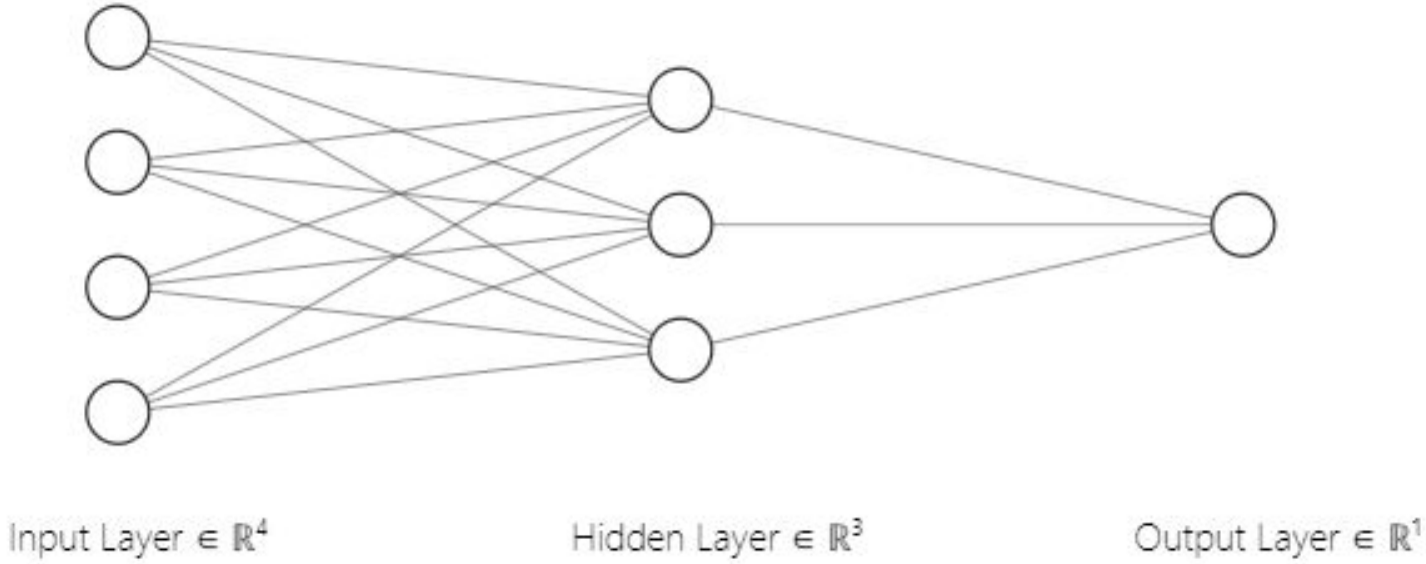
- Neural networks come in all shapes and sizes!



- 2 layer network with 3 inputs, 3 neurons in the hidden layer, and 1 output

Neural Networks

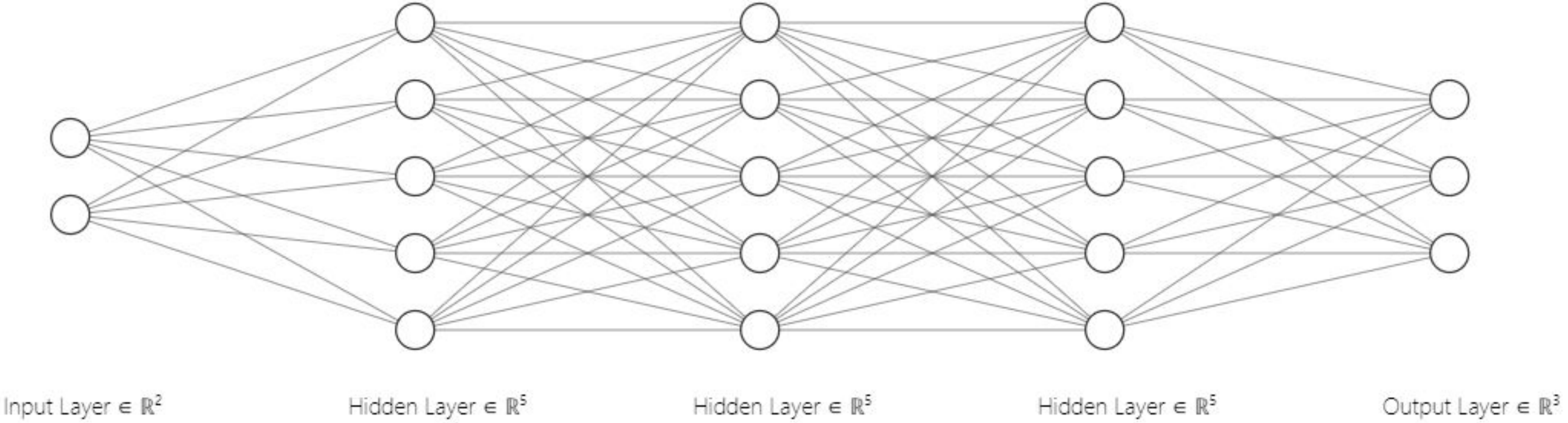
- Neural networks come in all shapes and sizes!



- 2 layer network with 4 inputs, 3 neurons in the hidden layer, and 1 output

Neural Networks

- Neural networks come in all shapes and sizes!

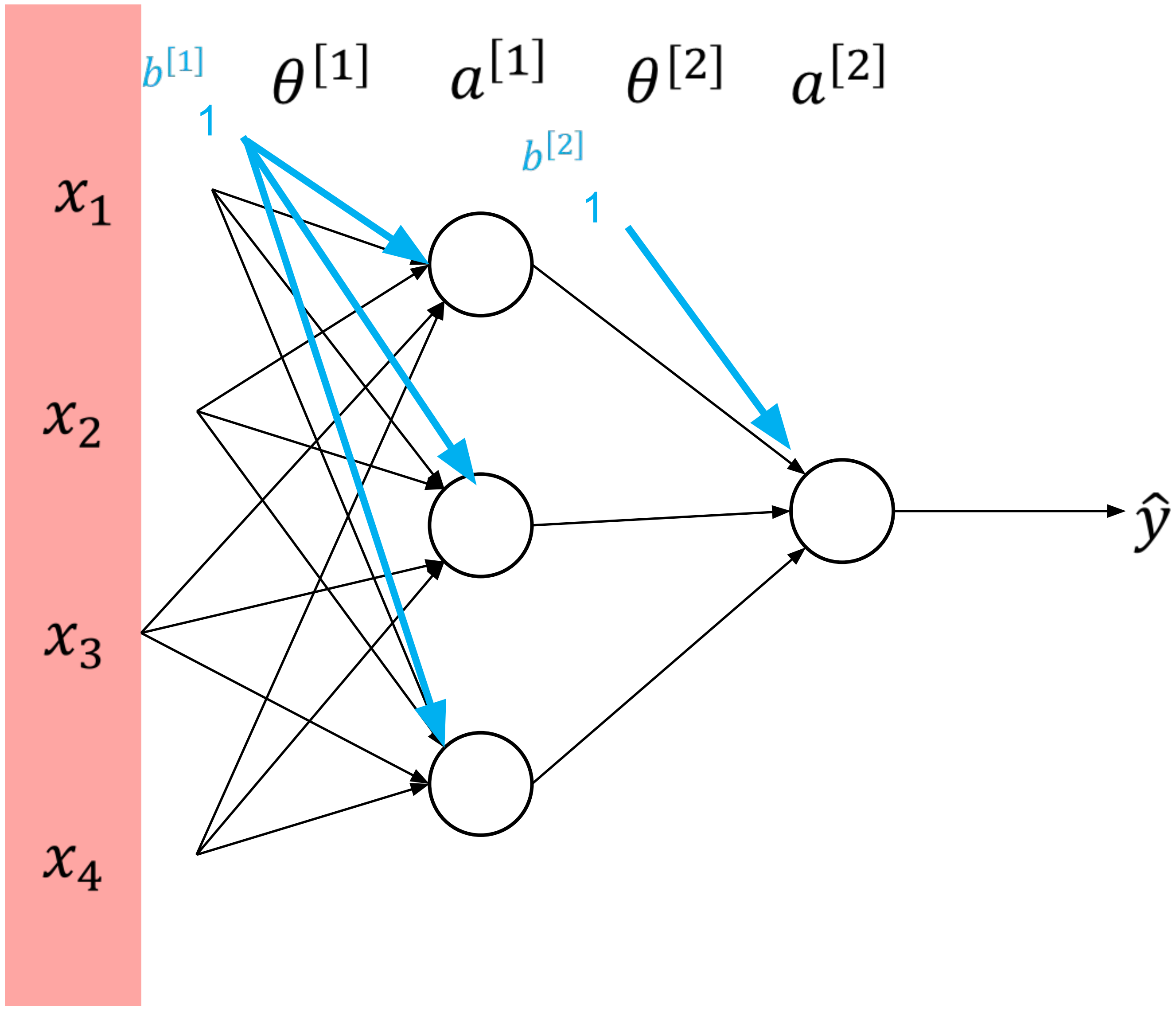


- 4 layer network with 2 inputs, 5 neurons in each hidden layer, and 3 outputs

Anatomy of a Neural Network

- **Input Layer**

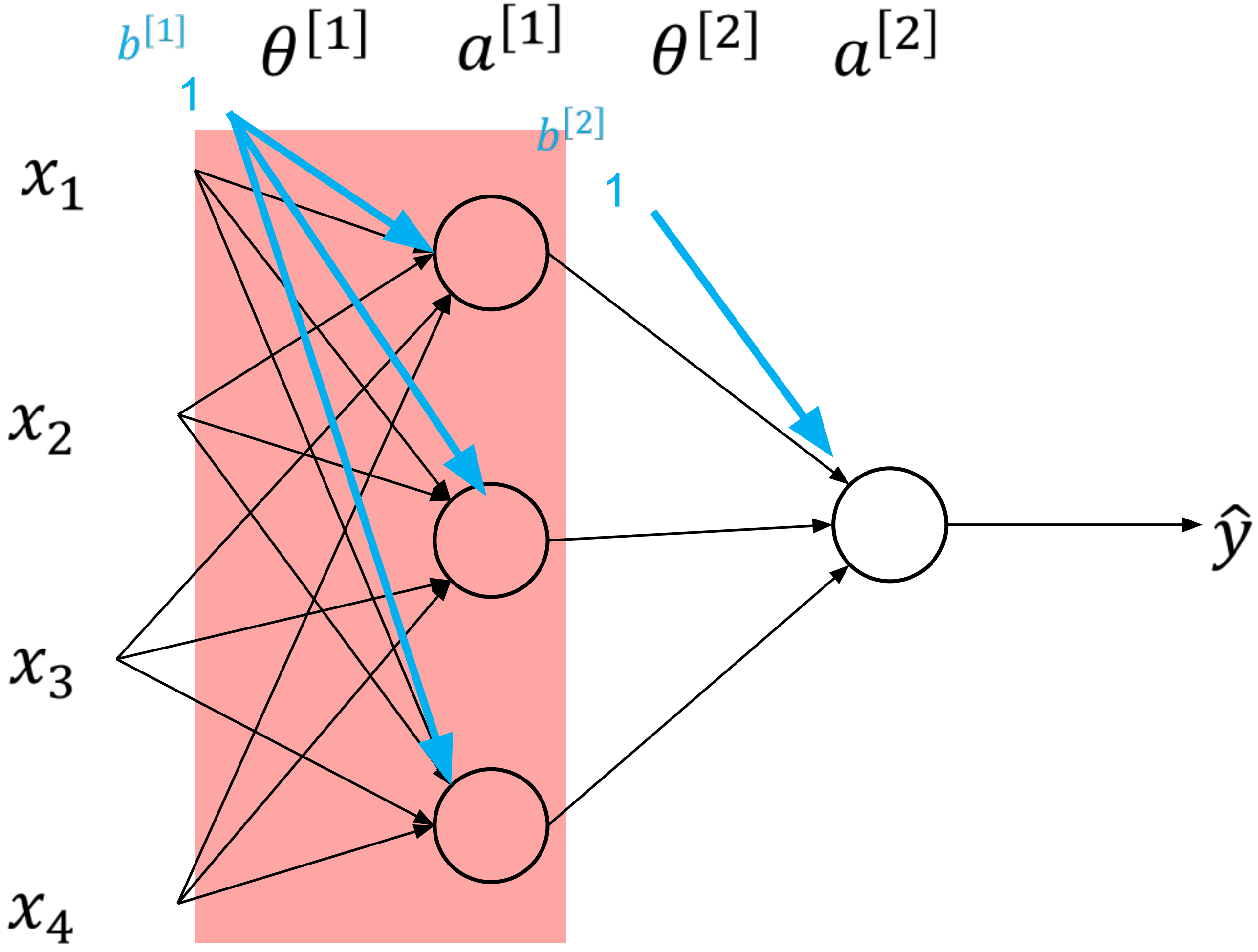
- This is our data X.
- Has the shape $n \times d$
- n = no. of instances
- d = no. of features



Anatomy of a Neural Network

- **Hidden Layer**

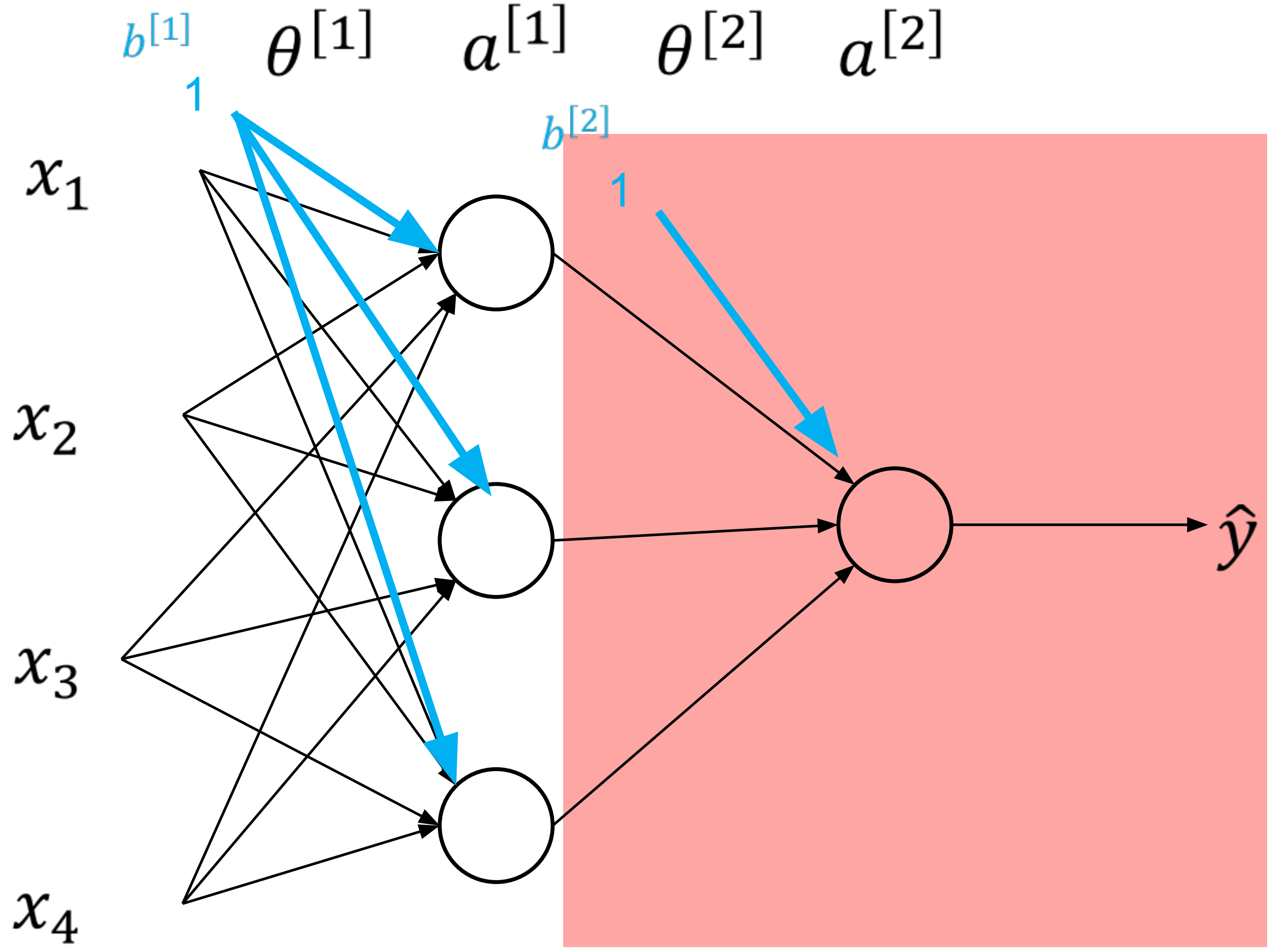
- Gives room for the network to learn other “representations” of your data to lower loss.
- Intermediate representation.



Anatomy of a Neural Network

- **Output Layer**

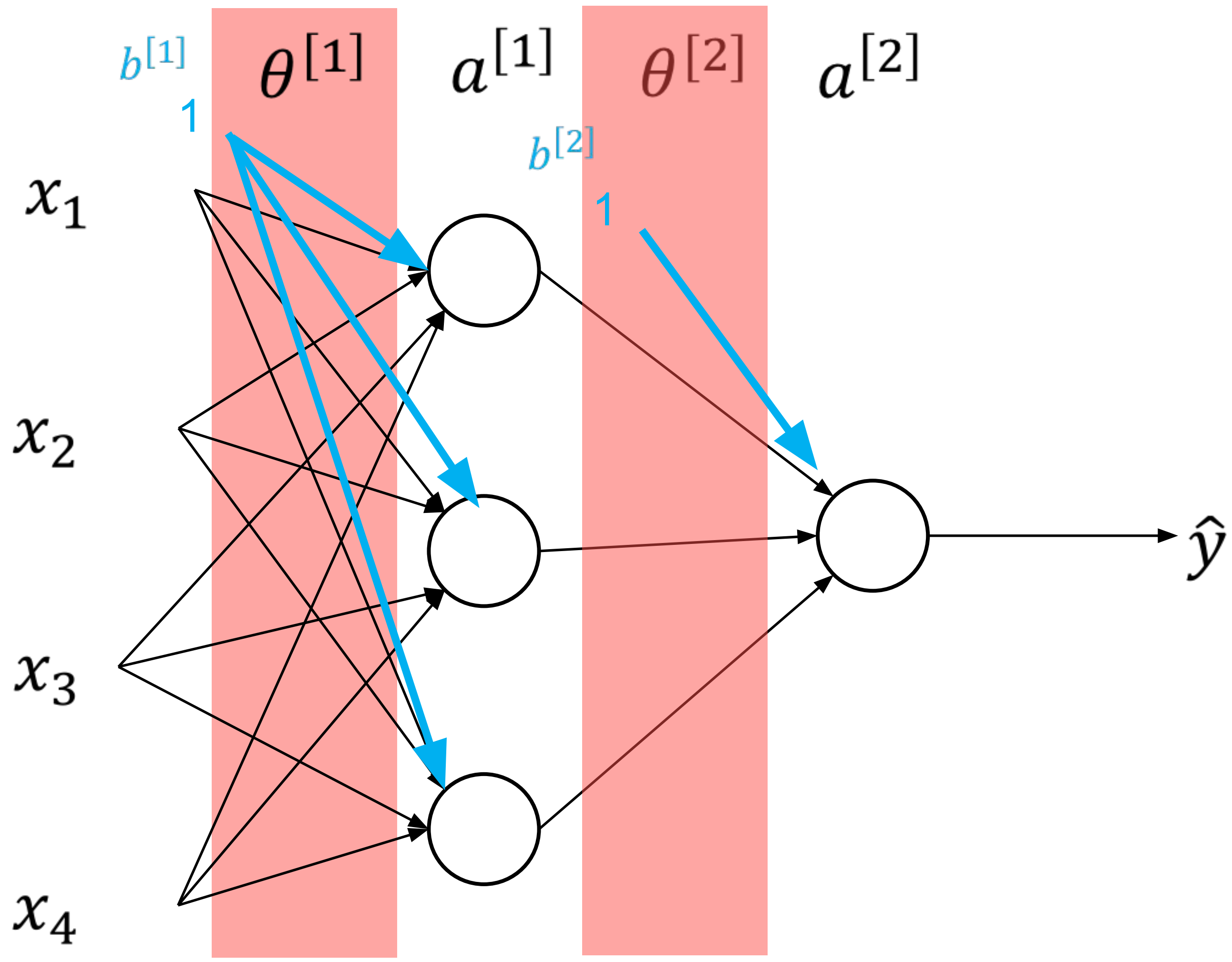
- Outputs the “answer” to your task.
- May consist of one or more neurons.
- Can be designed for regression or classification.



Anatomy of a Neural Network

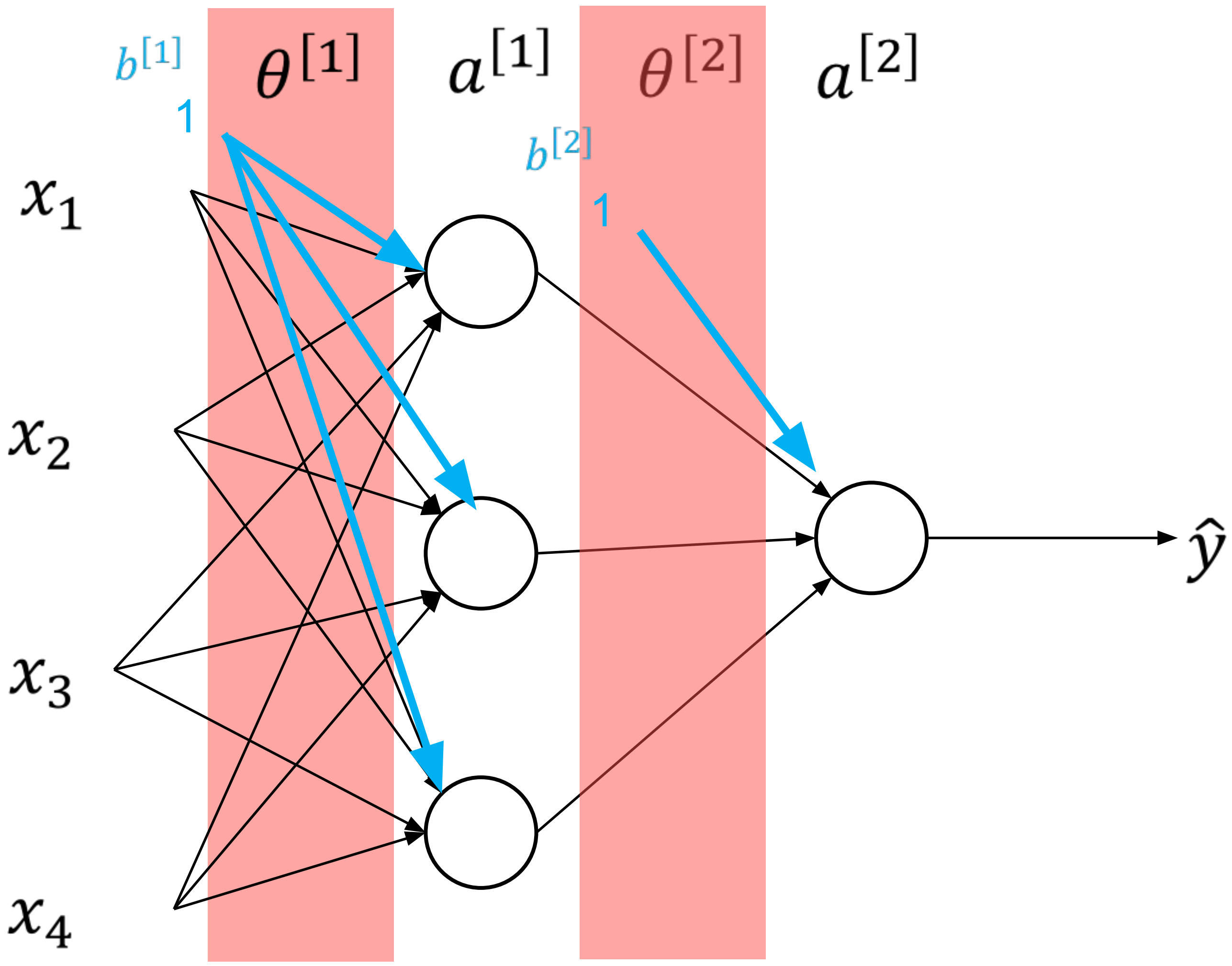
- **Weights**
- Same as the weights in logistic / linear regression, except now we have separate sets for each layer!

- Layer 1 ($\theta^{[1]}$) has 12 parameters
- Layer 2 ($\theta^{[2]}$) has 3 parameters



Anatomy of a Neural Network

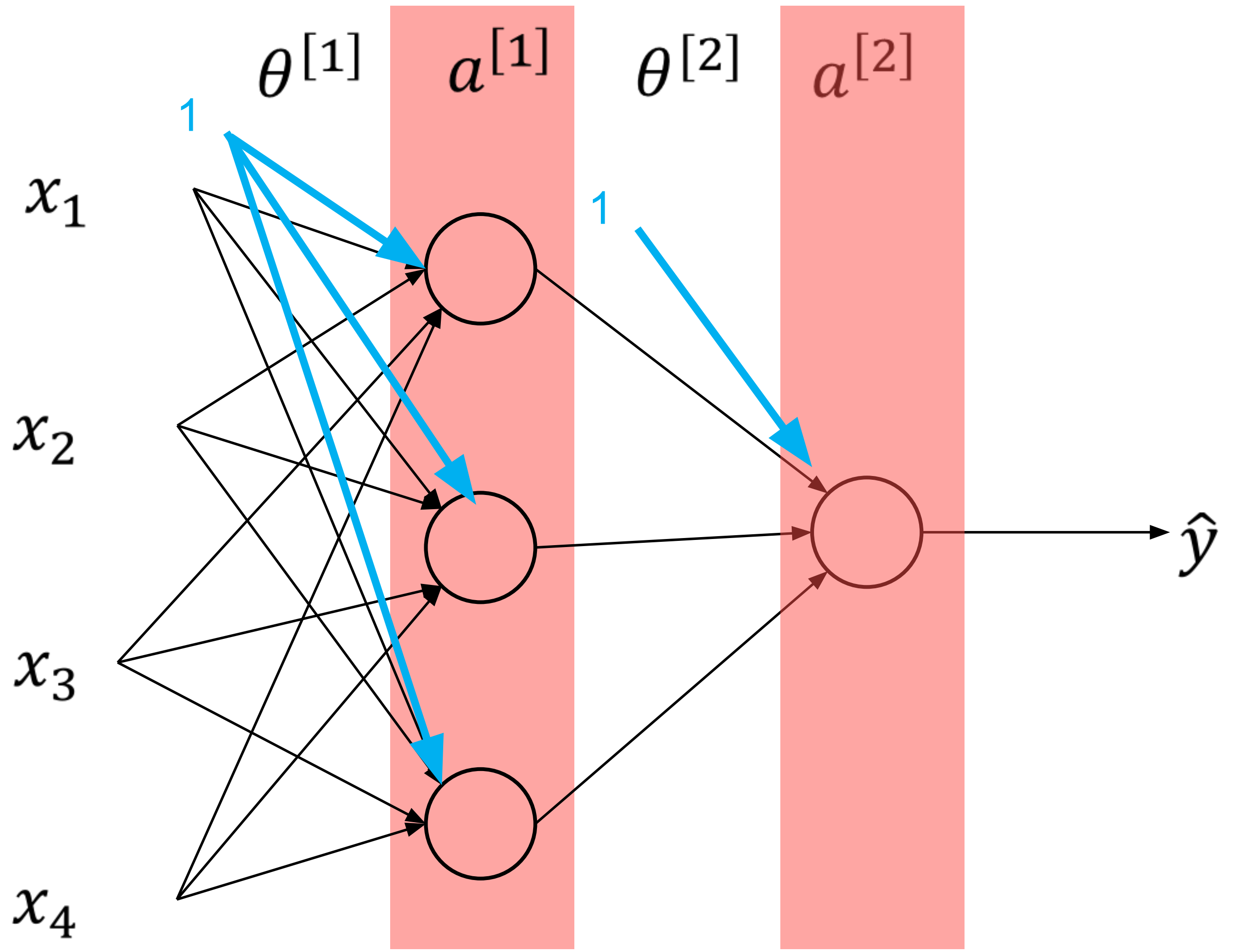
- **Biases (shown in blue)**
- Provides the "intercept" for each layer. Input to these lines is always 1.



Anatomy of a Neural Network

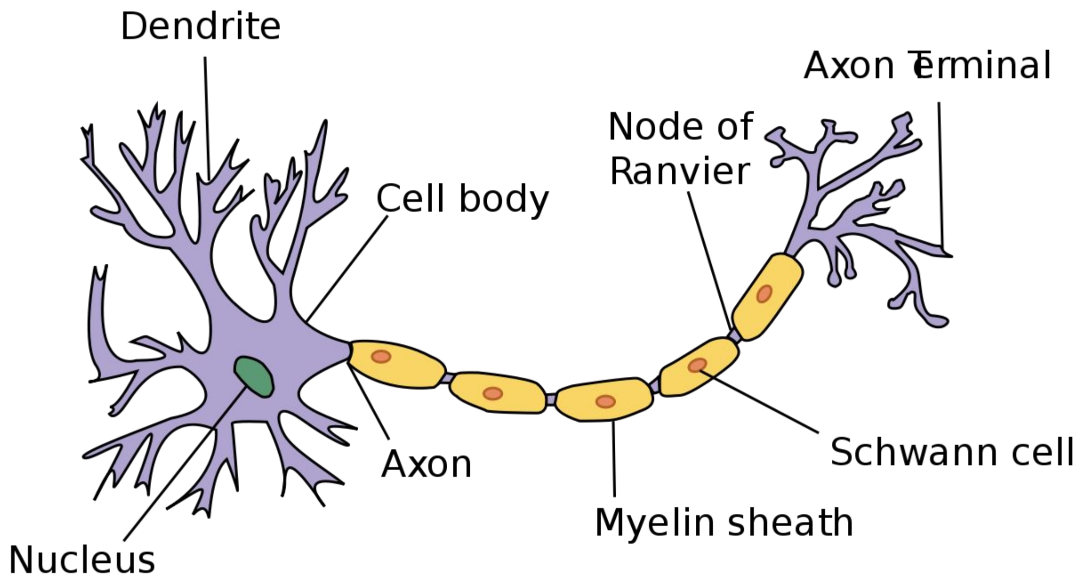
- **Activations**

- Adds non-linearity to the network.
- (example: sigmoid, tanh, ReLU)

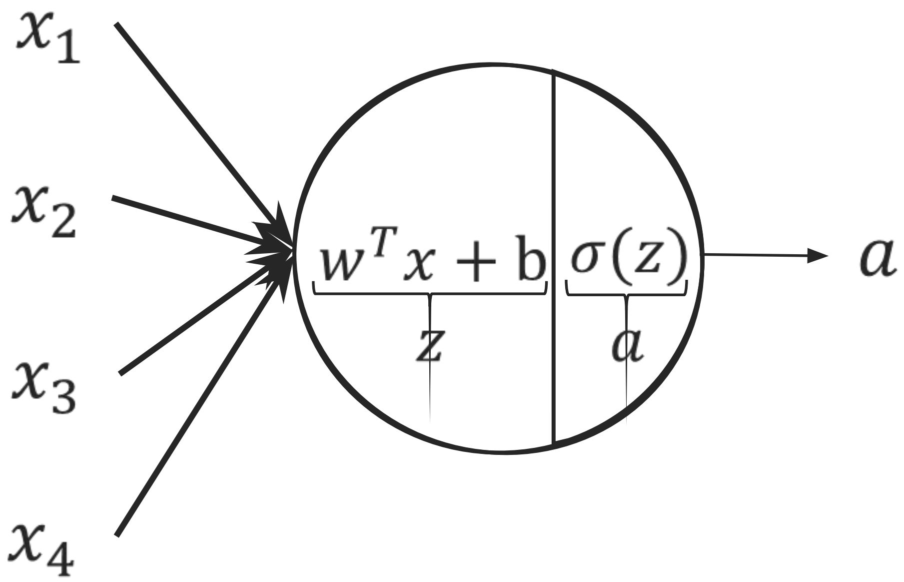


Why is it Called Neural Networks?

A neuron in our brain



A neuron in our neural network



But neurons in our brain are far more complex, and neural networks aren't accurate representations of the brain

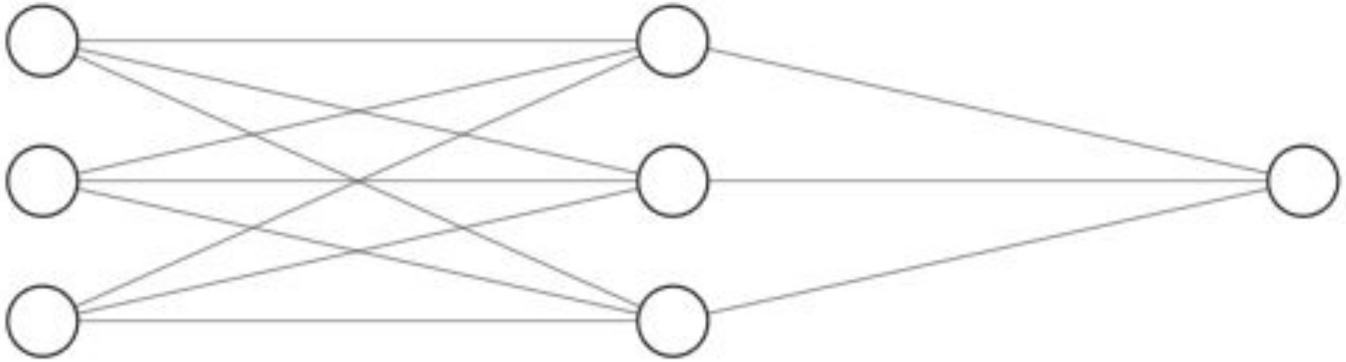
Training a Neural Network

while i < iterations:

    sample training data
    [Forward propagation]   get sample data's predictions
    compute the loss by comparing predictions to labels
    [Backpropagation]   compute the gradients
    Adjust each weight and bias based on the gradient
    i++

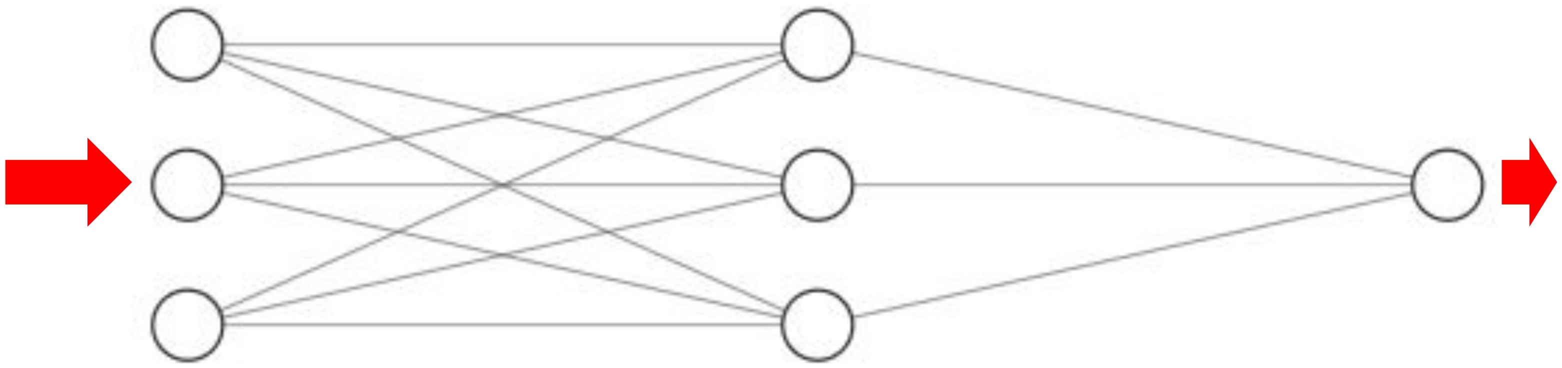
Training a Neural Network

1	9	1
7	8	8
4	8	9
7	1	3
2	3	7
7	6	7
0	4	9



Training a Neural Network

1	9	1
7	8	8
4	8	9
7	1	3
2	3	7
7	6	7
0	4	9



8
9
7
8
4
3
8

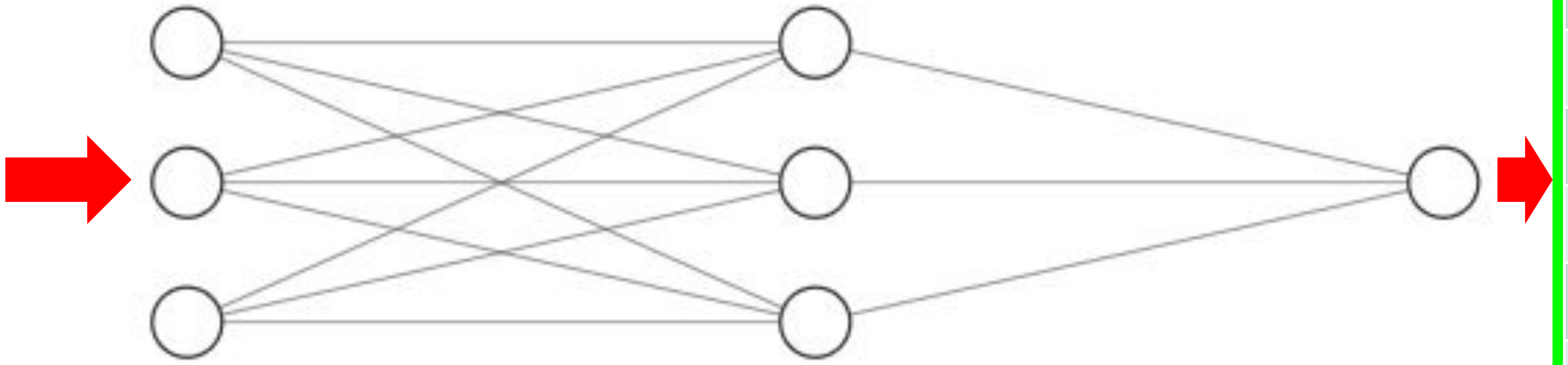
FORWARD PROPAGATION

STEP 1: Feed the input data to the network. Compute the predictions.

Training a Neural Network

LOSS:
Some value

1	9	1
7	8	8
4	8	9
7	1	3
2	3	7
7	6	7
0	4	9



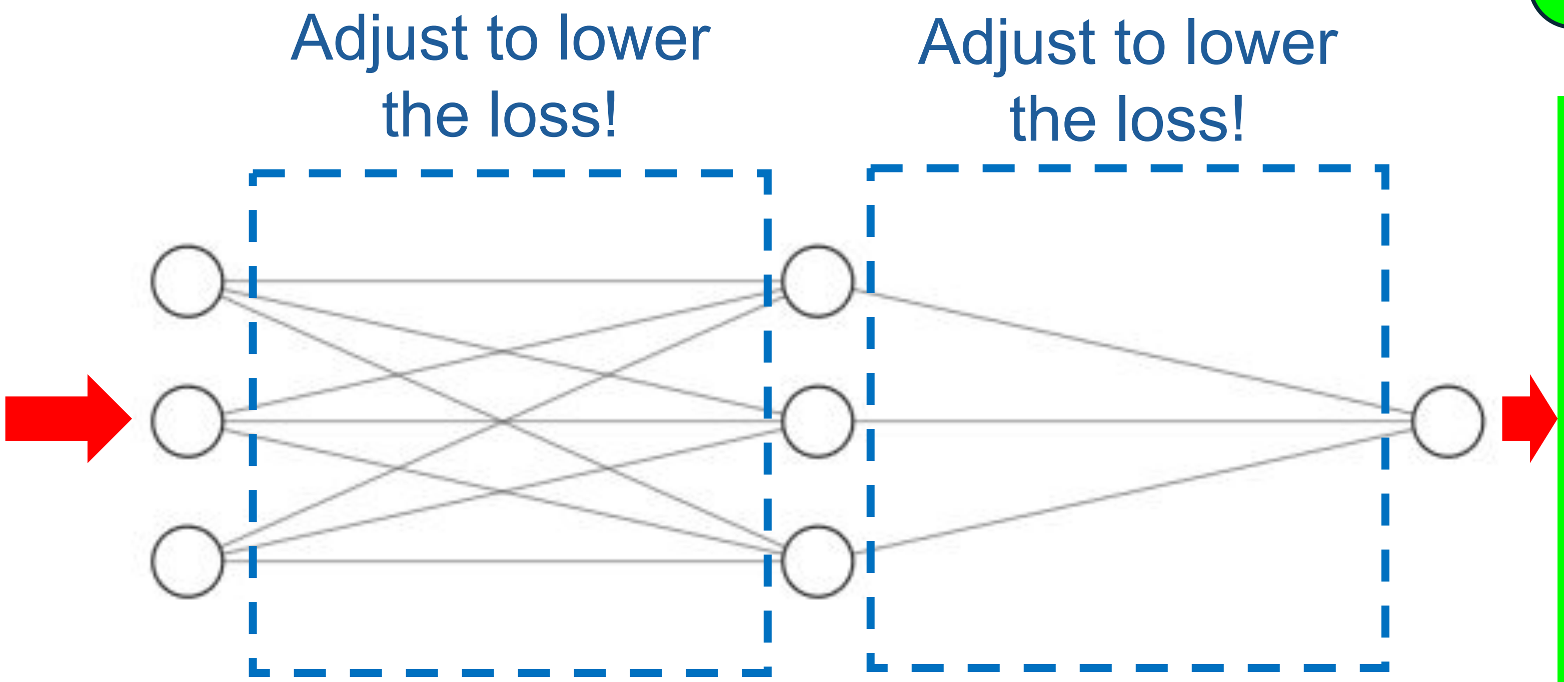
8	9
9	6
7	9
8	8
4	2
3	9
8	7

STEP 2: Compute the **loss**, based on how good the predictions are.

Training a Neural Network

LOSS:
Some value

1	9	1
7	8	8
4	8	9
7	1	3
2	3	7
7	6	7
0	4	9



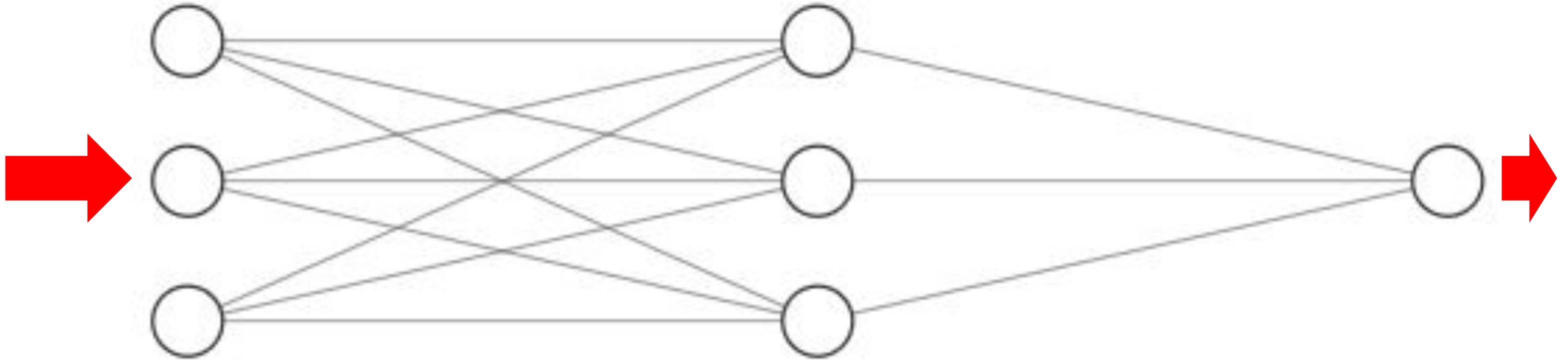
8	9
9	6
7	9
8	8
4	2
3	9
8	7

BACKWARD PROPAGATION

STEP 3: Adjust every weight of the network to the direction that would lower the loss. To know the direction, we need to **compute the derivatives**!

Training a Neural Network

1	9	1
7	8	8
4	8	9
7	1	3
2	3	7
7	6	7
0	4	9

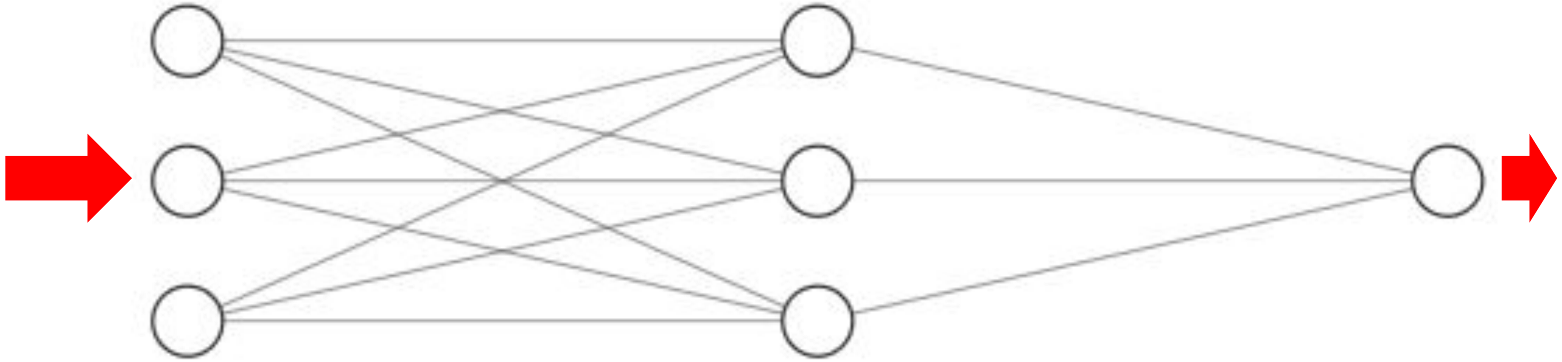


8
7
8
7
4
6
8

STEP 4: Repeat Step 1 again. Feed the input and get the predictions. Hopefully they are a bit better this time!

Training a Neural Network

1	9	1
7	8	8
4	8	9
7	1	3
2	3	7
7	6	7
0	4	9

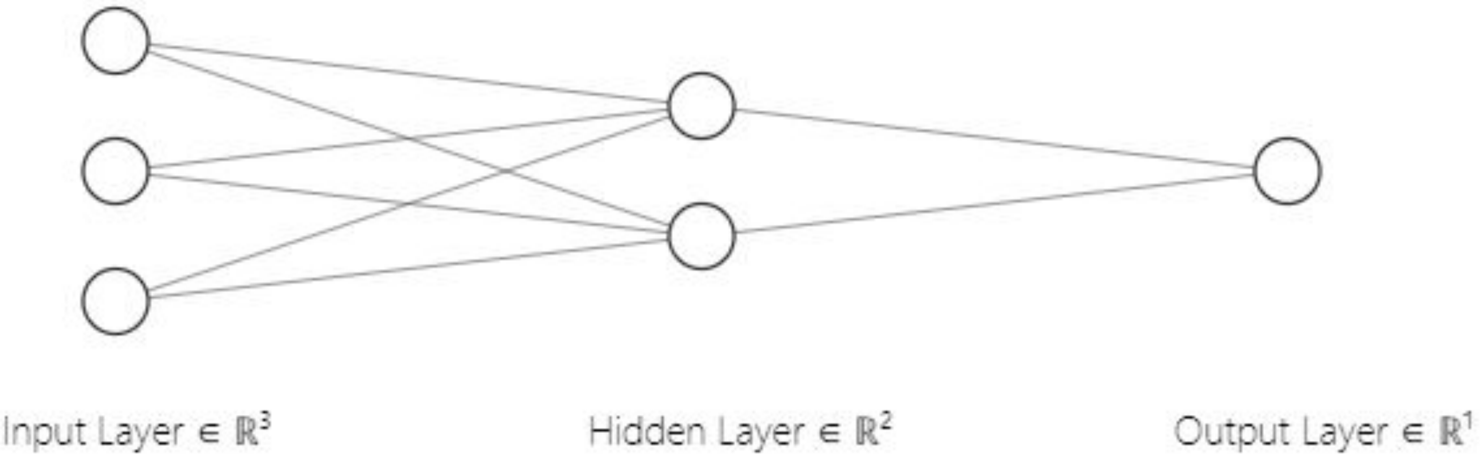


8
7
8
7
4
6
8

Repeat until the desired performance is achieved.

Forward Propagation

0.1	-0.3	1
-1	-0.2	-0.4
0.5	0.1	0.2
-0.2	0.6	0.8



0
0
1
1

$\theta^{[1]}$

0.2	0.5	-0.1
0.1	-0.2	0.3

$b^{[1]}$

0.3
0.6

$\theta^{[2]}$

1.2	0.2

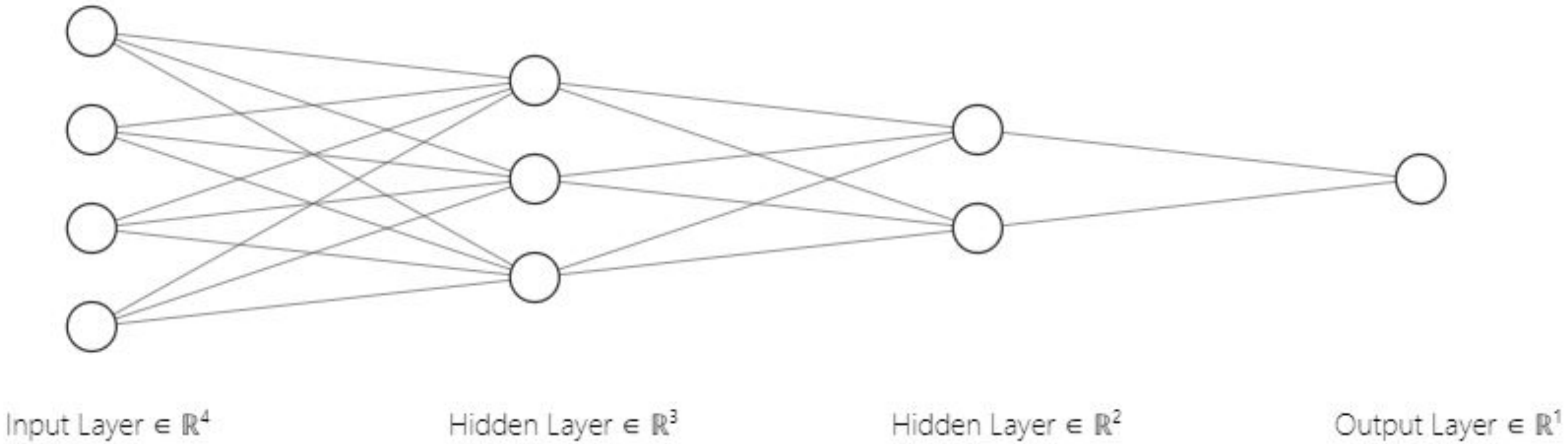
$b^{[2]}$

0

Just let the numbers flow! (can be vectorized)

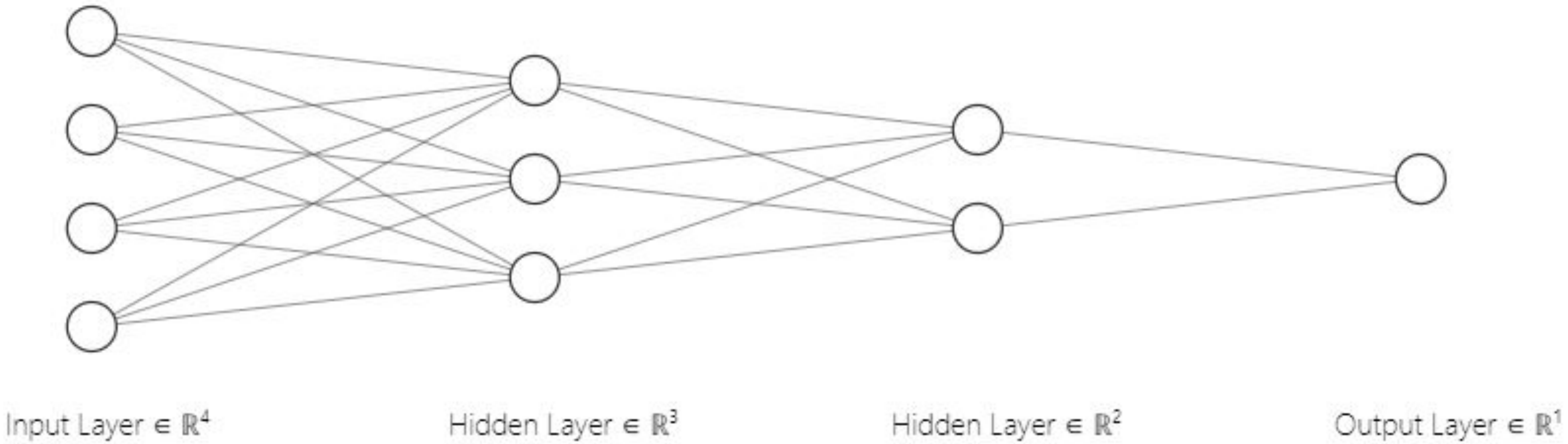
Backward Propagation

- Goal: Just like in linear / logistic regression – **how do we adjust the parameters to make the loss function smaller?**



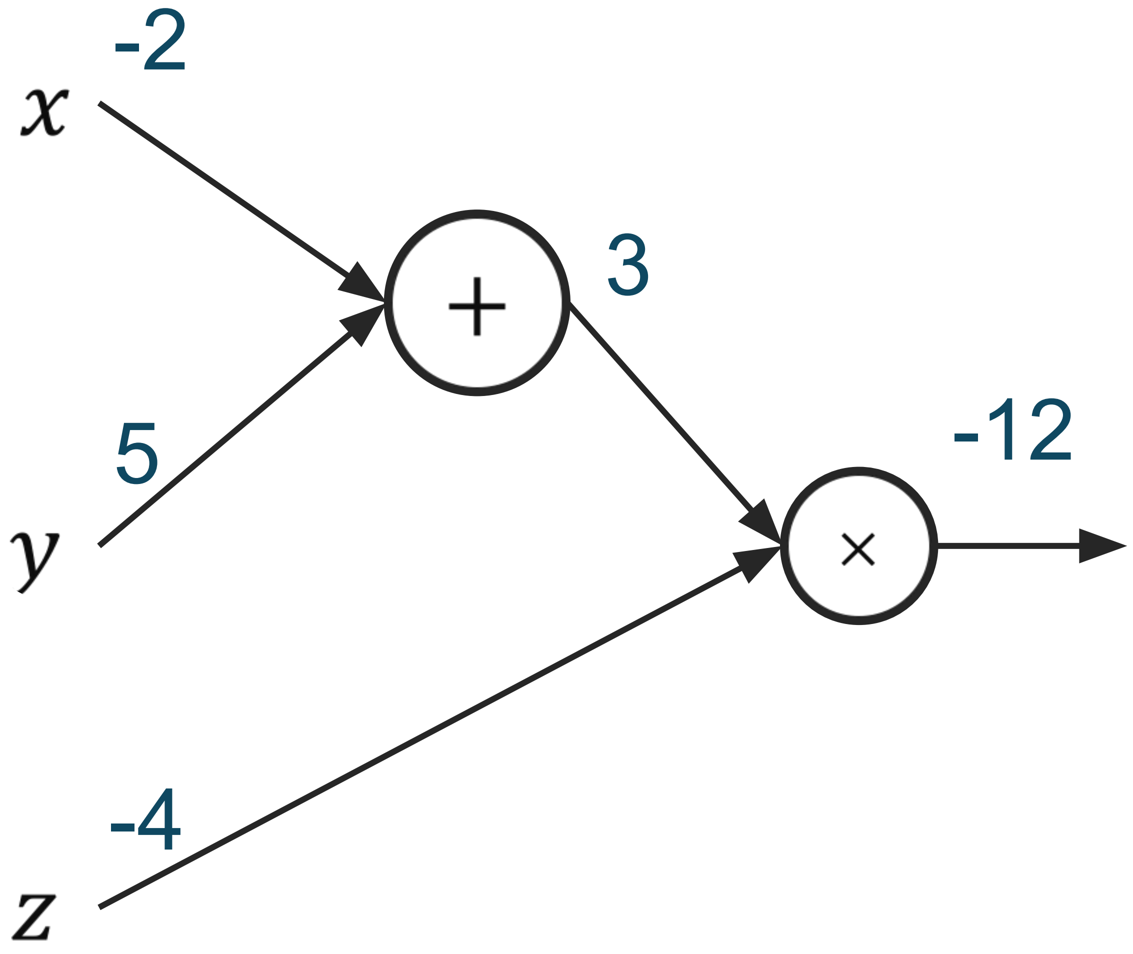
Backward Propagation

- Approach: Use derivatives to compute the **effect of each parameter to the loss function**!



Intuition on Derivatives
(Review from Math)

$$f(x,y,z) = (x+y)z$$

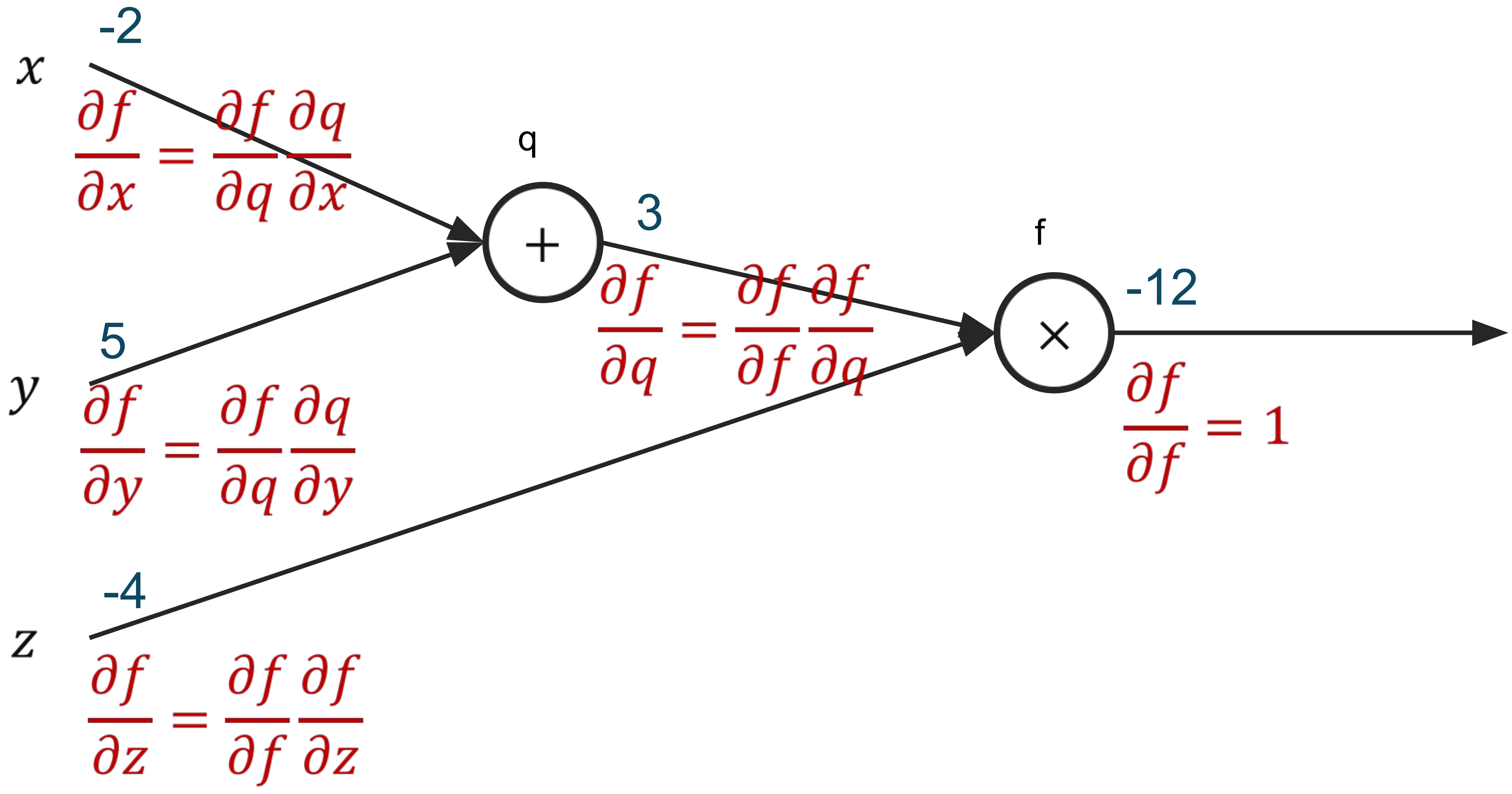


If I add 1 to x, how much will that change f(x,y,z)?

Will it make f lower, or higher?

What if I change y, or z?

Calculate for the gradients! $f(x, y, z) = (x + y)z$



Some Derivative Rules (Recap)

$$f(x) = e^x \qquad \frac{df}{dx} = e^x$$

$$f(x) = \frac{1}{x} \qquad \frac{df}{dx} = -\frac{1}{x^2}$$

$$f(x) = ax \qquad \frac{df}{dx} = a$$

$$f(x) = c + x \qquad \frac{df}{dx} = 1$$

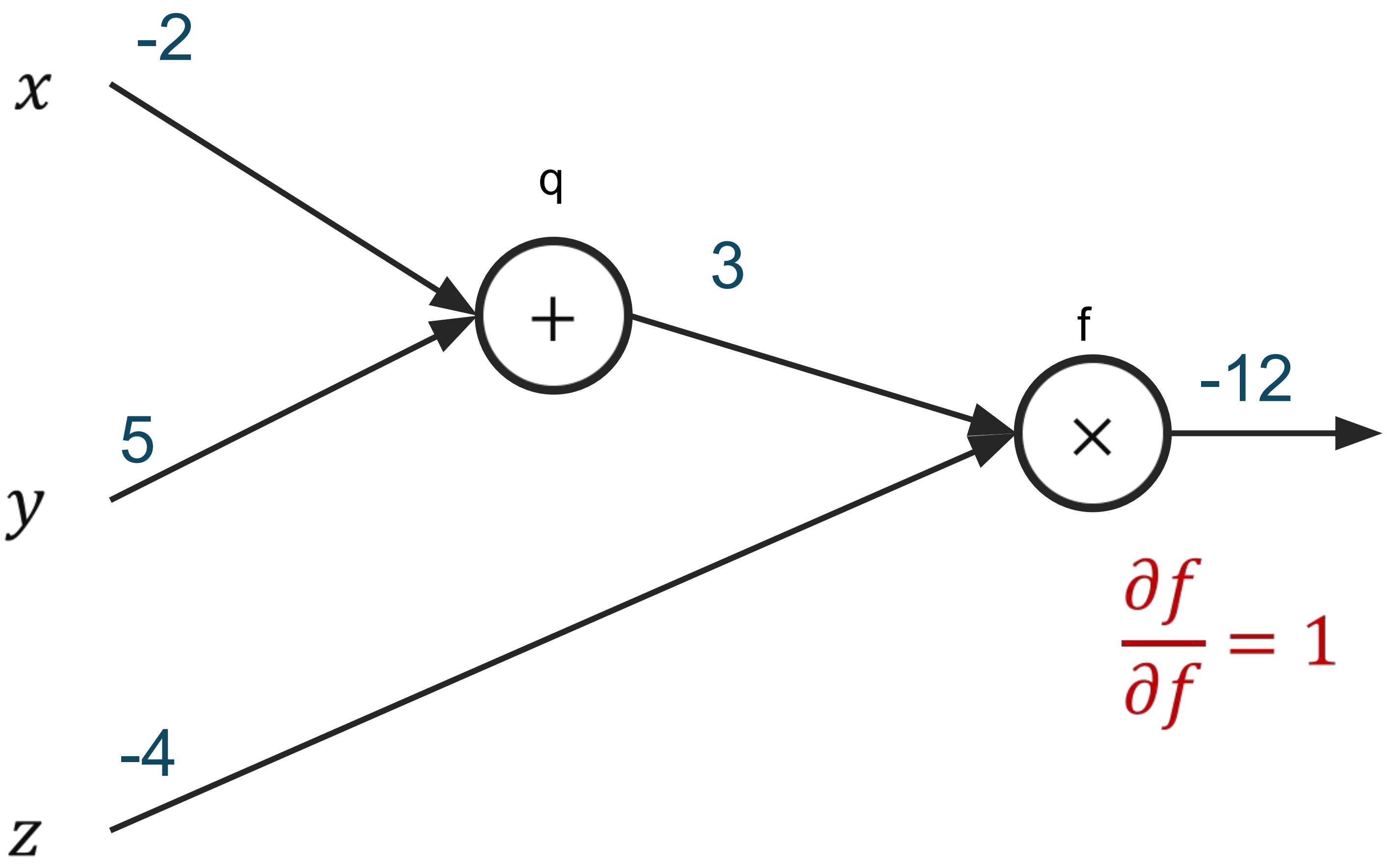
Calculate for the gradients! $f(x, y, z) = (x + y)z$

$q = (x + y)$

$$\frac{\partial q}{\partial x} = 1 , \quad \frac{\partial q}{\partial y} = 1$$

$f = qz$

$$\frac{\partial f}{\partial z} = q , \quad \frac{\partial f}{\partial q} = z$$



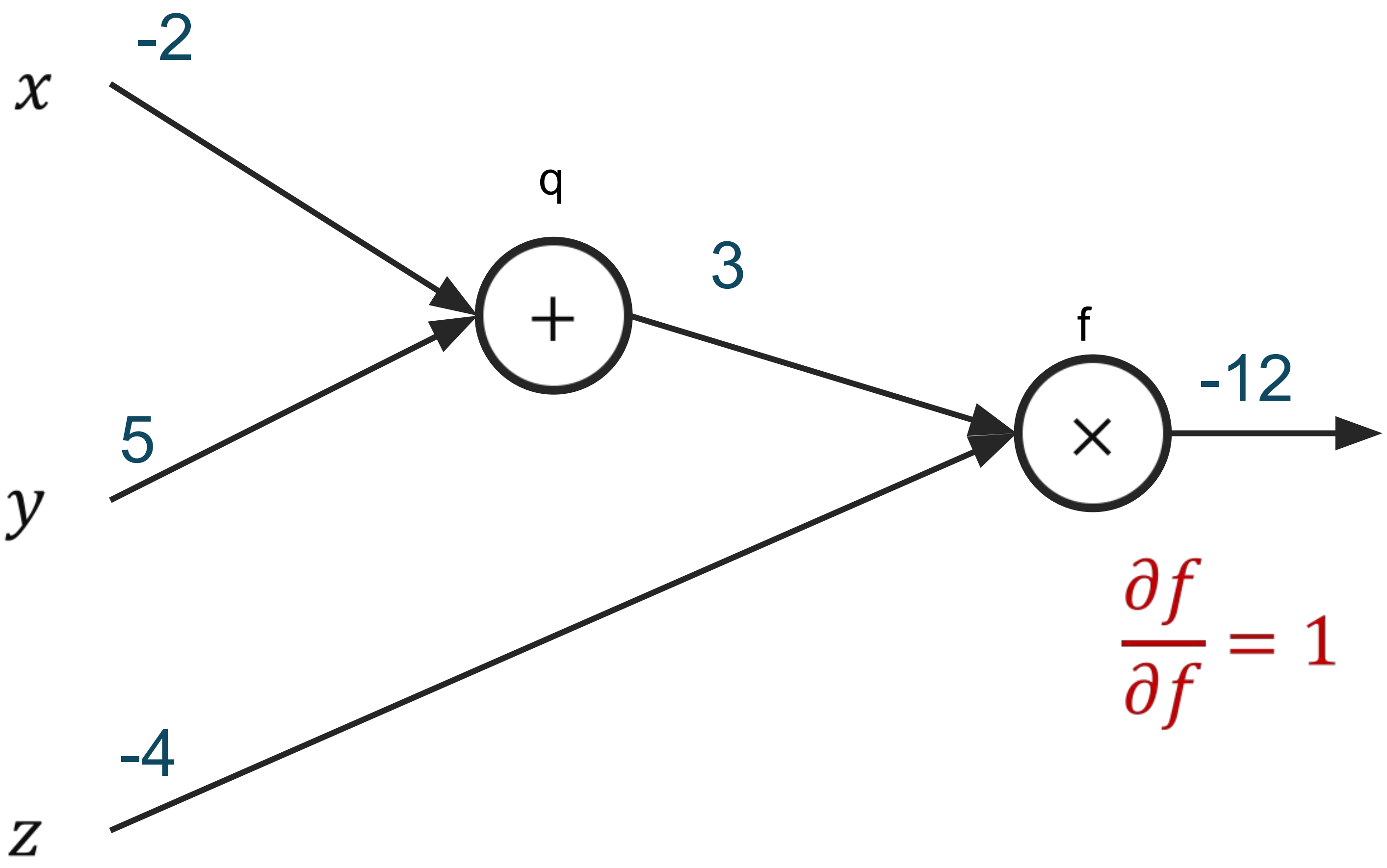
Calculate for the gradients! $f(x, y, z) = (x + y)z$

$$q = (x + y)$$

$$\frac{\partial q}{\partial x} = 1\,, \qquad \frac{\partial q}{\partial y} = 1$$

$$f = qz$$

$$\frac{\partial f}{\partial z} = q\,, \qquad \frac{\partial f}{\partial q} = z$$



$$\frac{\partial f}{\partial f} = 1$$

$$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial f}\frac{\partial f}{\partial z} = 1 * q = 1 * 3 = 3$$

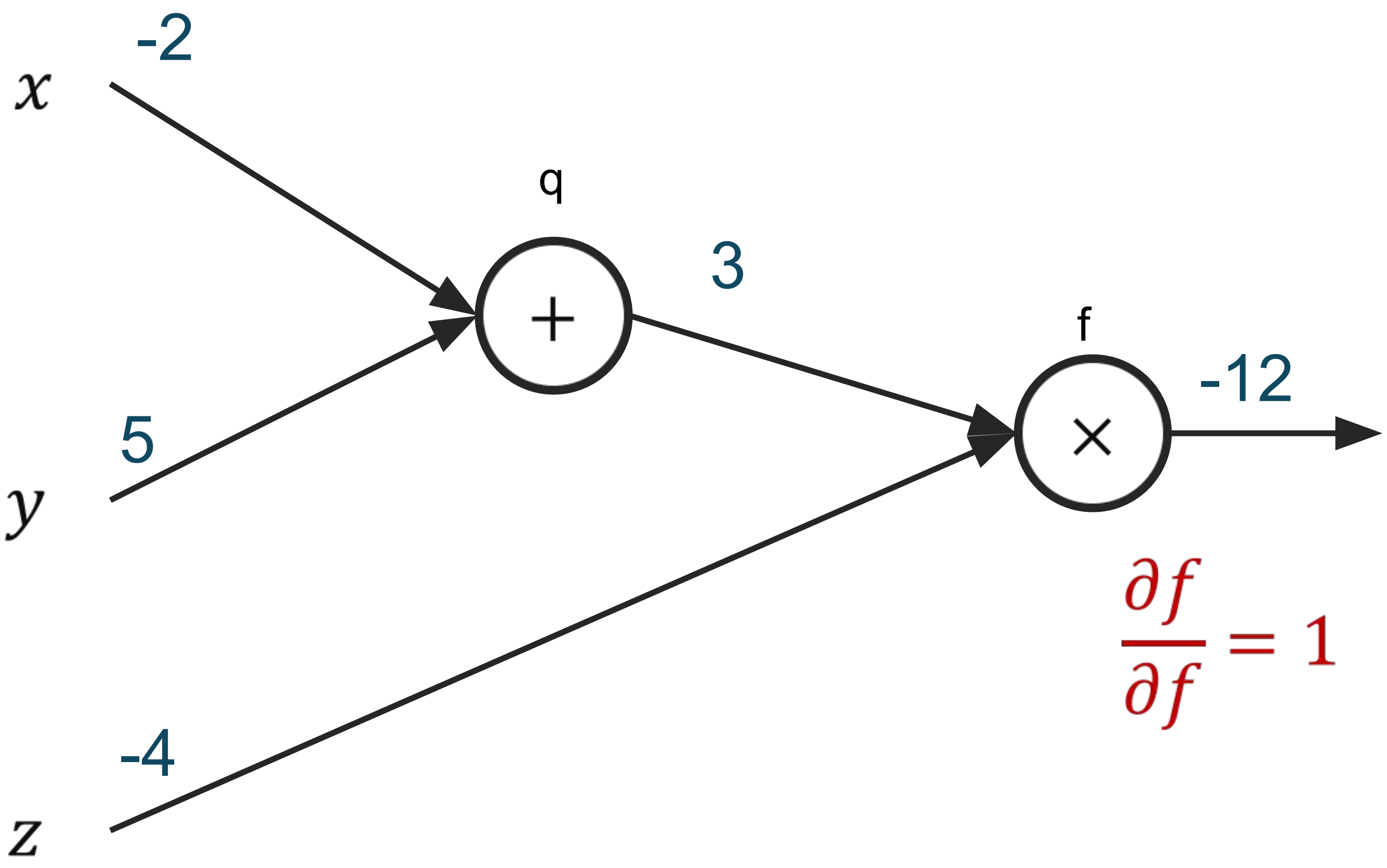
Calculate for the gradients! $f(x, y, z) = (x + y)z$

$$q = (x + y)$$

$$\frac{\partial q}{\partial x} = 1\,, \qquad \frac{\partial q}{\partial y} = 1$$

$$f = qz$$

$$\frac{\partial f}{\partial z} = q\,, \qquad \frac{\partial f}{\partial q} = z$$



$$\frac{\partial f}{\partial f} = 1$$

$$\frac{\partial f}{\partial z} = 3$$

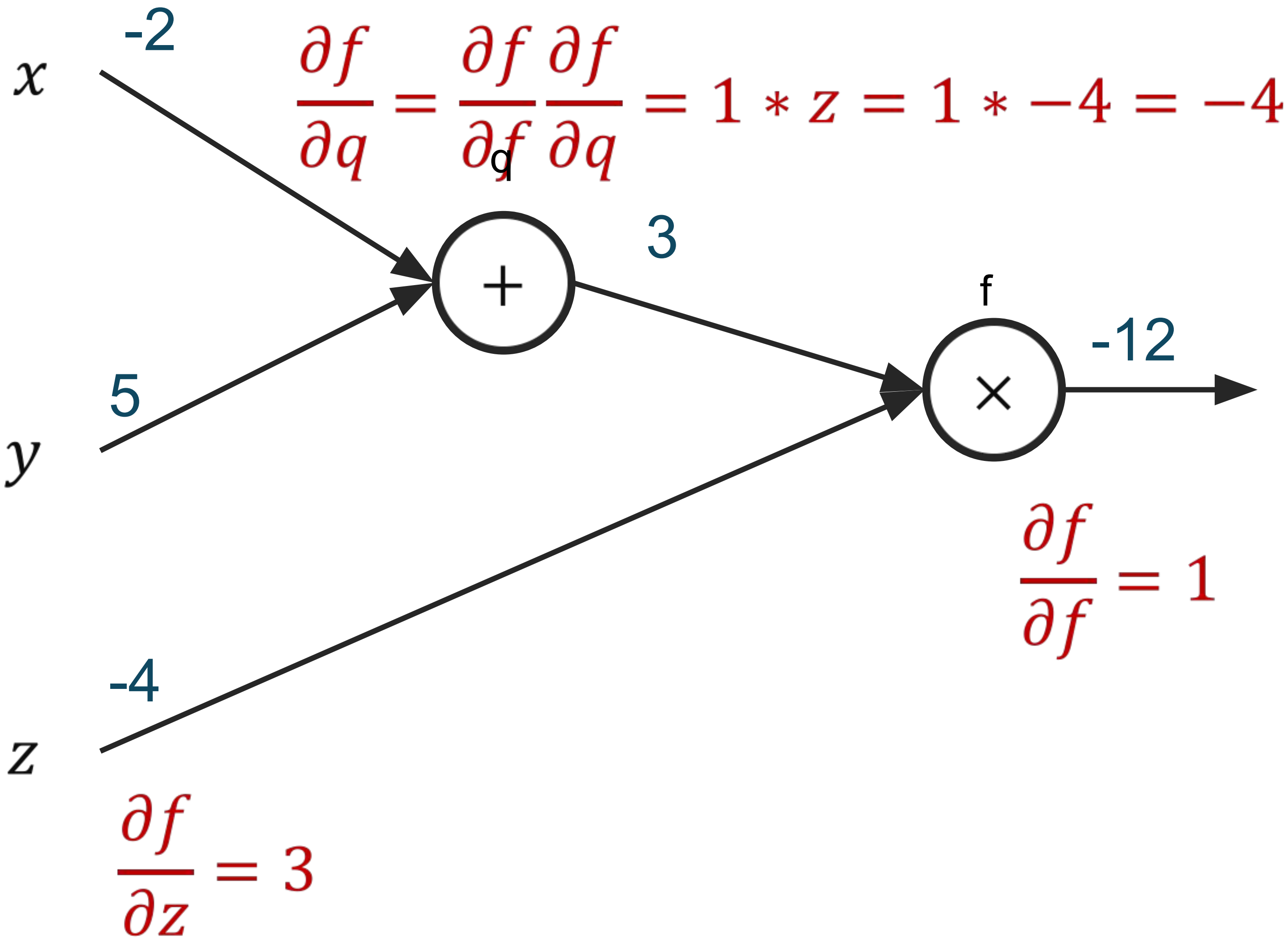
Calculate for the gradients! $f(x, y, z) = (x + y)z$

$$q = (x + y)$$

$$\frac{\partial q}{\partial x} = 1\,, \qquad \frac{\partial q}{\partial y} = 1$$

$$f = qz$$

$$\frac{\partial f}{\partial z} = q\,, \qquad \frac{\partial f}{\partial q} = z$$



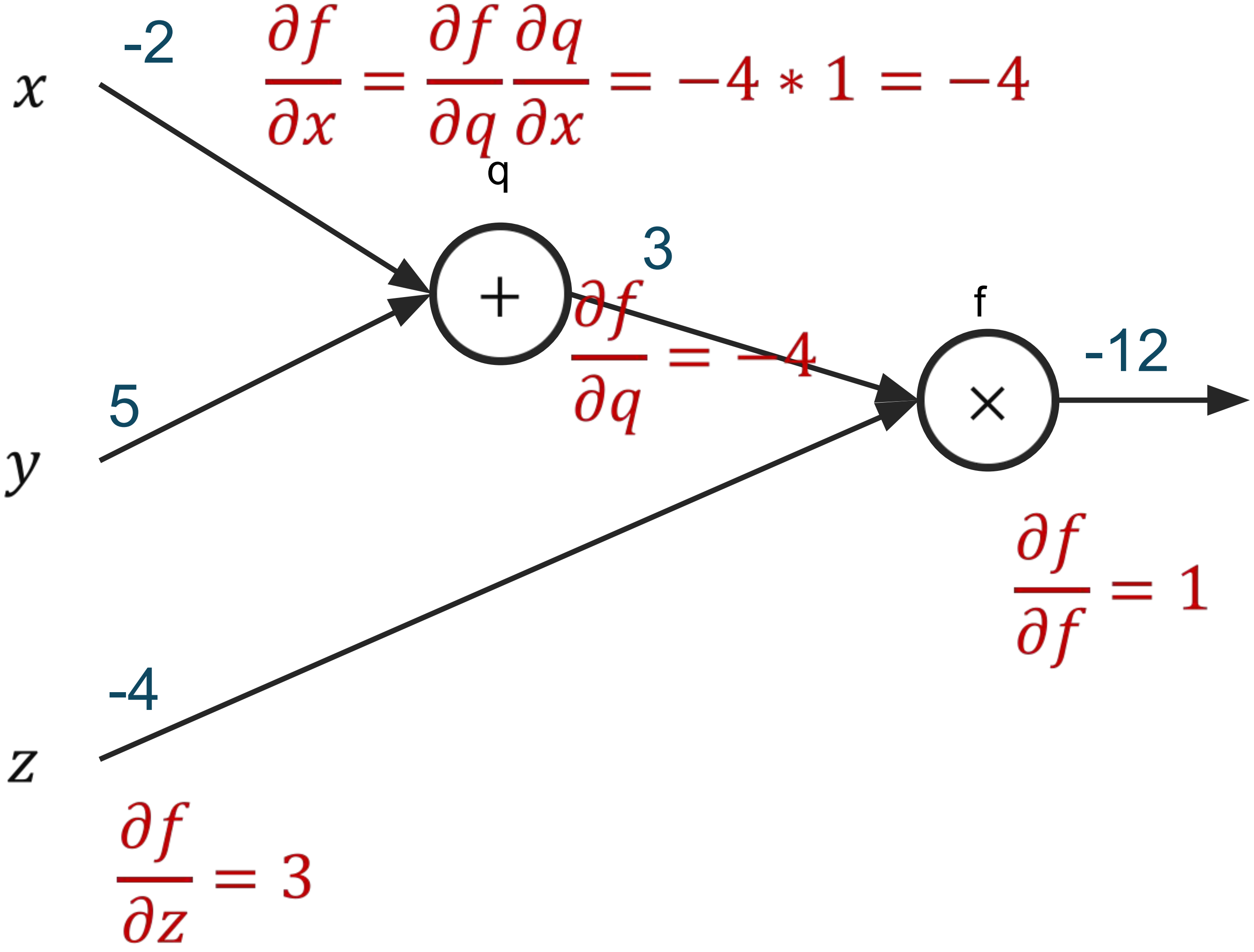
Calculate for the gradients! $f(x, y, z) = (x + y)z$

$$q = (x + y)$$

$$\frac{\partial q}{\partial x} = 1\,, \qquad \frac{\partial q}{\partial y} = 1$$

$$f = qz$$

$$\frac{\partial f}{\partial z} = q\,, \qquad \frac{\partial f}{\partial q} = z$$



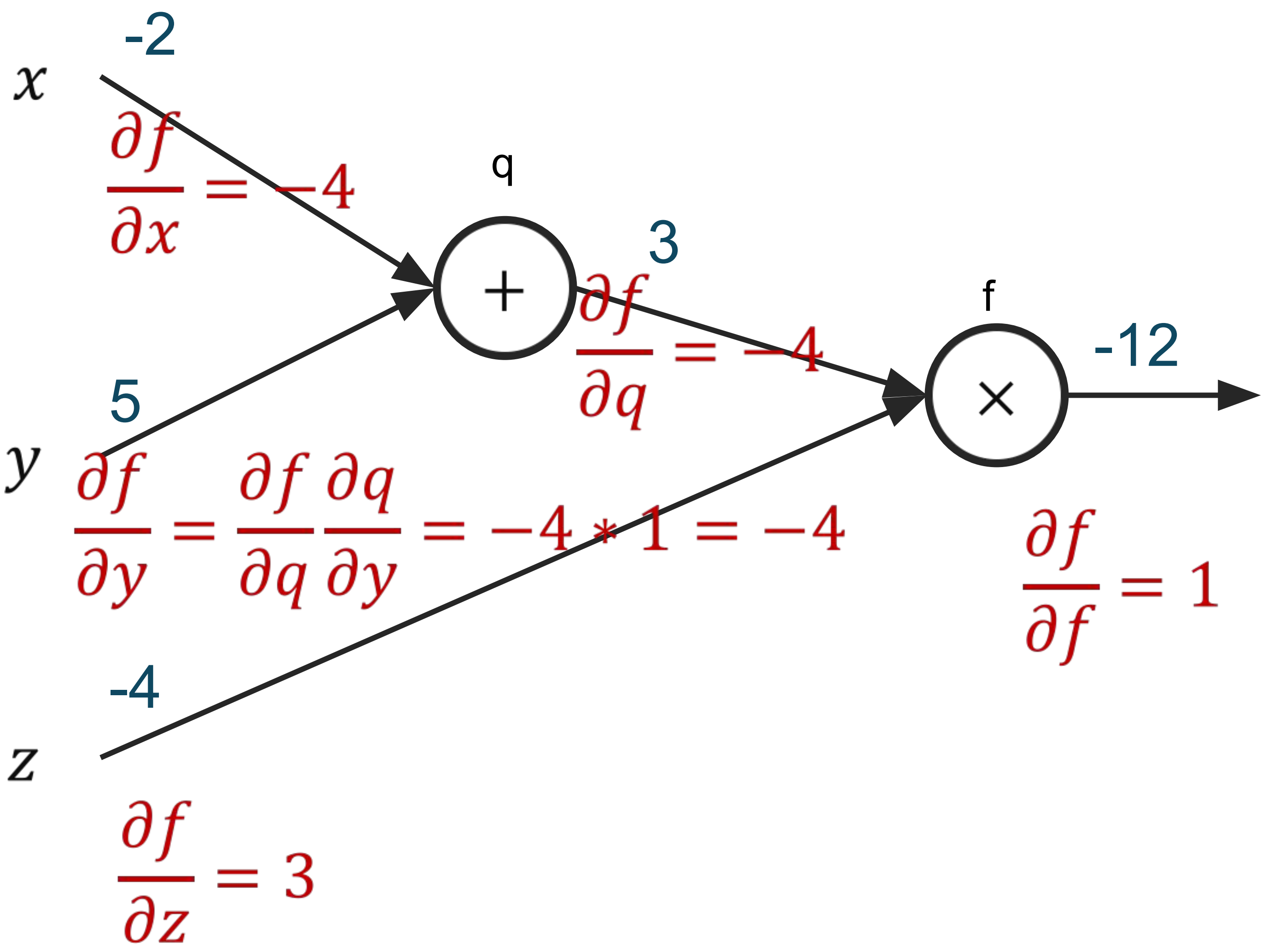
Calculate for the gradients! $f(x,y,z) = (x+y)z$

$$q = (x+y)$$

$$\frac{\partial q}{\partial x} = 1, \quad \frac{\partial q}{\partial y} = 1$$

$$f = qz$$

$$\frac{\partial f}{\partial z} = q, \quad \frac{\partial f}{\partial q} = z$$



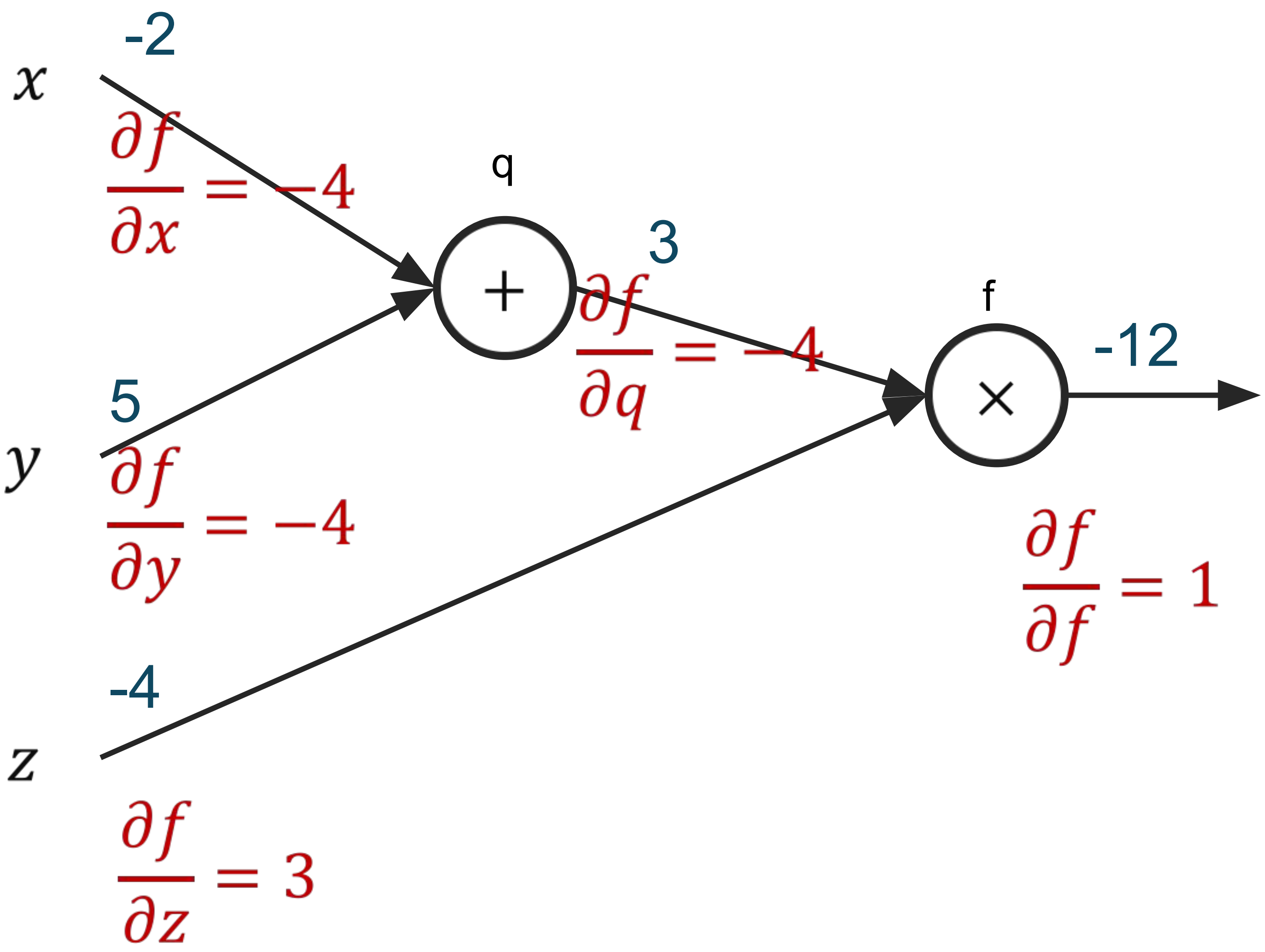
Calculate for the gradients! $f(x,y,z) = (x+y)z$

$$q = (x+y)$$

$$\frac{\partial q}{\partial x} = 1, \quad \frac{\partial q}{\partial y} = 1$$

$$f = qz$$

$$\frac{\partial f}{\partial z} = q, \quad \frac{\partial f}{\partial q} = z$$



$$f(x,y,z) = (x+y)z$$

if $x := x + h$,

$$f' = f + \frac{\partial f}{\partial x} * h$$

Example:

$h = 1, x = -1,$

$f = (-1+5) * -4 = -16$

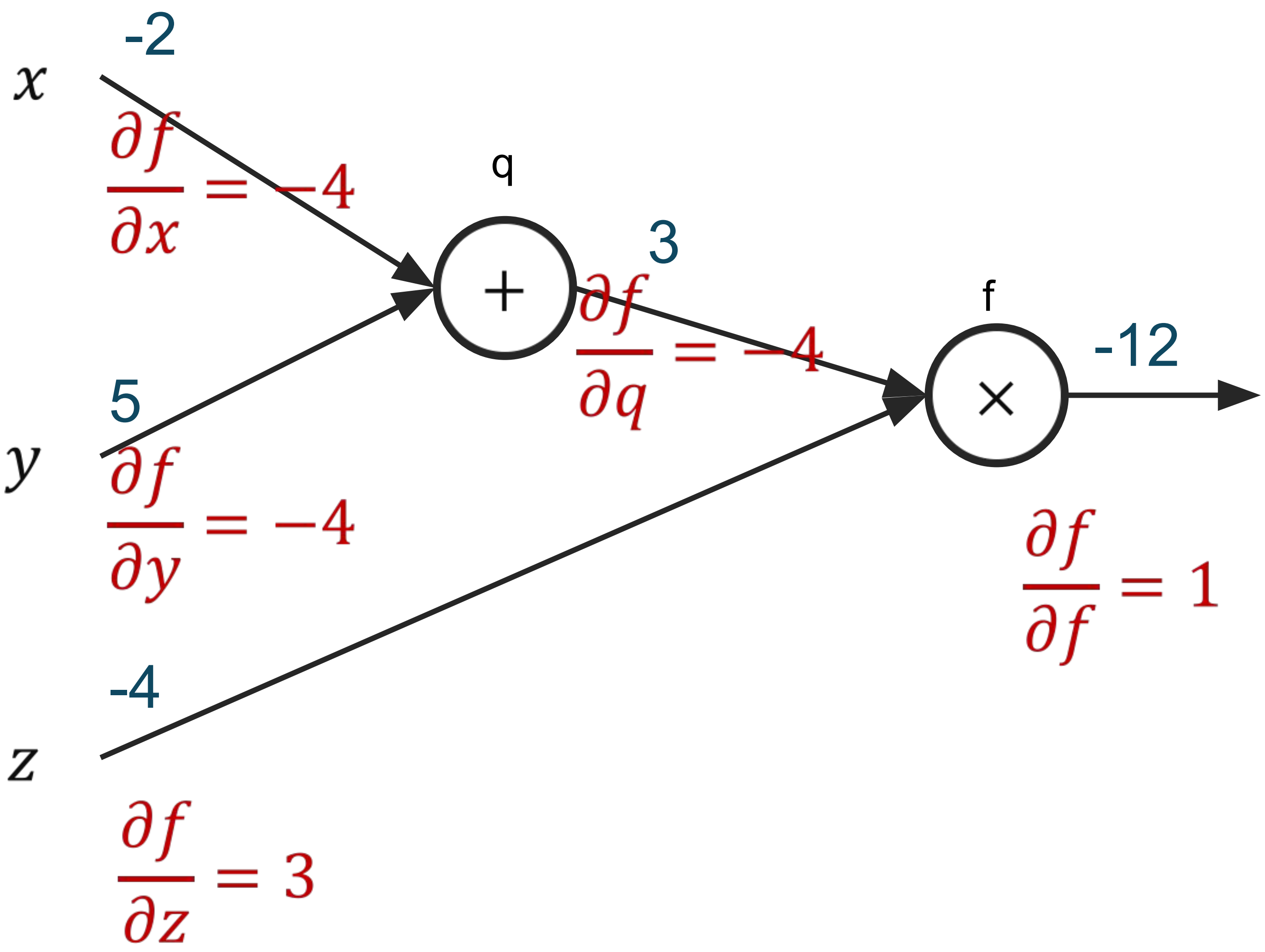
$$f' = f + \frac{\partial f}{\partial x} * h$$

$$f' = f + -4 * 1$$

$$f' = -12 + -4 = -16$$

An increase in x decreases f!

$$f(x,y,z) = (x+y)z = -12$$



$$f(x,y,z) = (x+y)z$$

If $y := y + h$,

$$f' = f + \frac{\partial f}{\partial y} * h$$

Example:

$h = 1, y = 6,$

$f = (-2+6) * -4 = -16$

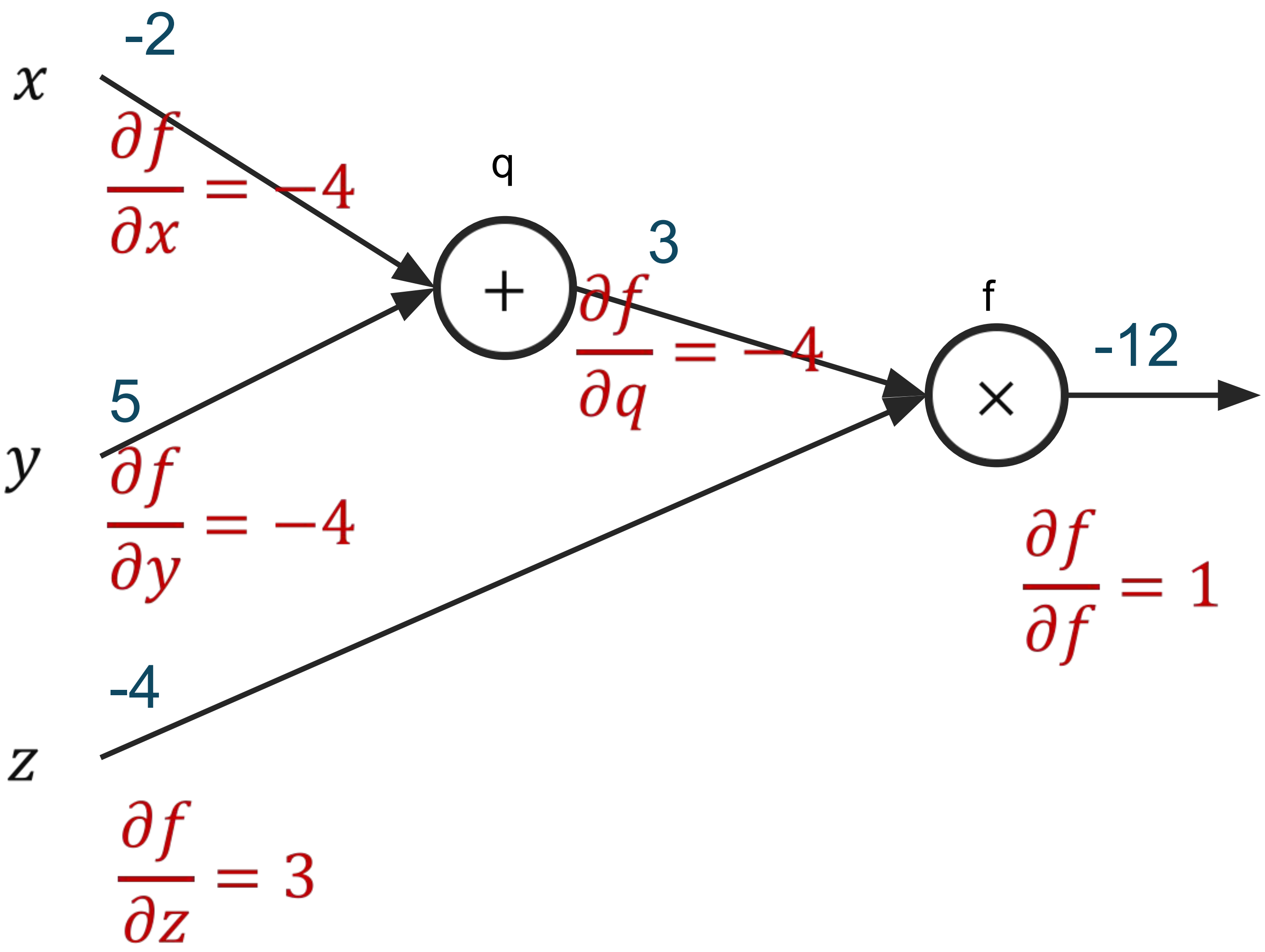
$$f' = f + \frac{\partial f}{\partial y} * h$$

$$f' = f + -4 * 1$$

$$f' = -12 + -4 = -16$$

An increase in y decreases f!

$$f(x,y,z) = (x+y)z = -12$$



$$f(x, y, z) = (x + y)z$$

If $z := z + h$,

$$f' = f + \frac{\partial f}{\partial z} * h$$

Example:

$h = 1, z = -3,$

$f = (-2 + 5) * -3 = -9$

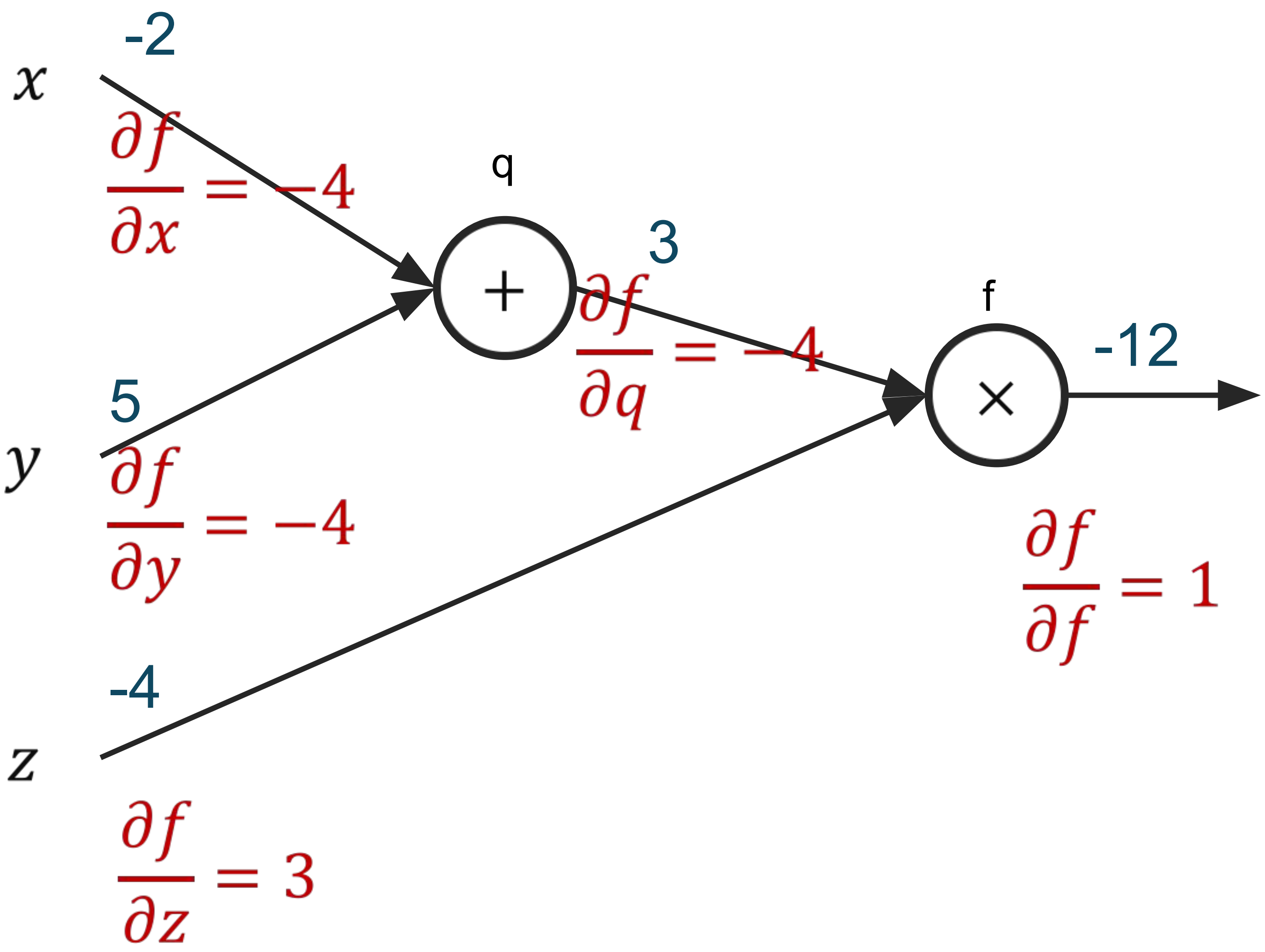
$$f' = f + \frac{\partial f}{\partial z} * h$$

$$f' = f + 3 * 1$$

$$f' = -12 + 3 = -9$$

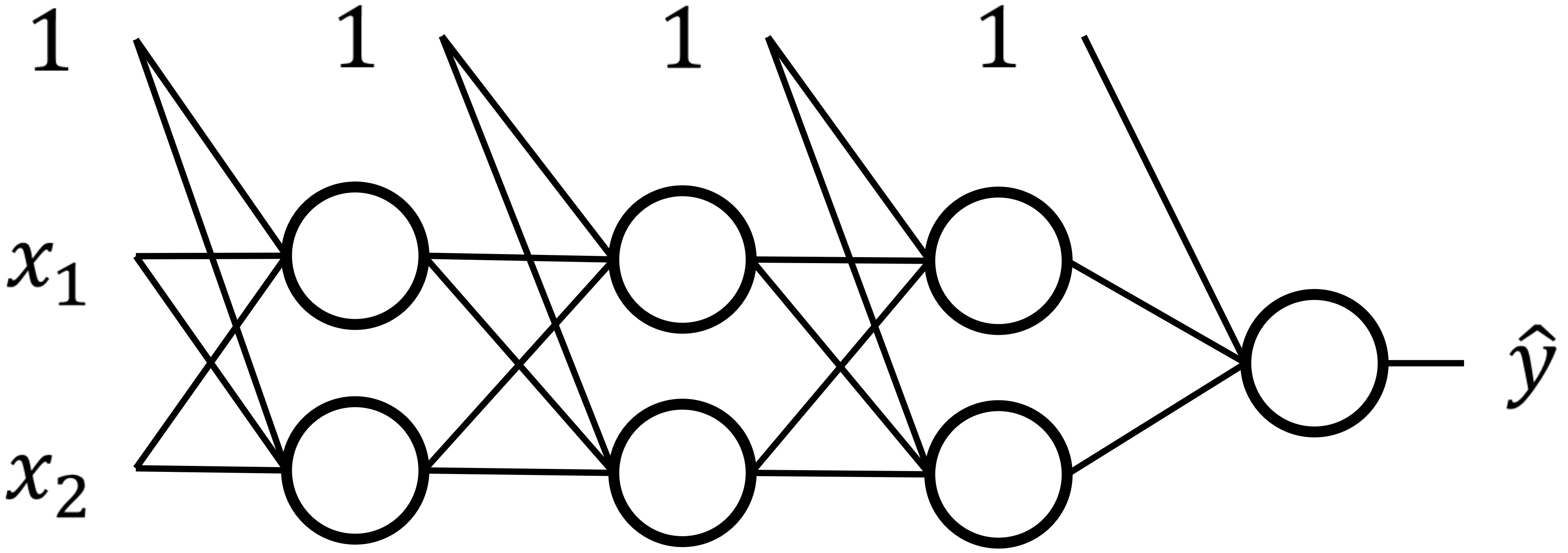
An increase in z increases f!

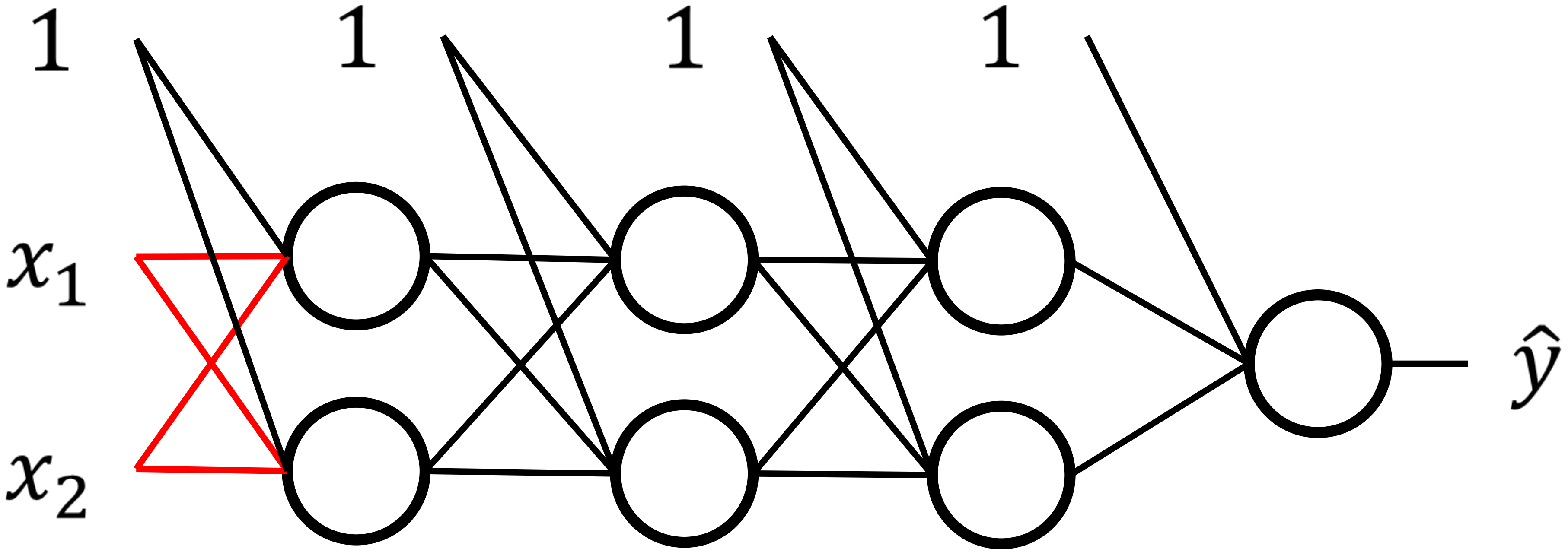
$$f(x, y, z) = (x + y)z = -12$$



Now Let's Apply Derivatives For Backpropagation

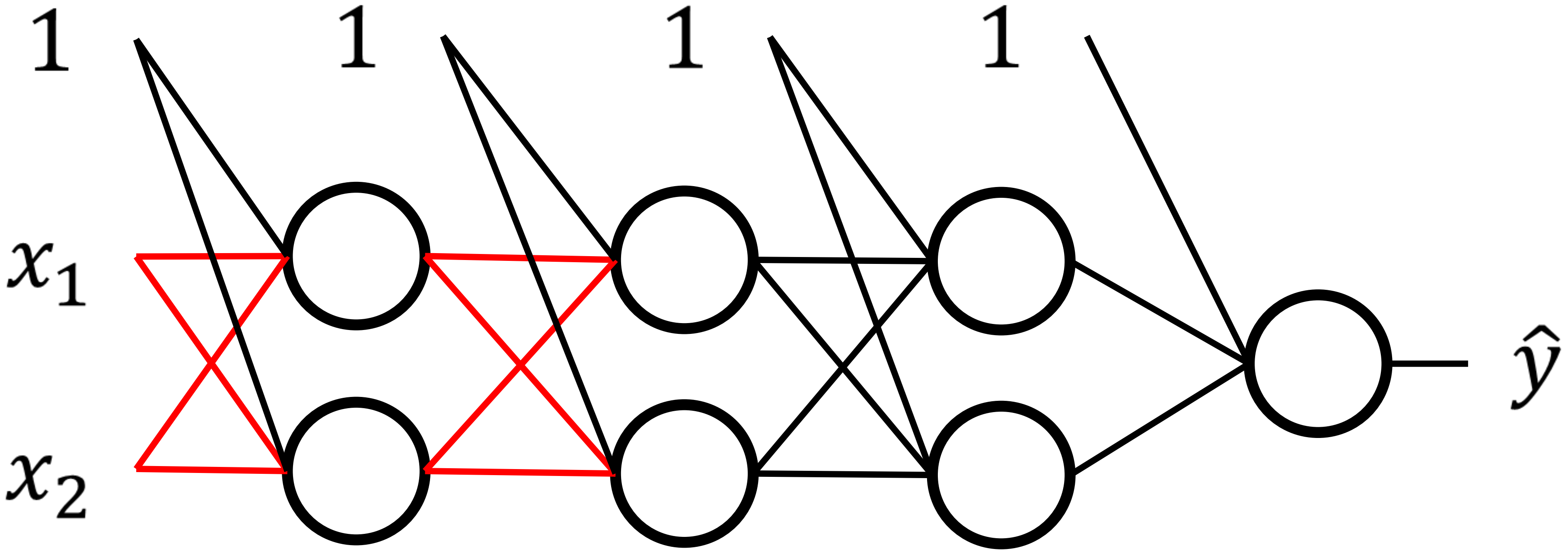
1. The NN Layout and Terms We’ll Use


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



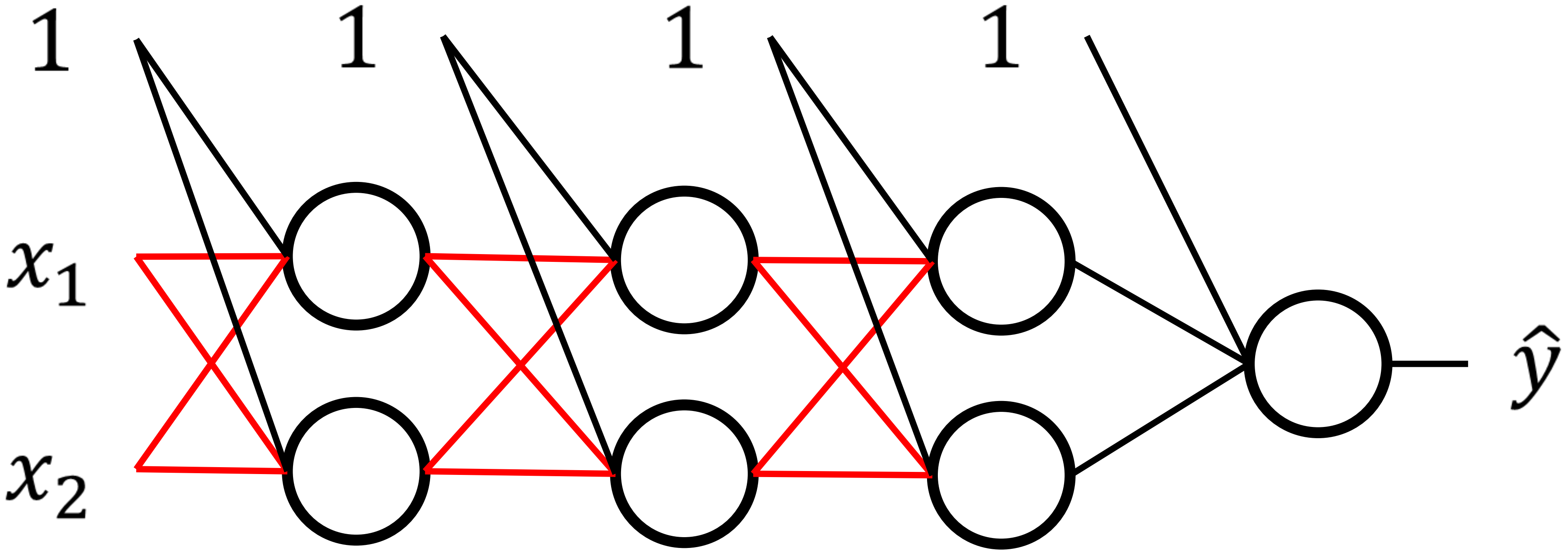
$$W^{[1]}$$

$$\begin{bmatrix} W_{11}^{[1]} & W_{12}^{[1]} \\ W_{21}^{[1]} & W_{22}^{[1]} \end{bmatrix}$$



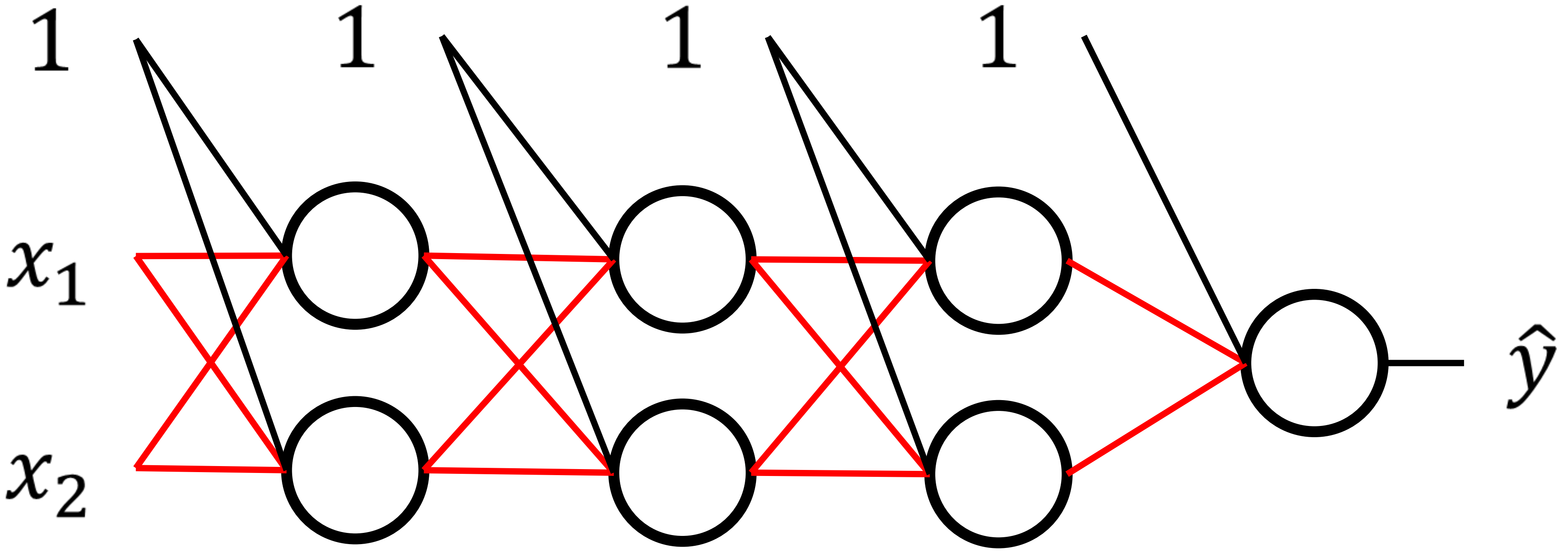
$$W^{[1]} \qquad W^{[2]}$$

$$\begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{bmatrix} \begin{bmatrix} w_{11}^{[2]} & w_{12}^{[2]} \\ w_{21}^{[2]} & w_{22}^{[2]} \end{bmatrix}$$



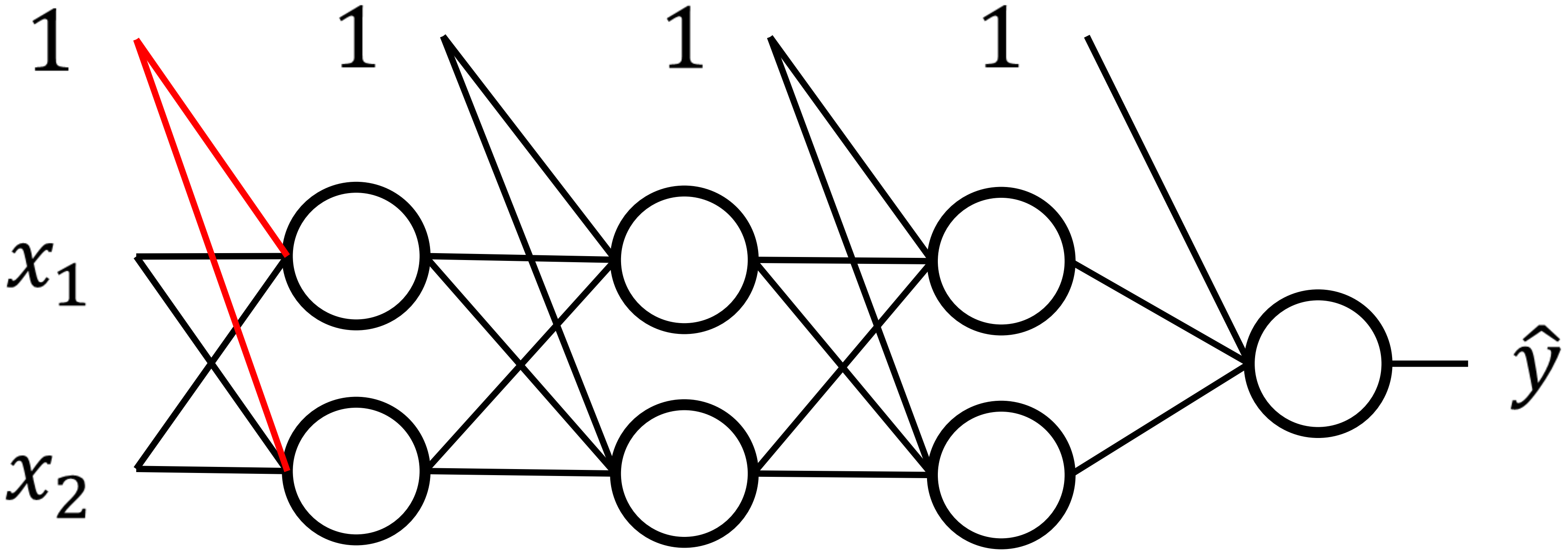
$$W^{[1]} \qquad W^{[2]} \qquad W^{[3]}$$

$$\begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{bmatrix} \begin{bmatrix} w_{11}^{[2]} & w_{12}^{[2]} \\ w_{21}^{[2]} & w_{22}^{[2]} \end{bmatrix} \begin{bmatrix} w_{11}^{[3]} & w_{12}^{[3]} \\ w_{21}^{[3]} & w_{22}^{[3]} \end{bmatrix}$$



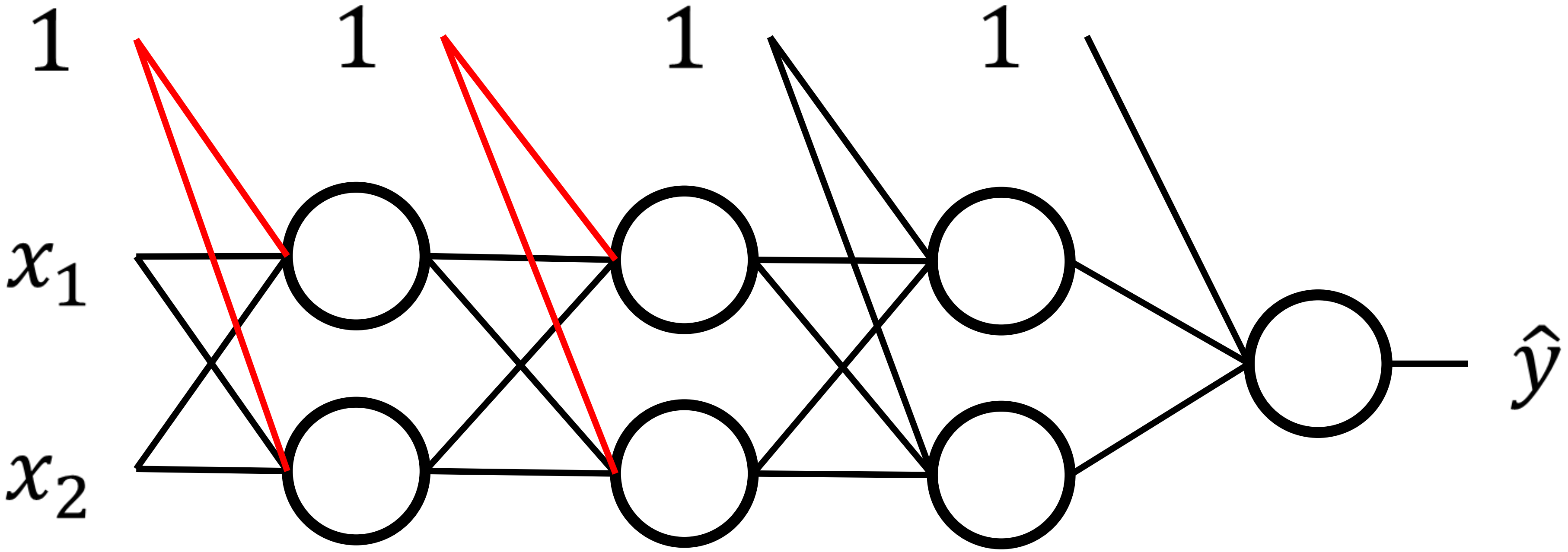
$$W^{[1]} \qquad W^{[2]} \qquad W^{[3]} \qquad W^{[4]}$$

$$\begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{bmatrix} \begin{bmatrix} w_{11}^{[2]} & w_{12}^{[2]} \\ w_{21}^{[2]} & w_{22}^{[2]} \end{bmatrix} \begin{bmatrix} w_{11}^{[3]} & w_{12}^{[3]} \\ w_{21}^{[3]} & w_{22}^{[3]} \end{bmatrix} \begin{vmatrix} w_{11}^{[4]} & w_{12}^{[4]} \end{vmatrix}$$



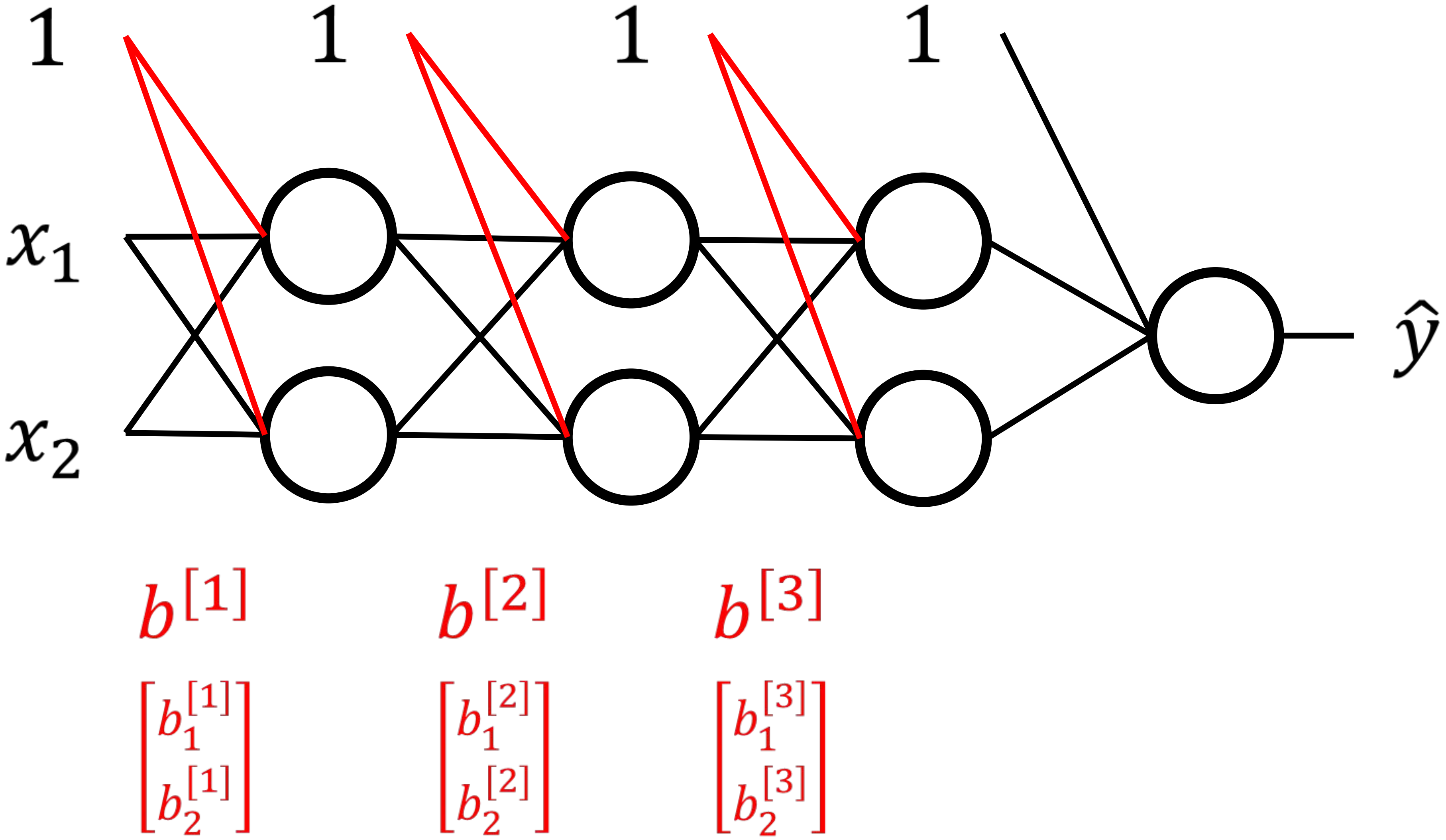
$$b^{[1]}$$

$$\begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \end{bmatrix}$$

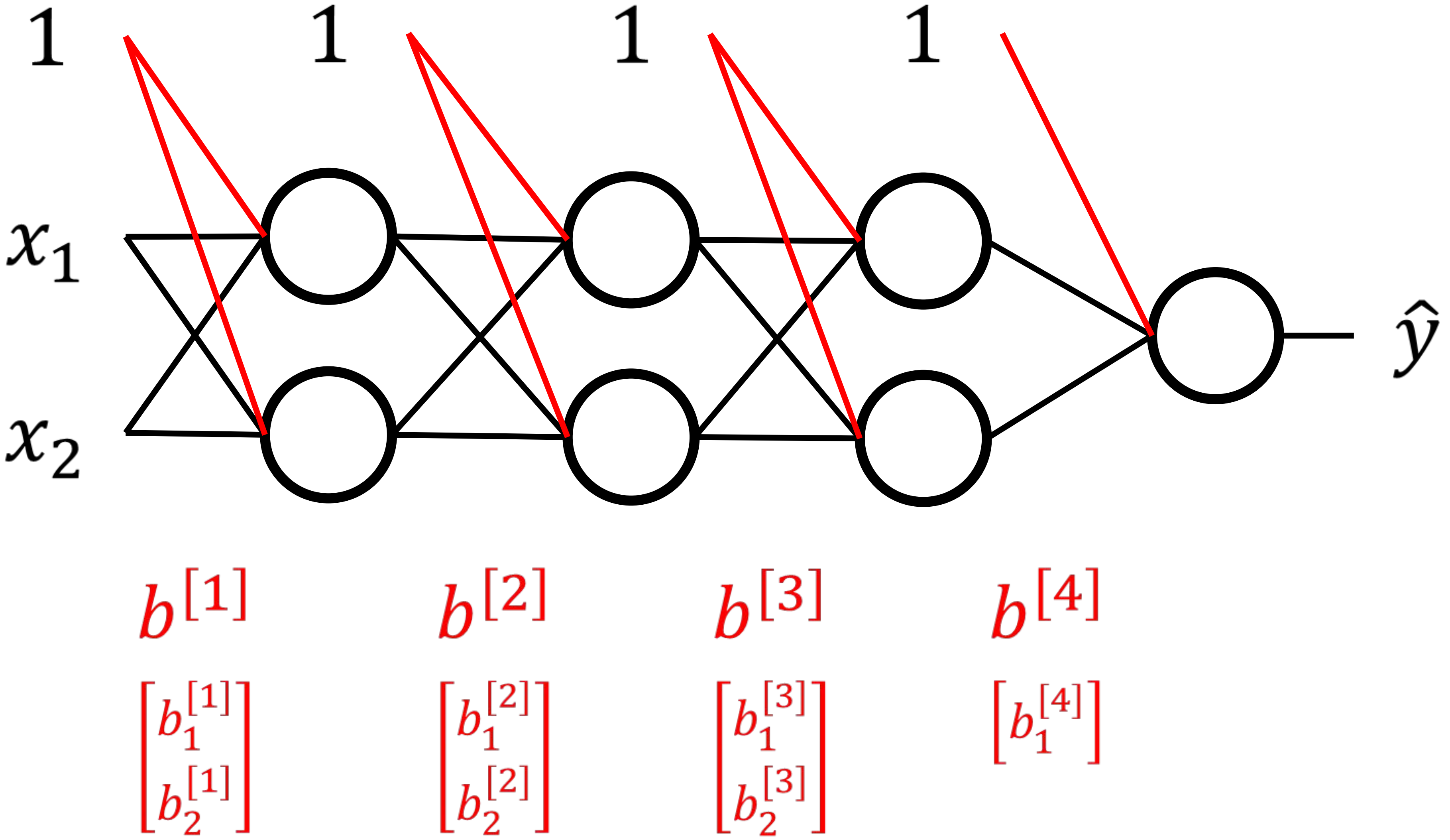


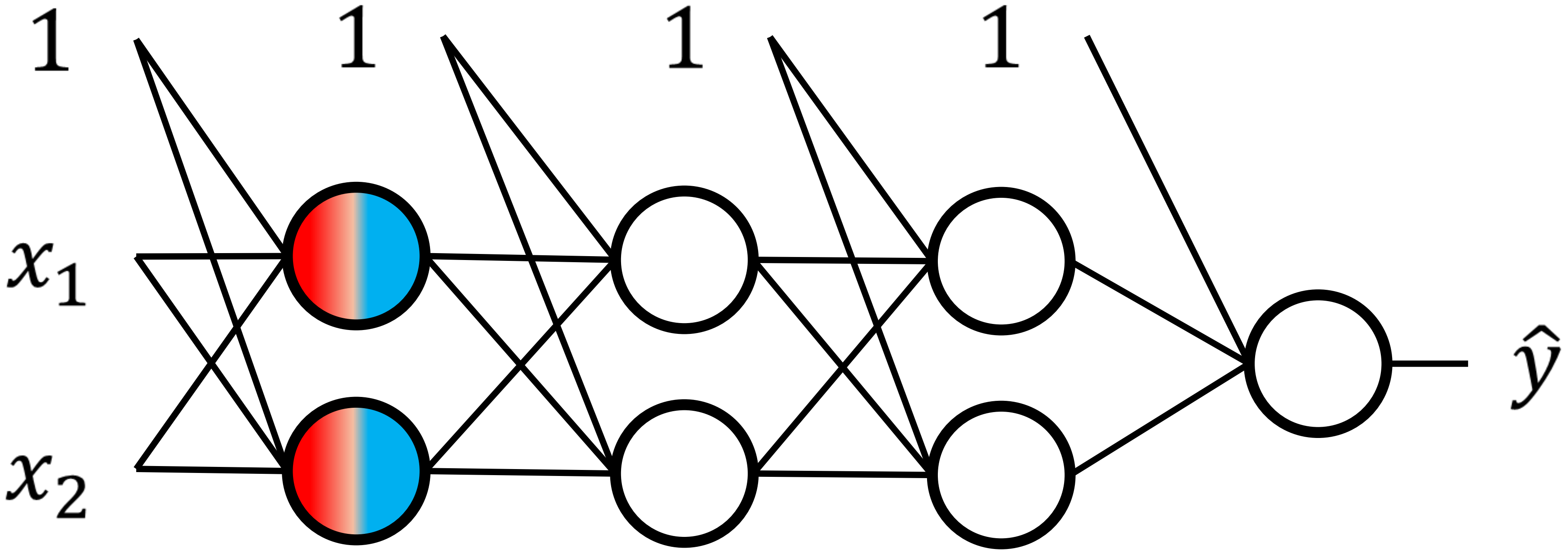
$b^{[1]}$ $\quad$ $b^{[2]}$

$$\begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \end{bmatrix} \qquad \begin{bmatrix} b_1^{[2]} \\ b_2^{[2]} \end{bmatrix}$$



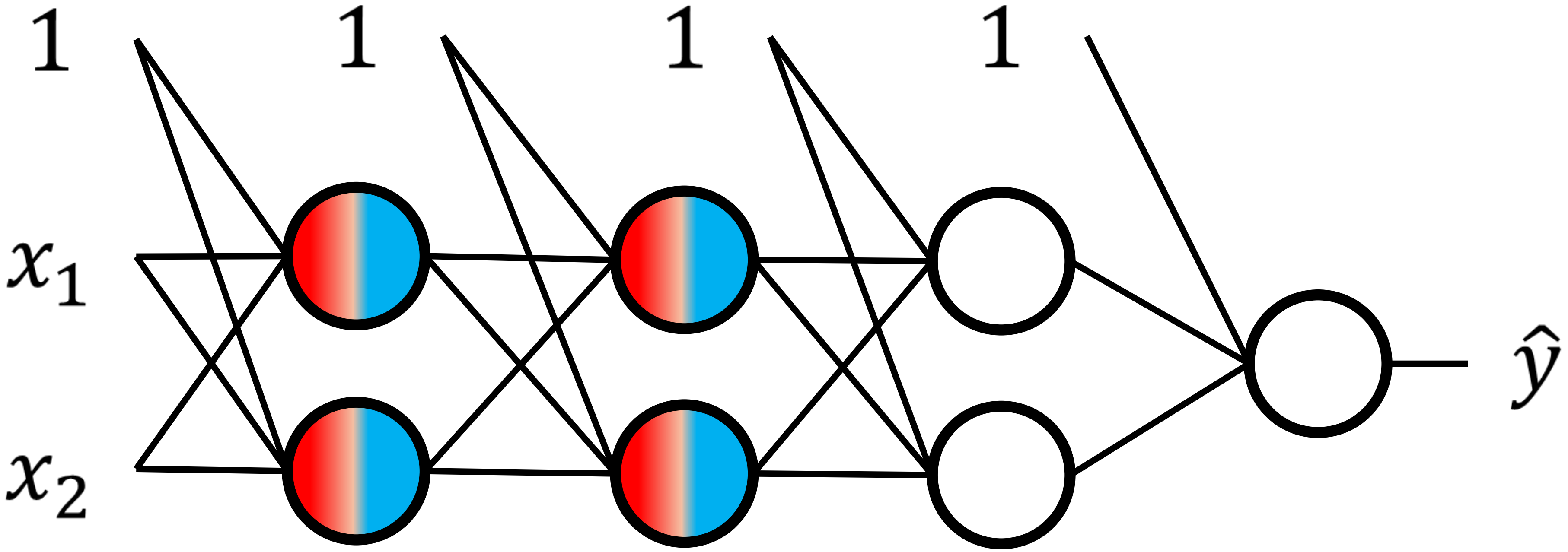
1
1
1
1
x_1
x_2
$\hat{y}$
$b^{[1]}$
$b^{[2]}$
$b^{[3]}$
$\begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \end{bmatrix}$
$\begin{bmatrix} b_1^{[2]} \\ b_2^{[2]} \end{bmatrix}$
$\begin{bmatrix} b_1^{[3]} \\ b_2^{[3]} \end{bmatrix}$


1
1
1
1
x_1
x_2
$\hat{y}$
$b^{[1]}$
$b^{[2]}$
$b^{[3]}$
$b^{[4]}$
$\begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \end{bmatrix}$
$\begin{bmatrix} b_1^{[2]} \\ b_2^{[2]} \end{bmatrix}$
$\begin{bmatrix} b_1^{[3]} \\ b_2^{[3]} \end{bmatrix}$
$\begin{bmatrix} b_1^{[4]} \end{bmatrix}$



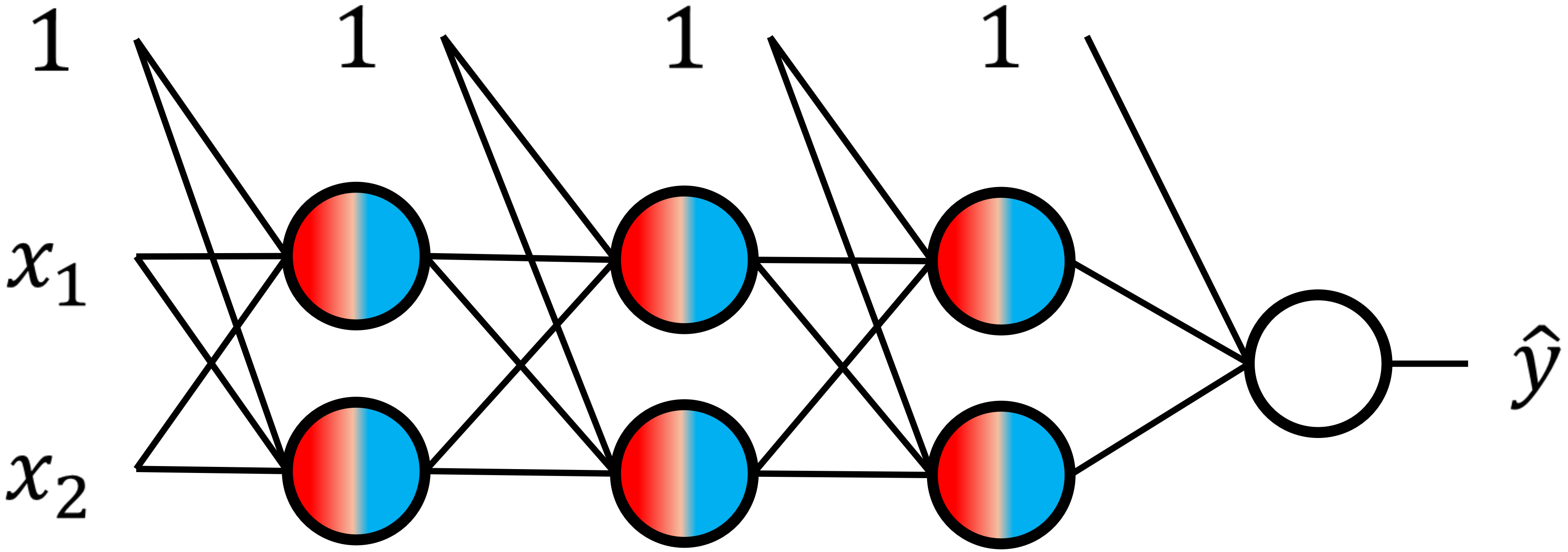
$$z^{[1]} \quad a^{[1]}$$

$$\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix}$$



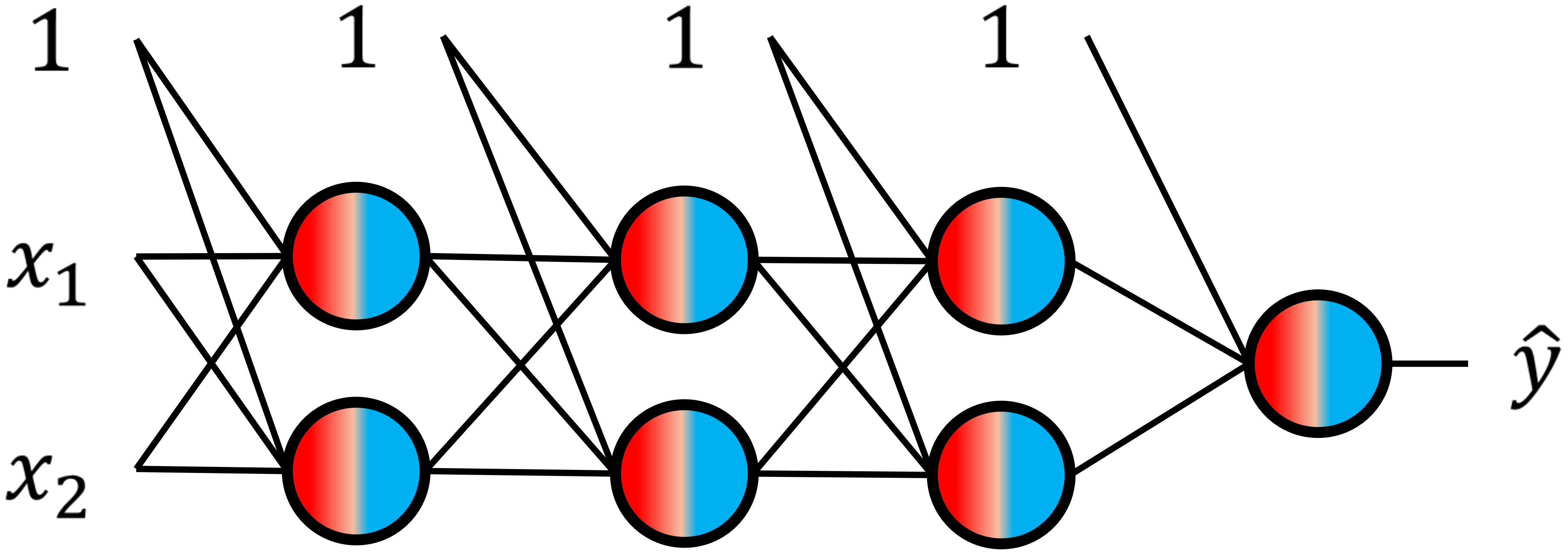
$$z^{[1]} \quad a^{[1]} \quad z^{[2]} \quad a^{[2]}$$

$$\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} \quad \begin{bmatrix} z_1^{[2]} \\ z_2^{[2]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[2]} \\ a_2^{[2]} \end{bmatrix}$$



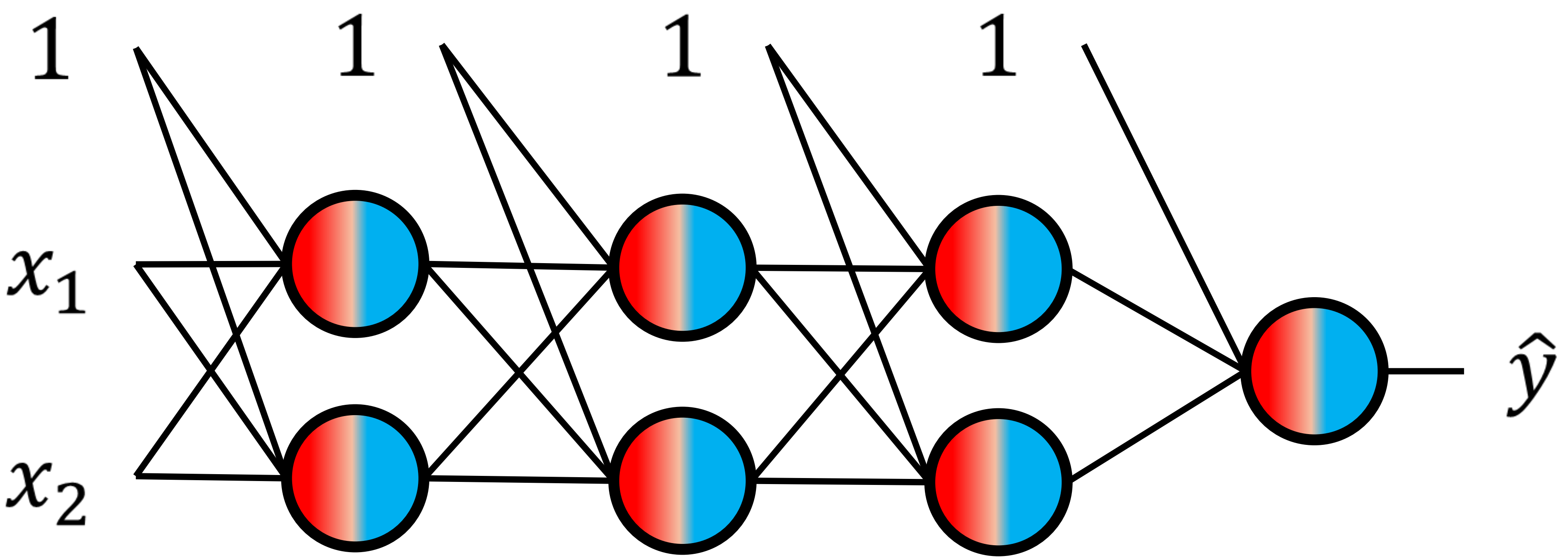
$$z^{[1]} \quad a^{[1]} \quad z^{[2]} \quad a^{[2]} \quad z^{[3]} \quad a^{[3]}$$

$$\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix} \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} \begin{bmatrix} z_1^{[2]} \\ z_2^{[2]} \end{bmatrix} \begin{bmatrix} a_1^{[2]} \\ a_2^{[2]} \end{bmatrix} \begin{bmatrix} z_1^{[3]} \\ z_2^{[3]} \end{bmatrix} \begin{bmatrix} a_1^{[3]} \\ a_2^{[3]} \end{bmatrix}$$



$$z^{[1]} \quad a^{[1]} \quad z^{[2]} \quad a^{[2]} \quad z^{[3]} \quad a^{[3]} \quad z^{[4]} \quad a^{[4]}$$

$$\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} \quad \begin{bmatrix} z_1^{[2]} \\ z_2^{[2]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[2]} \\ a_2^{[2]} \end{bmatrix} \quad \begin{bmatrix} z_1^{[3]} \\ z_2^{[3]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[3]} \\ a_2^{[3]} \end{bmatrix} \quad \begin{bmatrix} z_1^{[4]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[4]} \end{bmatrix}$$

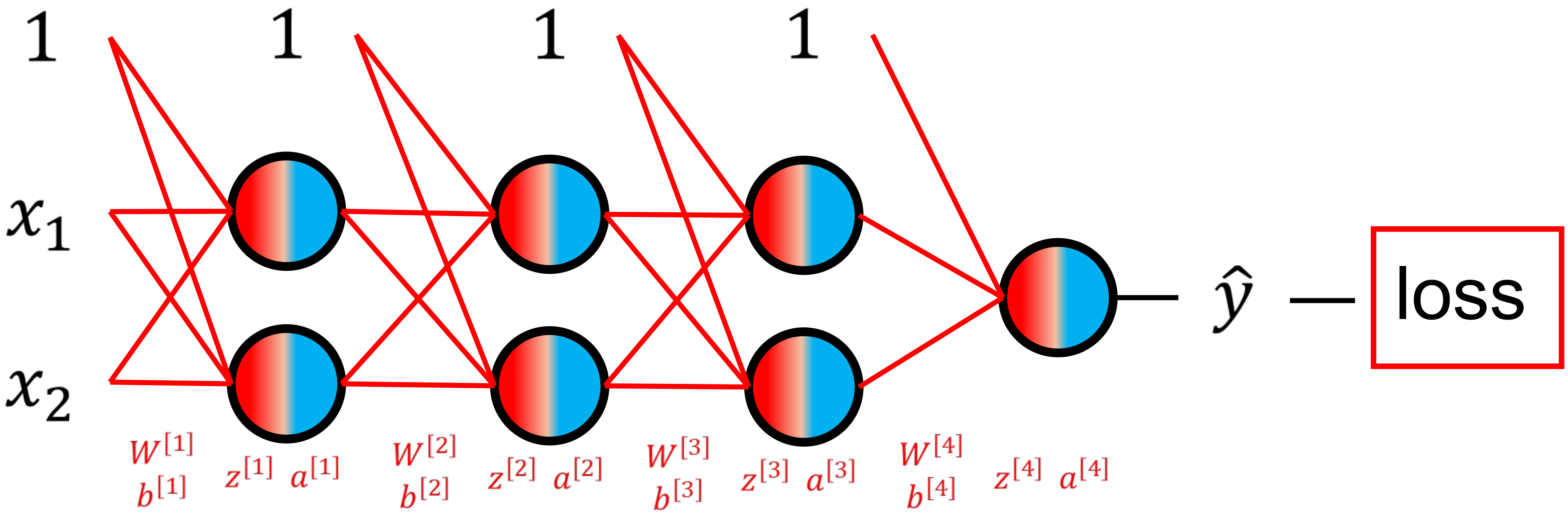


$$z^{[1]} \quad a^{[1]} \quad z^{[2]} \quad a^{[2]} \quad z^{[3]} \quad a^{[3]} \quad z^{[4]} \quad a^{[4]}$$

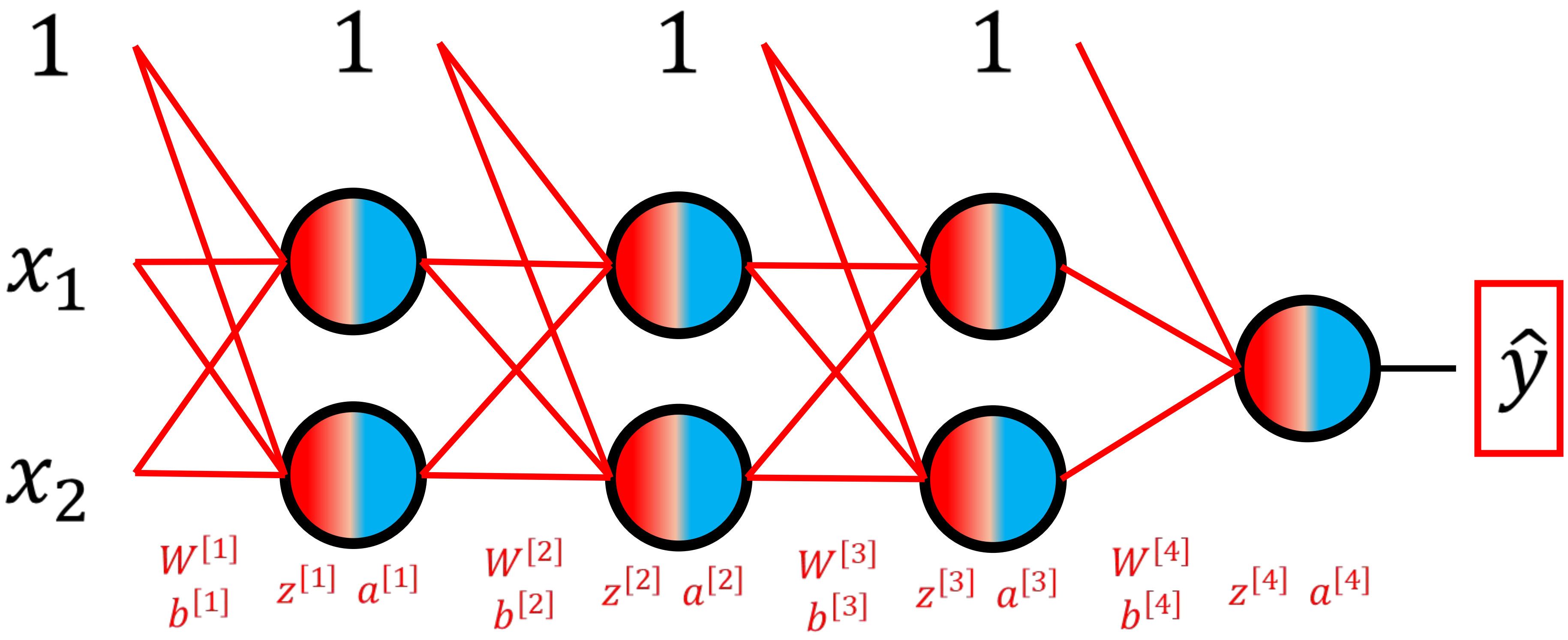
$$\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} \quad \begin{bmatrix} z_1^{[2]} \\ z_2^{[2]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[2]} \\ a_2^{[2]} \end{bmatrix} \quad \begin{bmatrix} z_1^{[3]} \\ z_2^{[3]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[3]} \\ a_2^{[3]} \end{bmatrix} \quad \begin{bmatrix} z_1^{[4]} \end{bmatrix} \quad \begin{bmatrix} a_1^{[4]} \end{bmatrix}$$

z pertain to score received by each neuron **BEFORE** the activation function is applied.
a pertain to the activation of each neuron, **AFTER** the activation function is applied to score.

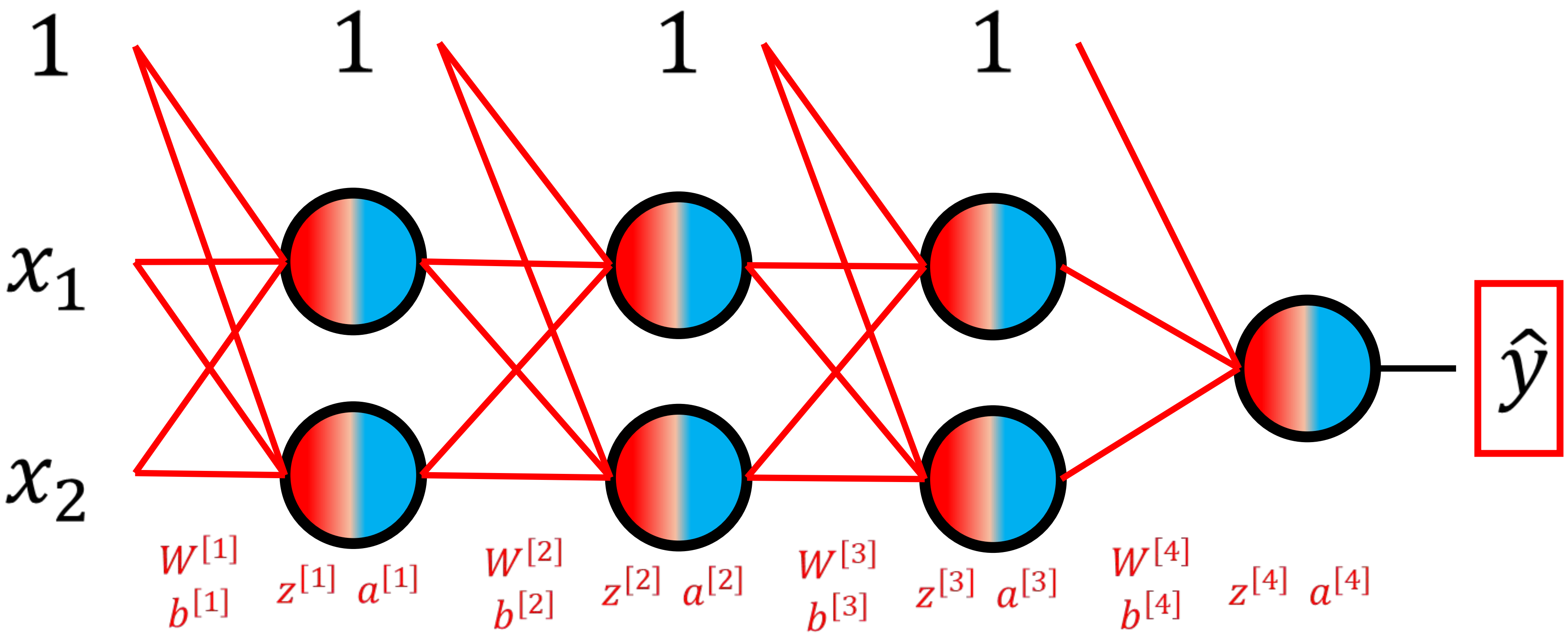
2. Stating Our Goal



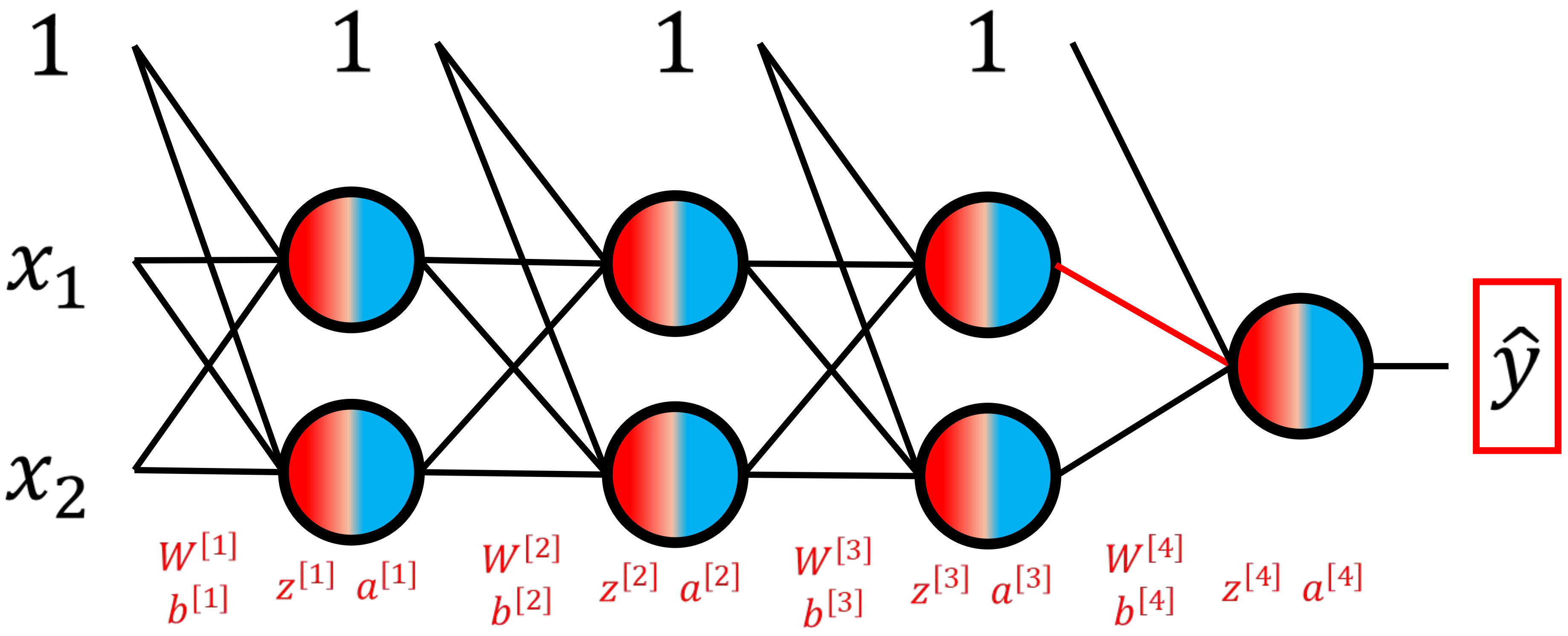
We want to know, if we change each **weight**, how does it affect the **loss**?



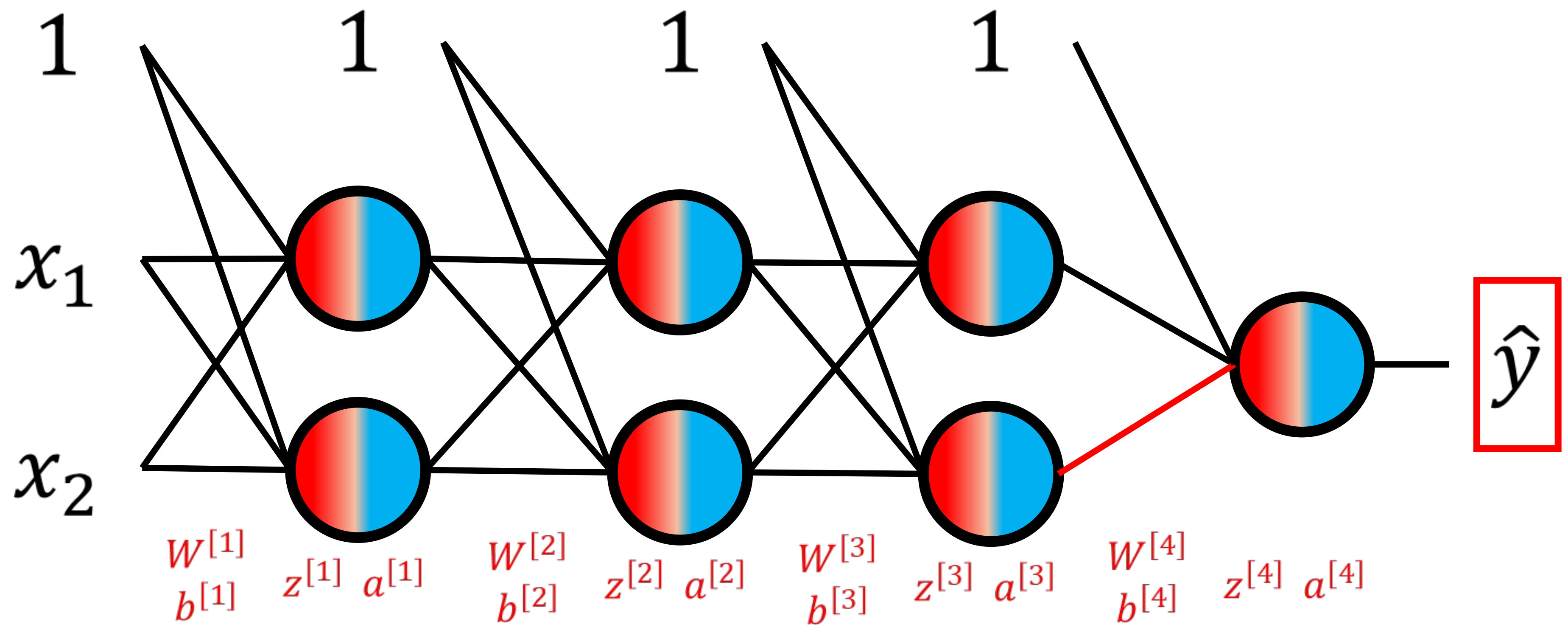
To make things simpler, let's ignore the loss first and ask: how does each **weight** affect the $\hat{y}$ (the prediction)?



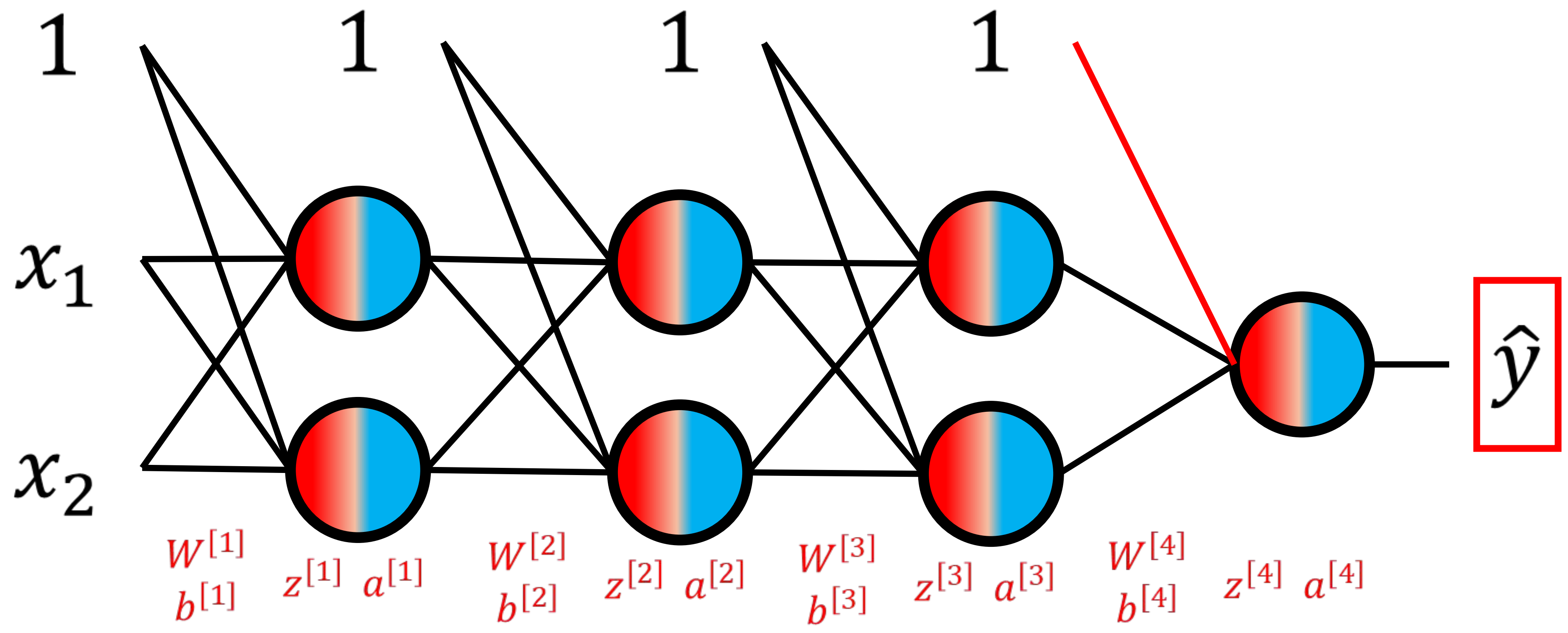
Therefore, we need to compute all of the following…



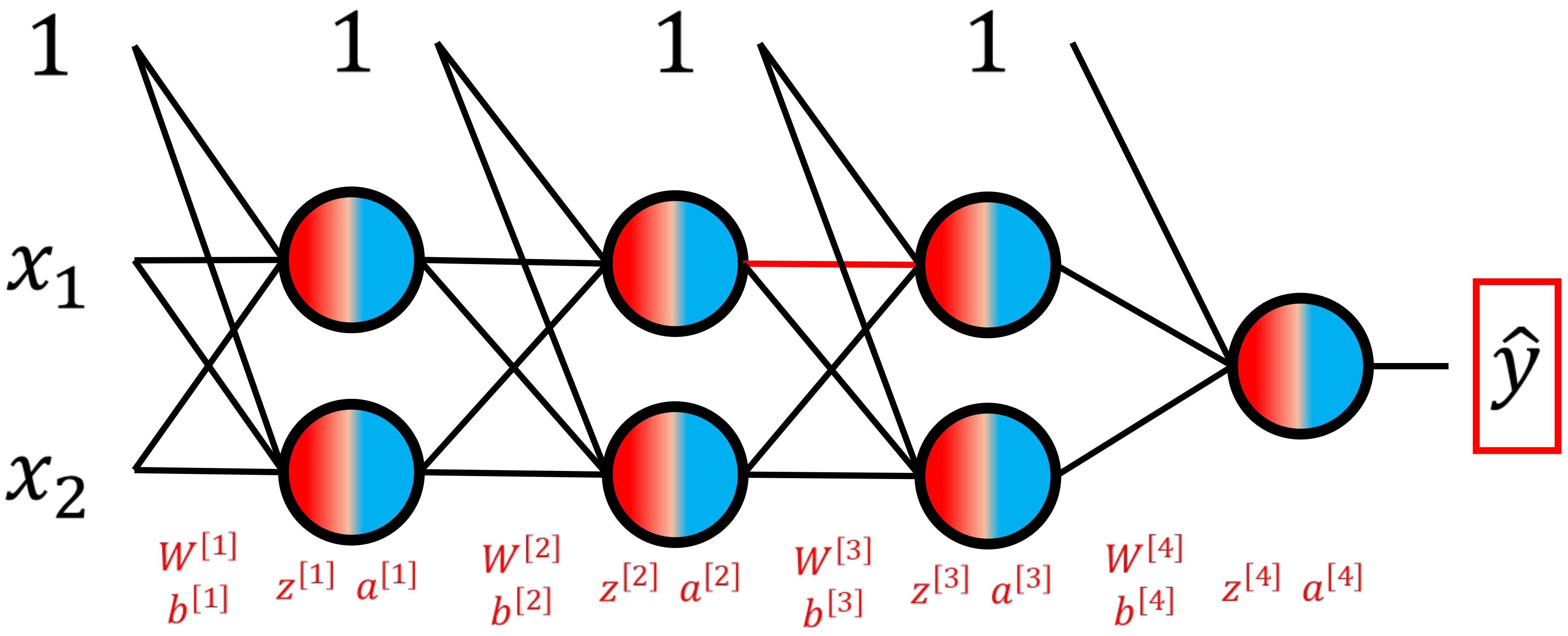
Derivative of $\hat{y}$ with respect to $W_{11}^{[4]}$



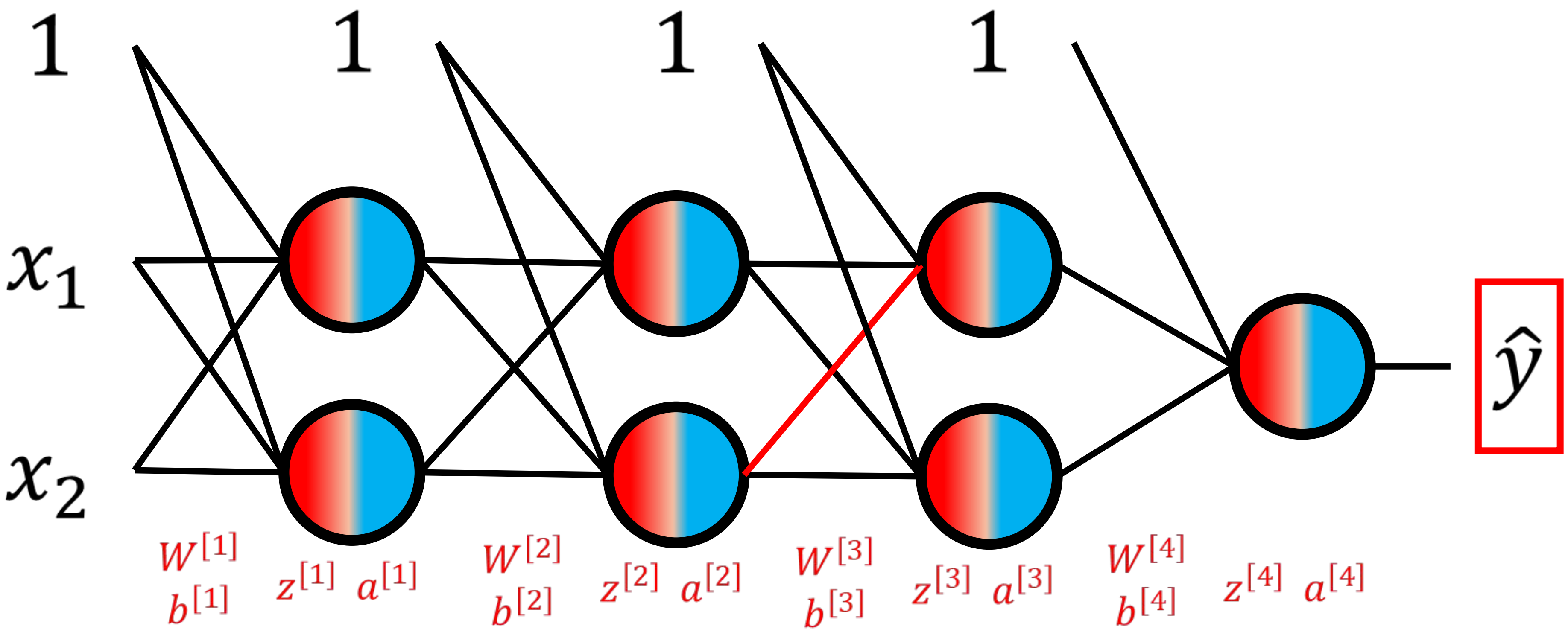
Derivative of $\hat{y}$ with respect to $W_{12}^{[4]}$



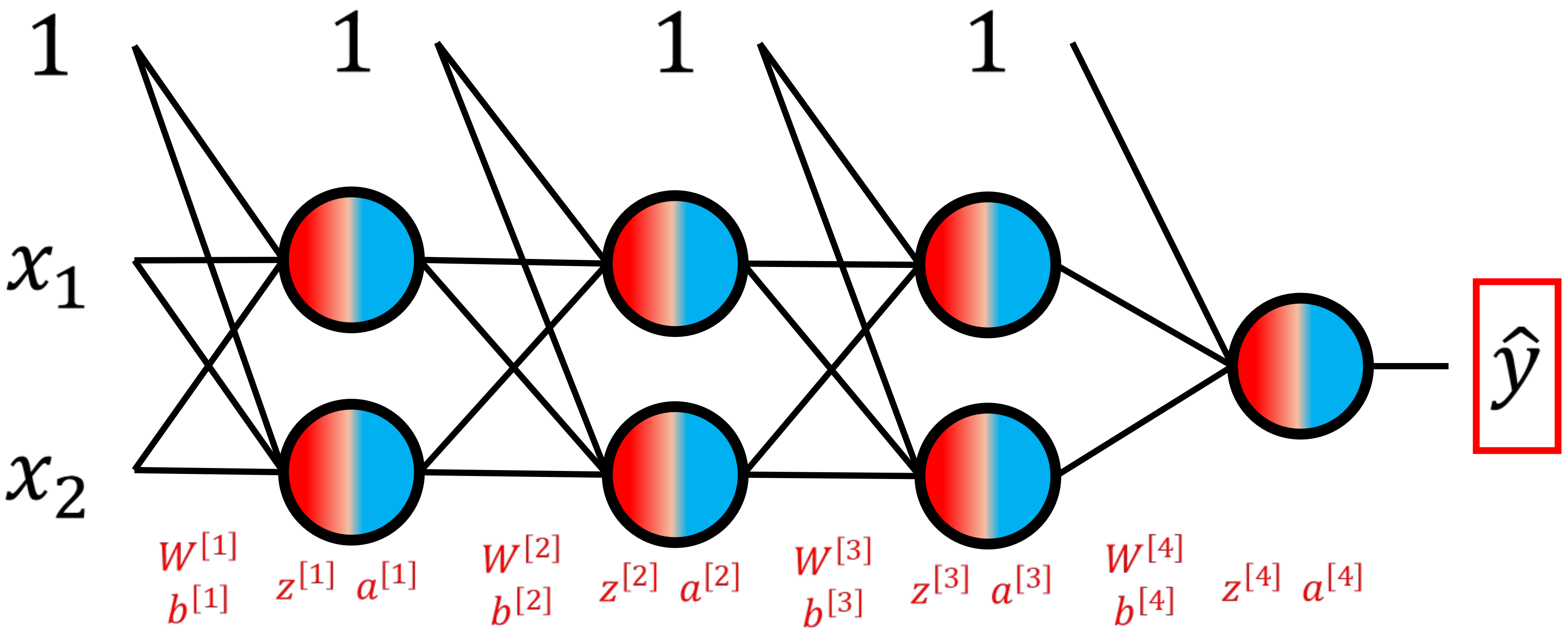
Derivative of $\hat{y}$ with respect to $b_1^{[4]}$



Derivative of $\hat{y}$ with respect to $W_{11}^{[3]}$

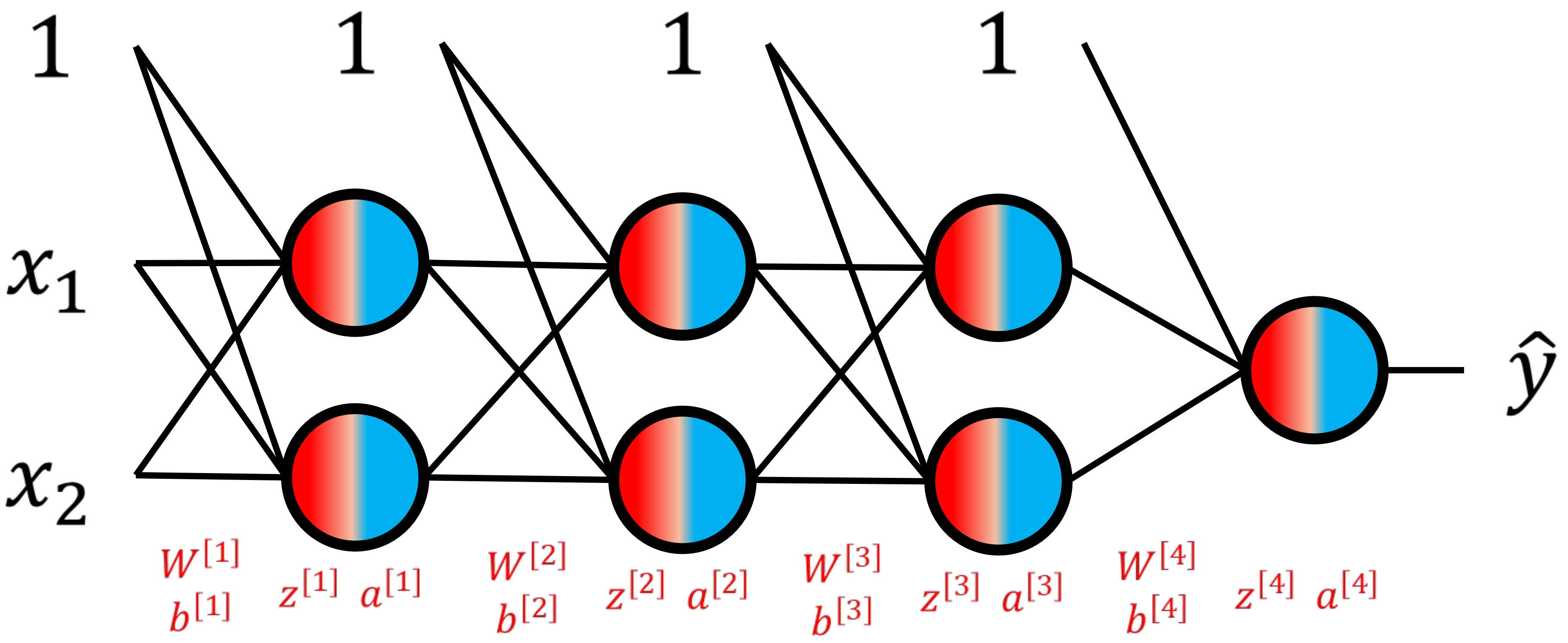


Derivative of $\hat{y}$ with respect to $W_{12}^{[3]}$

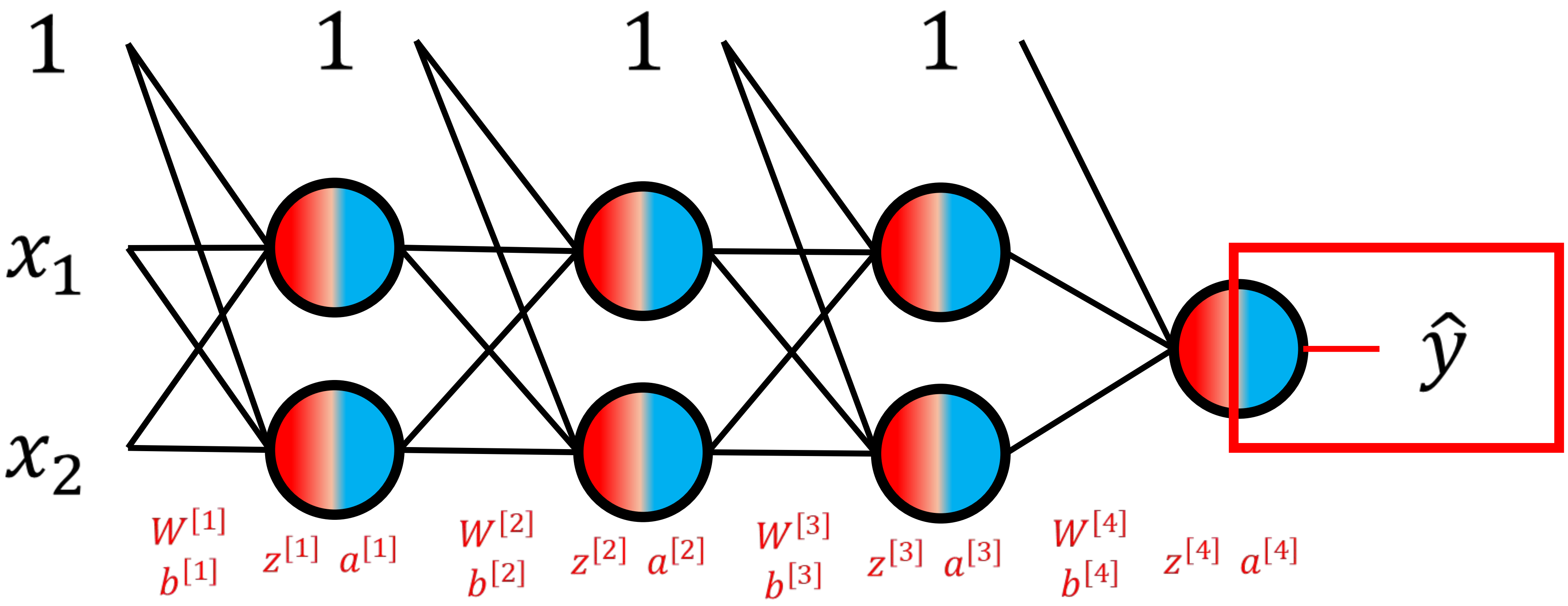


And so on and so forth until we reach the first layer…

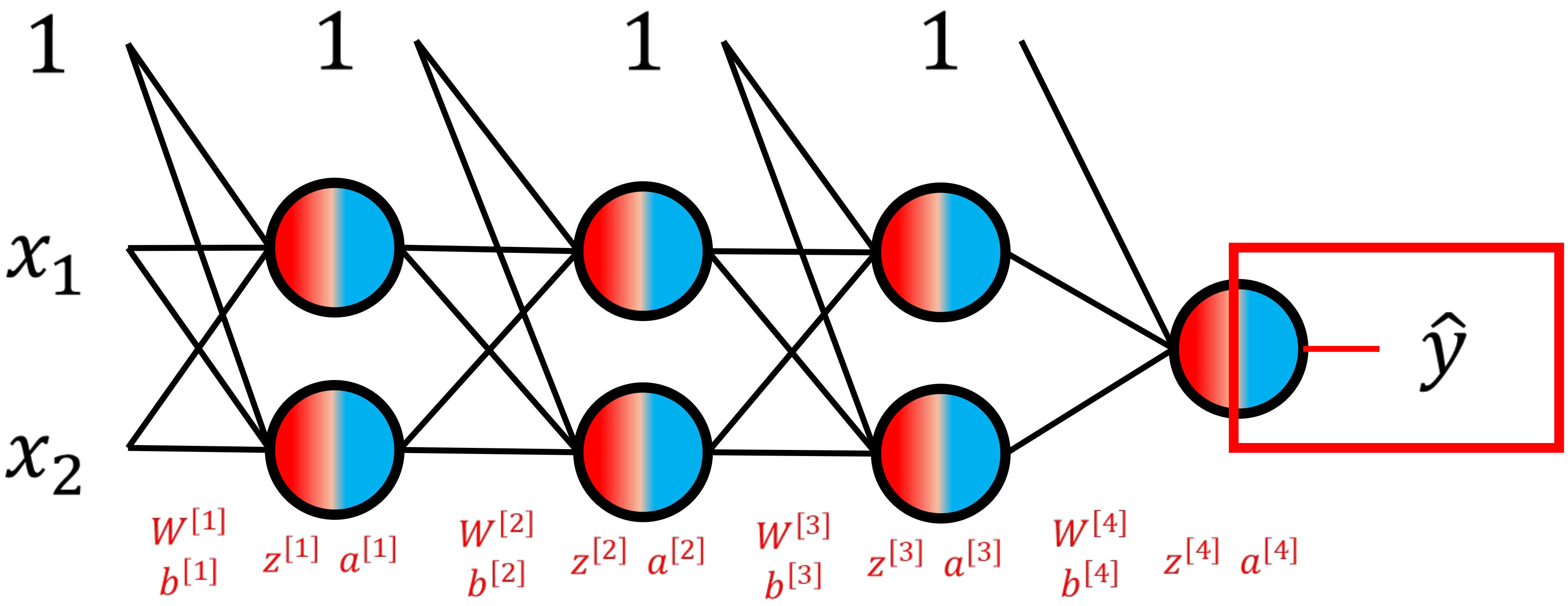
3. Computing the Derivatives



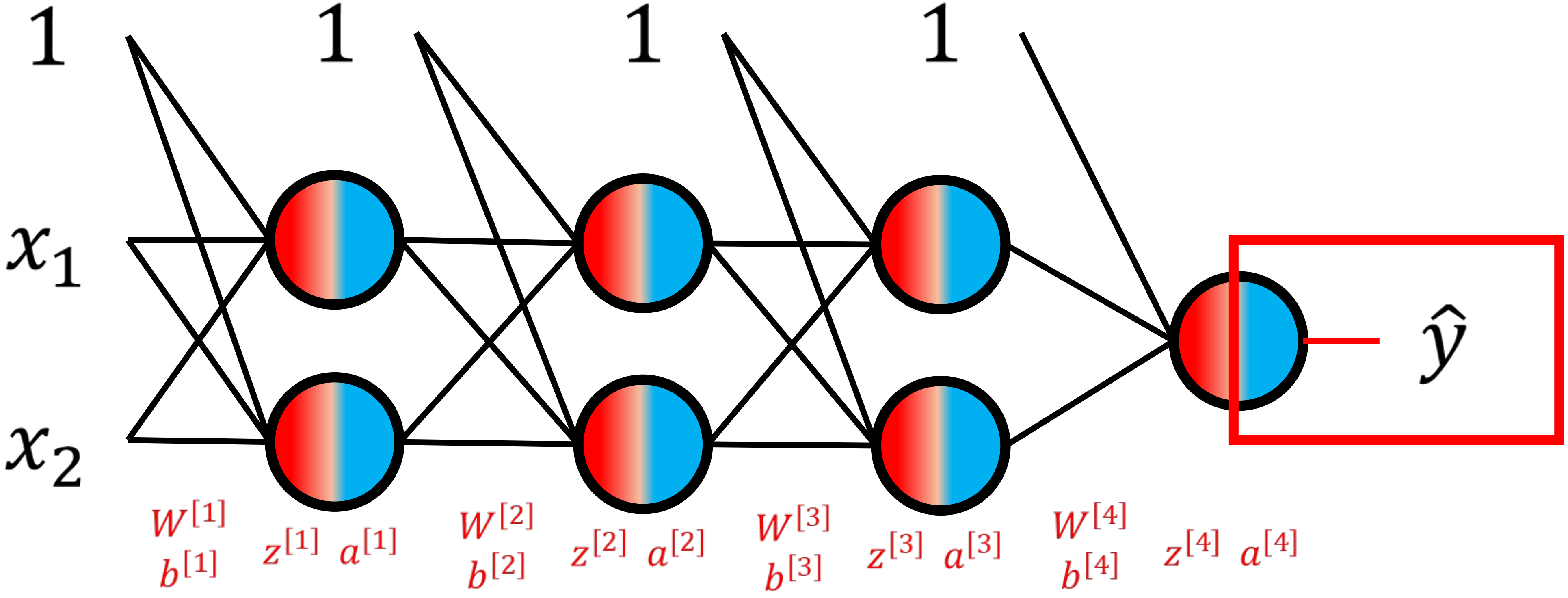
Strategy: We compute the derivatives of neighboring operations first, then we can just use the chain rule later!



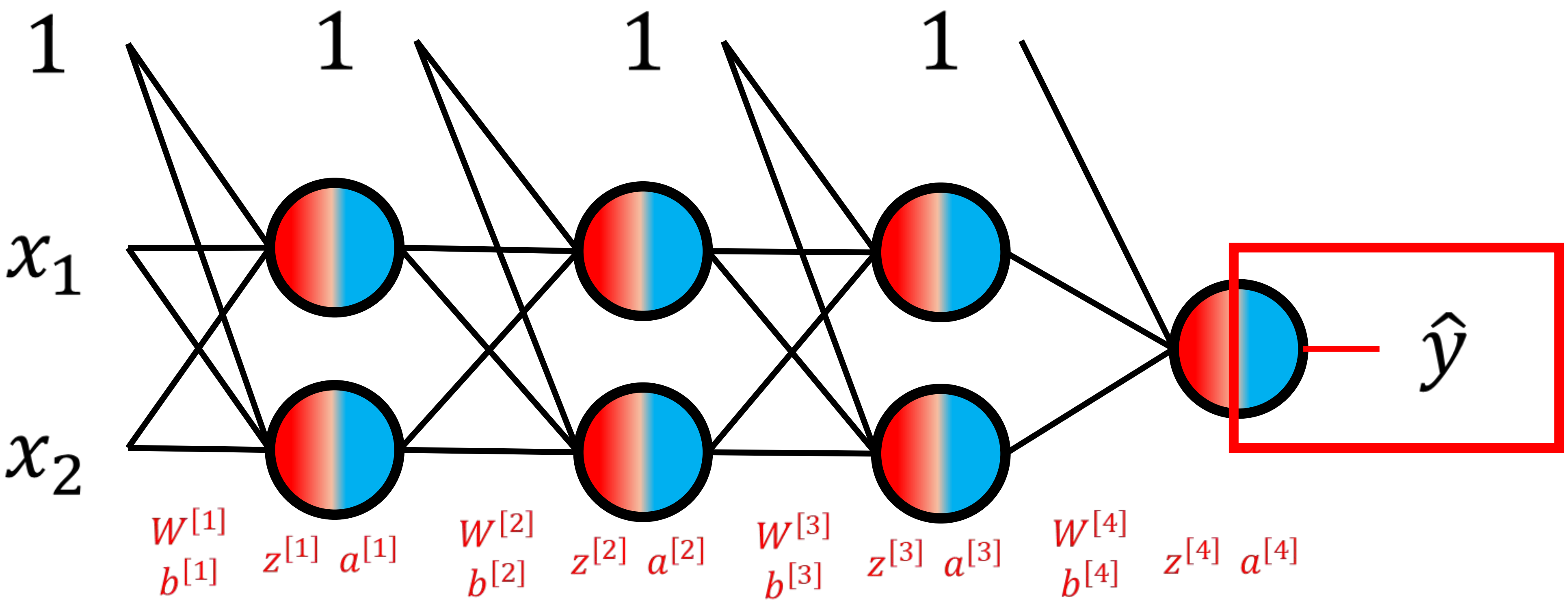
How does $a_1^{[4]}$ affect $\hat{y}$?



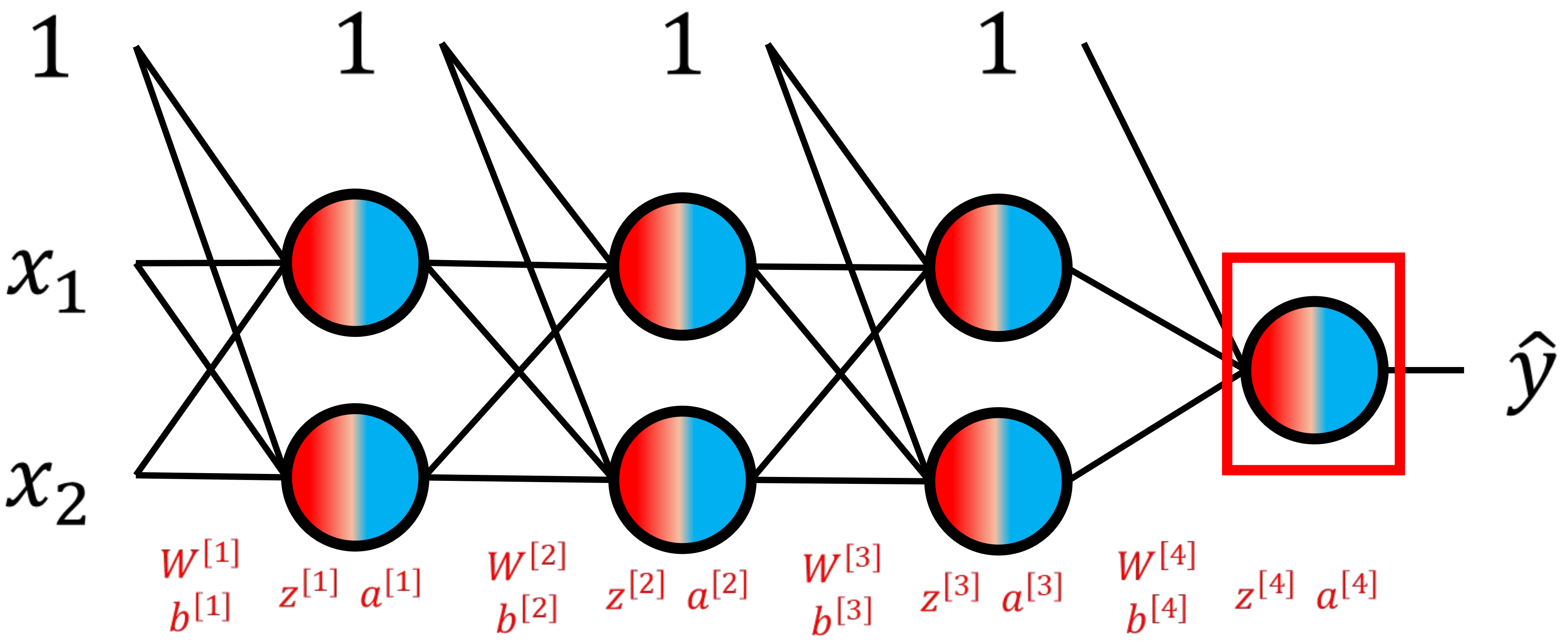
$$\frac{\partial \hat{y}}{\partial a_1^{[4]}}$$



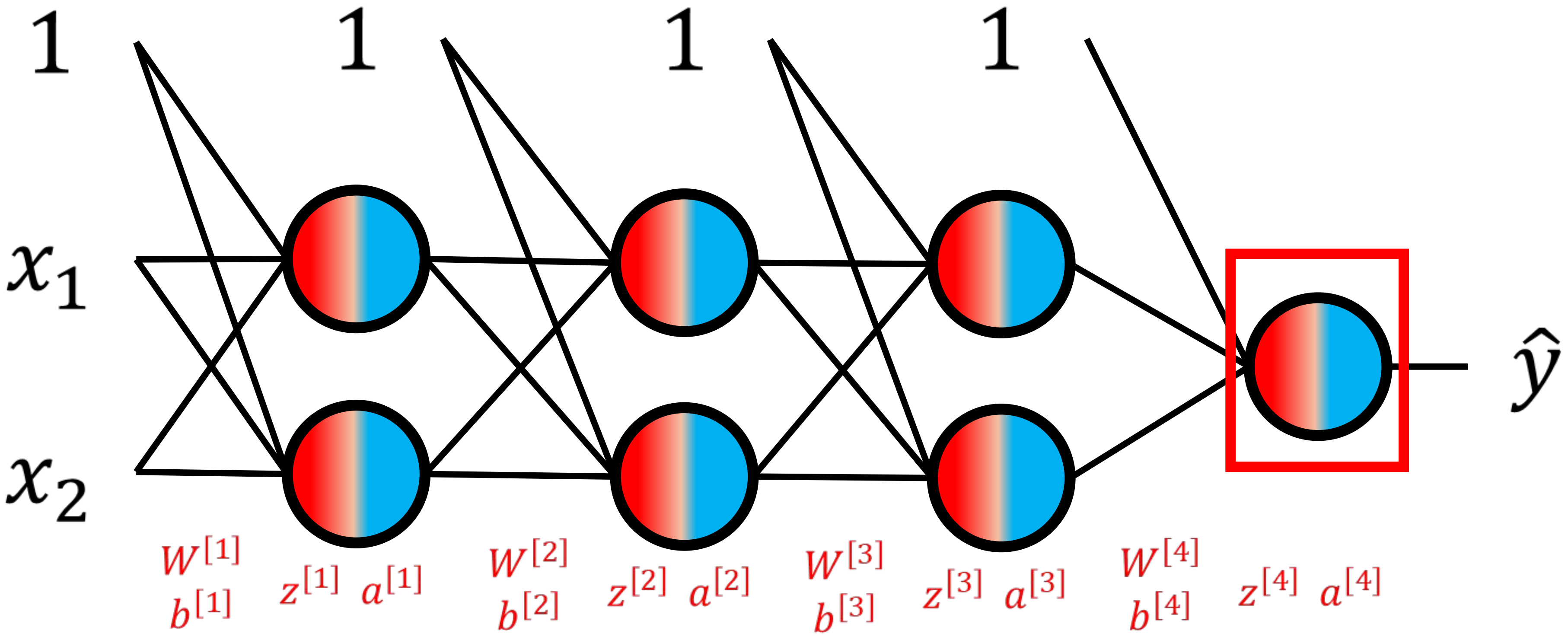
$$\frac{\partial a_1^{[4]}}{\partial a_1^{[4]}}$$



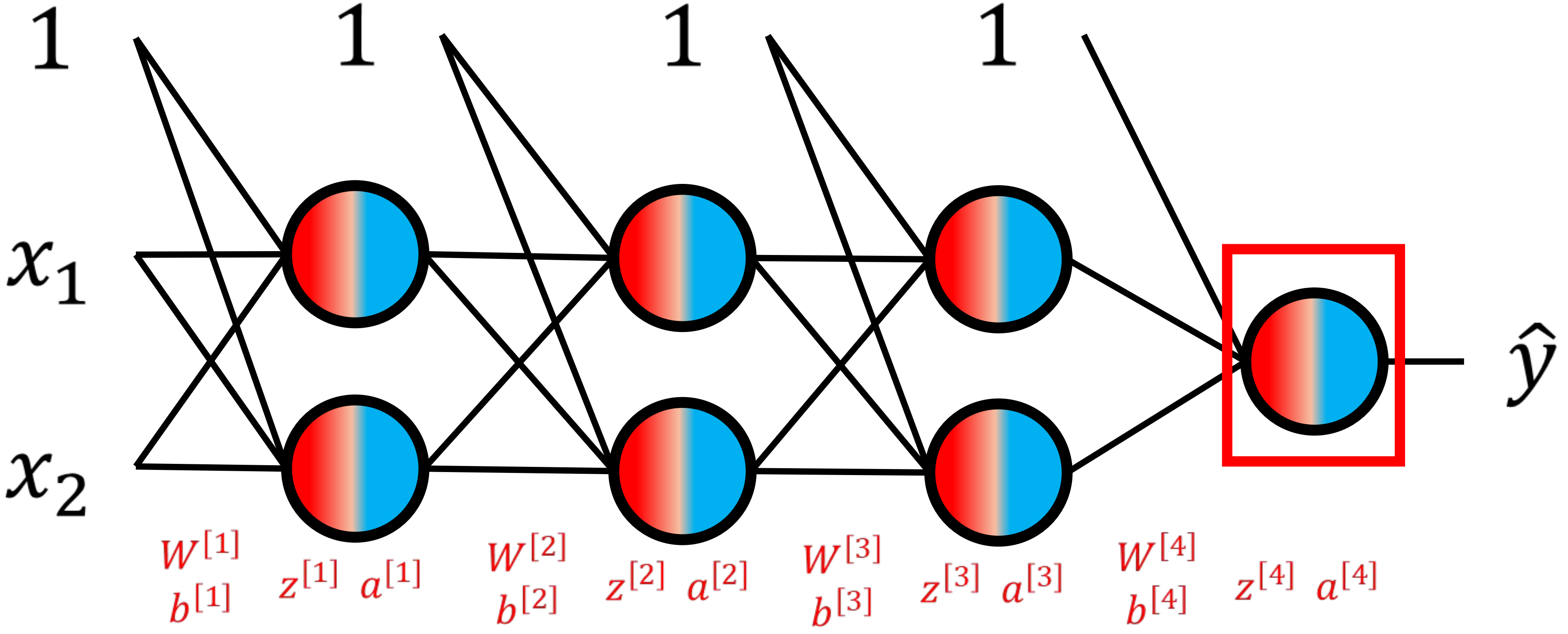
$$\frac{\partial a_1^{[4]}}{\partial a_1^{[4]}} = 1$$



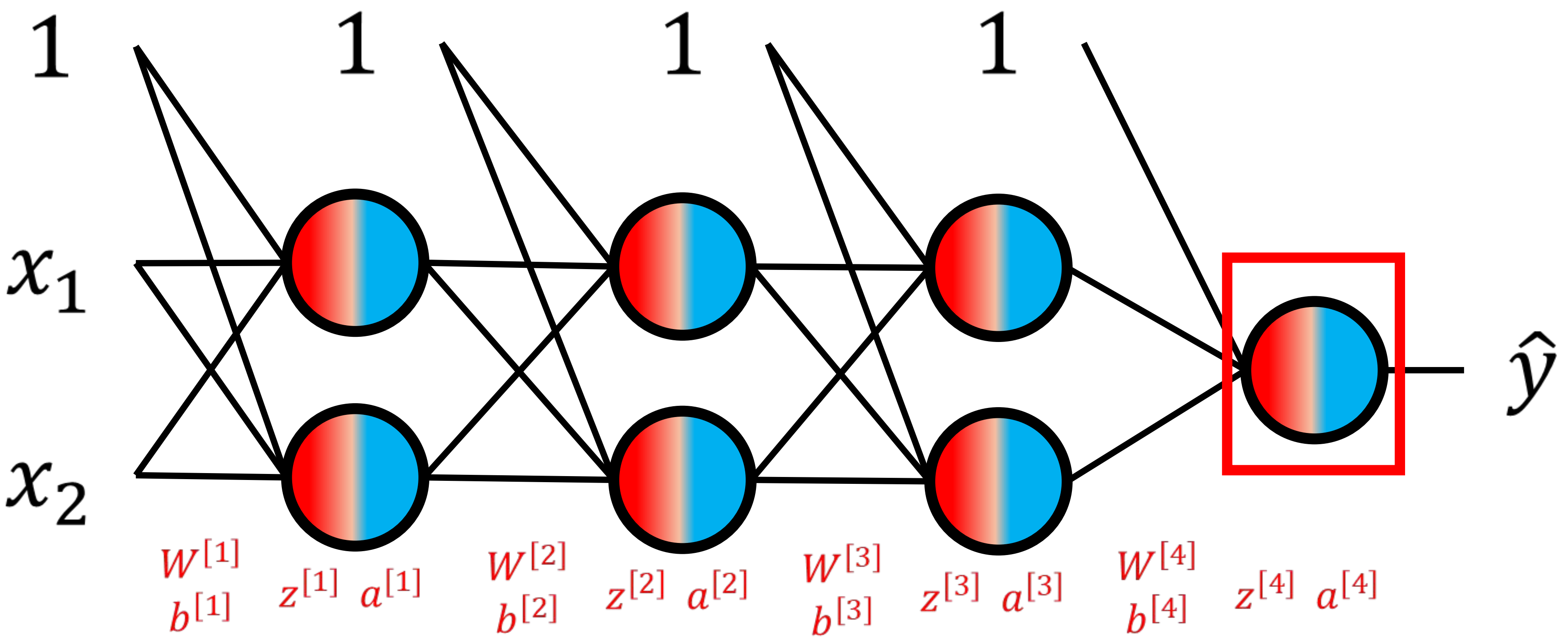
How does $z_1^{[4]}$ affect $a_1^{[4]}$?



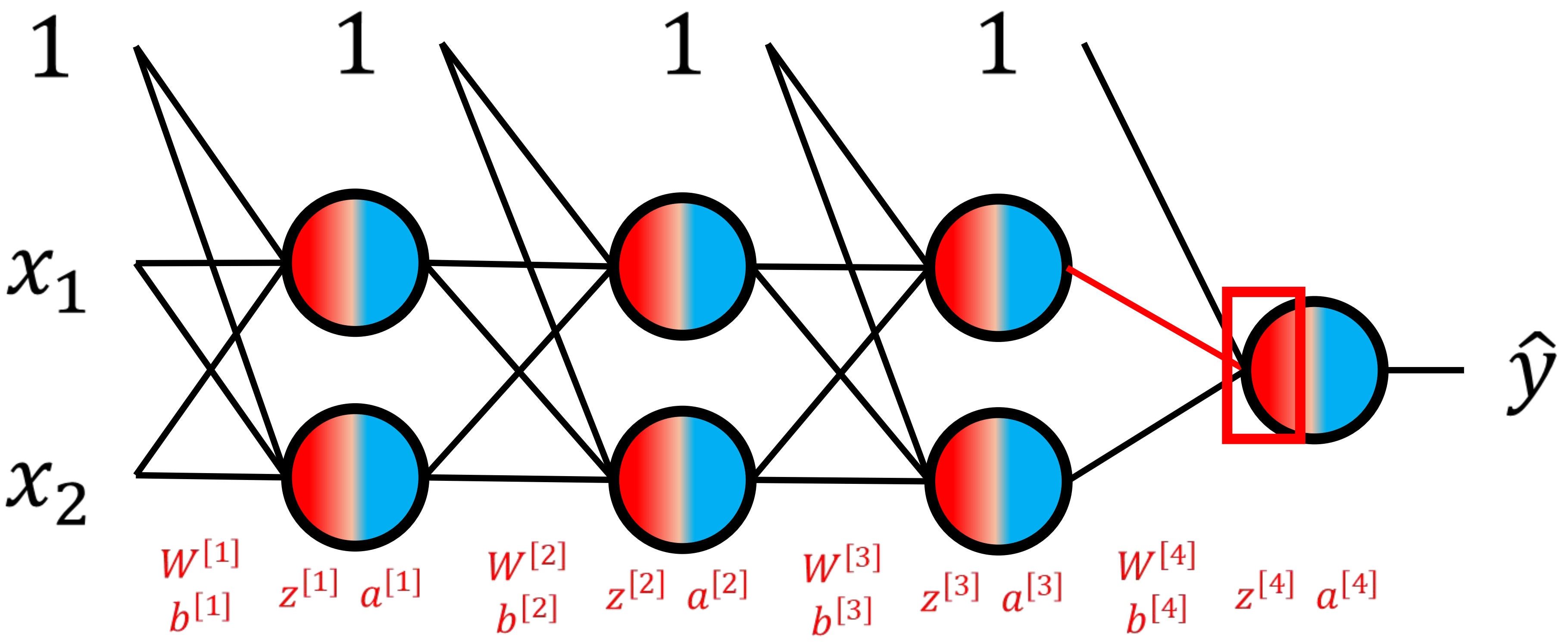
$$\frac{\partial a_1^{[4]}}{\partial z_1^{[4]}}$$



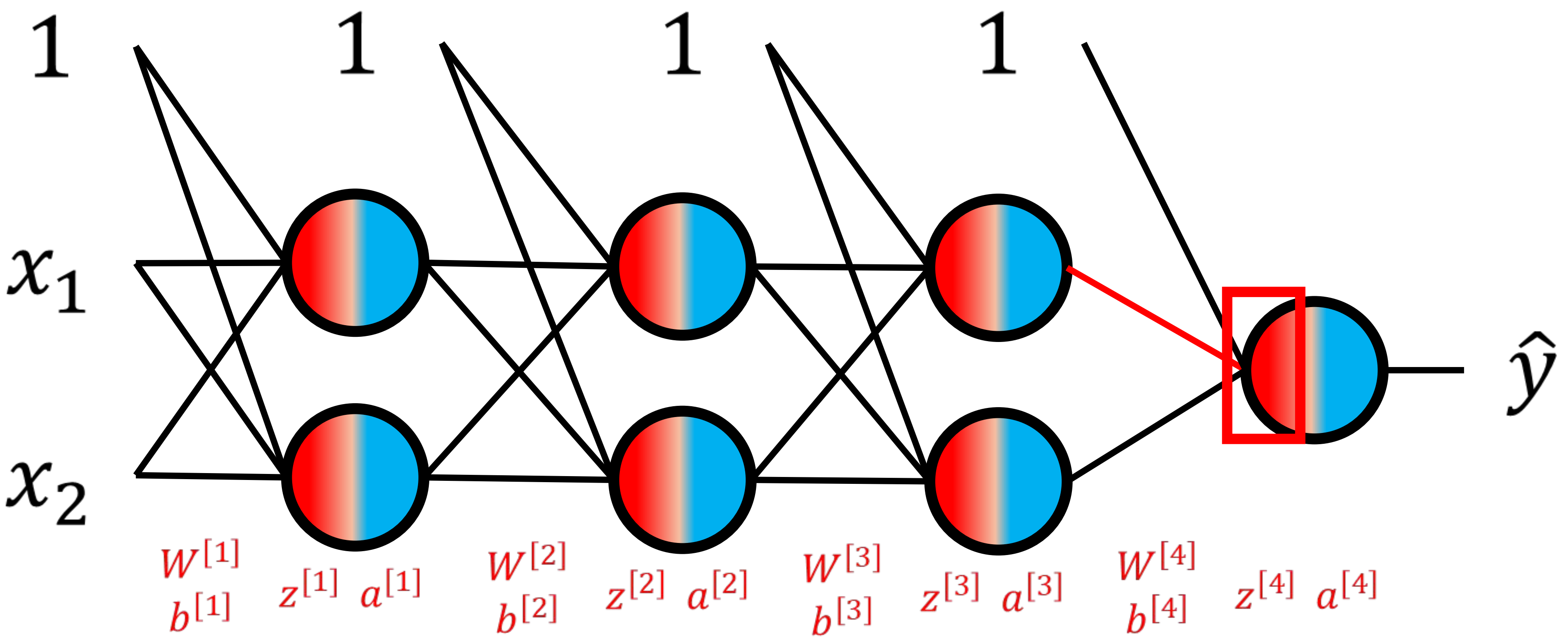
$$\frac{\partial \sigma(z_1^{[4]})}{\partial z_1^{[4]}}$$



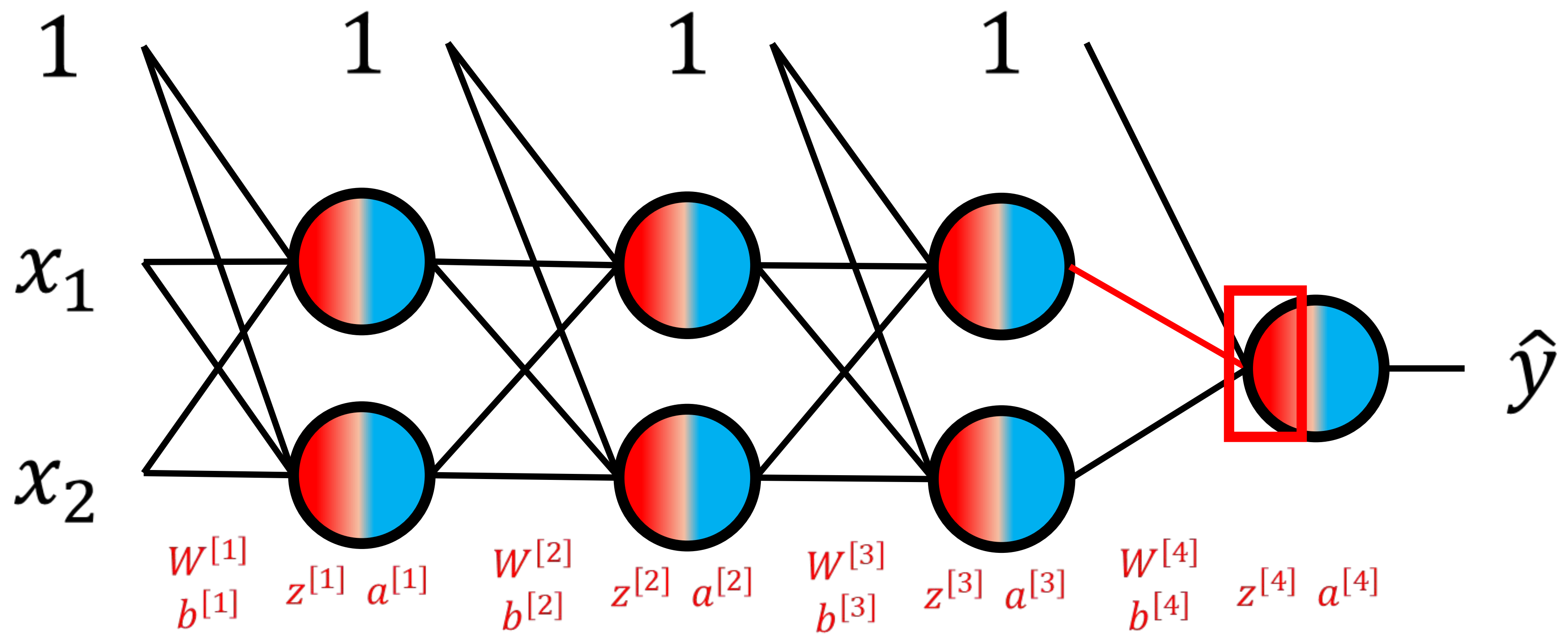
$$\frac{\partial \sigma(z_1^{[4]})}{\partial z_1^{[4]}} = \sigma\left(z_1^{[4]}\right)\left(1 - \sigma\left(z_1^{[4]}\right)\right)$$



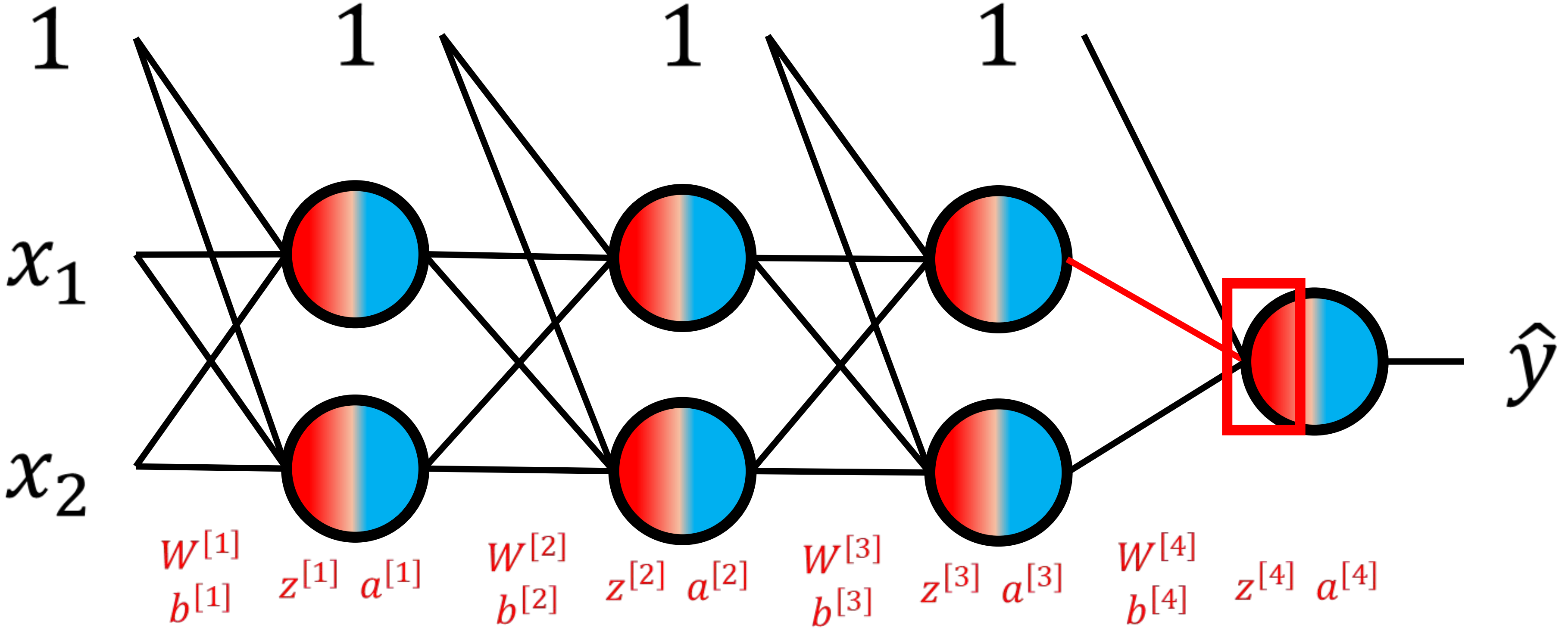
How does $W_{11}^{[4]}$ affect $z_1^{[4]}$?



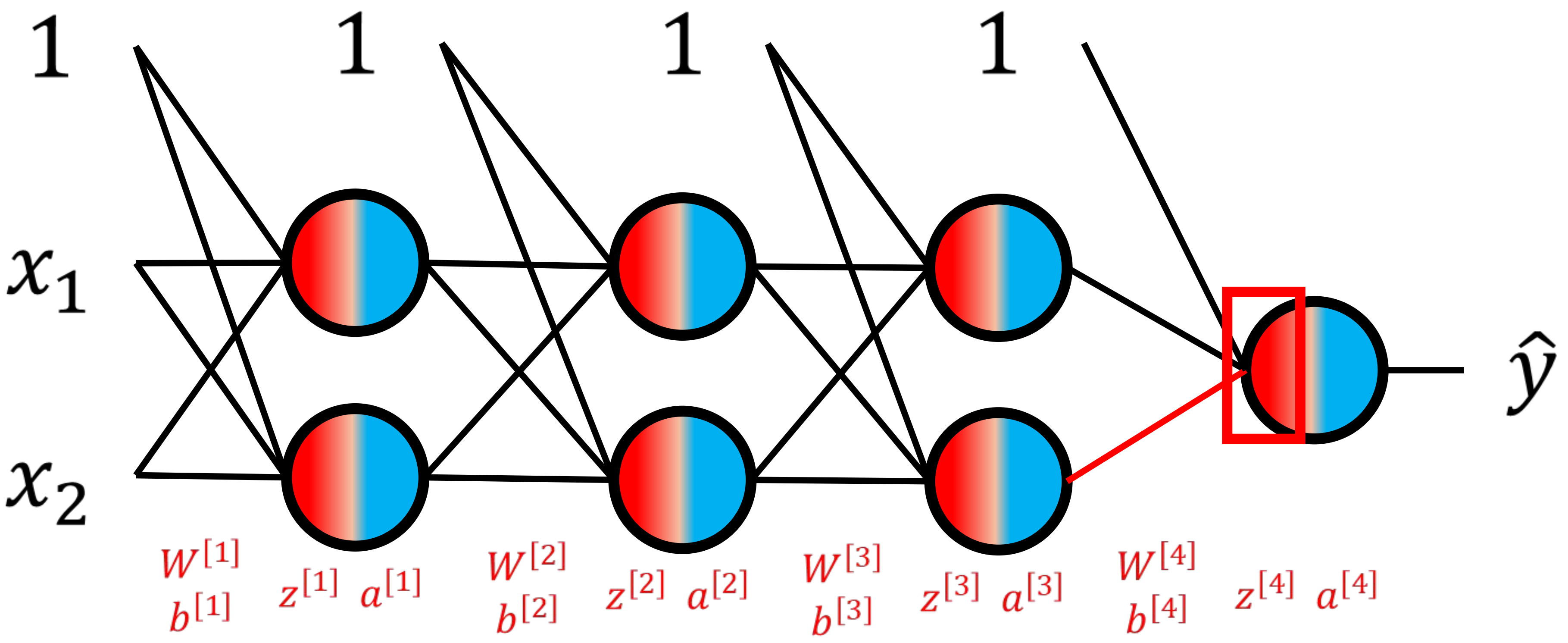
$$\frac{\partial z_1^{[4]}}{\partial W_{11}^{[4]}}$$



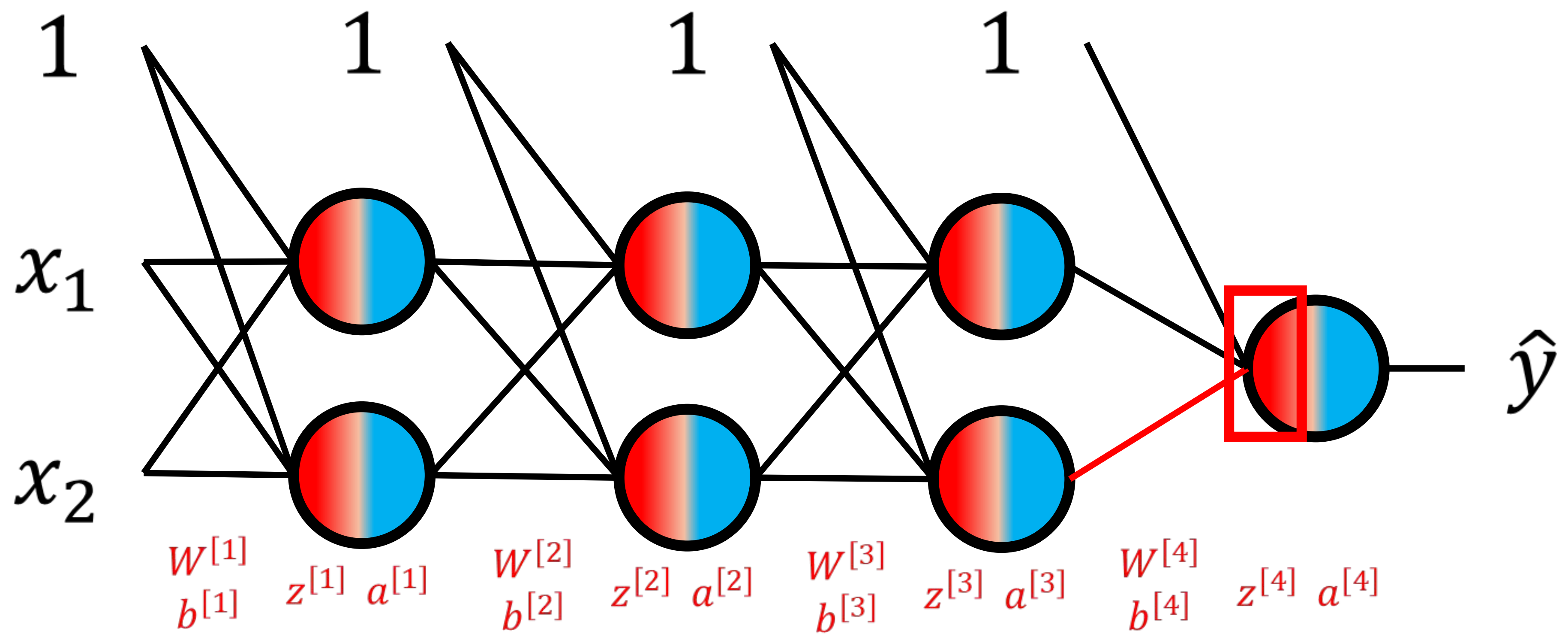
$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial W_{11}^{[4]}}$$



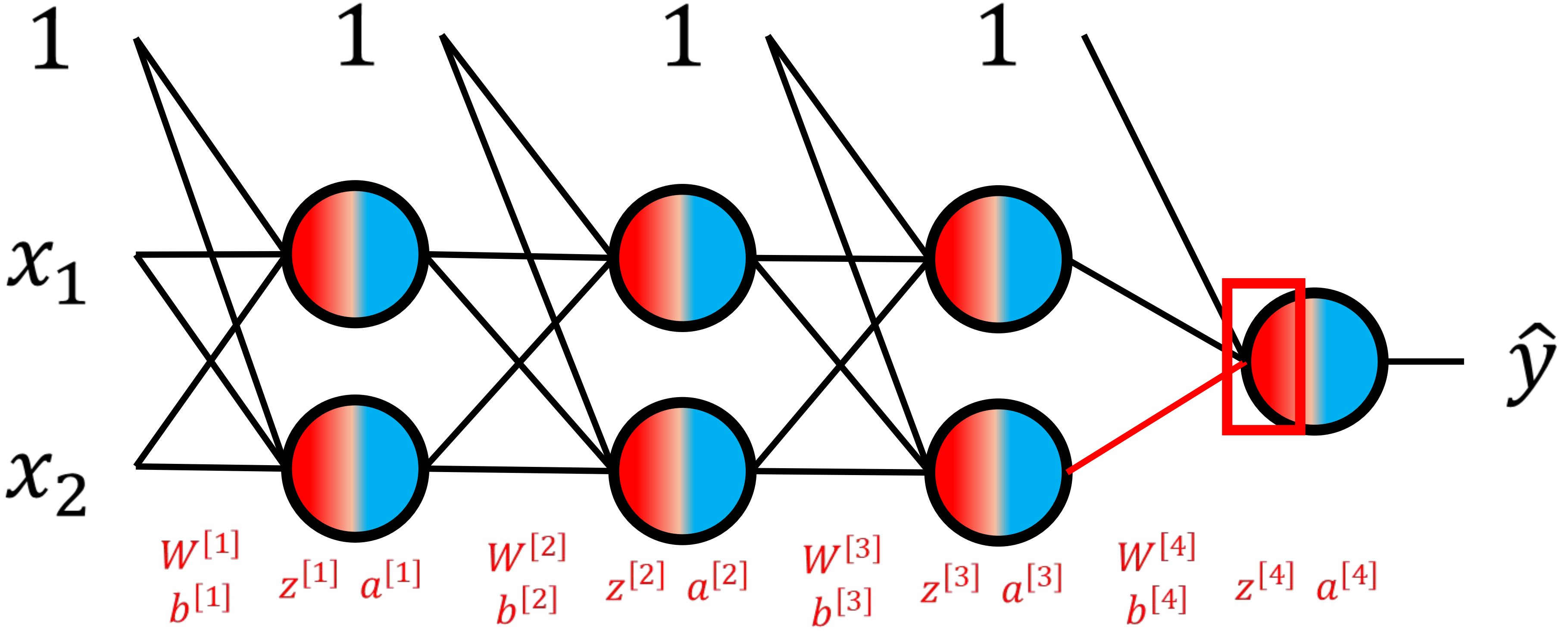
$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial W_{11}^{[4]}} = a_1^{[3]}$$



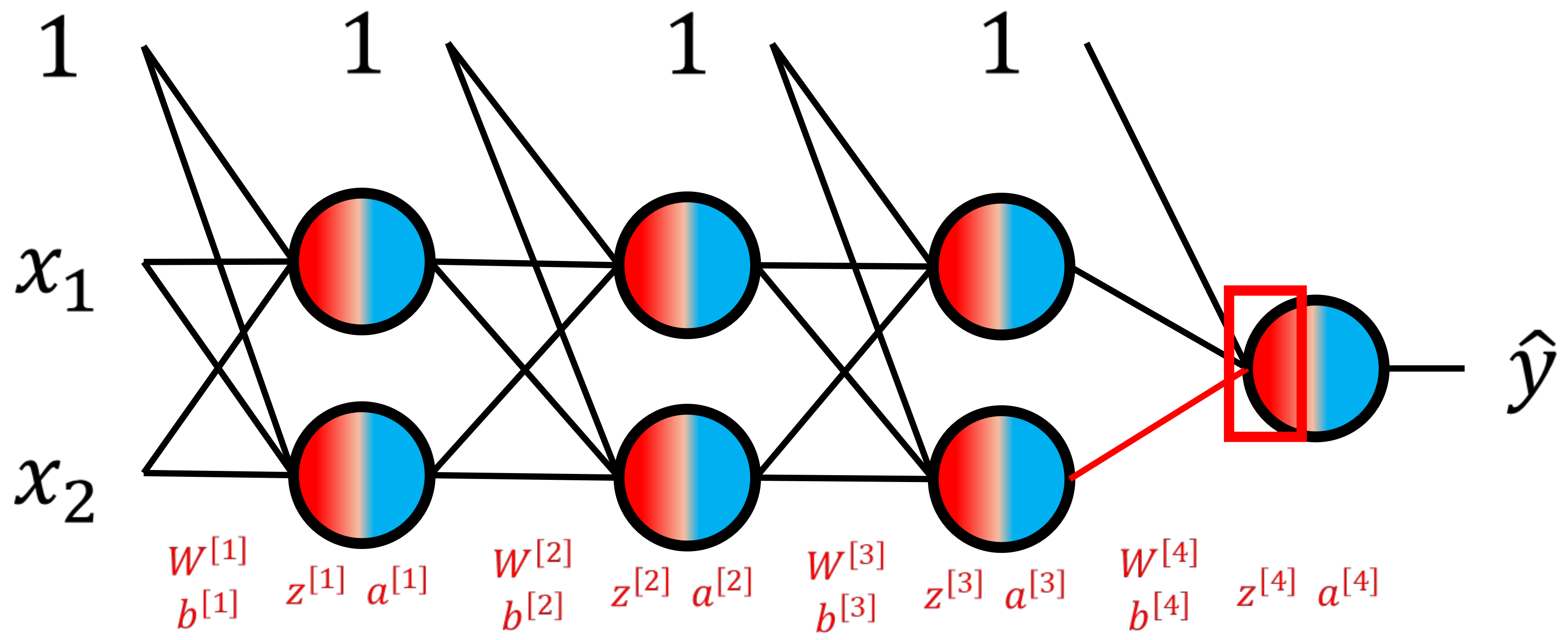
How does $W_{12}^{[4]}$ affect $z_1^{[4]}$?



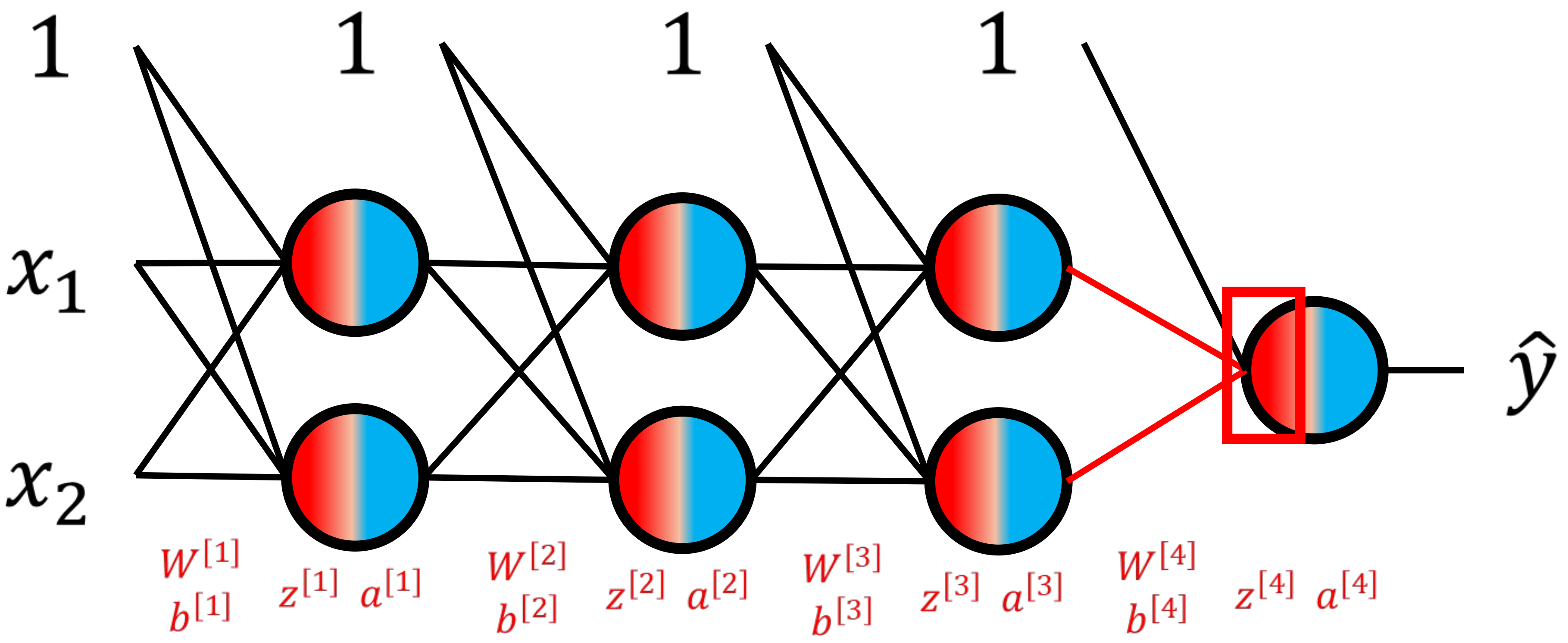
$$\frac{\partial z_1^{[4]}}{\partial W_{12}^{[4]}}$$



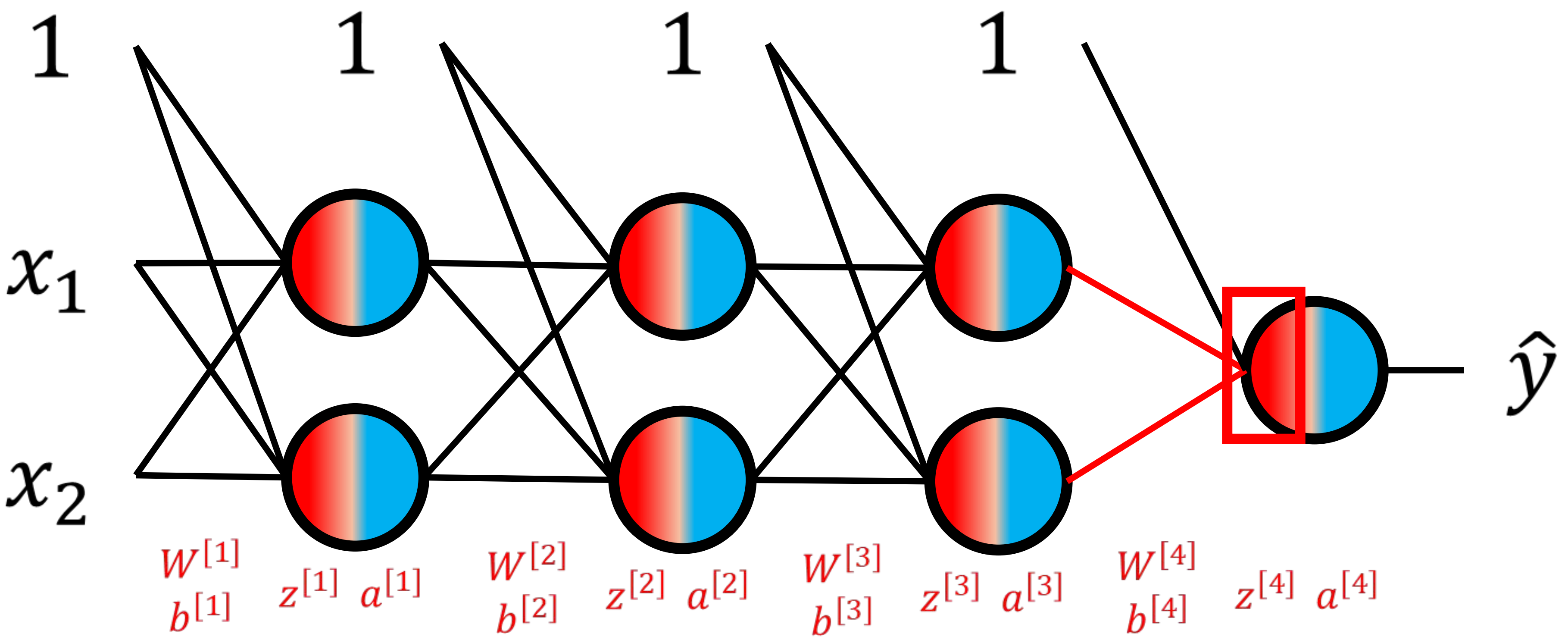
$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial W_{12}^{[4]}}$$



$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial W_{12}^{[4]}} = a_2^{[3]}$$

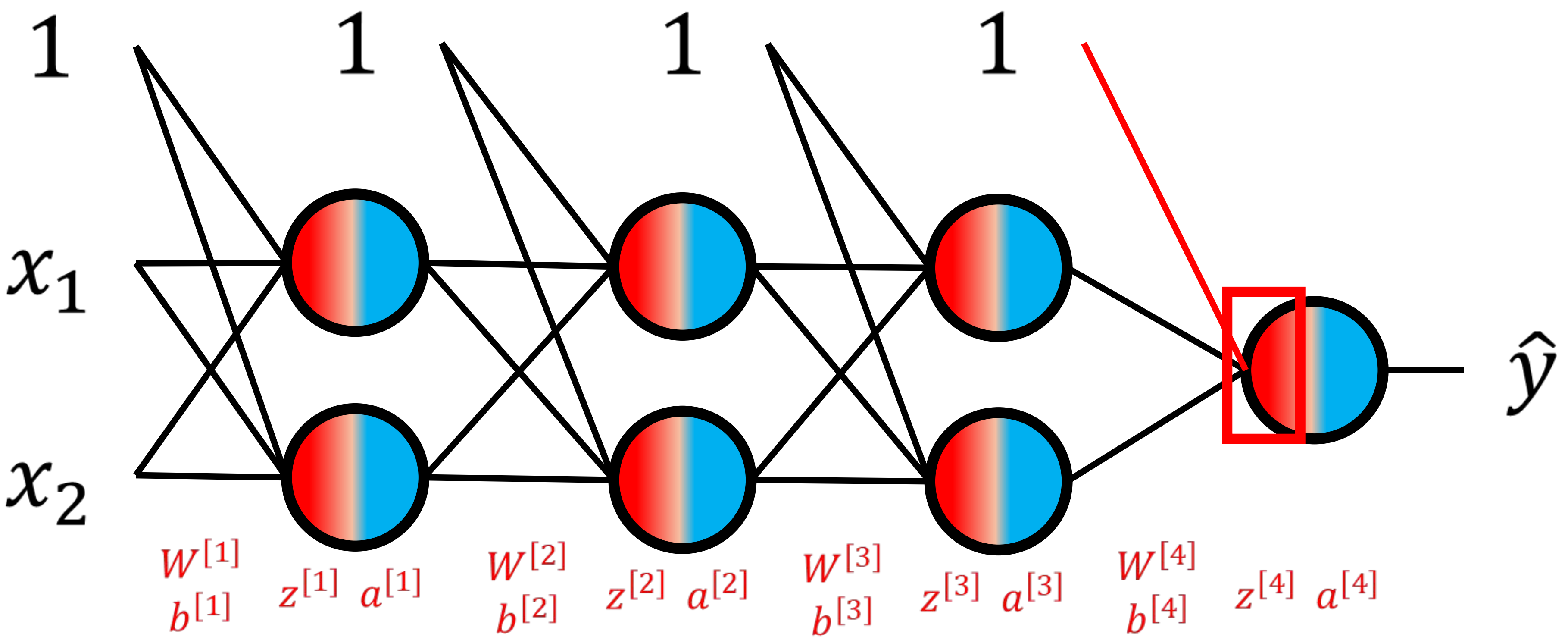


$$\frac{\partial z_1^{[4]}}{\partial W_{11}^{[4]}} = a_1^{[3]} \qquad \frac{\partial z_1^{[4]}}{\partial W_{12}^{[4]}} = a_2^{[3]}$$

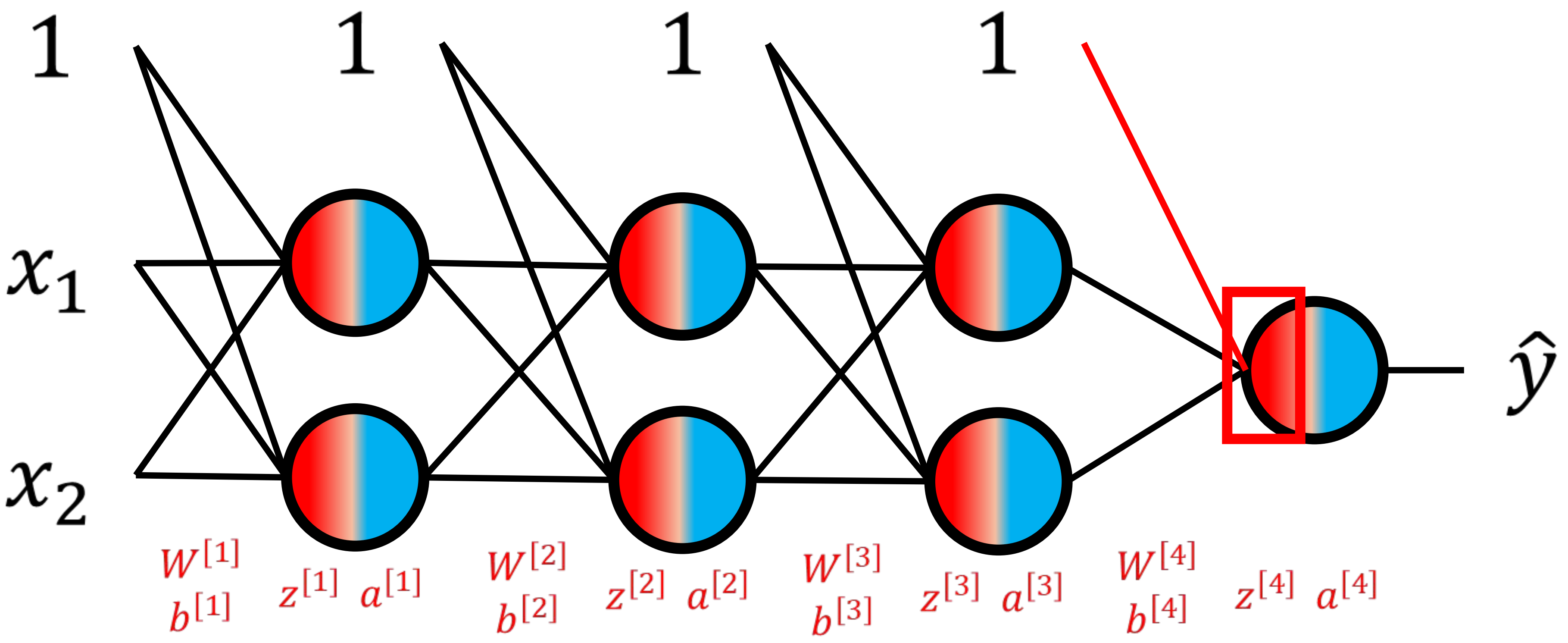


$$\frac{\partial z_1^{[4]}}{\partial W_{11}^{[4]}} = a_1^{[3]} \qquad \frac{\partial z_1^{[4]}}{\partial W_{12}^{[4]}} = a_2^{[3]} \qquad \frac{\partial z^{[4]}}{\partial W^{[4]}} = a^{[3]}$$

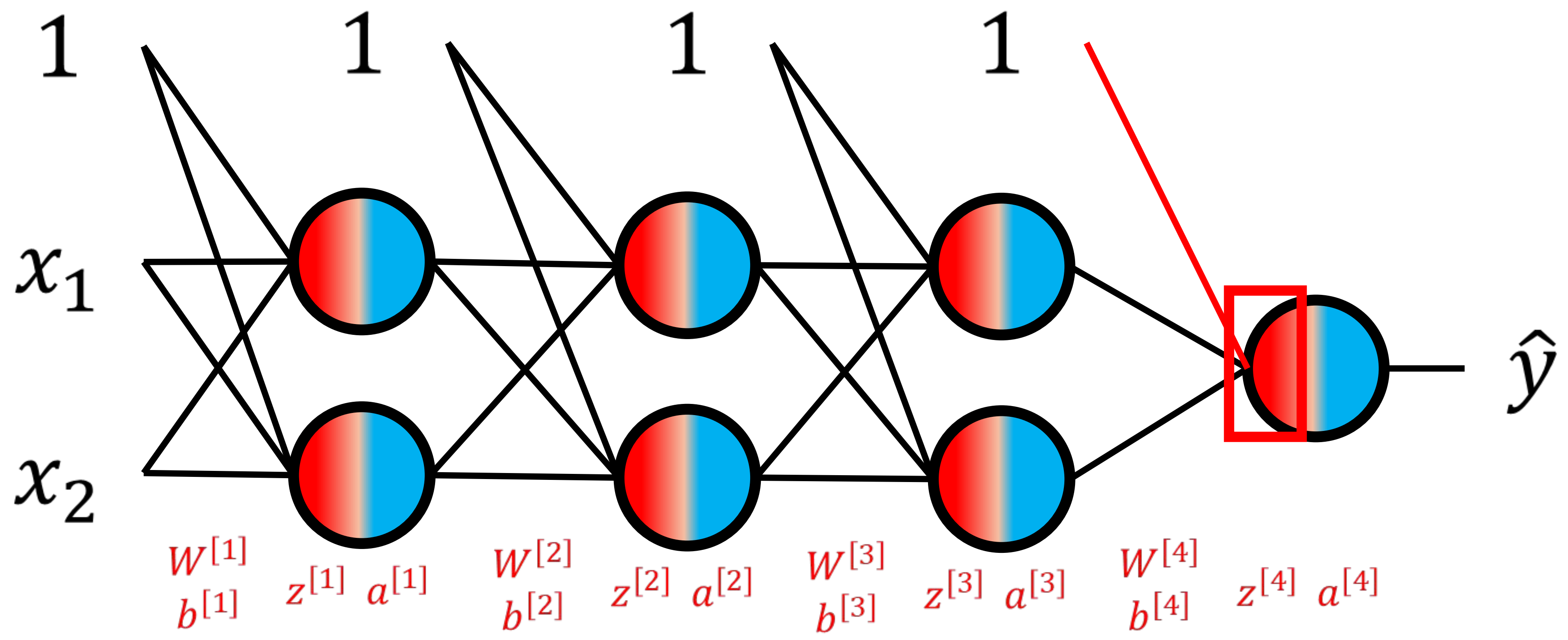
Solve all the gradients all at once!



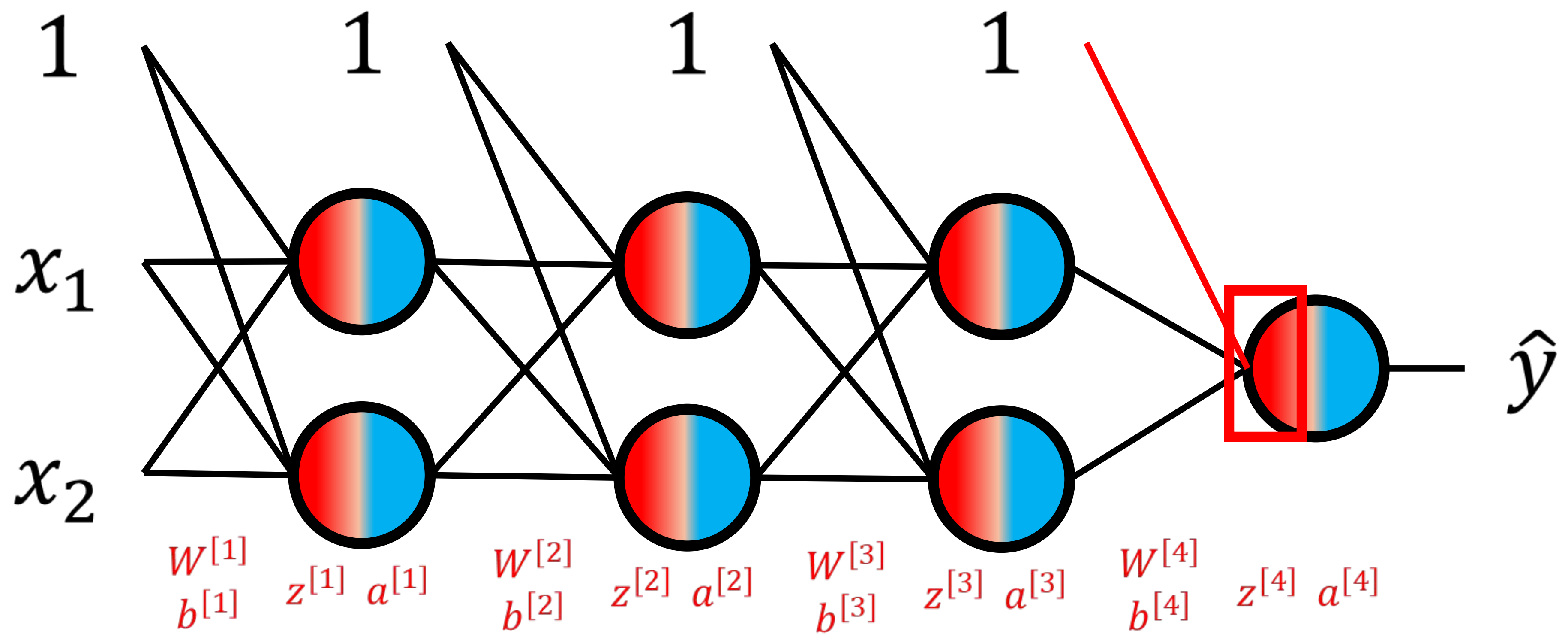
How does $b_1^{[4]}$ affect $z_1^{[4]}$?



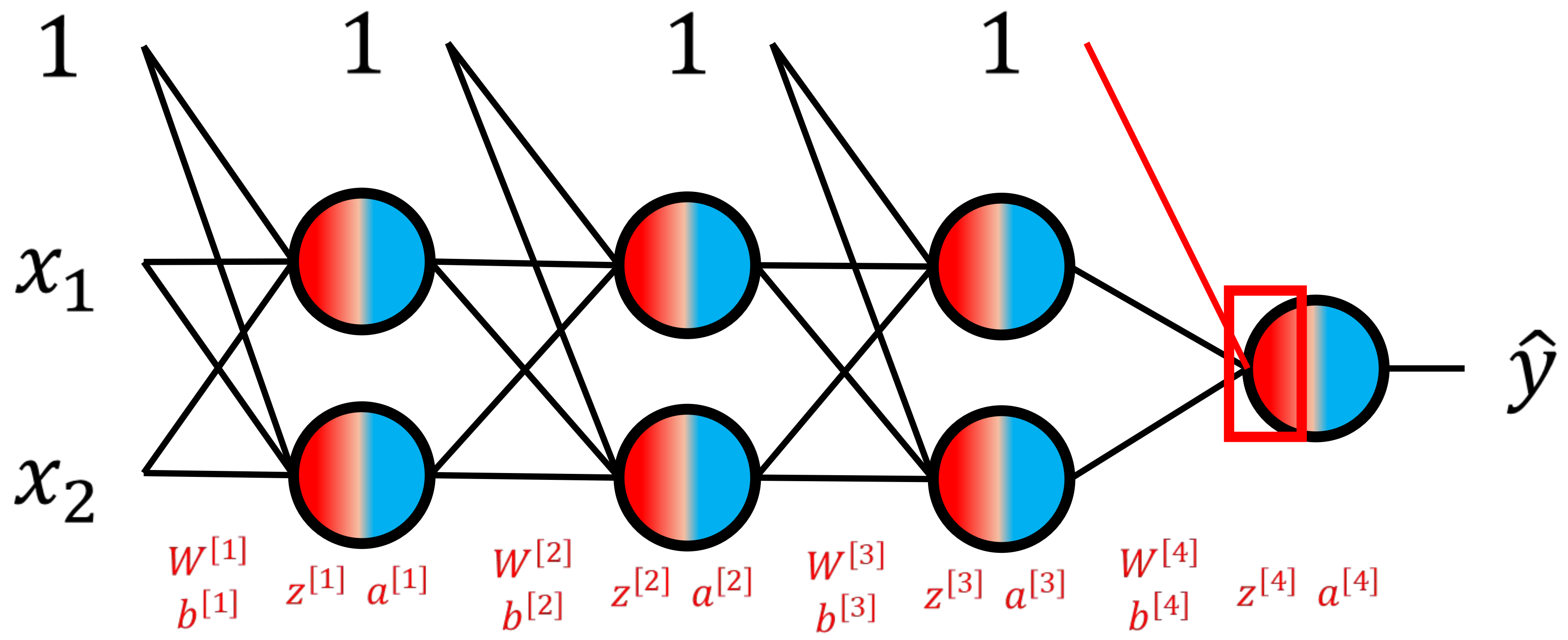
$$\frac{\partial z_1^{[4]}}{\partial b_1^{[4]}}$$



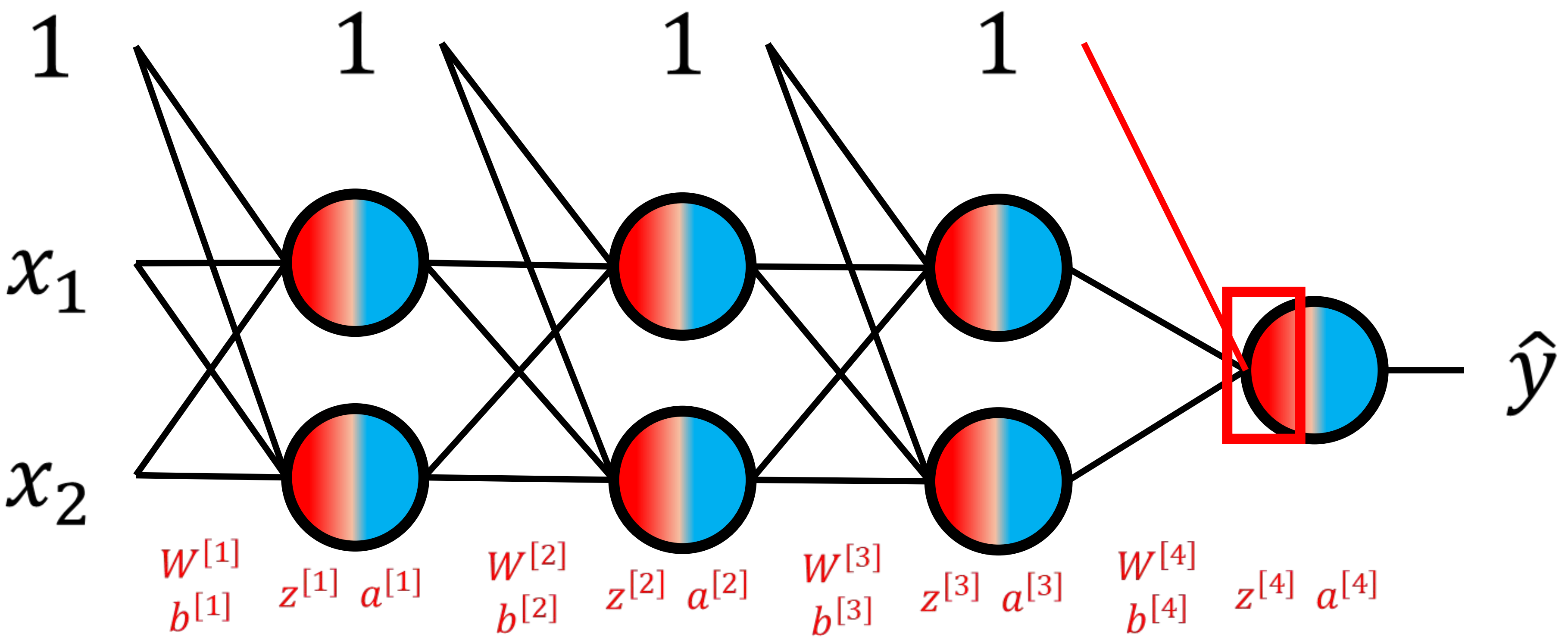
$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial b_1^{[4]}}$$



$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial b_1^{[4]}} = 1$$



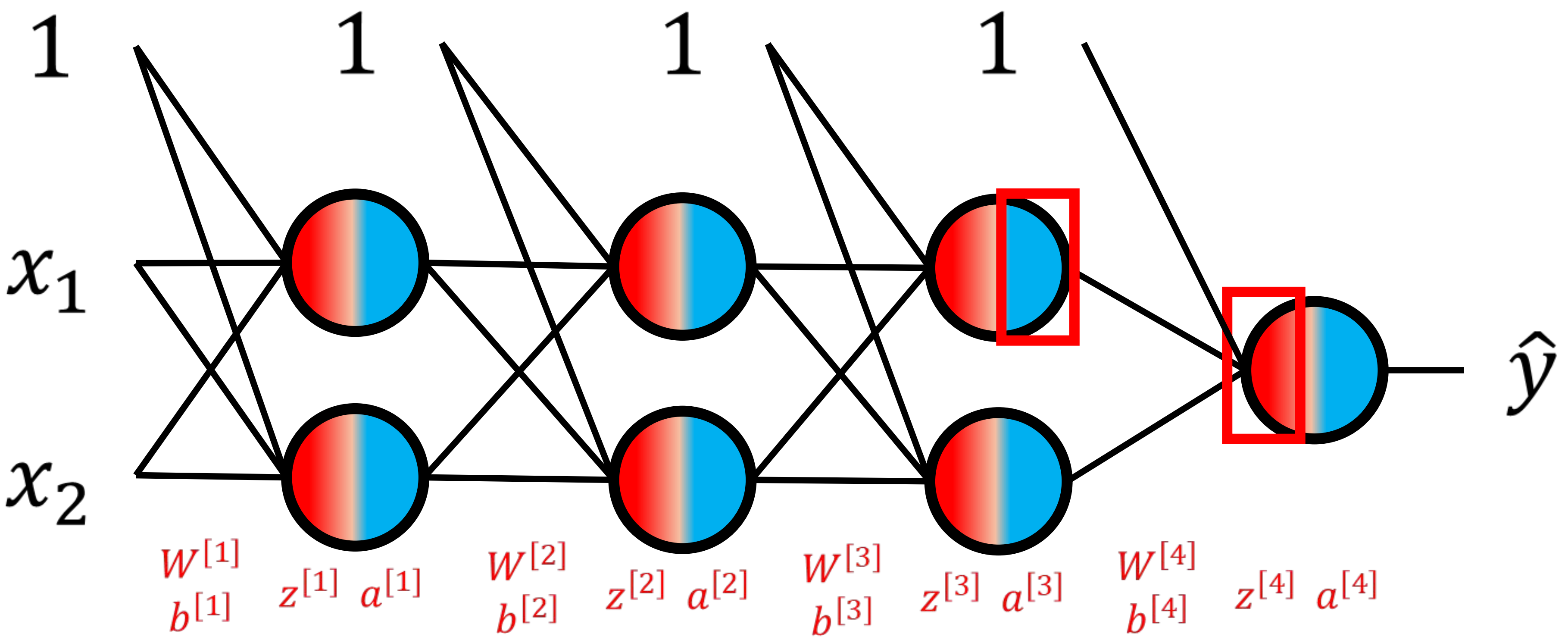
$$\frac{\partial z_1^{[4]}}{\partial b_1^{[4]}} = 1$$



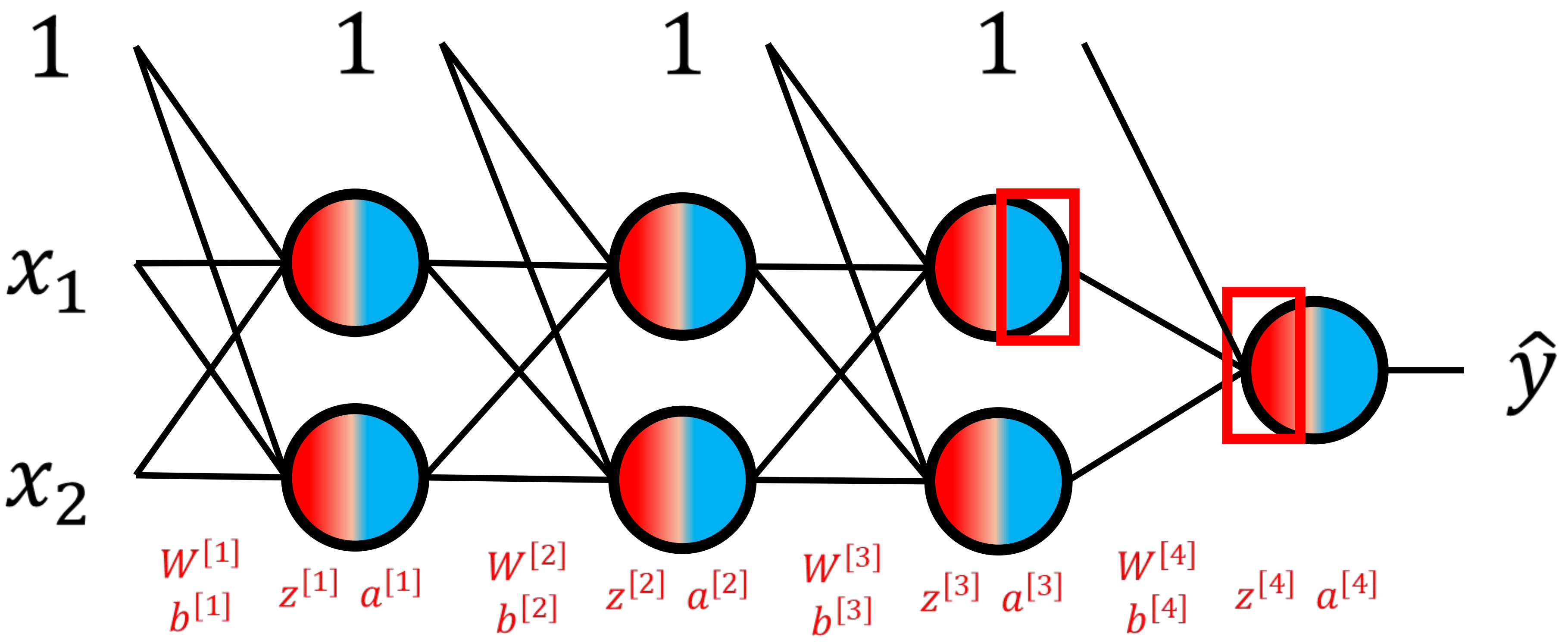
$$\frac{\partial z_1^{[4]}}{\partial b_1^{[4]}} = 1$$

$$\frac{\partial z^{[4]}}{\partial b^{[4]}} = 1$$

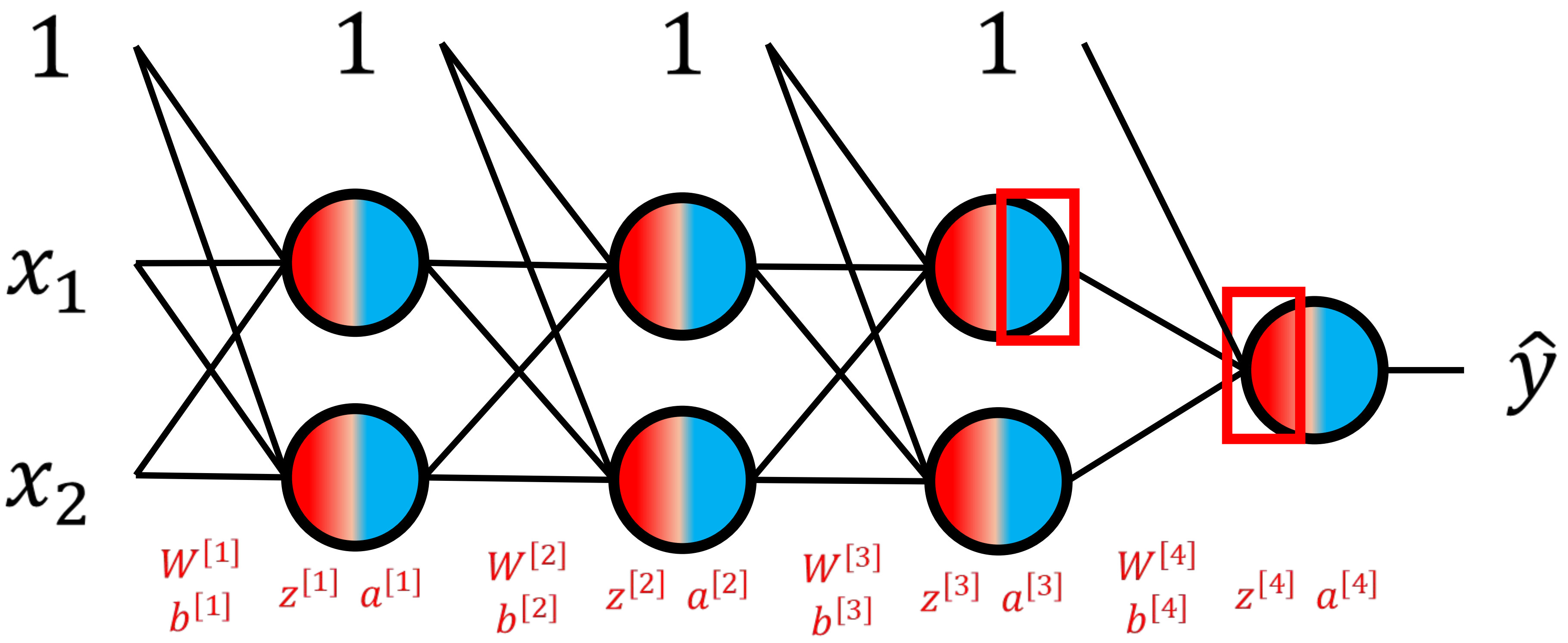
Solve for all the gradients at once!



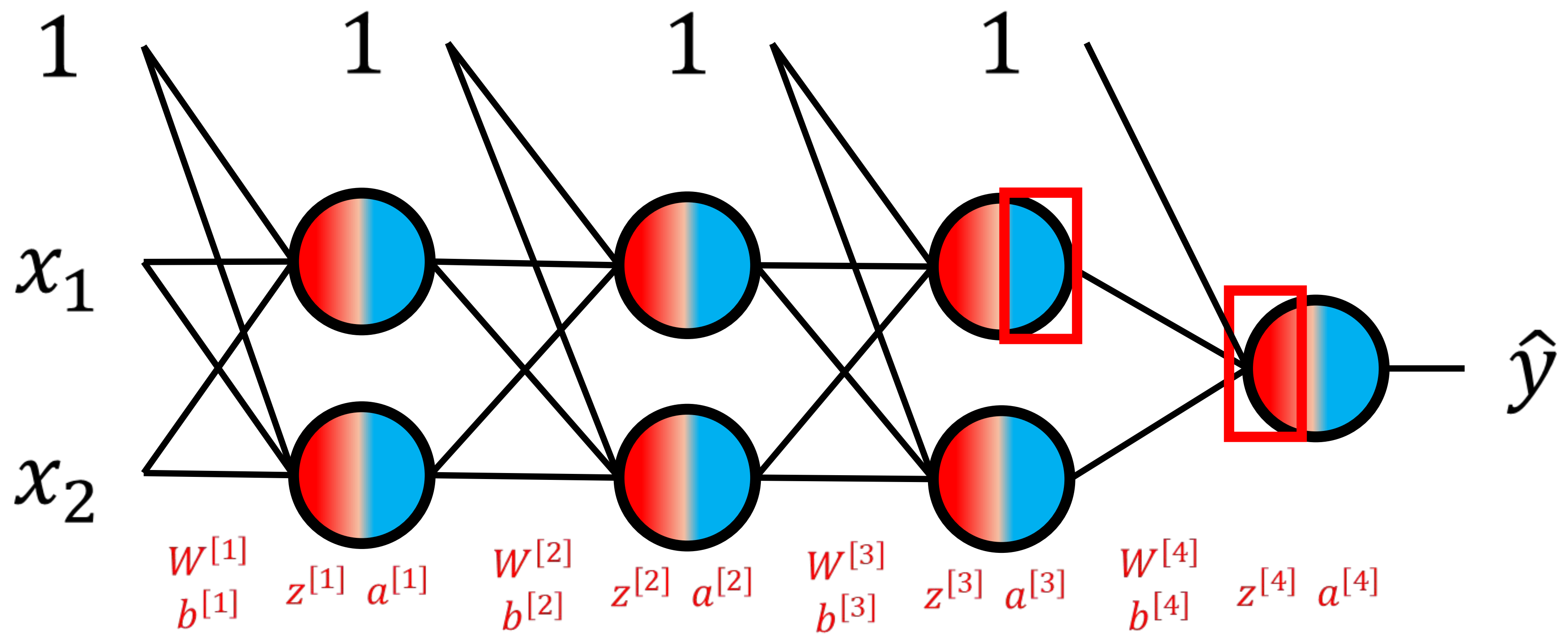
How does $a_1^{[3]}$ affect $z_1^{[4]}$?



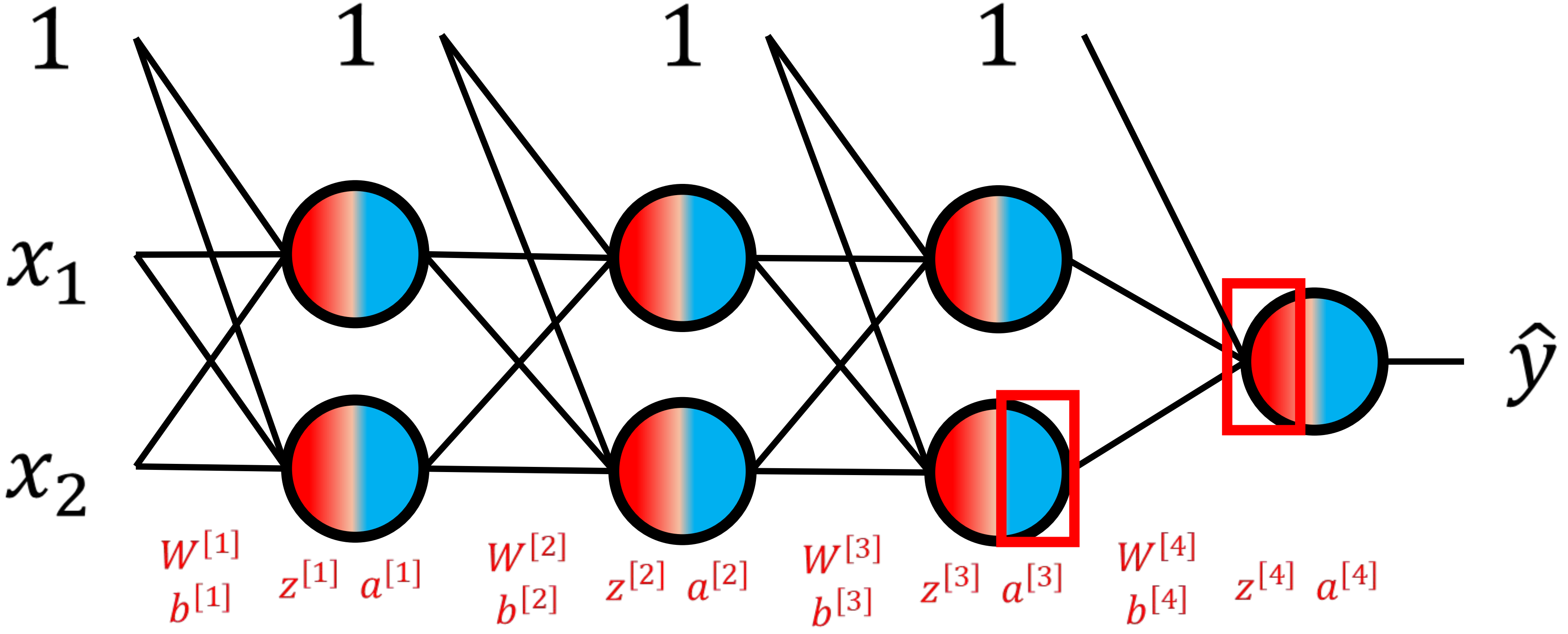
$$\frac{\partial z_1^{[4]}}{\partial a_1^{[3]}}$$



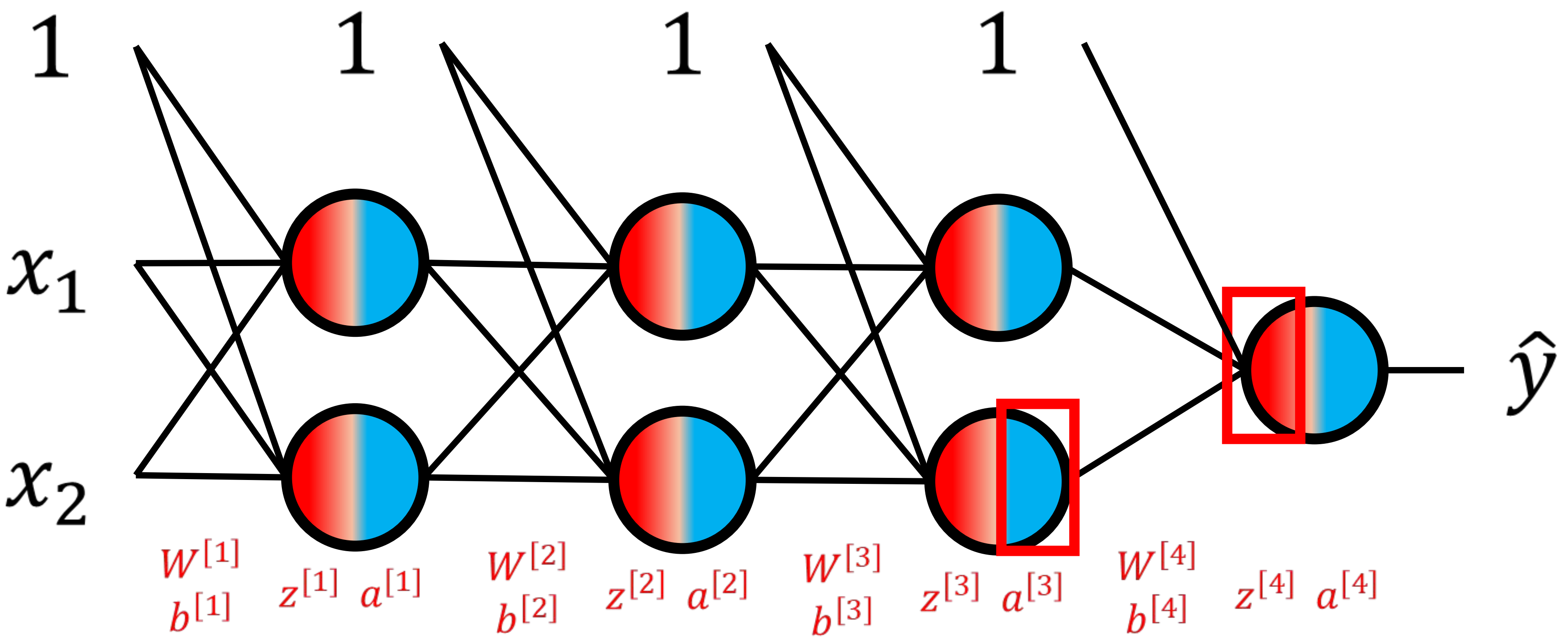
$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial a_1^{[3]}}$$



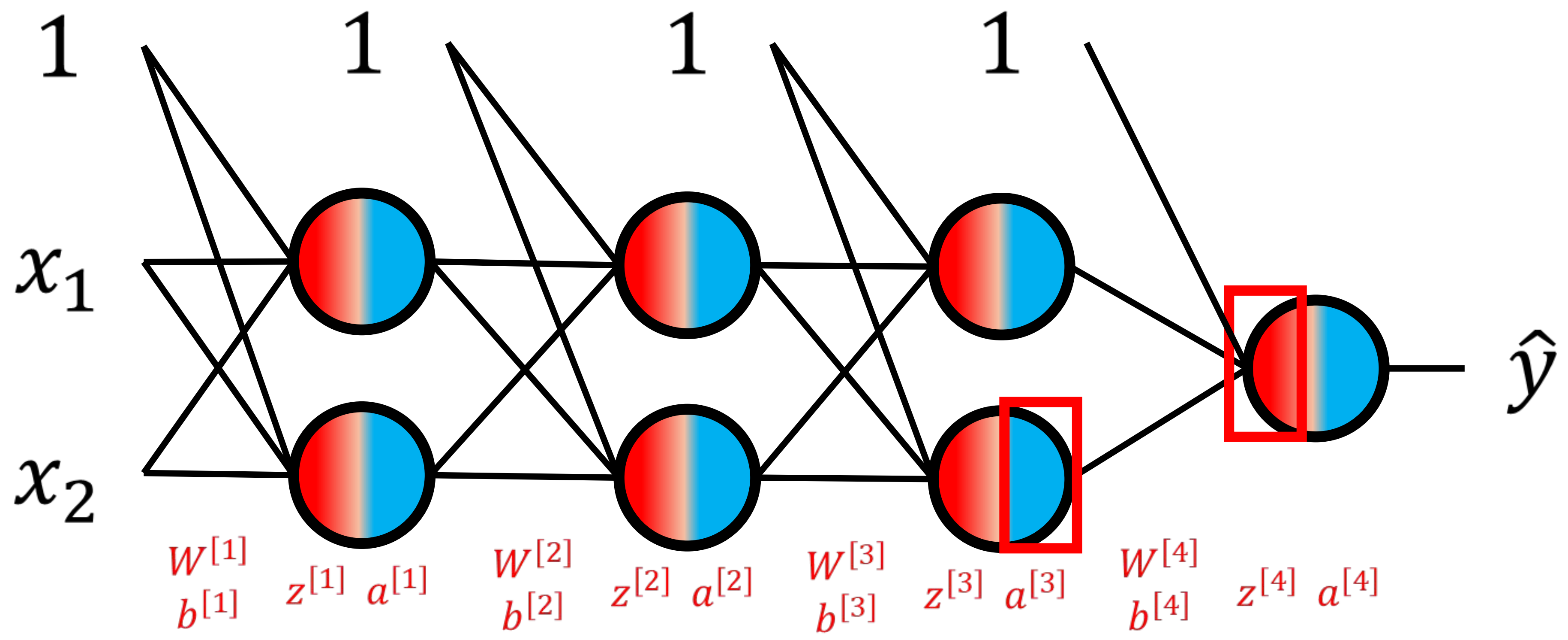
$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial a_1^{[3]}} = W_{11}^{[4]}$$



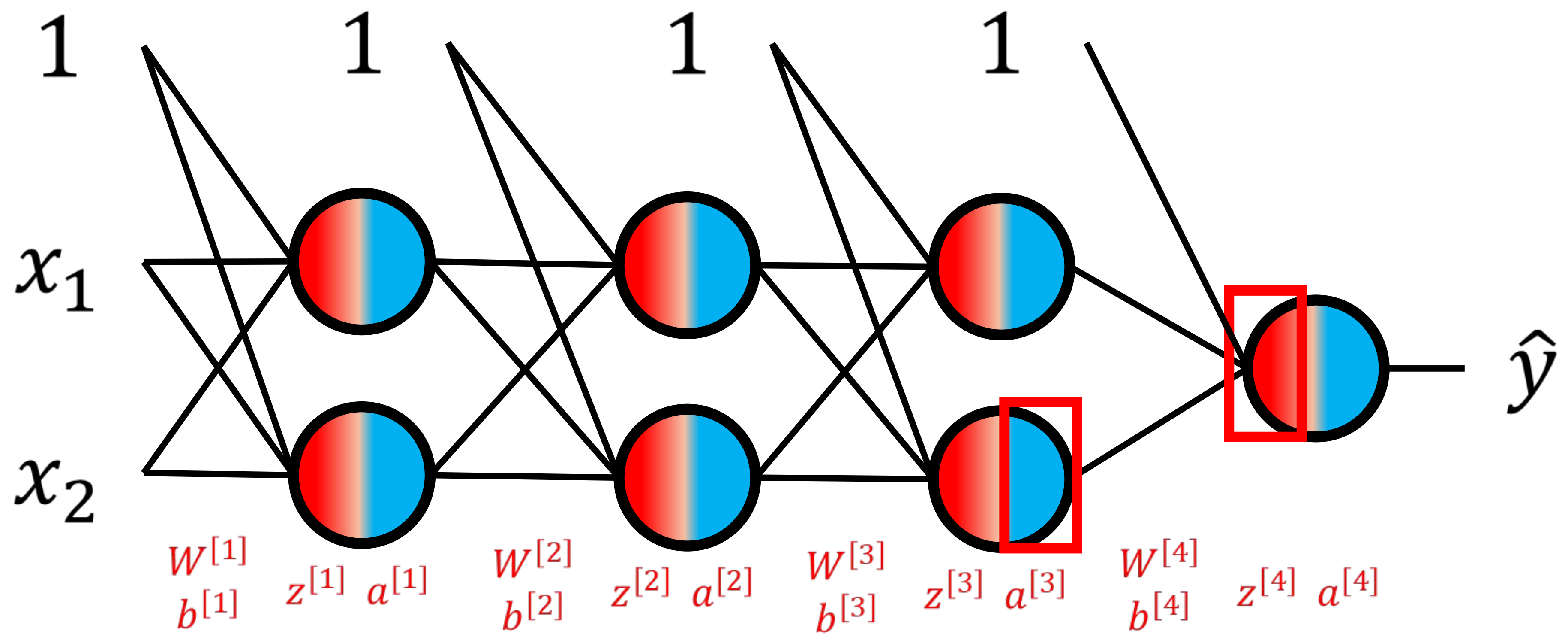
How does $a_2^{[3]}$ affect $z_1^{[4]}$?



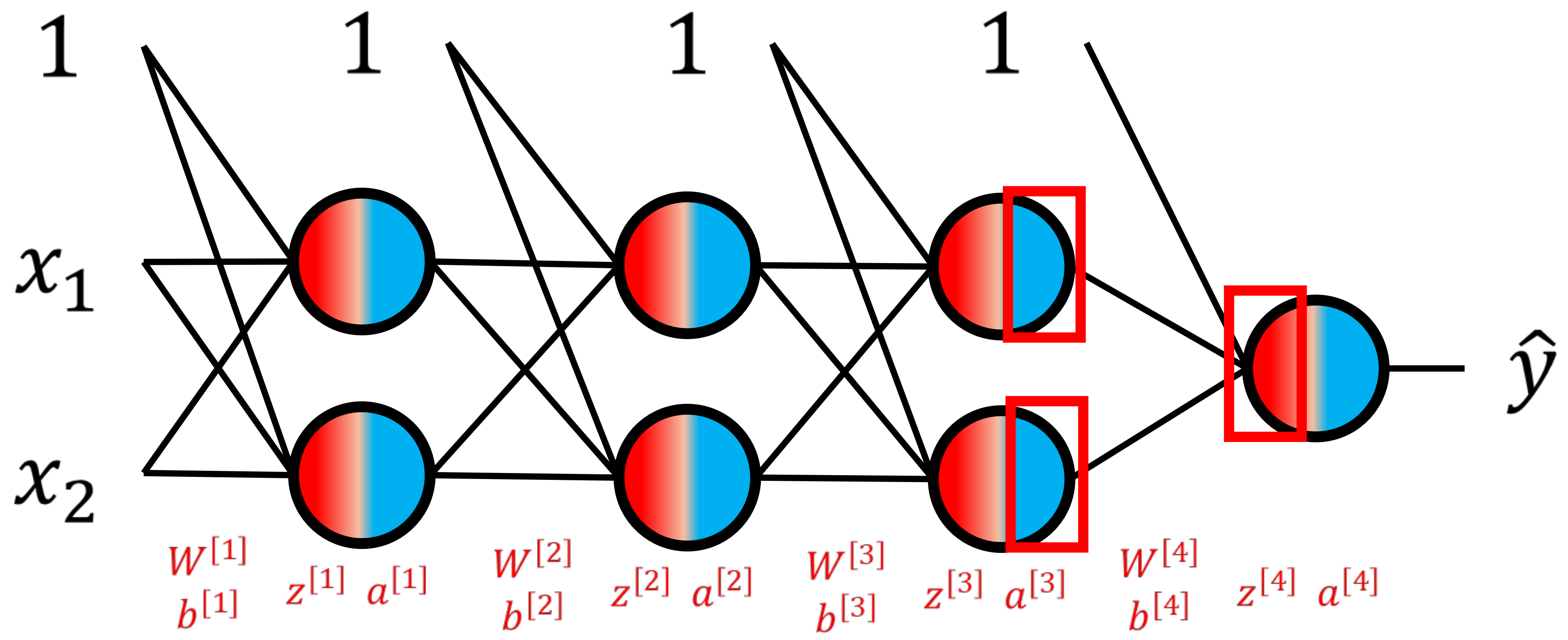
$$\frac{\partial z_1^{[4]}}{\partial a_2^{[3]}}$$



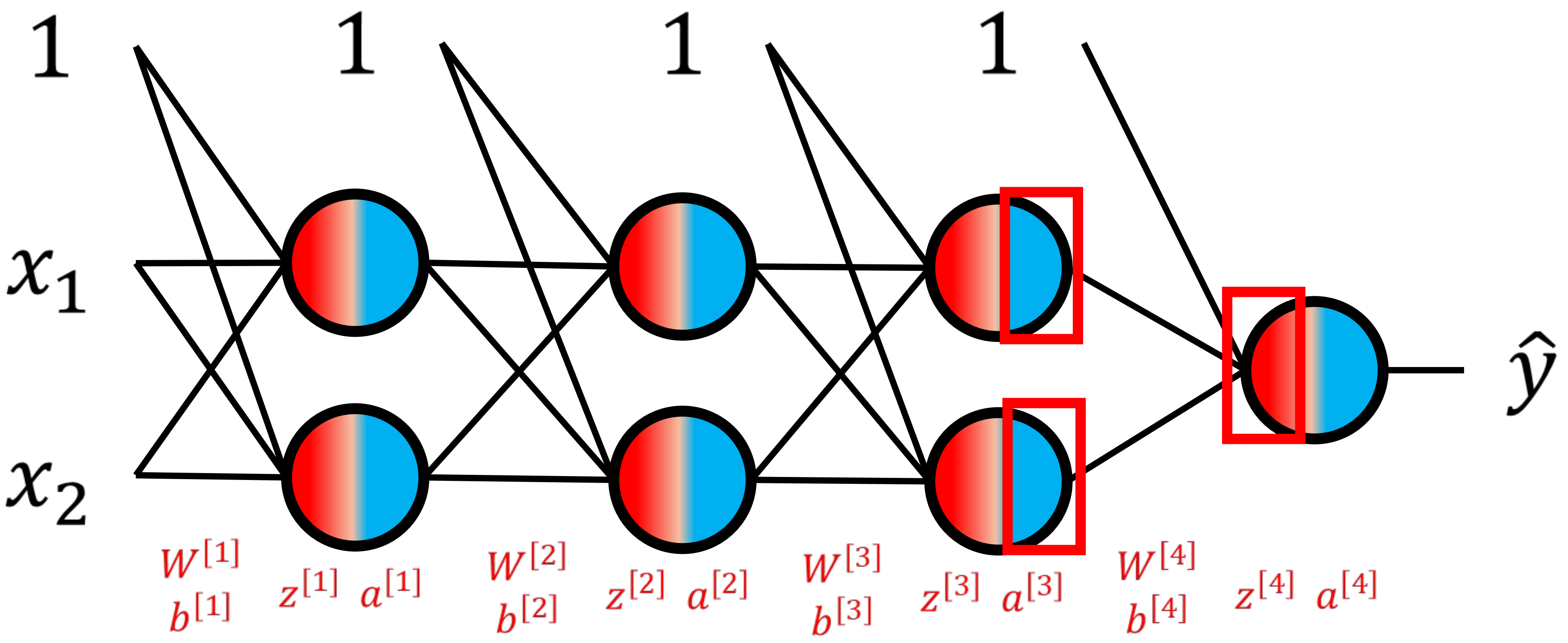
$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial a_2^{[3]}}$$



$$\frac{\partial W_{11}^{[4]} a_1^{[3]} + W_{12}^{[4]} a_2^{[3]} + b_1^{[4]}}{\partial a_2^{[3]}} = W_{12}^{[4]}$$

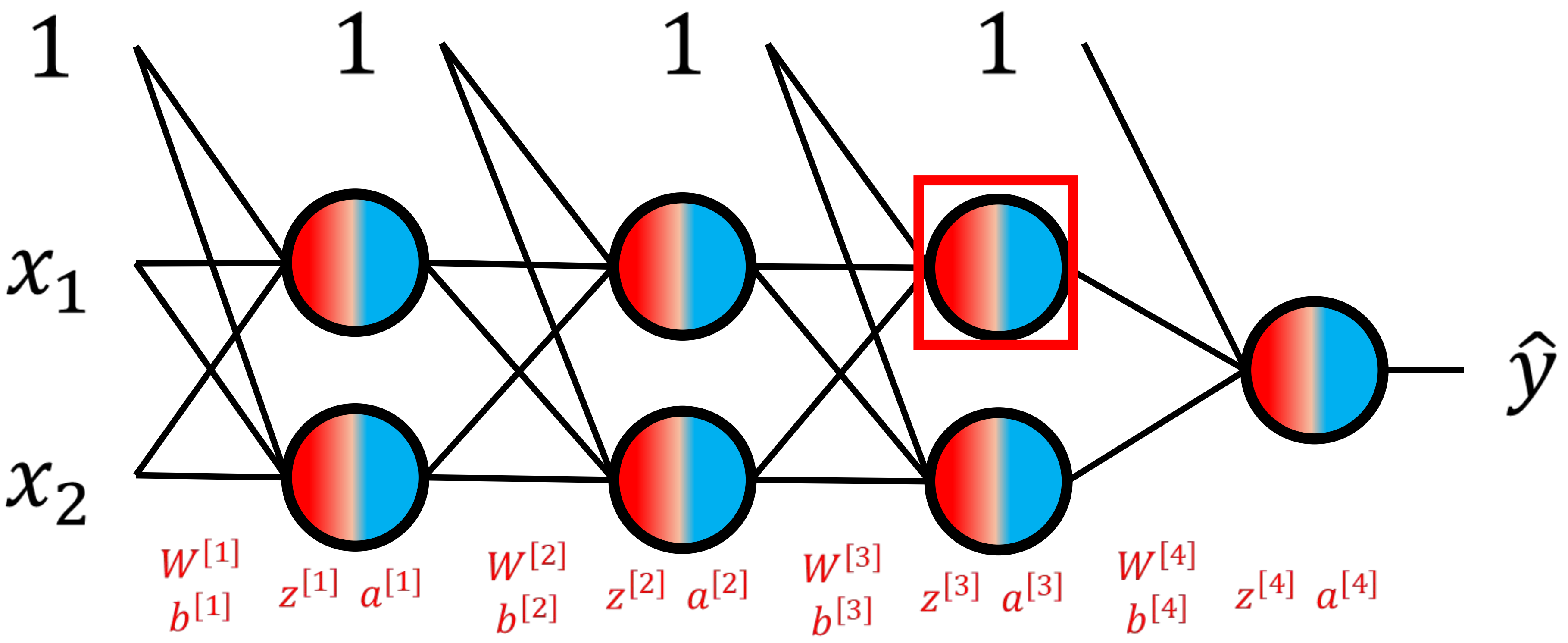


$$\frac{\partial z_1^{[4]}}{\partial a_1^{[3]}} = W_{11}^{[4]} \qquad \frac{\partial z_1^{[4]}}{\partial a_2^{[3]}} = W_{12}^{[4]}$$

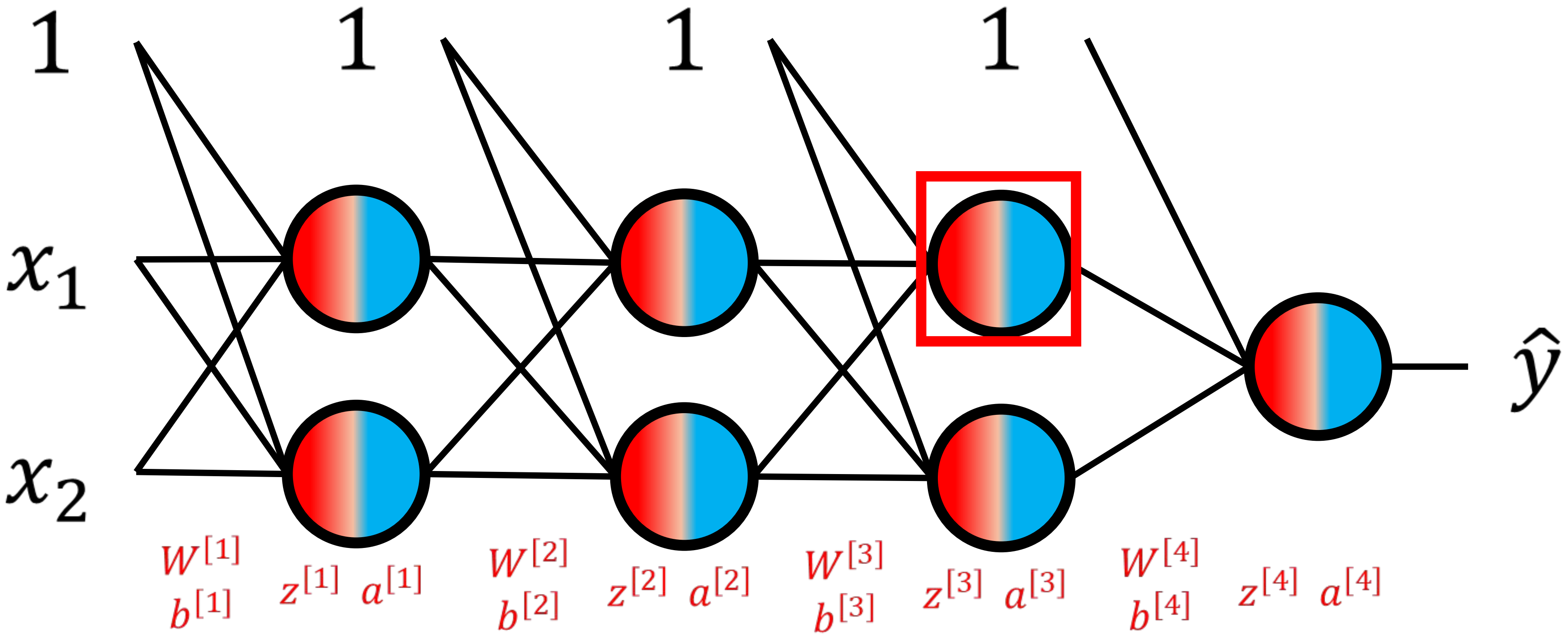


$$\frac{\partial z_1^{[4]}}{\partial a_1^{[3]}} = W_{11}^{[4]} \qquad \frac{\partial z_1^{[4]}}{\partial a_2^{[3]}} = W_{12}^{[4]} \qquad \frac{\partial z^{[4]}}{\partial a^{[3]}} = W^{[4]}$$

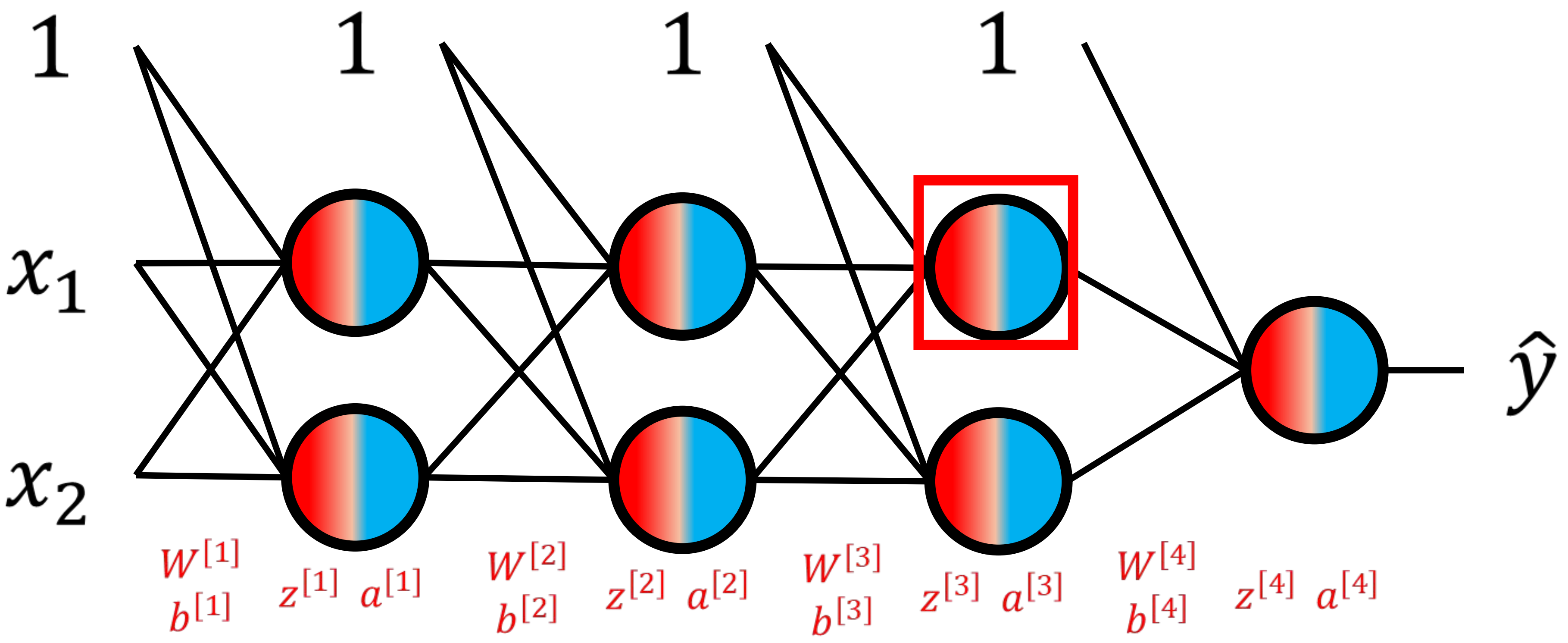
Solve all the gradients all at once!



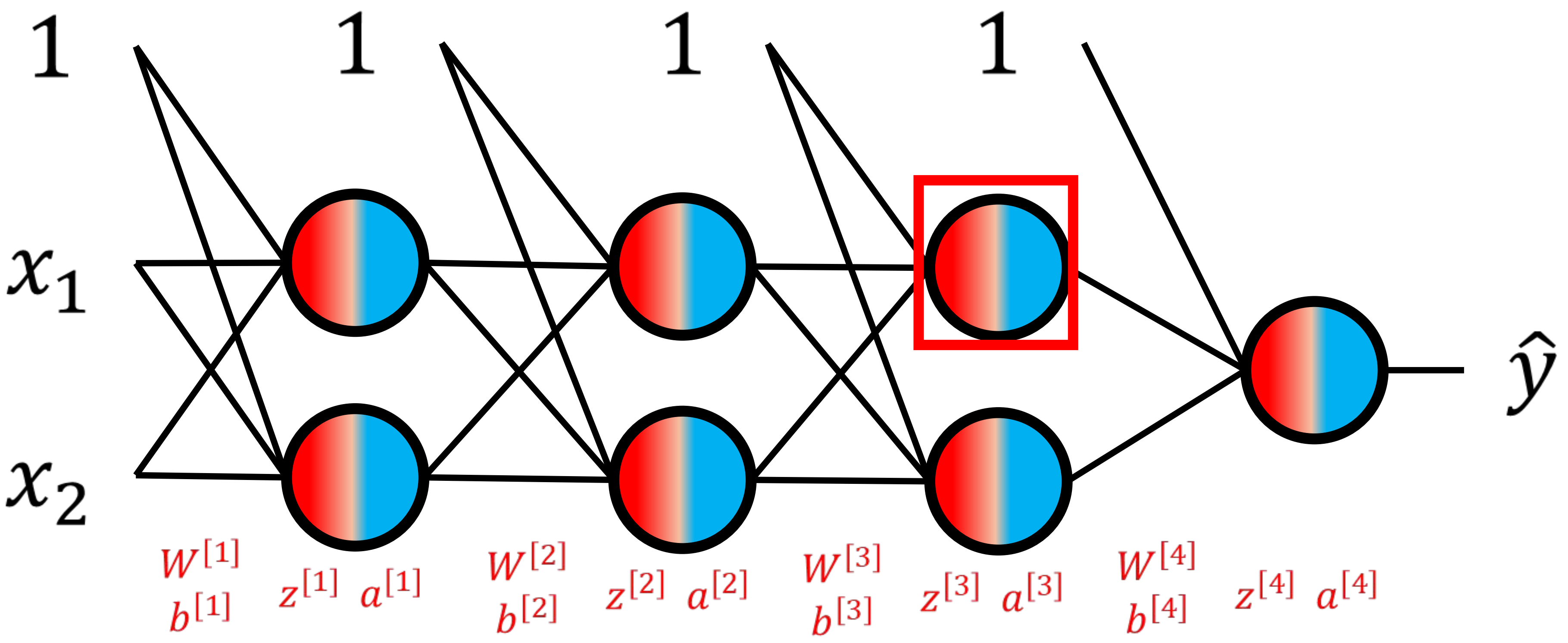
How does $z_1^{[3]}$ affect $a_1^{[3]}$?



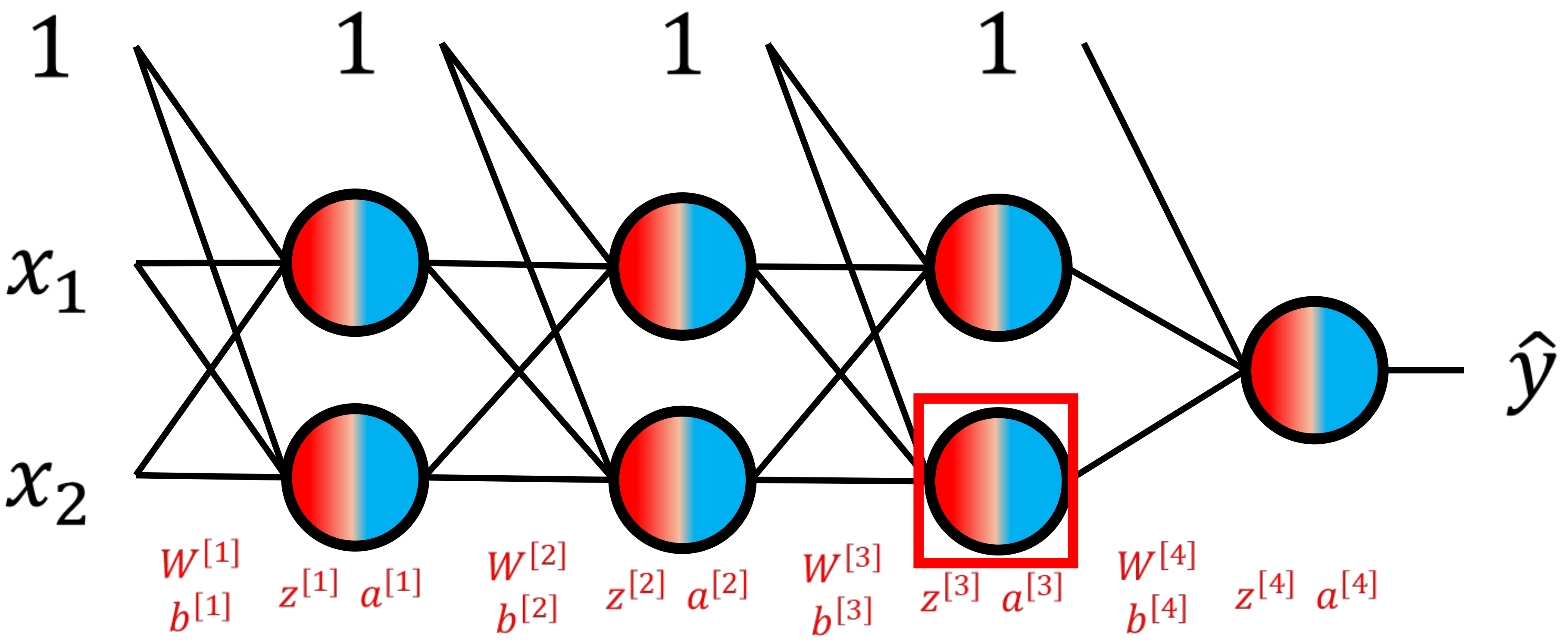
$$\frac{\partial a_1^{[3]}}{\partial z_1^{[3]}}$$



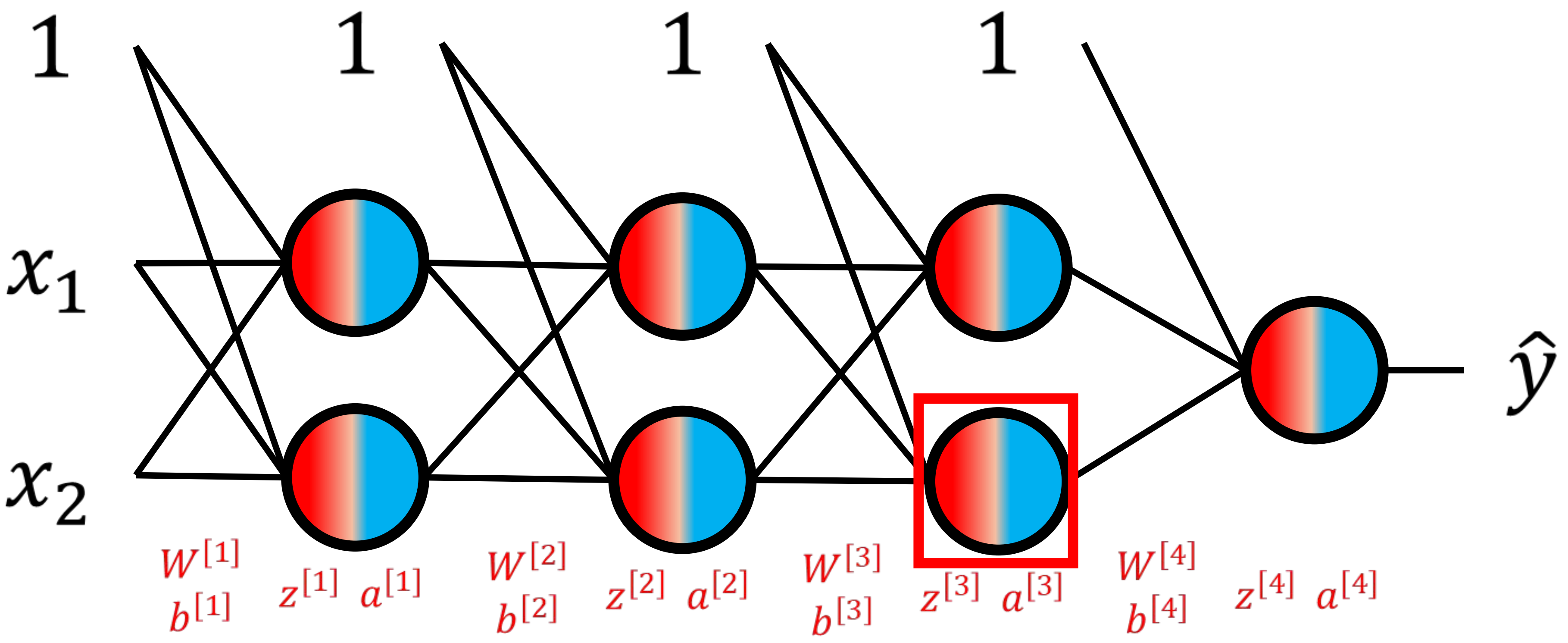
$$\frac{\partial \sigma\left(z_1^{[3]}\right)}{\partial z_1^{[3]}}$$



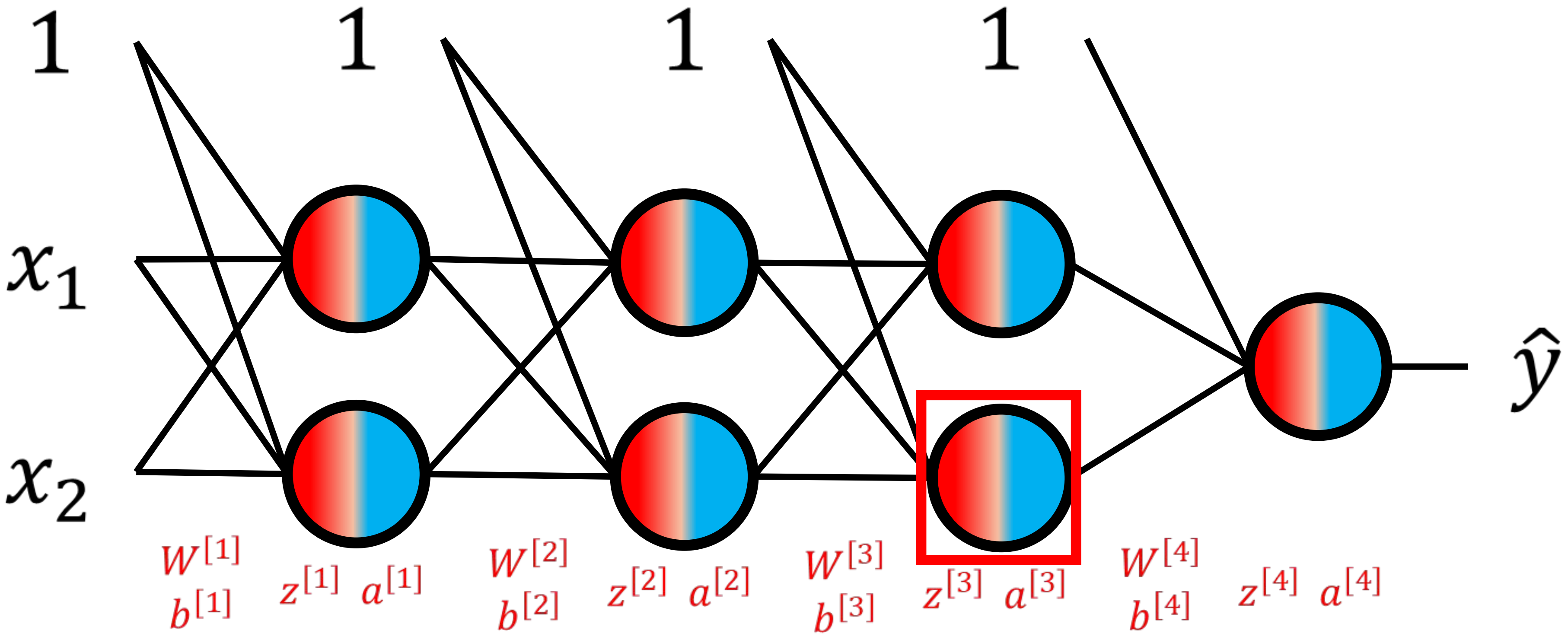
$$\frac{\partial \sigma\left(z_1^{[3]}\right)}{\partial z_1^{[3]}} = \sigma\left(z_1^{[3]}\right)\left(1 - \sigma\left(z_1^{[3]}\right)\right)$$



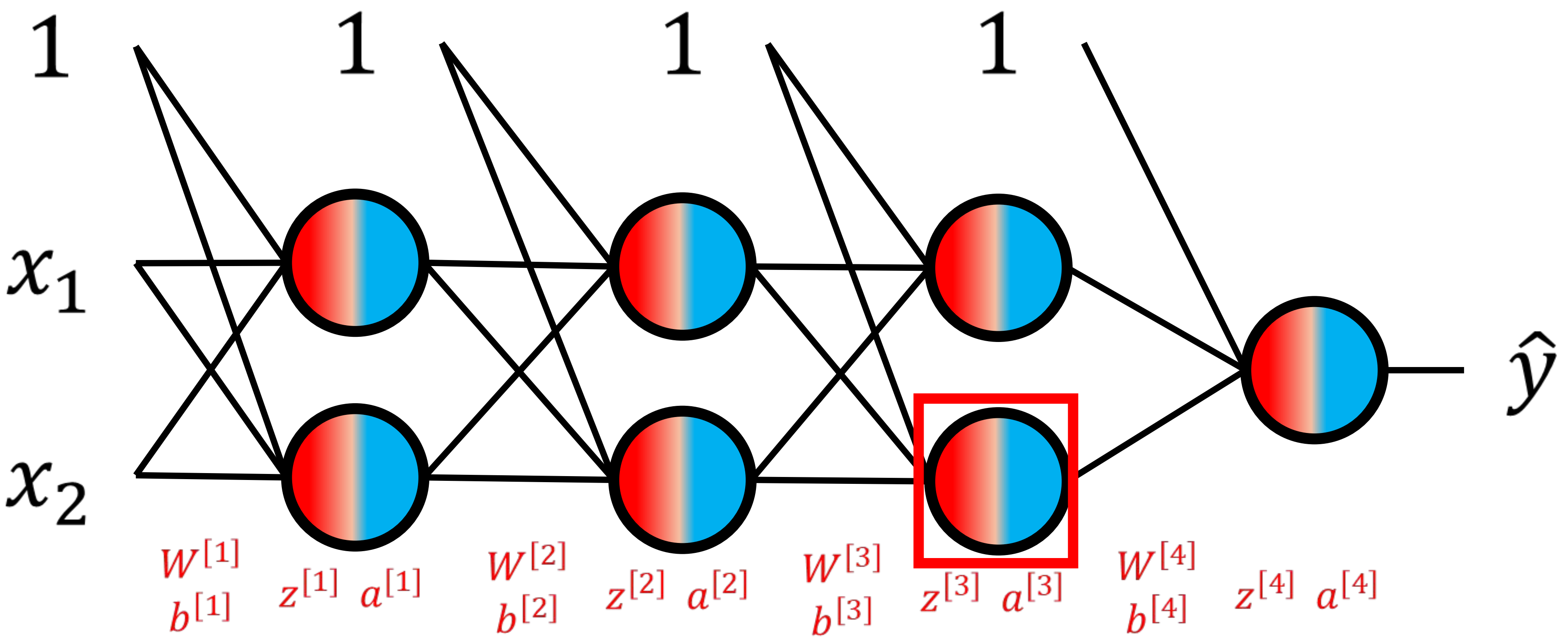
How does $z_2^{[3]}$ affect $a_2^{[3]}$?



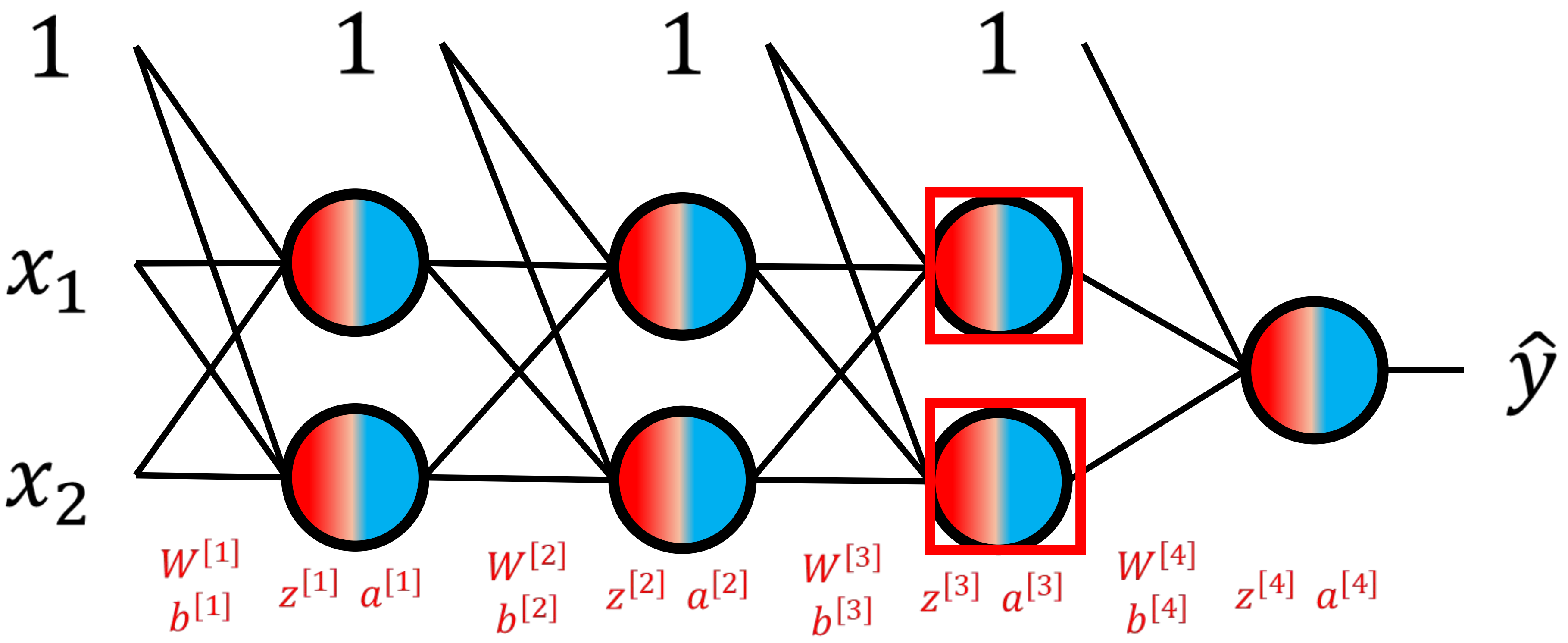
$$\frac{\partial a_2^{[3]}}{\partial z_2^{[3]}}$$



$$\frac{\partial \sigma\left(z_2^{[3]}\right)}{\partial z_2^{[3]}}$$

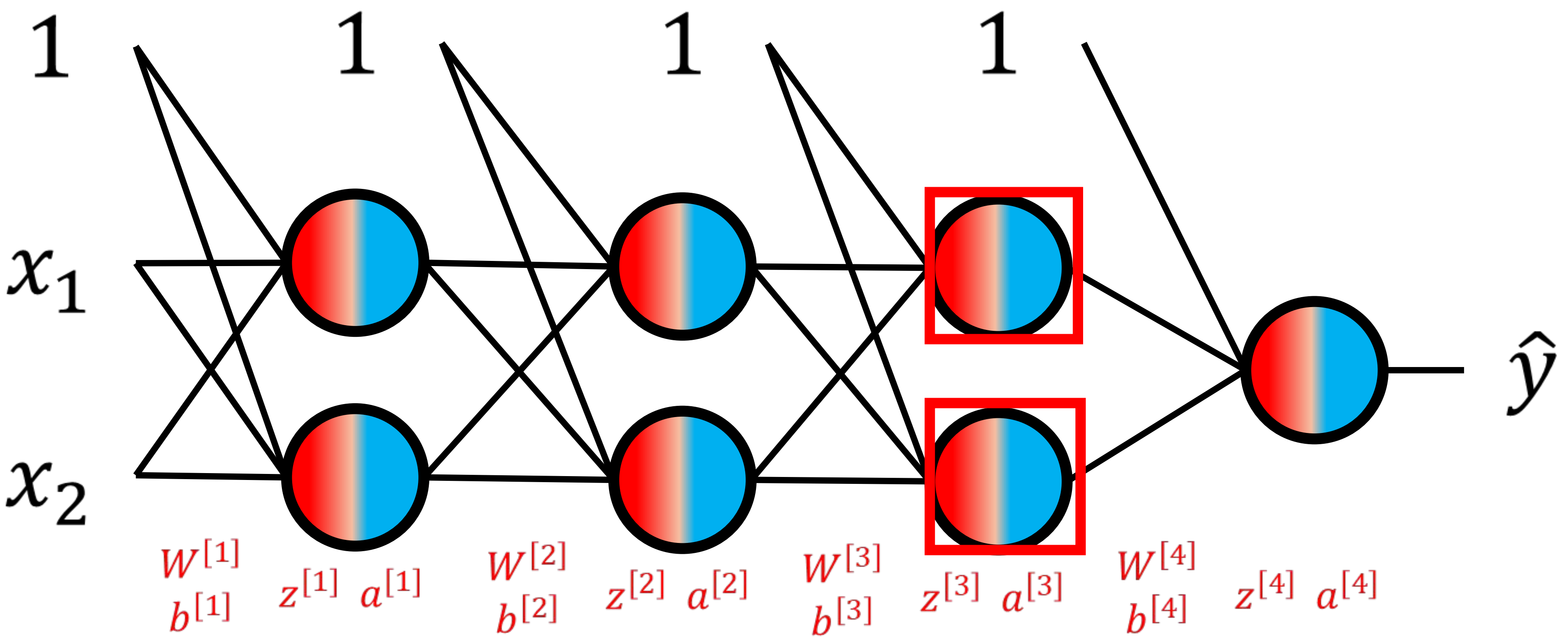


$$\frac{\partial \sigma\left(z_2^{[3]}\right)}{\partial z_2^{[3]}} = \sigma\left(z_2^{[3]}\right)\left(1 - \sigma\left(z_2^{[3]}\right)\right)$$



$$\frac{\partial a_1^{[3]}}{\partial z_1^{[3]}} = \sigma\left(z_1^{[3]}\right)\left(1 - \sigma\left(z_1^{[3]}\right)\right)$$

$$\frac{\partial a_2^{[3]}}{\partial z_2^{[3]}} = \sigma\left(z_2^{[3]}\right)\left(1 - \sigma\left(z_2^{[3]}\right)\right)$$



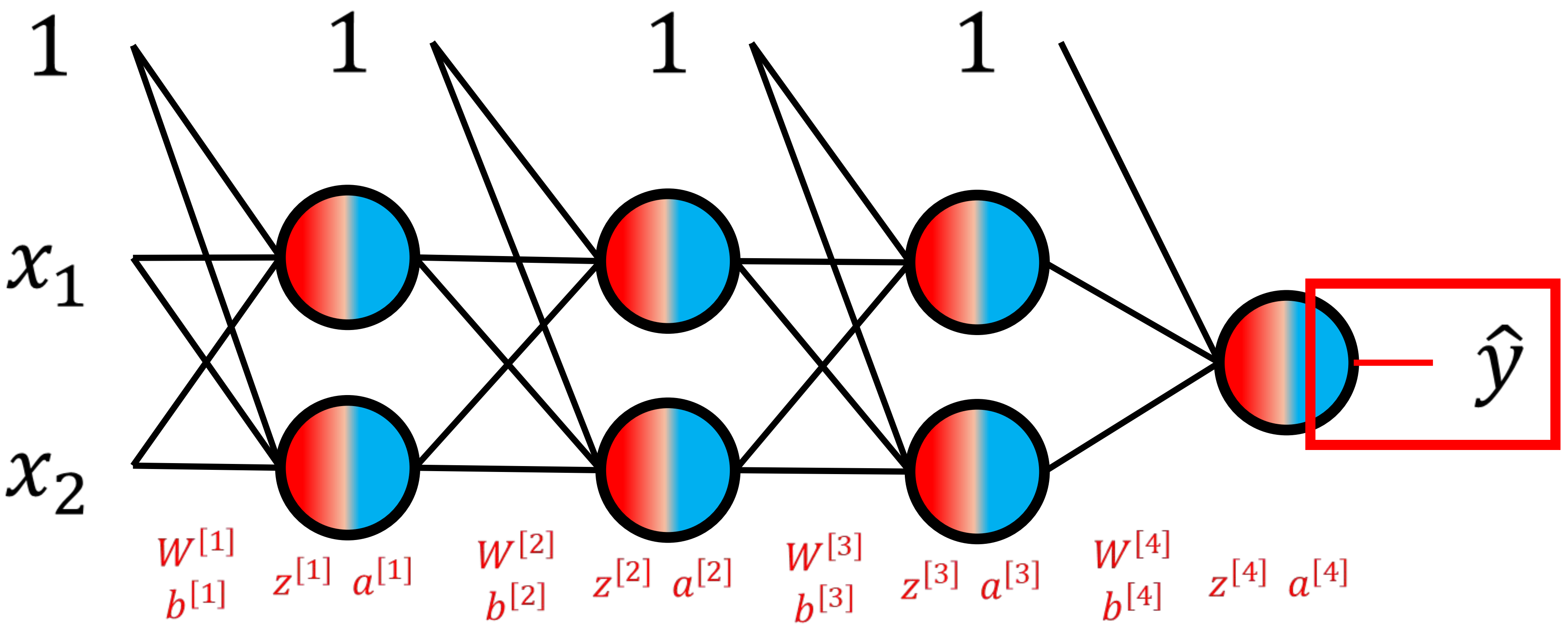
$$\frac{\partial a_1^{[3]}}{\partial z_1^{[3]}} = \sigma\left(z_1^{[3]}\right)\left(1 - \sigma\left(z_1^{[3]}\right)\right)$$

$$\frac{\partial a_2^{[3]}}{\partial z_2^{[3]}} = \sigma\left(z_2^{[3]}\right)\left(1 - \sigma\left(z_2^{[3]}\right)\right)$$

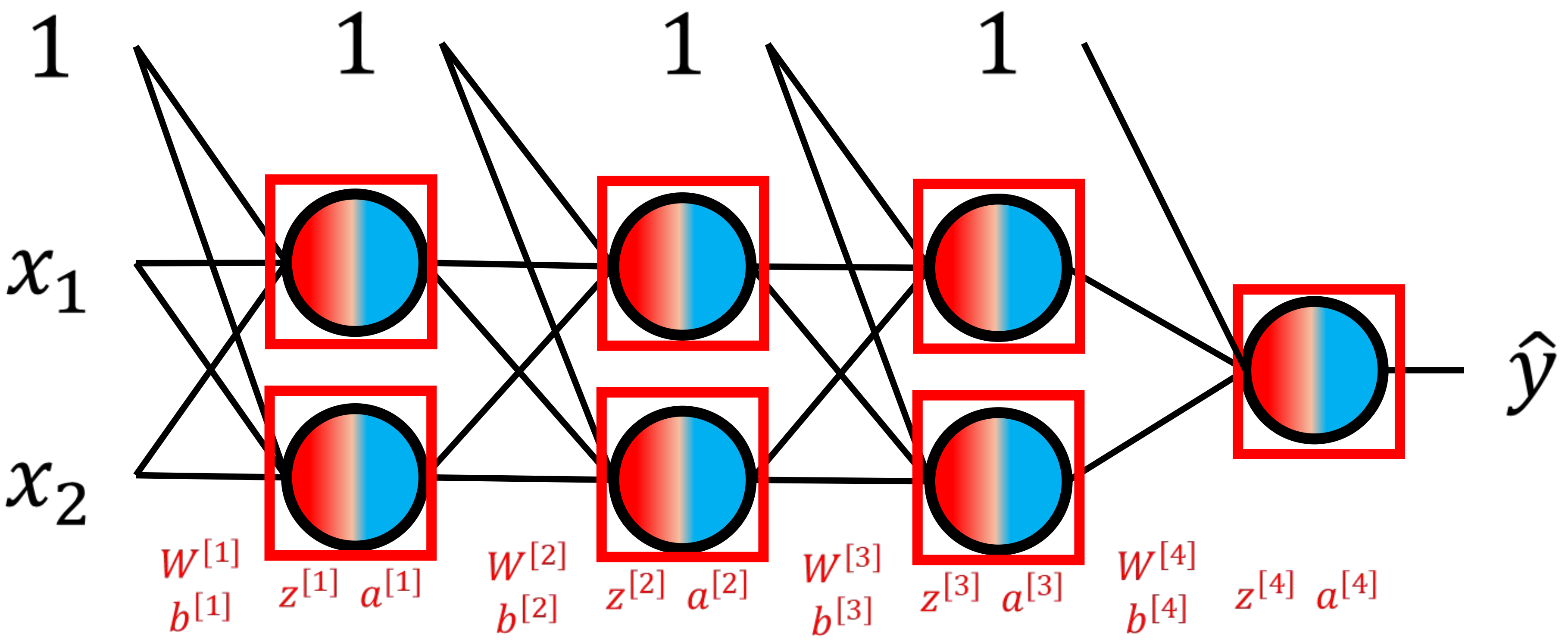
$$\frac{\partial a^{[3]}}{\partial z^{[3]}} = \sigma(z^{[3]})\left(1 - \sigma(z^{[3]})\right)$$

Solve all the gradients all at once!

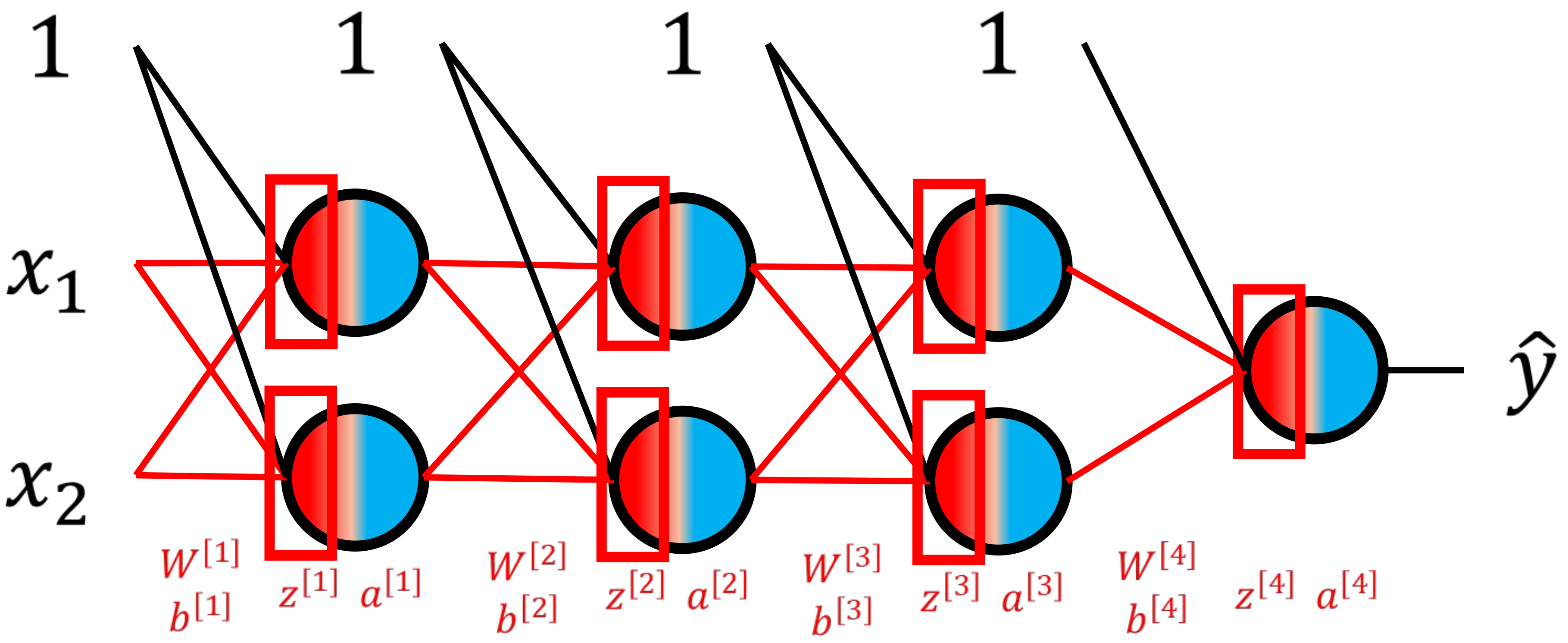
Summarizing the Derivatives that We Have So Far…



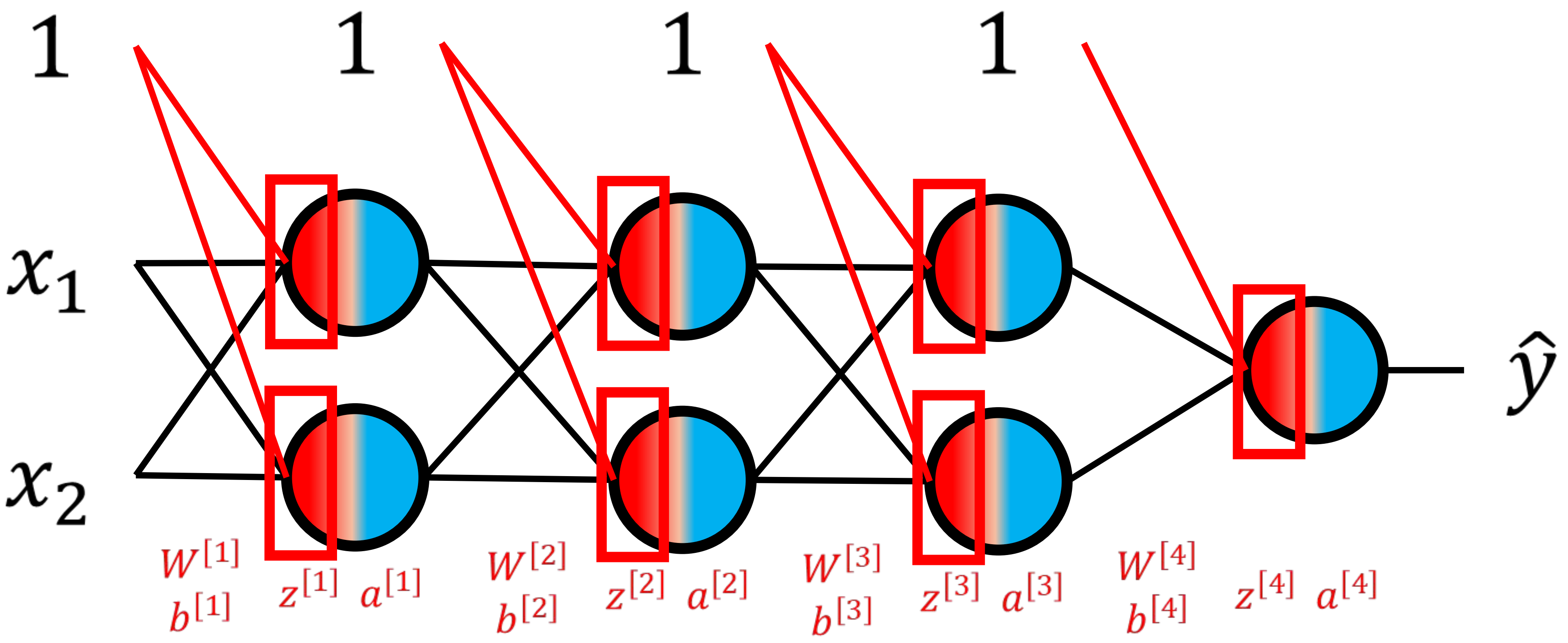
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$



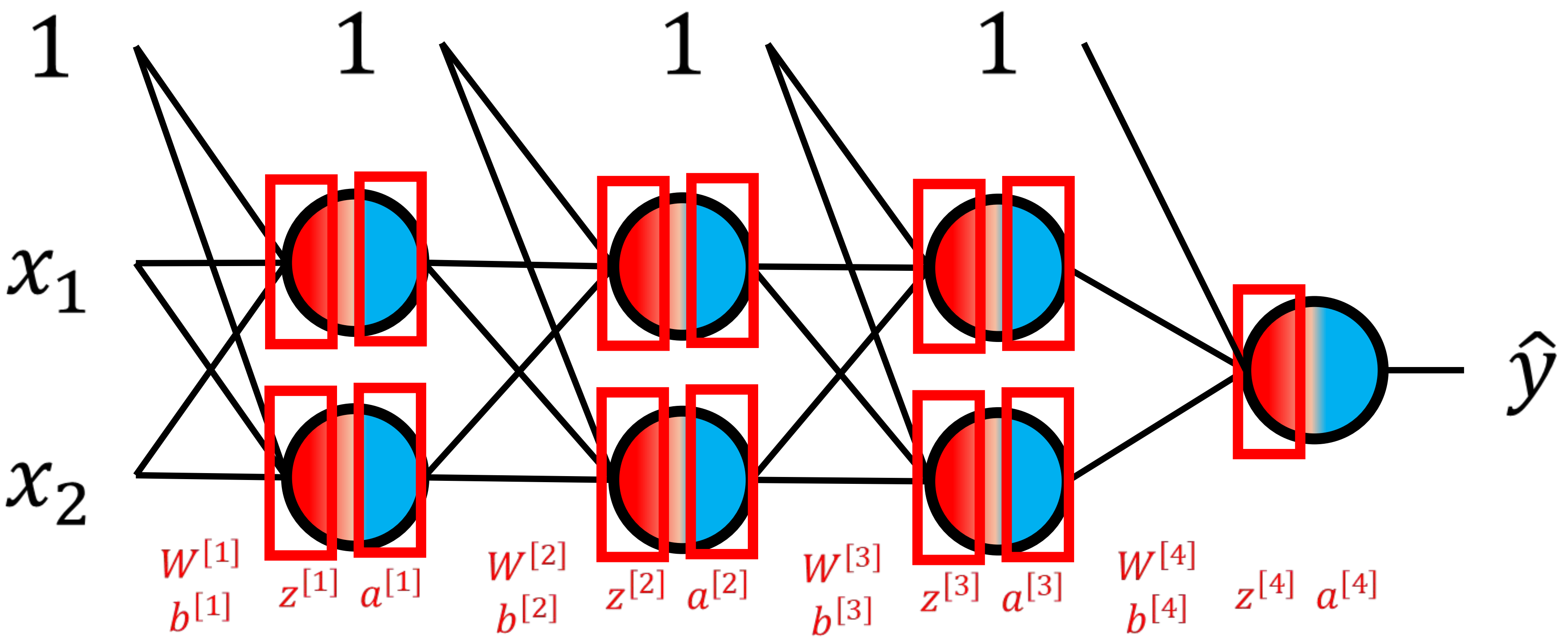
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1 \qquad \frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma\left(z^{[i]}\right)\left(1 - \sigma\left(z^{[i]}\right)\right)$$



$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1 \qquad \frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma\left(z^{[i]}\right)\left(1 - \sigma\left(z^{[i]}\right)\right) \qquad \frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$



$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1 \qquad \frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma\left(z^{[i]}\right)\left(1 - \sigma\left(z^{[i]}\right)\right) \qquad \frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]} \qquad \frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1 \qquad \frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma\left(z^{[i]}\right)\left(1 - \sigma\left(z^{[i]}\right)\right) \qquad \frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]} \qquad \frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

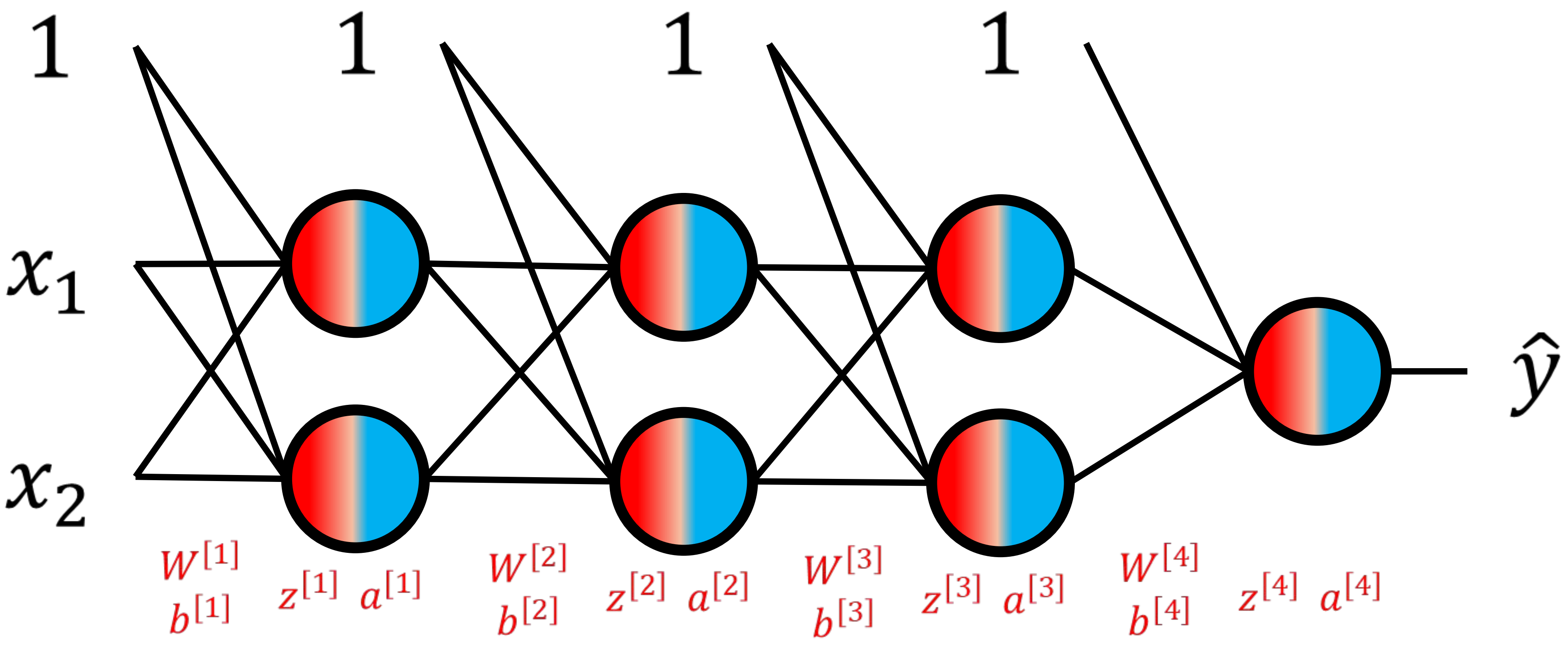
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



4. Computing for Our Goal

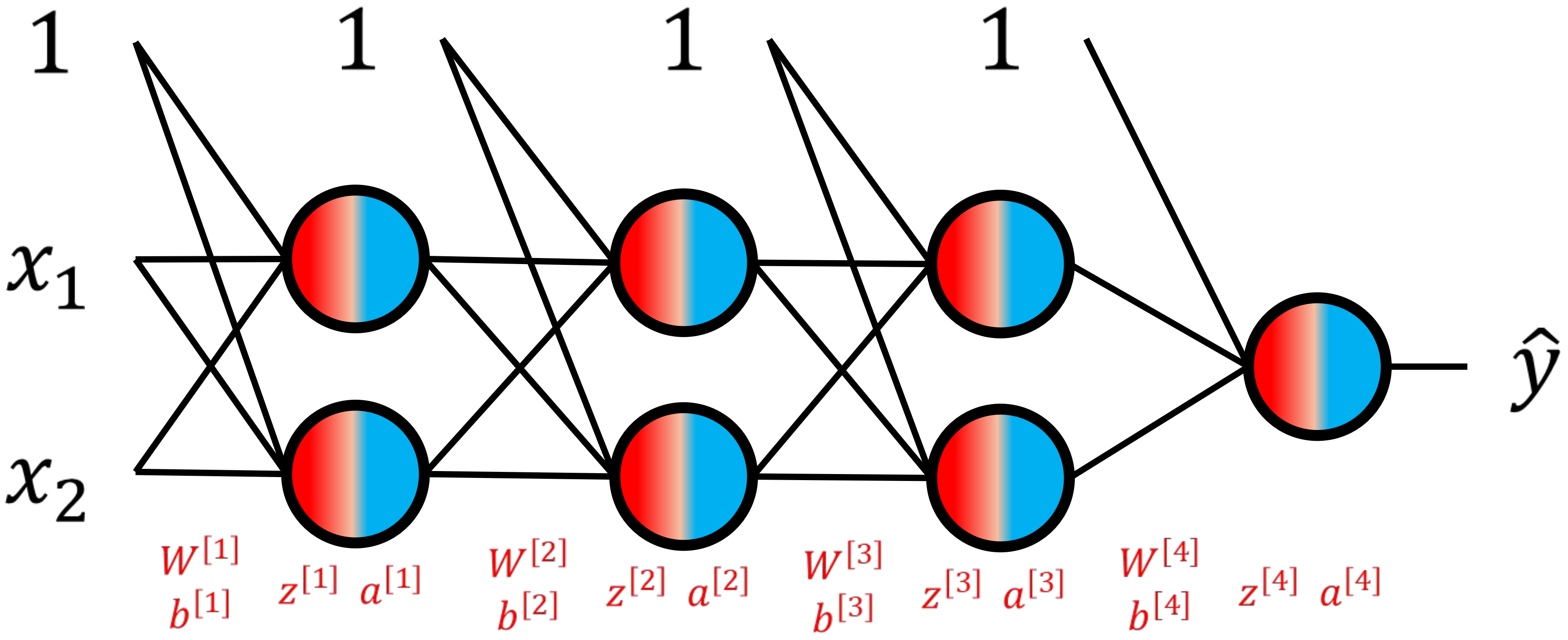
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



What we're *really* interested in:

How do the weights and biases affect $\hat{y}$?

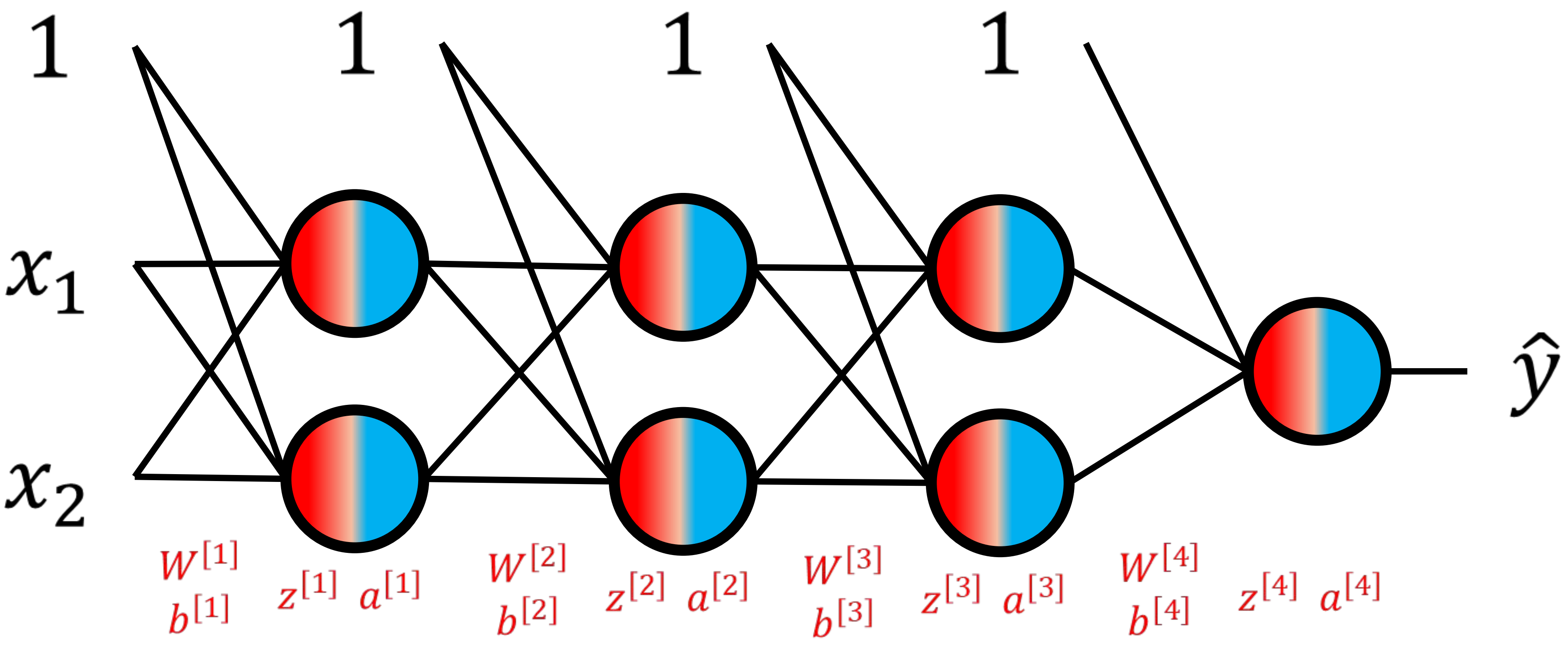
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



We can just re-use our answers here and put them in a **chain rule** to get what we want!

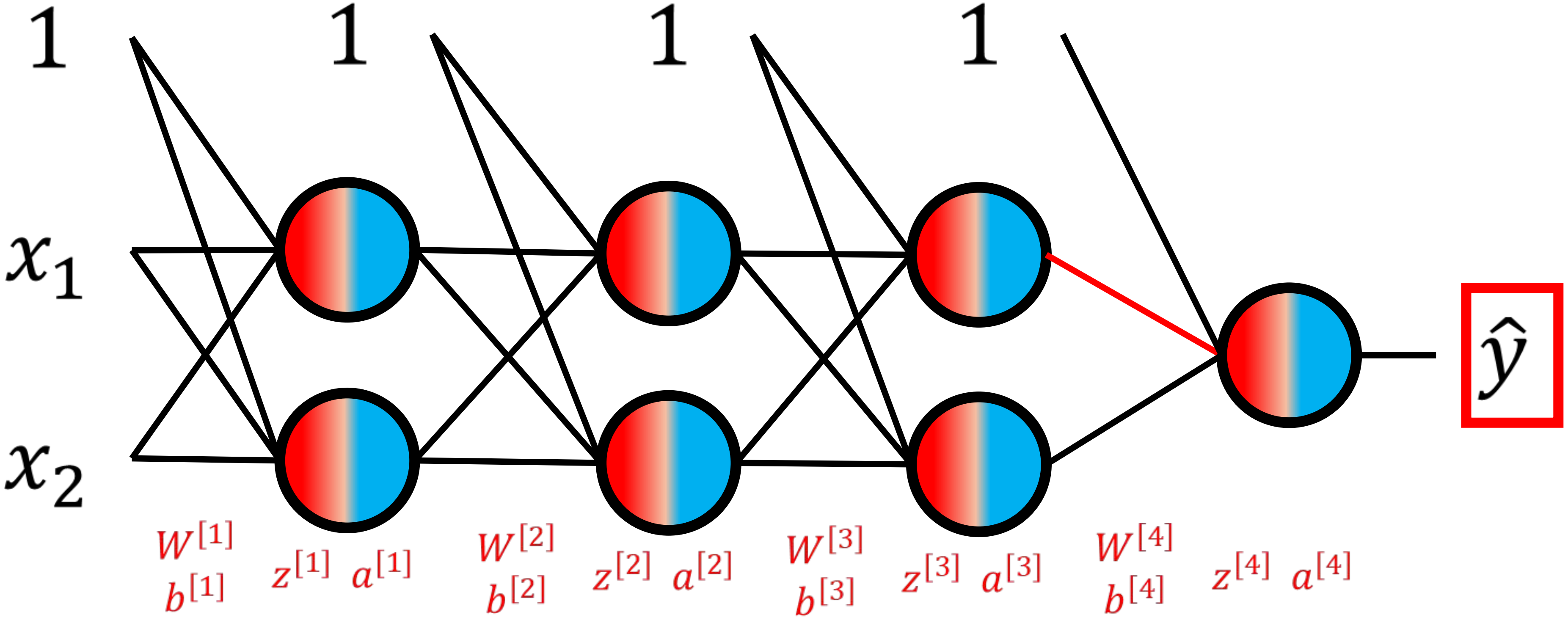
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $W_{11}^{[4]}$ affect $\hat{y}$?

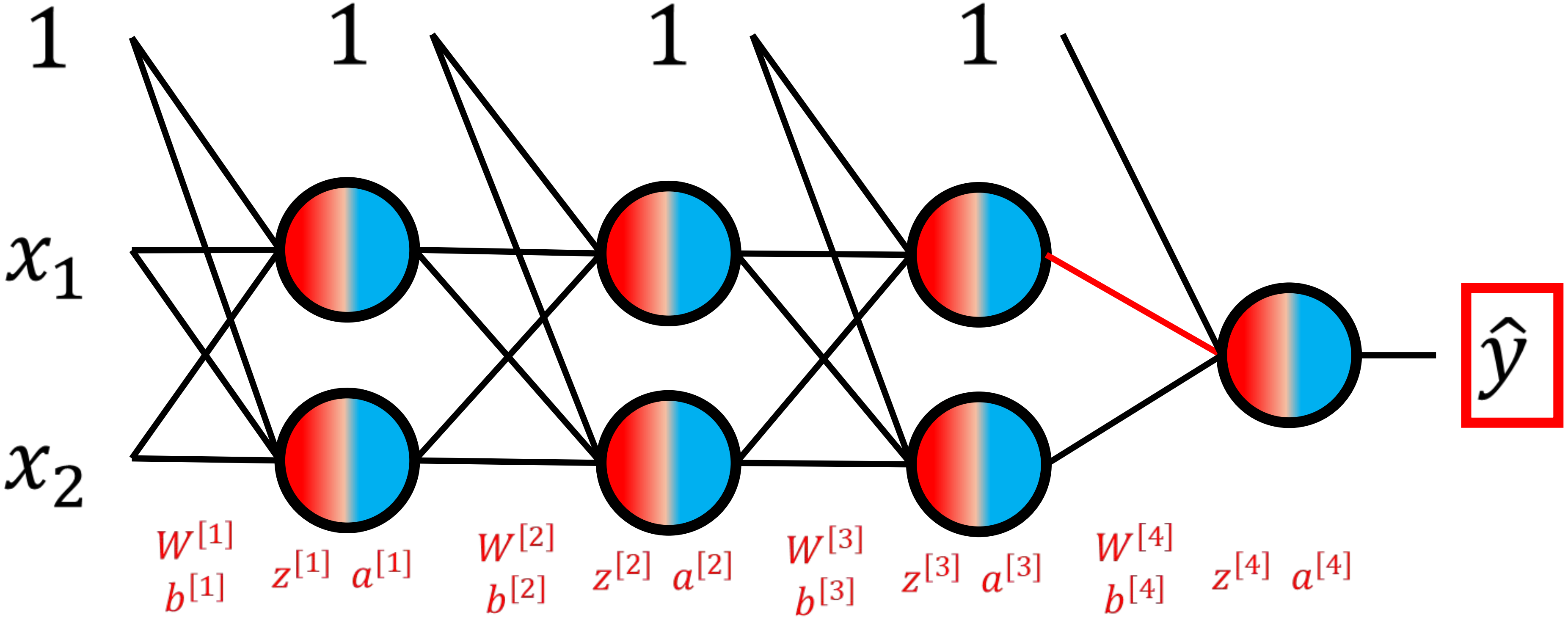
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{11}^{[4]}}$$

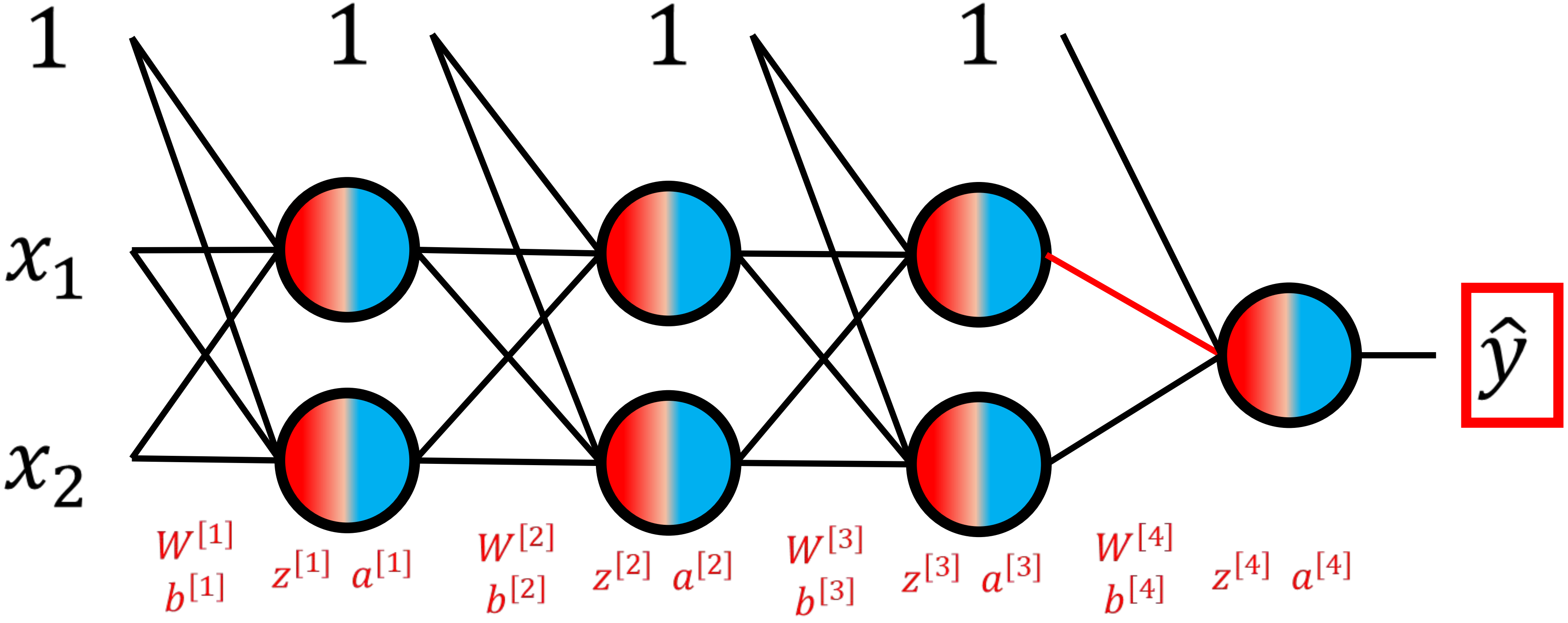
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{11}^{[4]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{[4]}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial W_{11}^{[4]}}$$

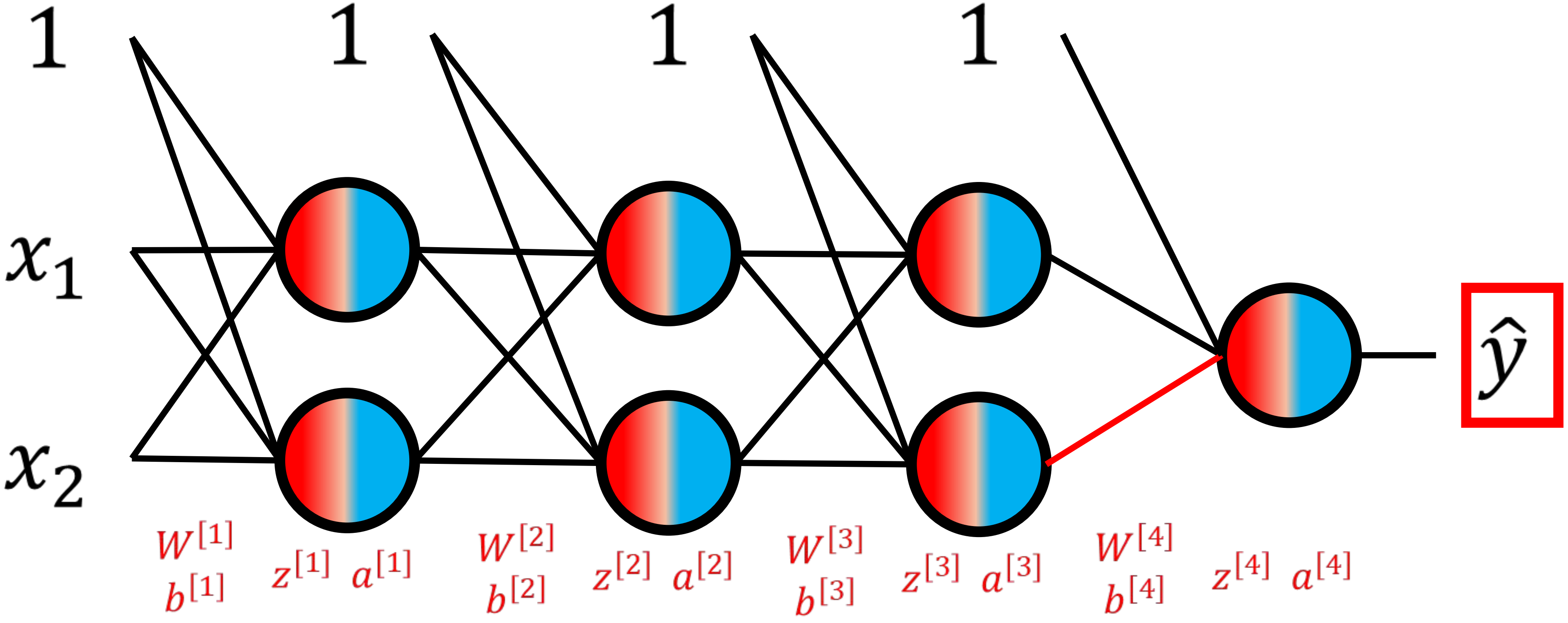
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $W_{12}^{[4]}$ affect $\hat{y}$?

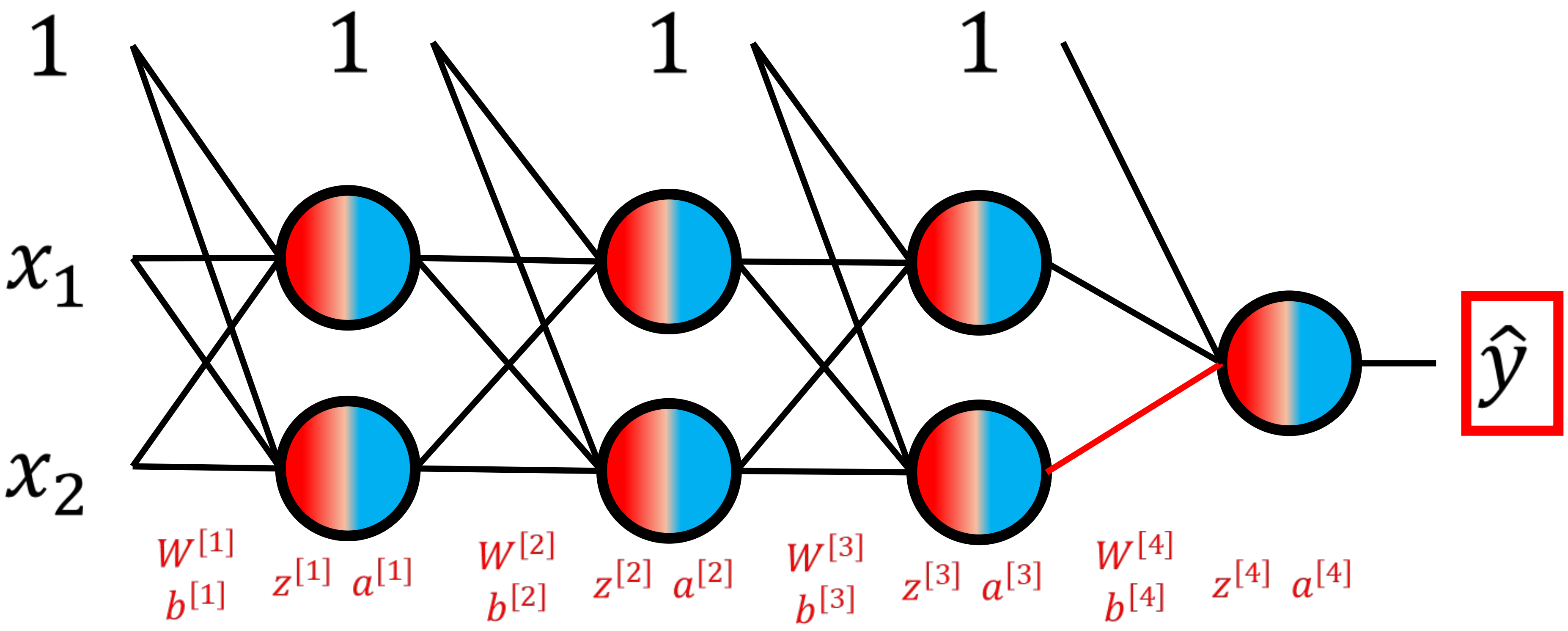
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{12}^{[4]}}$$

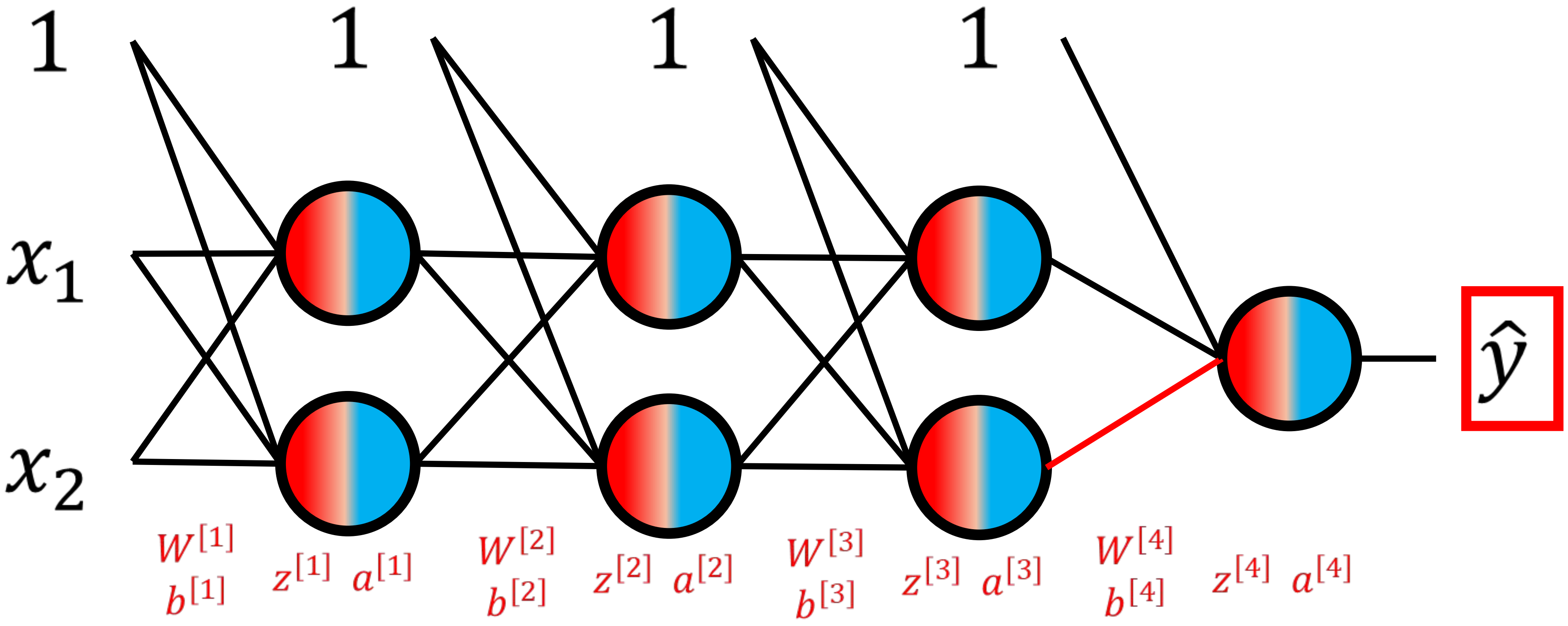
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{12}^{[4]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{[4]}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial W_{12}^{[4]}}$$

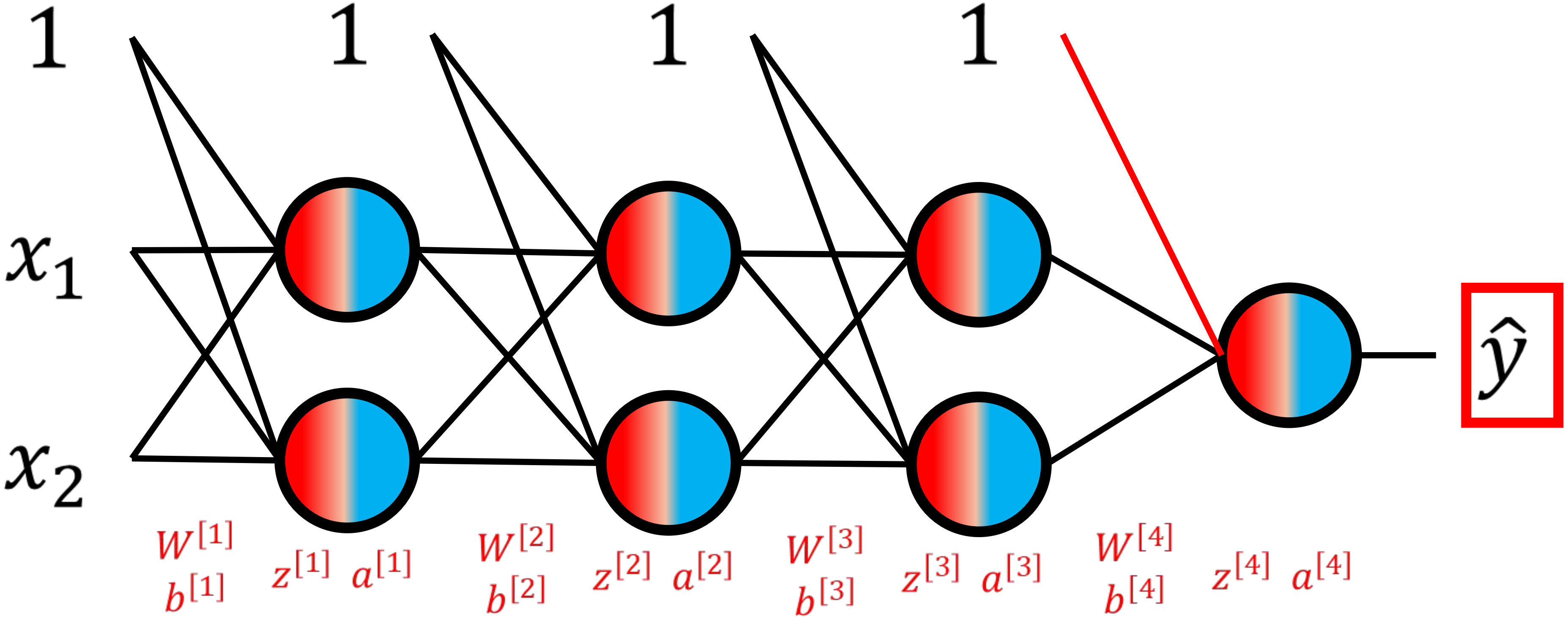
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $b_1^{[4]}$ affect $\hat{y}$?

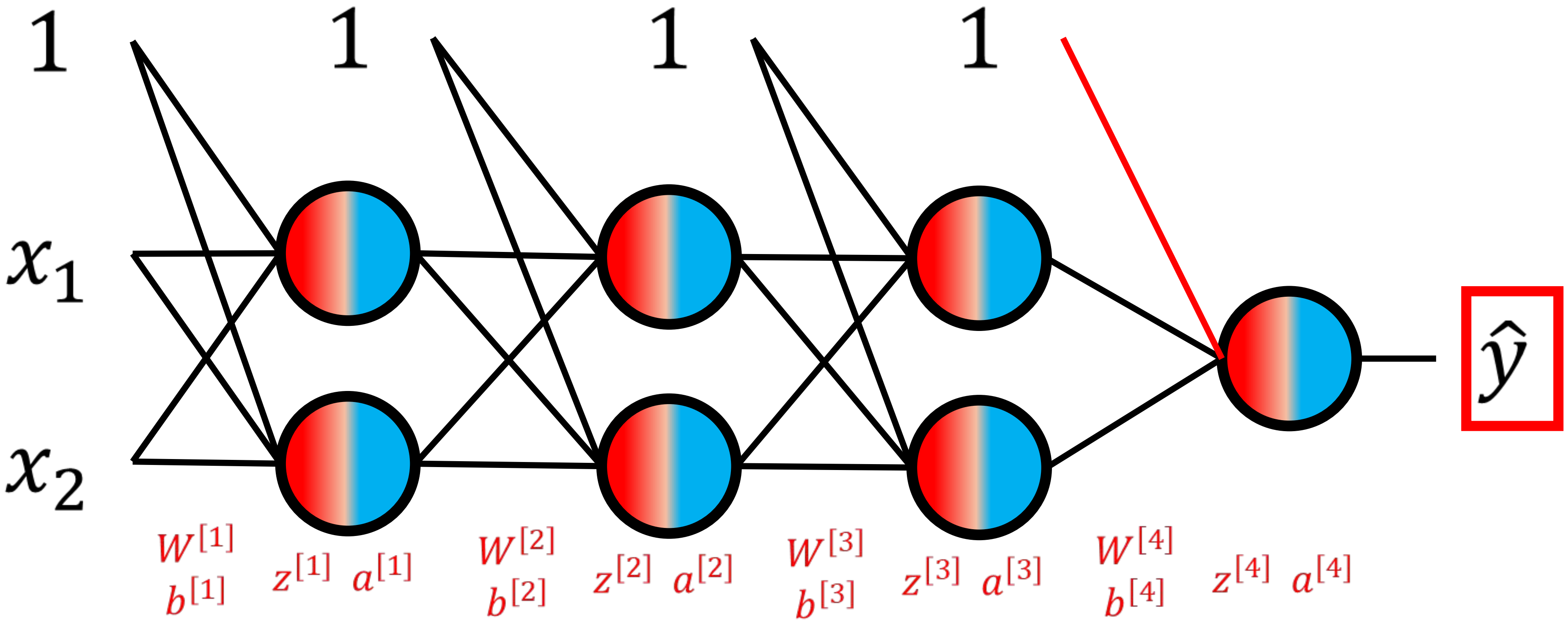
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial b_1^{[4]}}$$

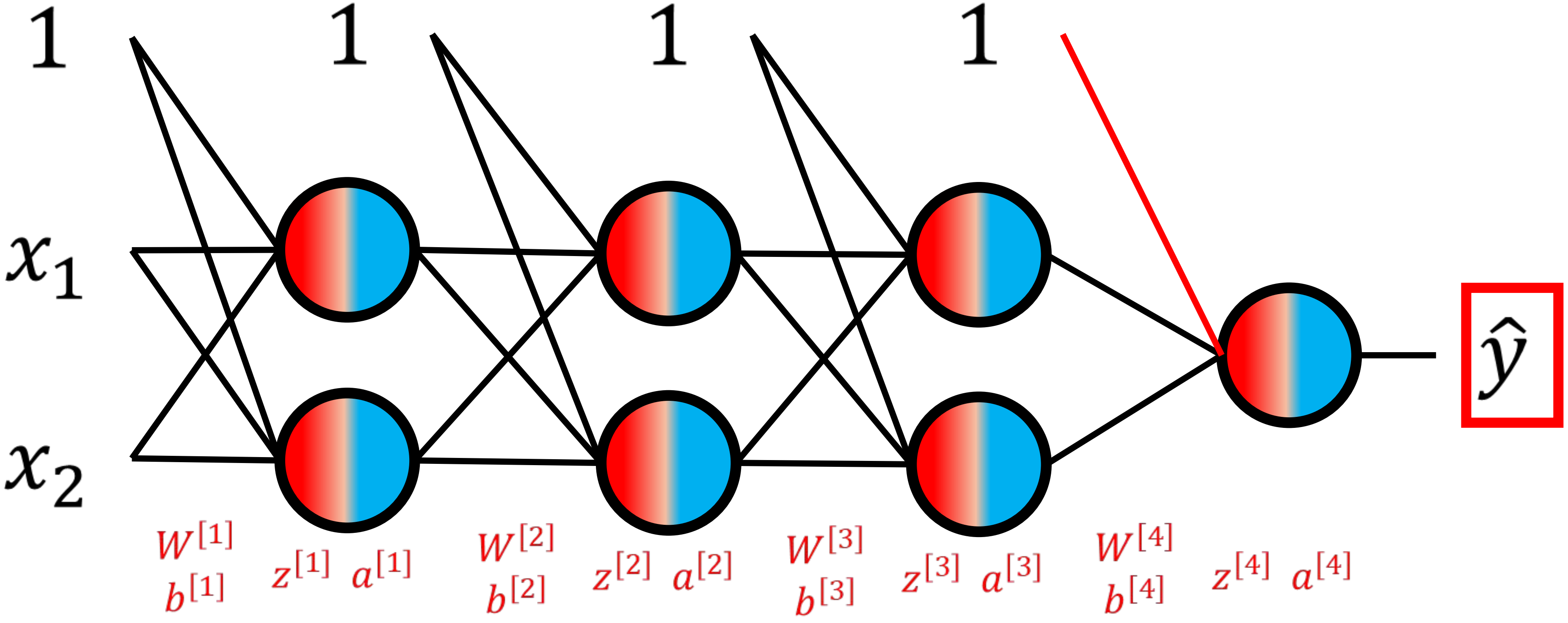
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma\left(z^{[i]}\right)\left(1 - \sigma\left(z^{[i]}\right)\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial b_1^{[4]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{[4]}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial b_1^{[4]}}$$

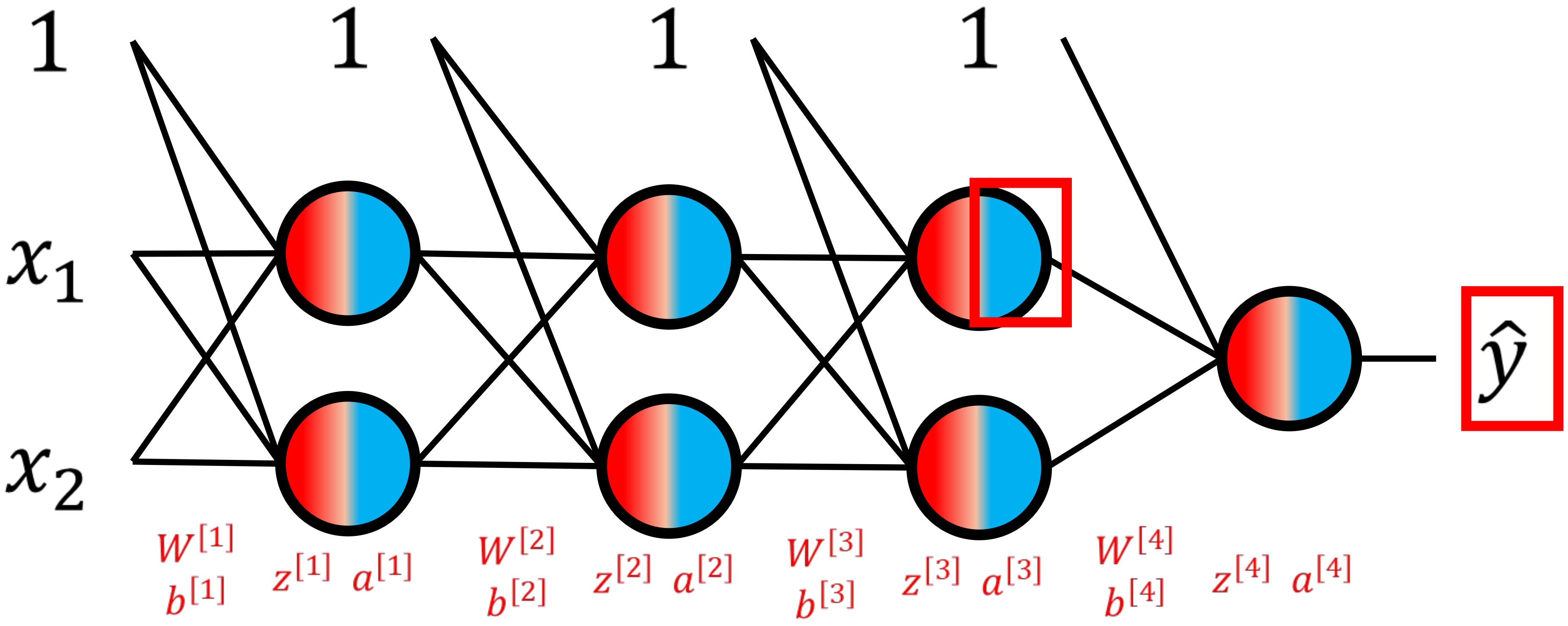
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $a_1^{[3]}$ affect $\hat{y}$?

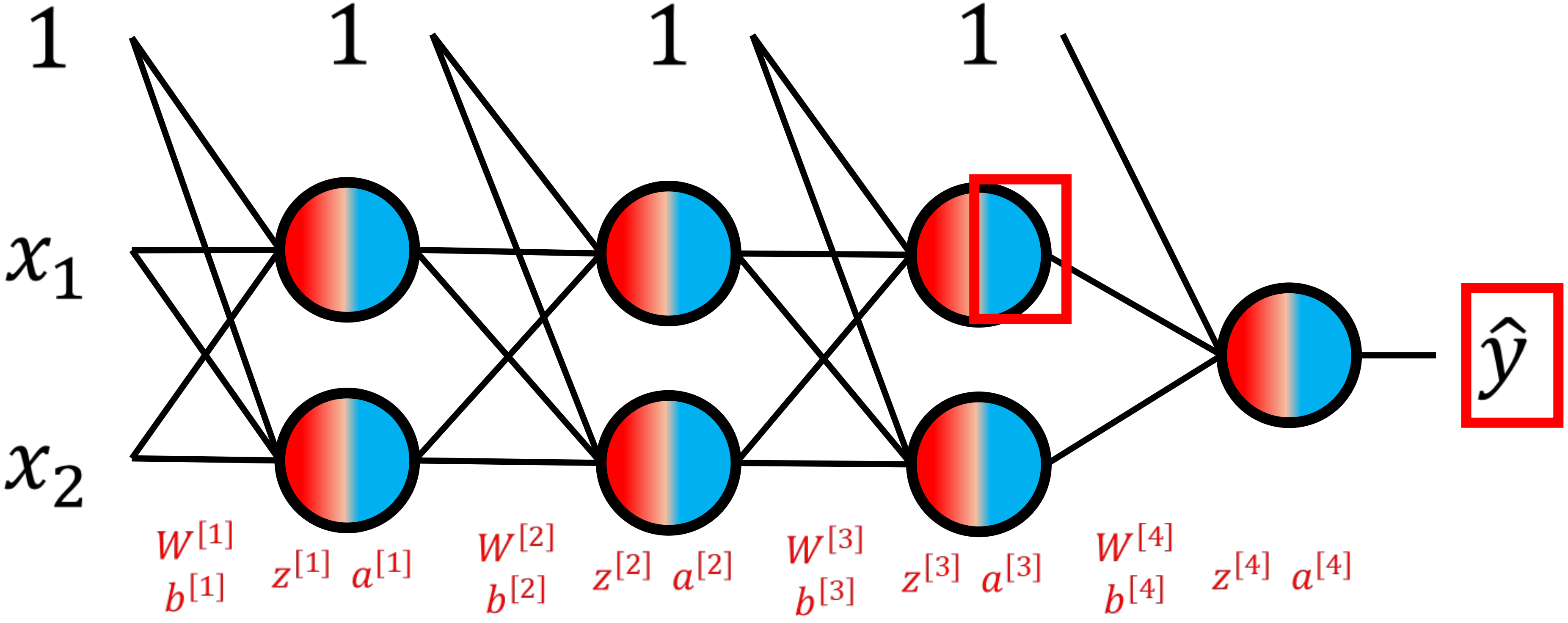
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial a_1^{[3]}}$$

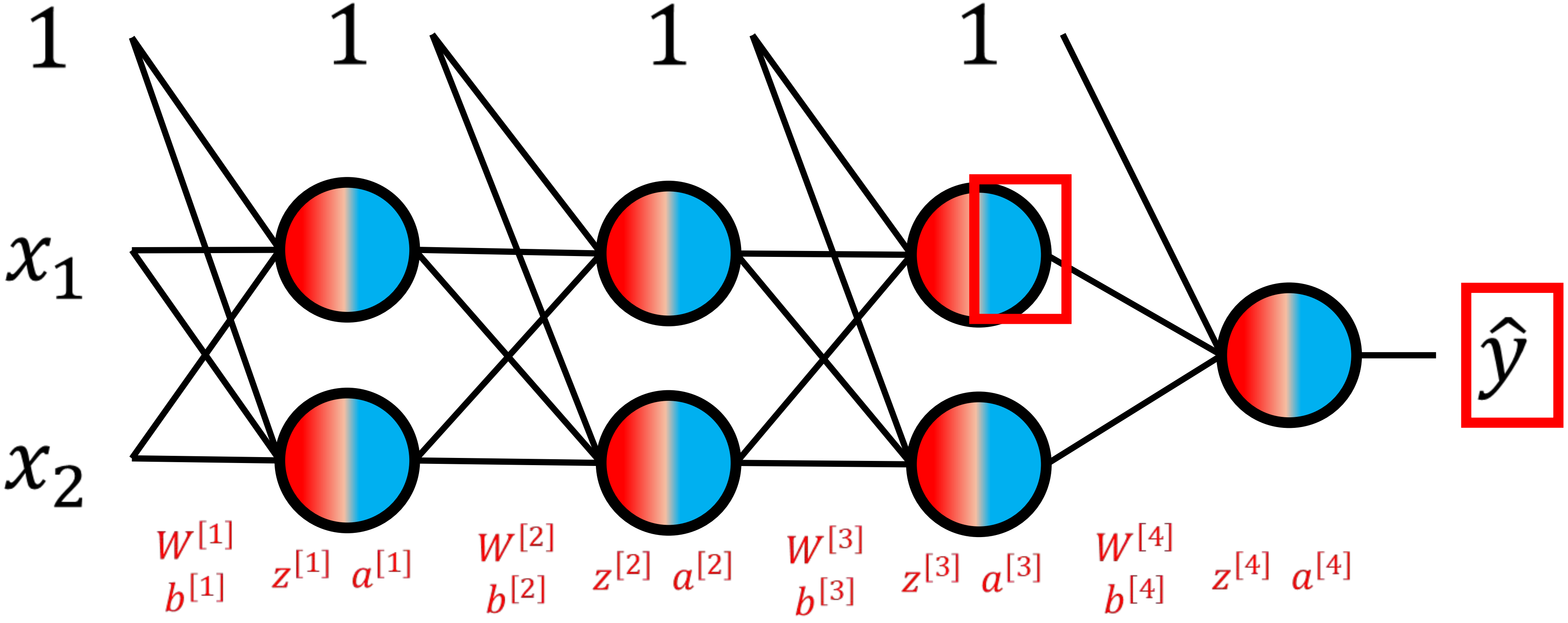
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial a_1^{[3]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{4}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial a_1^{[3]}}$$

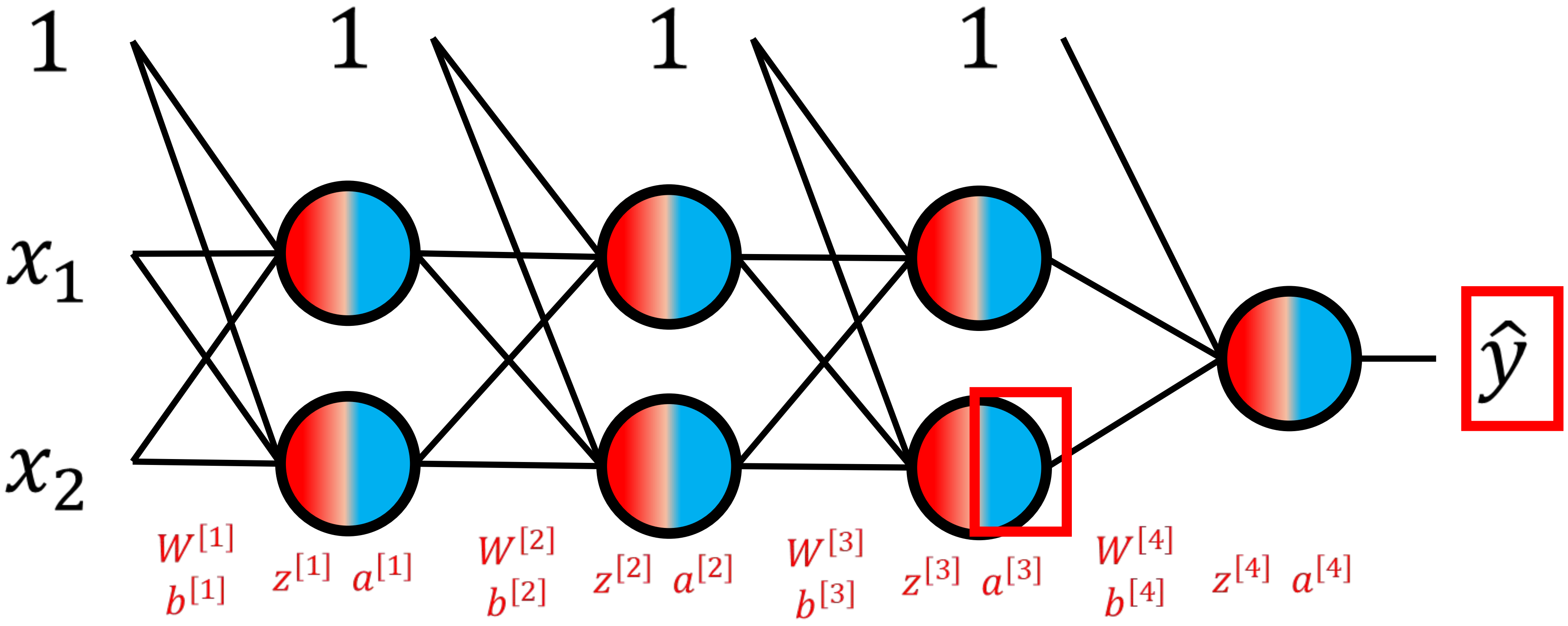
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $a_2^{[3]}$ affect $\hat{y}$?

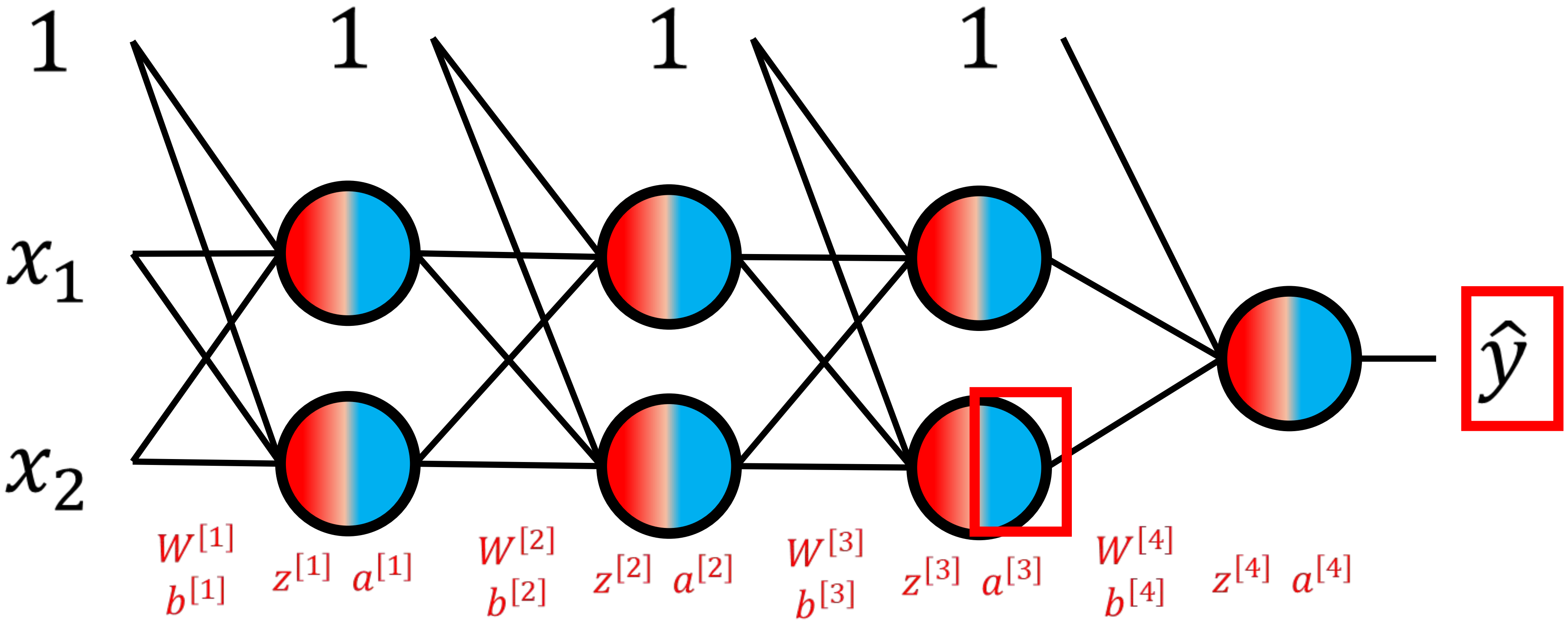
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial a_2^{[3]}}$$

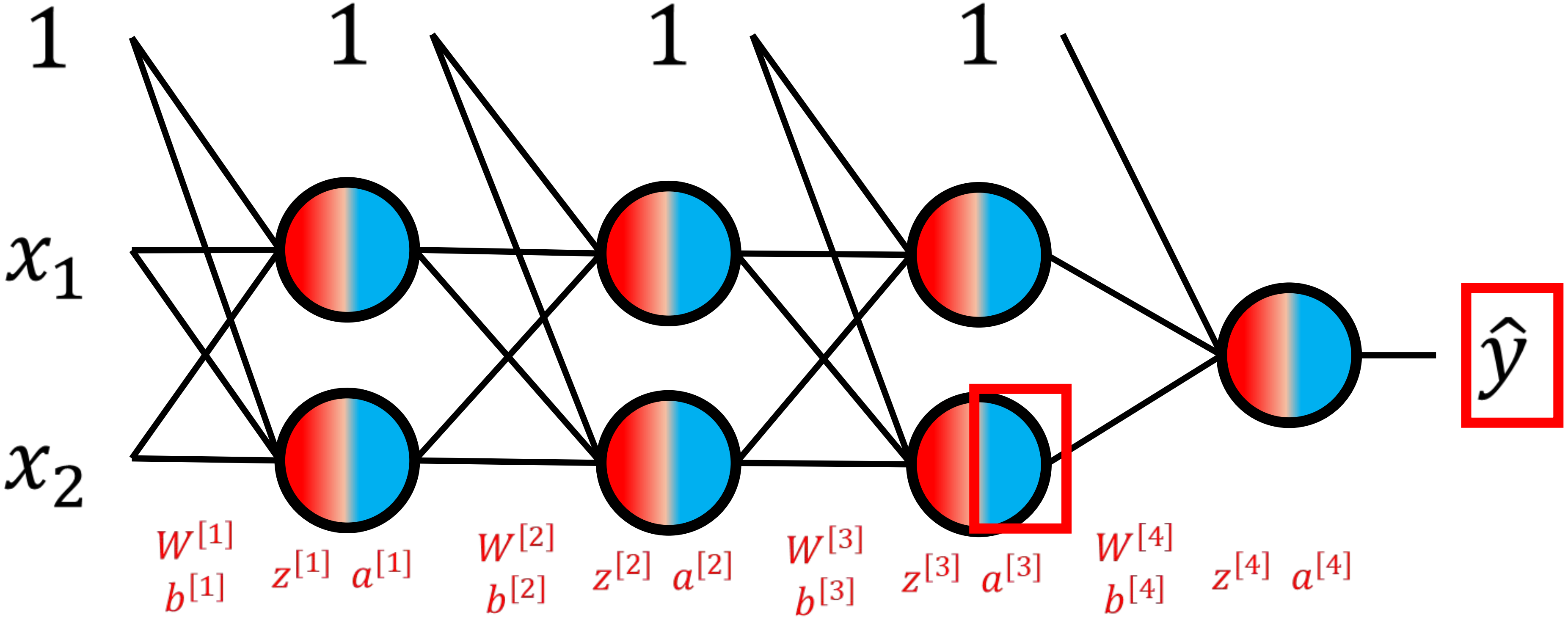
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial a_2^{[3]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{4}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial a_2^{[3]}}$$

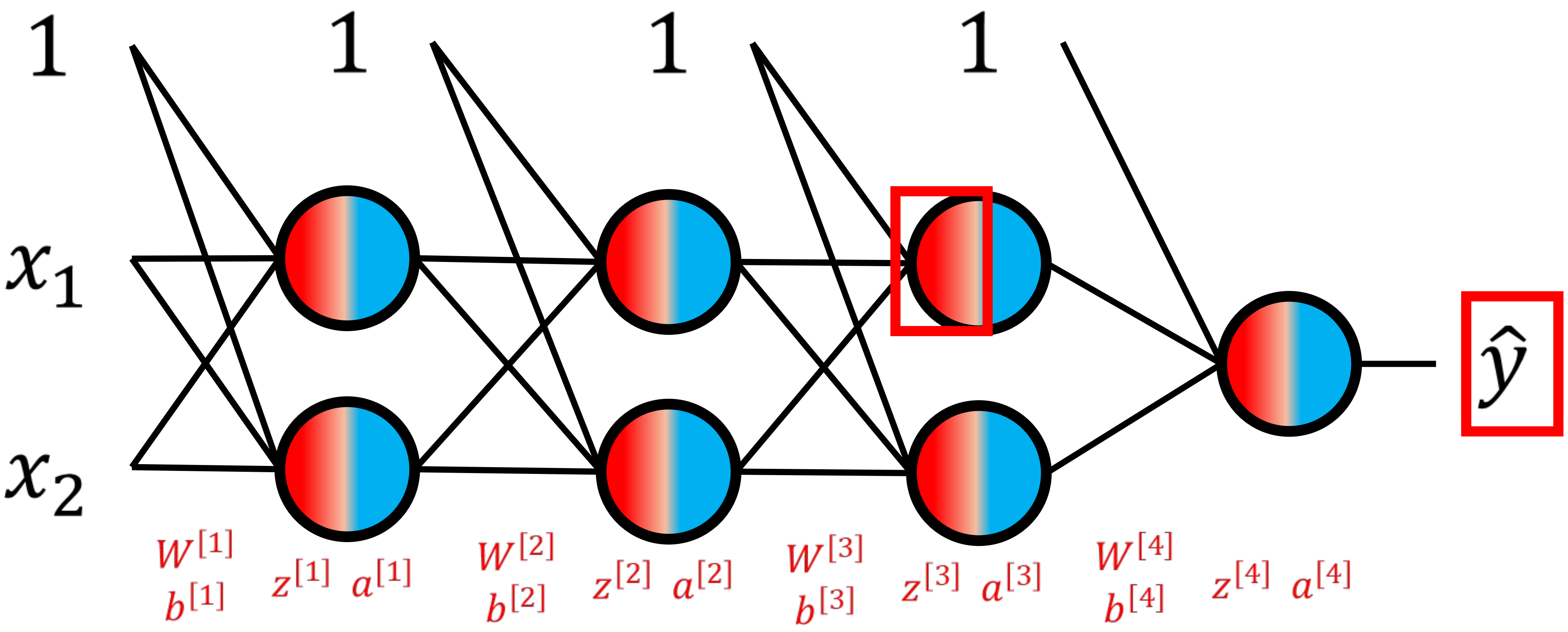
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $z_1^{[3]}$ affect $\hat{y}$?

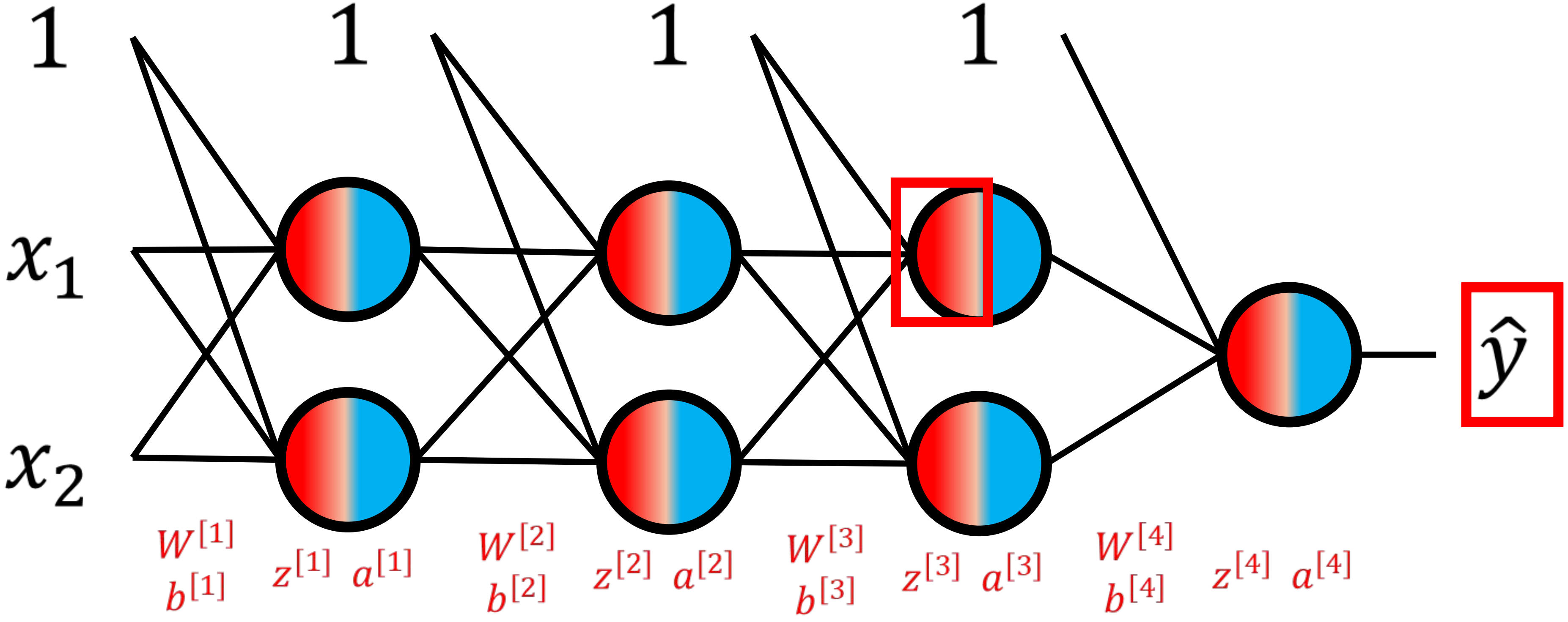
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial z_1^{[3]}}$$

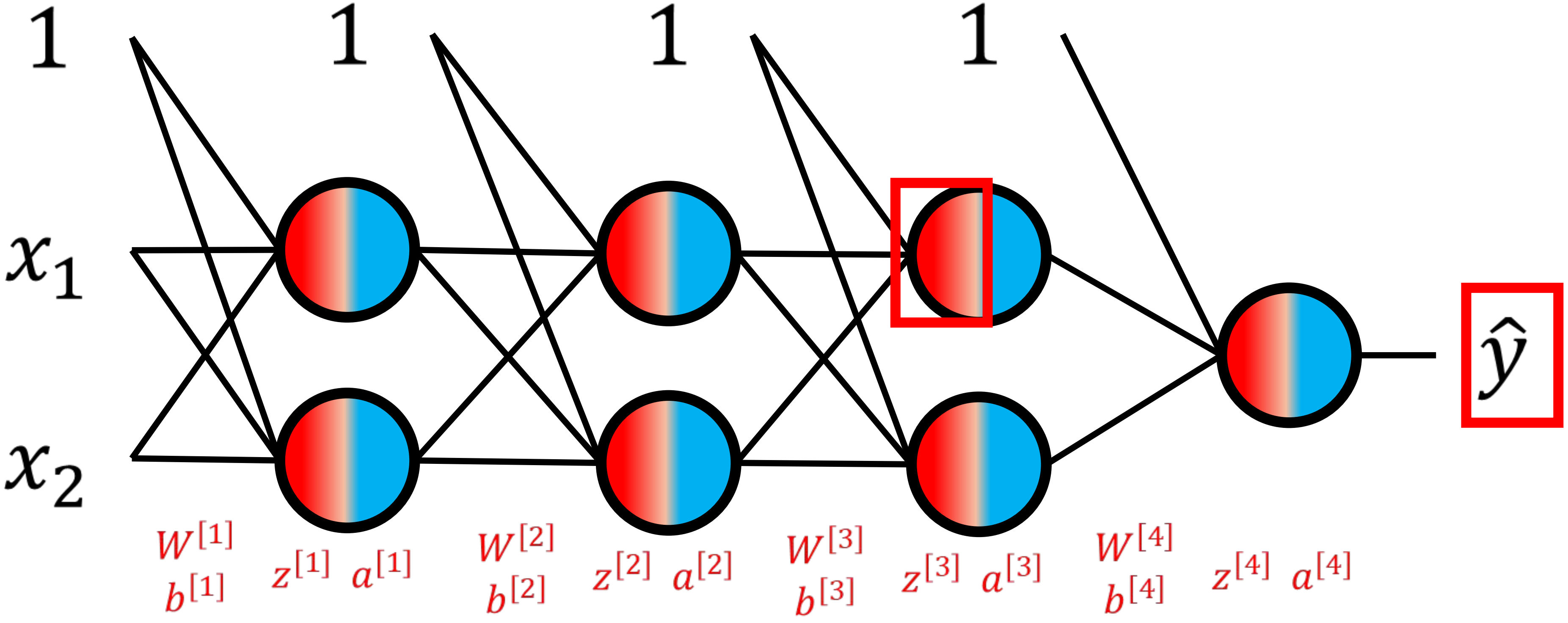
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial z_1^{[3]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{4}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial a_1^{[3]}} \times \frac{\partial a_1^{[3]}}{\partial z_1^{[3]}}$$

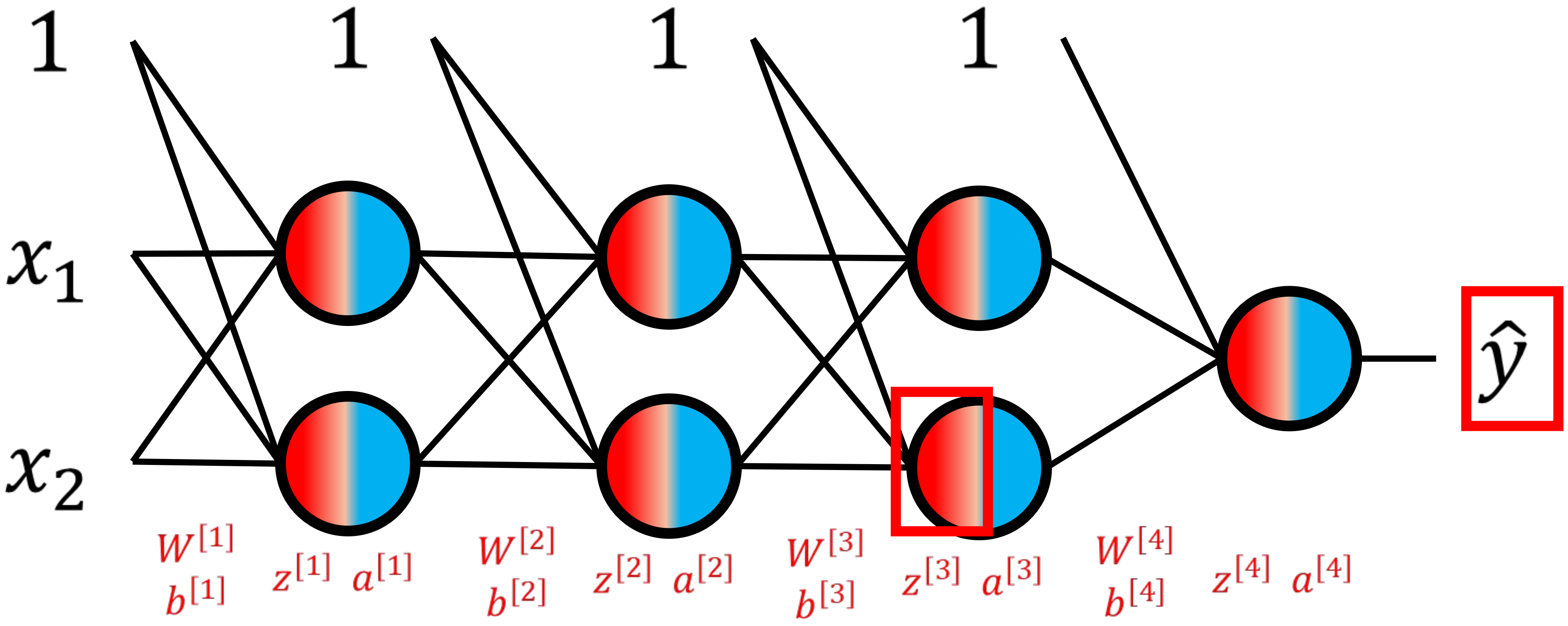
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $z_2^{[3]}$ affect $\hat{y}$?

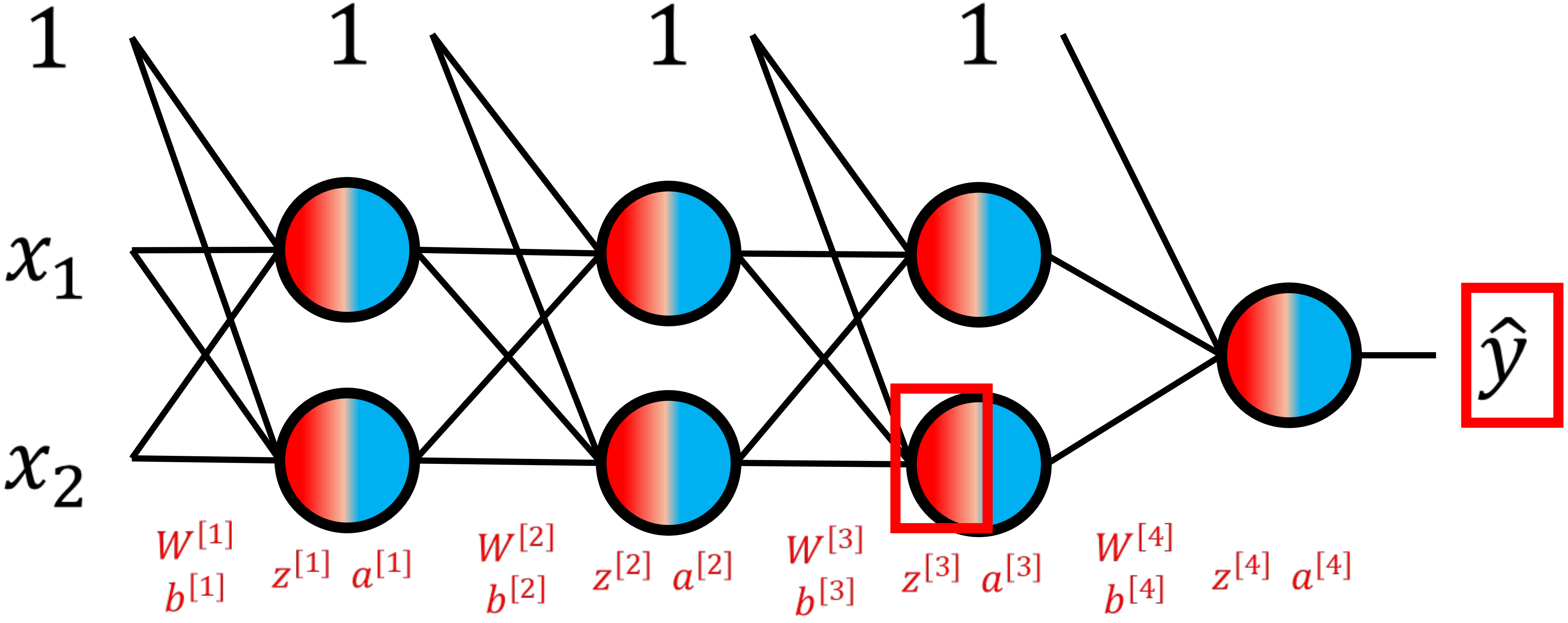
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial z_2^{[3]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{4}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial a_2^{[3]}} \times \frac{\partial a_2^{[3]}}{\partial z_2^{[3]}}$$

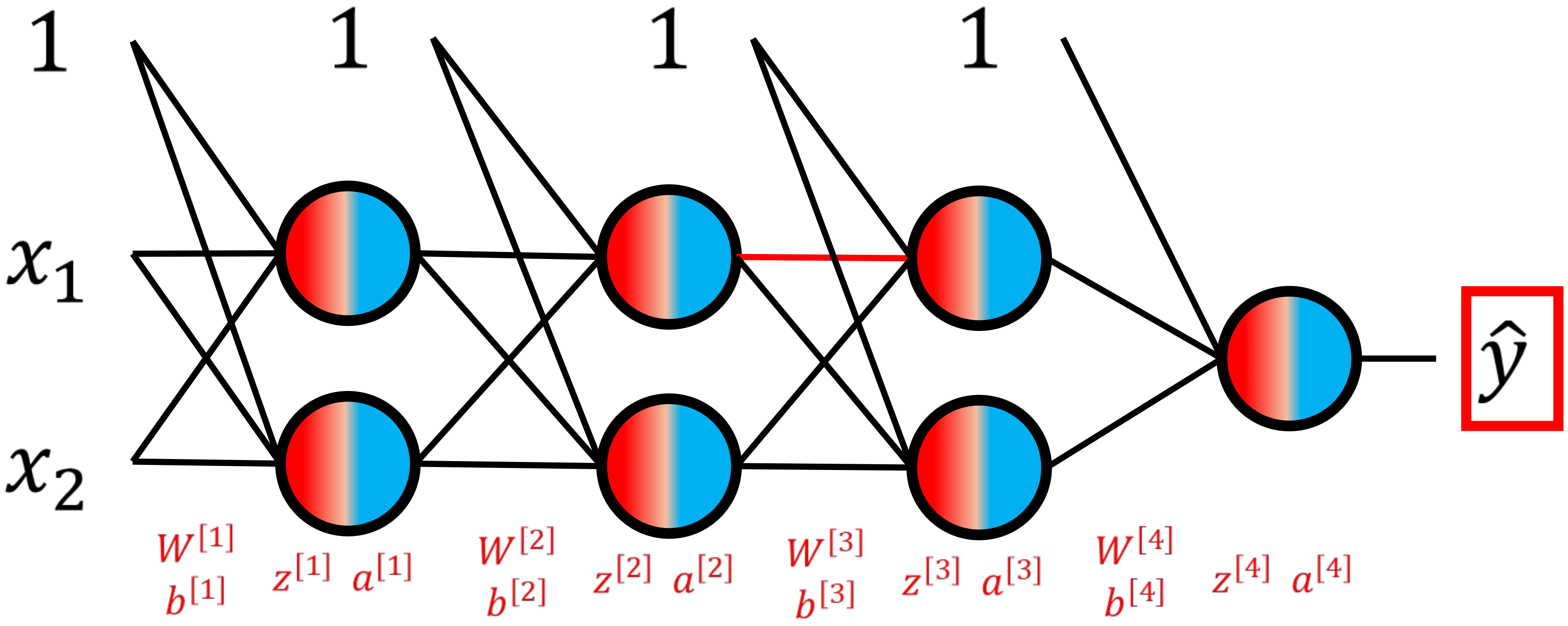
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $W_{11}^{[3]}$ affect $\hat{y}$?

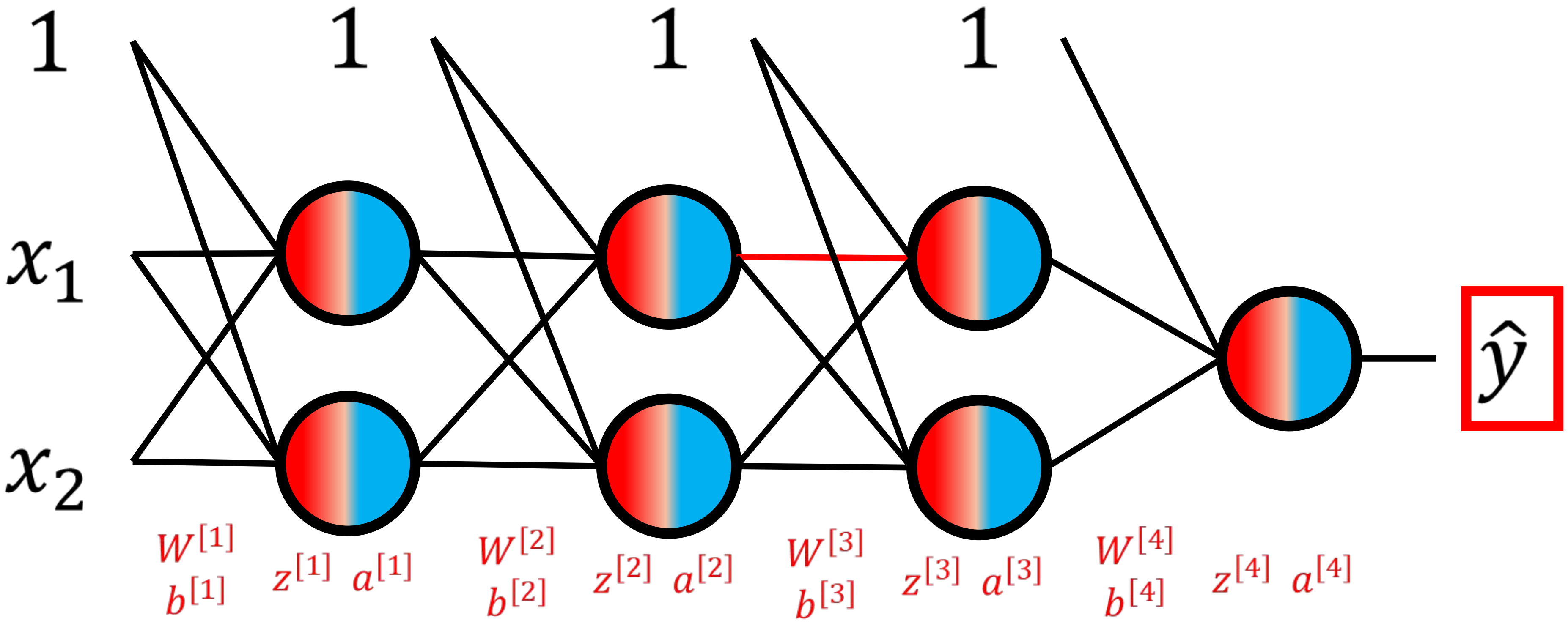
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{11}^{[3]}}$$

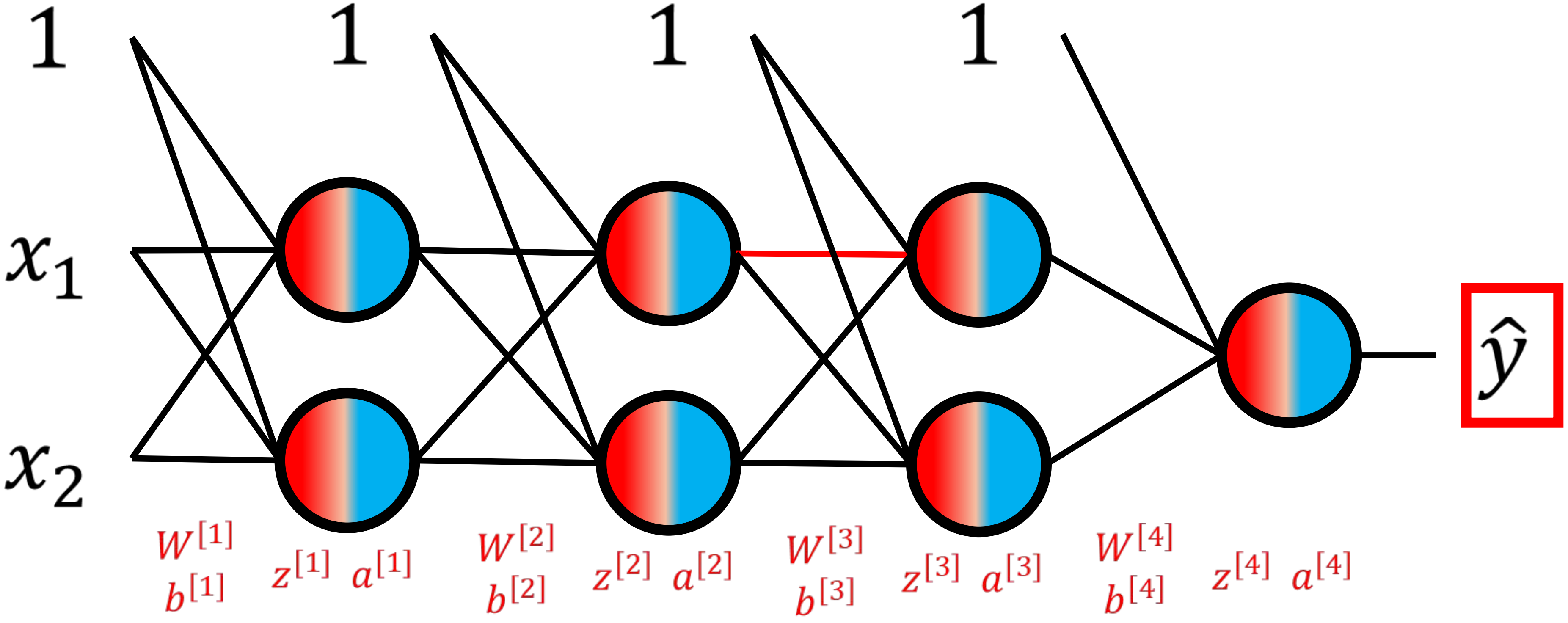
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{11}^{[3]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{[4]}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial a_1^{[3]}} \times \frac{\partial a_1^{[3]}}{\partial z_1^{[3]}} \times \frac{\partial z_1^{[3]}}{\partial W_{11}^{[3]}}$$

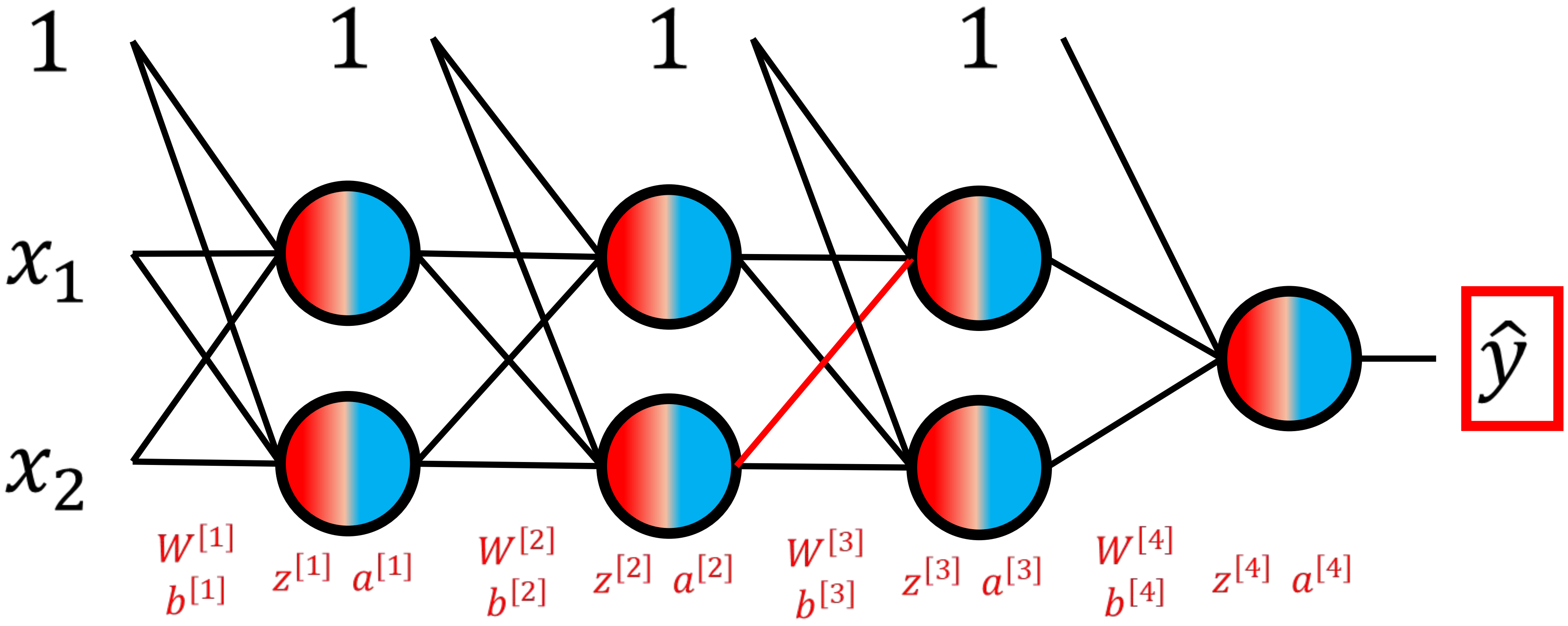
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $W_{12}^{[3]}$ affect $\hat{y}$?

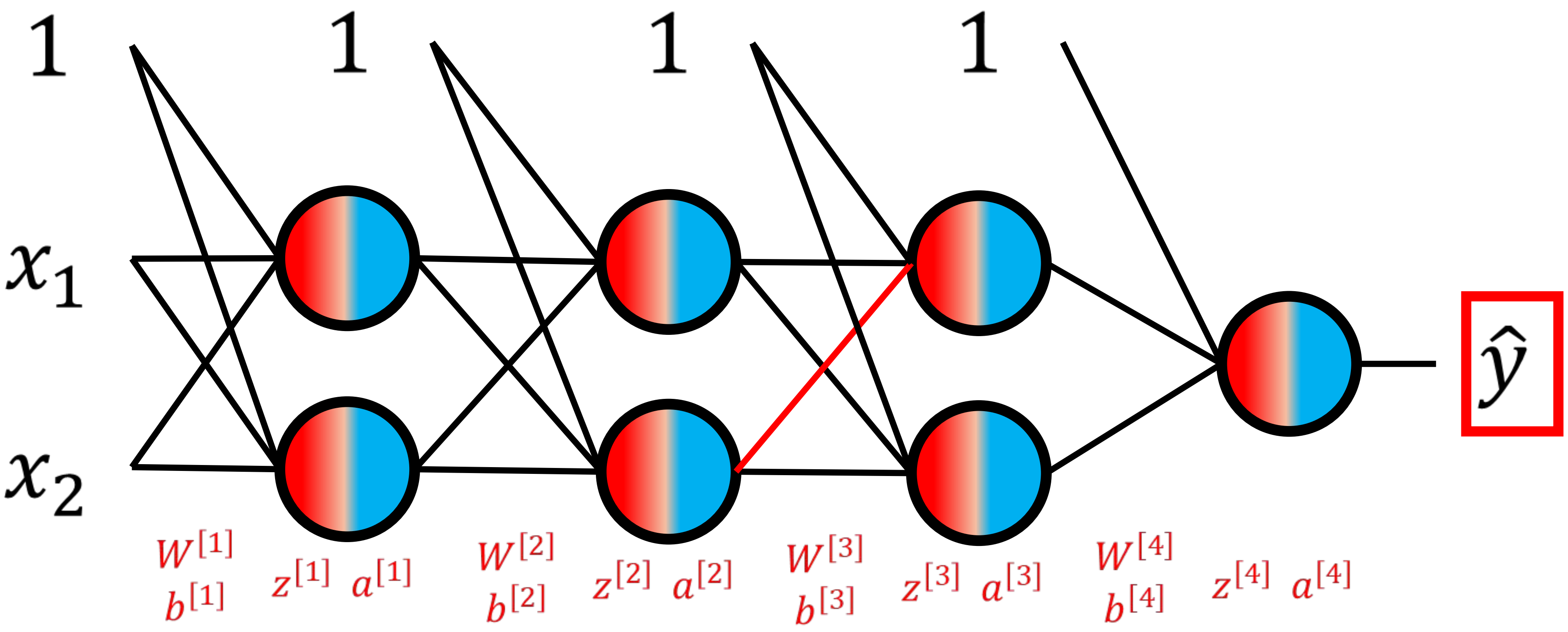
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W^{[3]}_{12}}$$

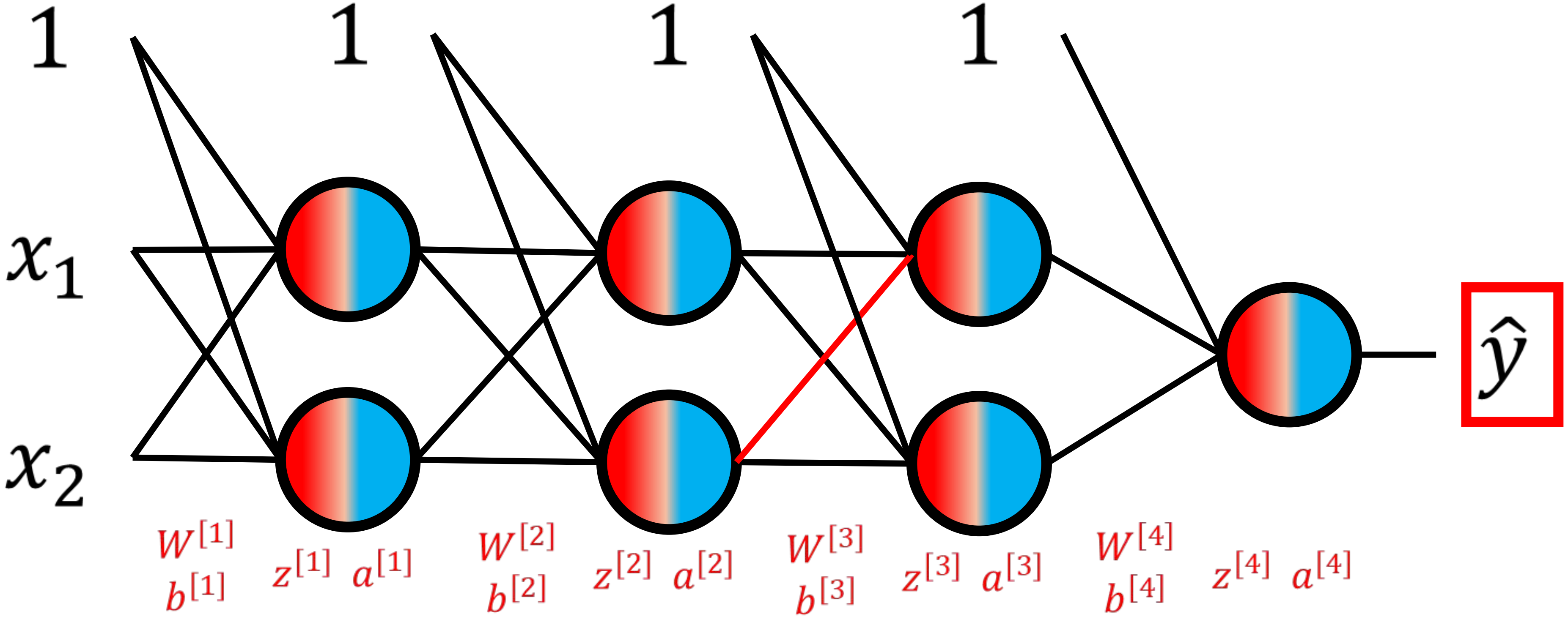
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma\left(z^{[i]}\right)\left(1 - \sigma\left(z^{[i]}\right)\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{12}^{[3]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{[4]}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial a_1^{[3]}} \times \frac{\partial a_1^{[3]}}{\partial z_1^{[3]}} \times \frac{\partial z_1^{[3]}}{\partial W_{12}^{[3]}}$$

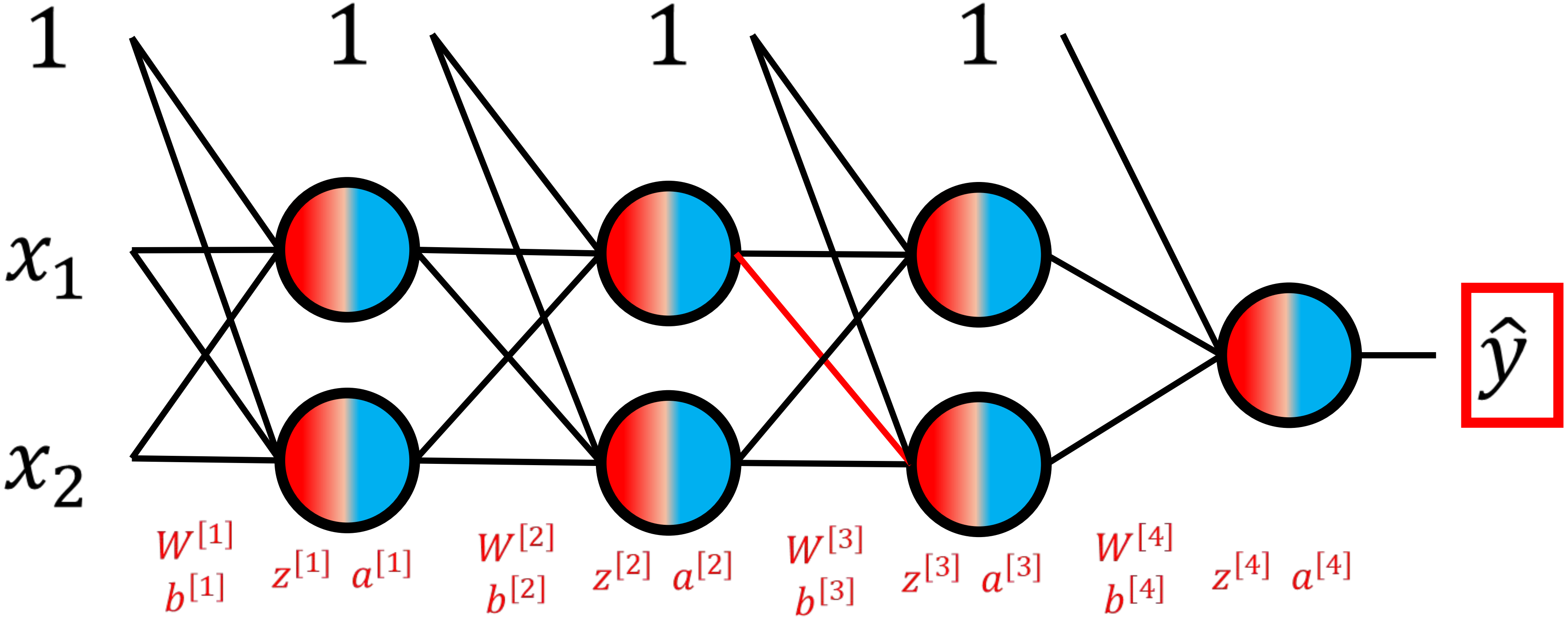
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $W_{21}^{[3]}$ affect $\hat{y}$?

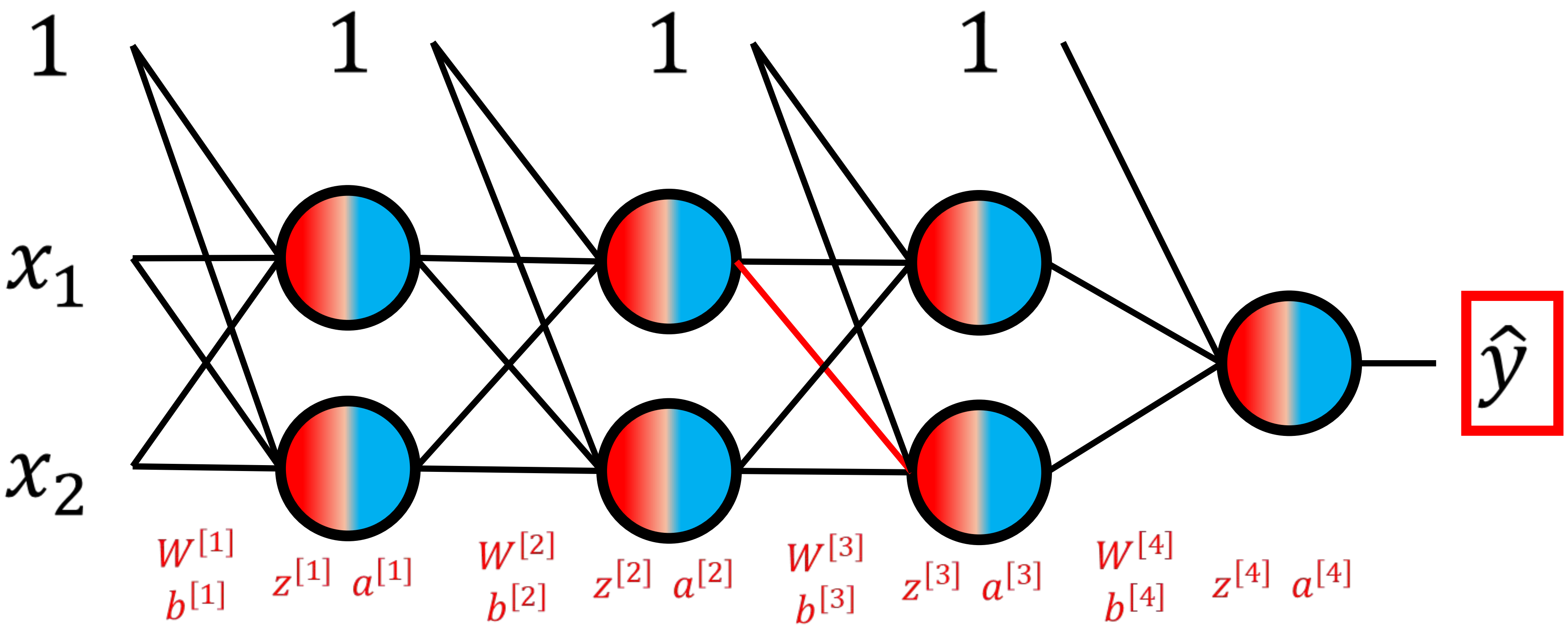
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{21}^{[3]}}$$

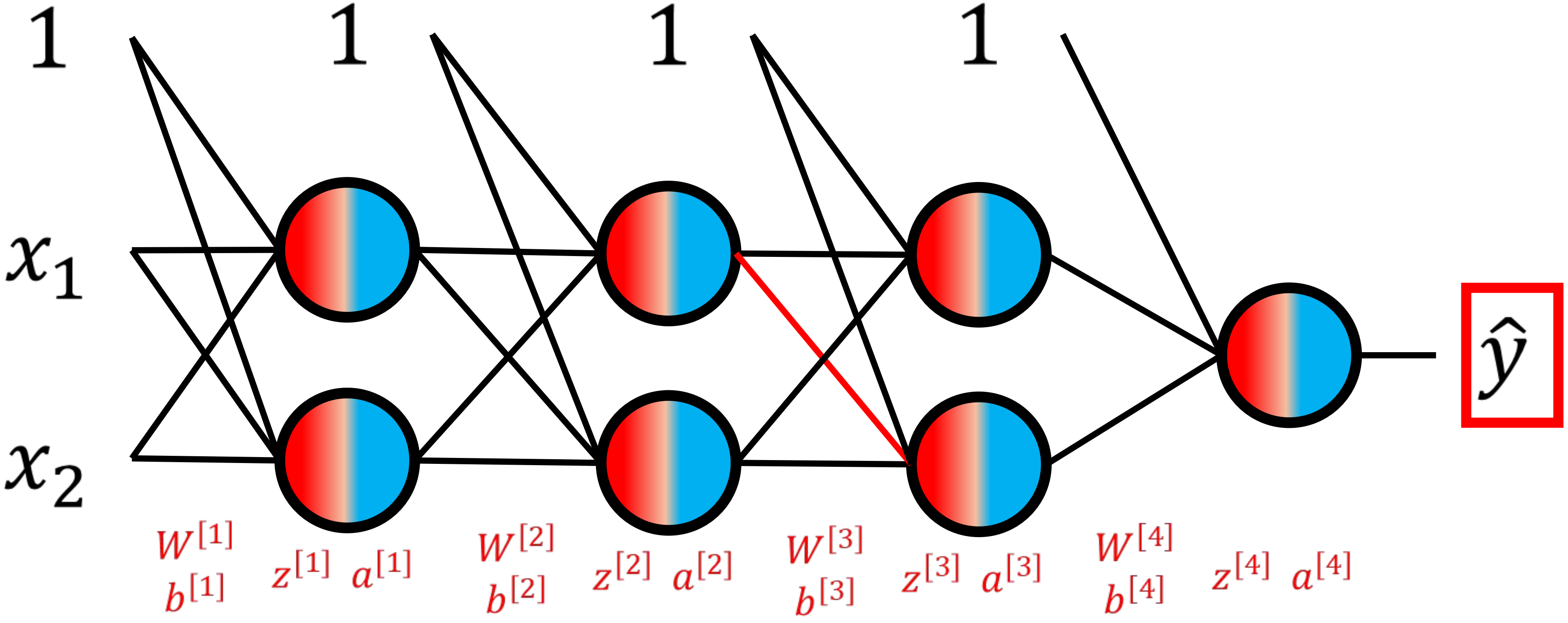
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{21}^{[3]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{[4]}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial a_2^{[3]}} \times \frac{\partial a_2^{[3]}}{\partial z_2^{[3]}} \times \frac{\partial z_2^{[3]}}{\partial W_{21}^{[3]}}$$

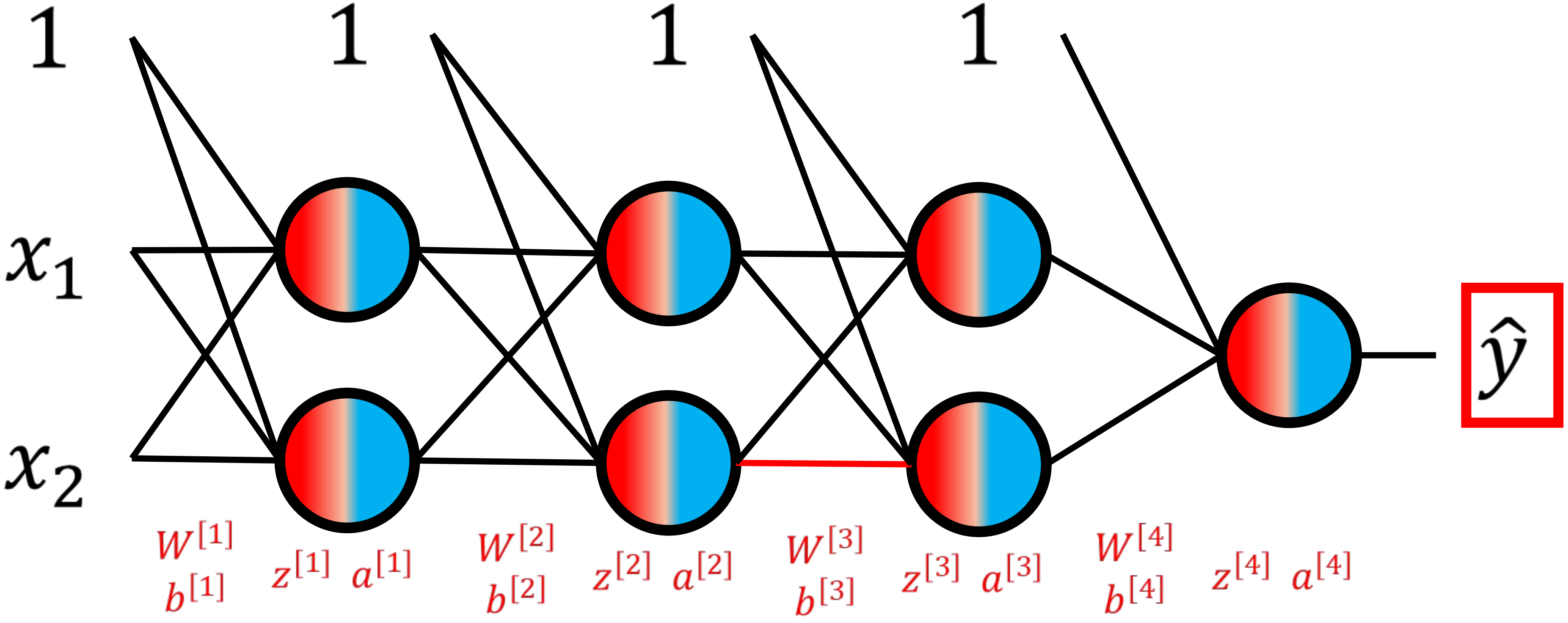
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $W^{[3]}_{22}$ affect $\hat{y}$?

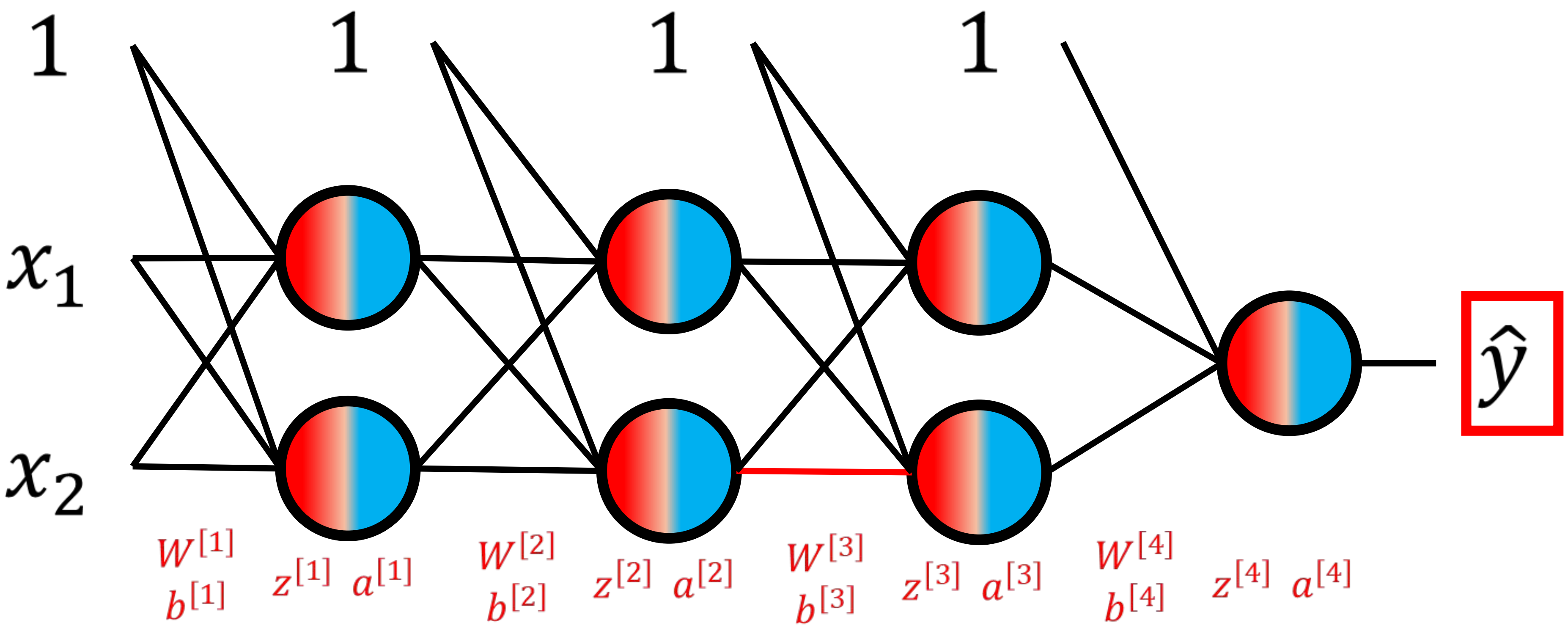
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{22}^{[3]}}$$

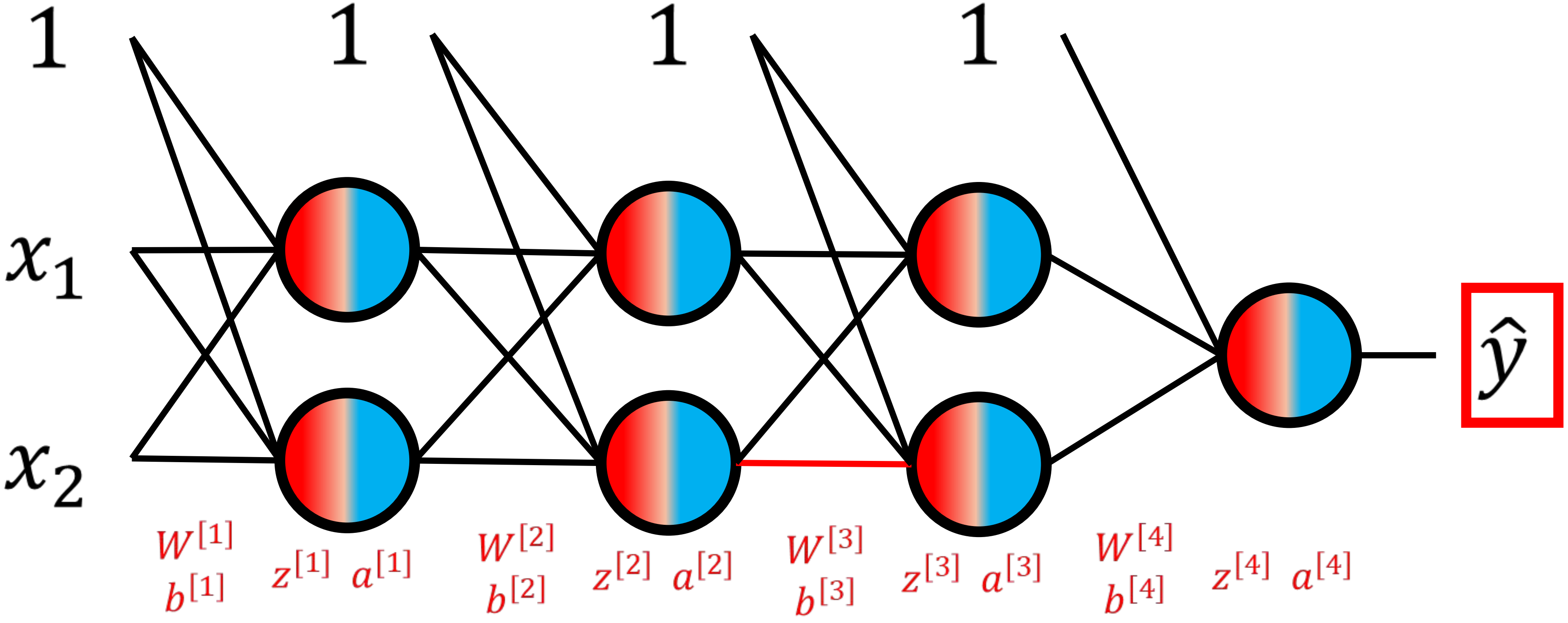
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial W_{22}^{[3]}} = \frac{\partial \hat{y}}{\partial a_1^{[4]}} \times \frac{\partial a_1^{[4]}}{\partial z_1^{[4]}} \times \frac{\partial z_1^{[4]}}{\partial a_2^{[3]}} \times \frac{\partial a_2^{[3]}}{\partial z_2^{[3]}} \times \frac{\partial z_2^{[3]}}{\partial W_{22}^{[3]}}$$

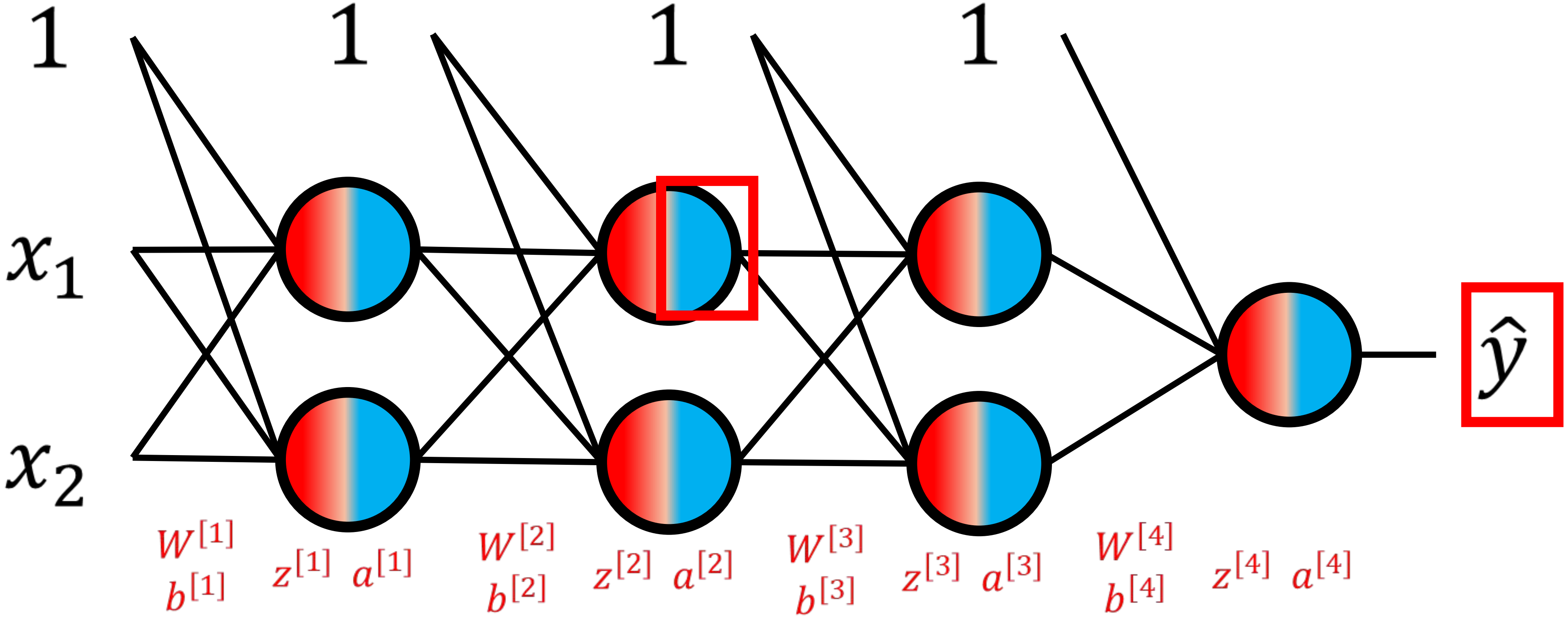
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $a_1^{[2]}$ affect $\hat{y}$?

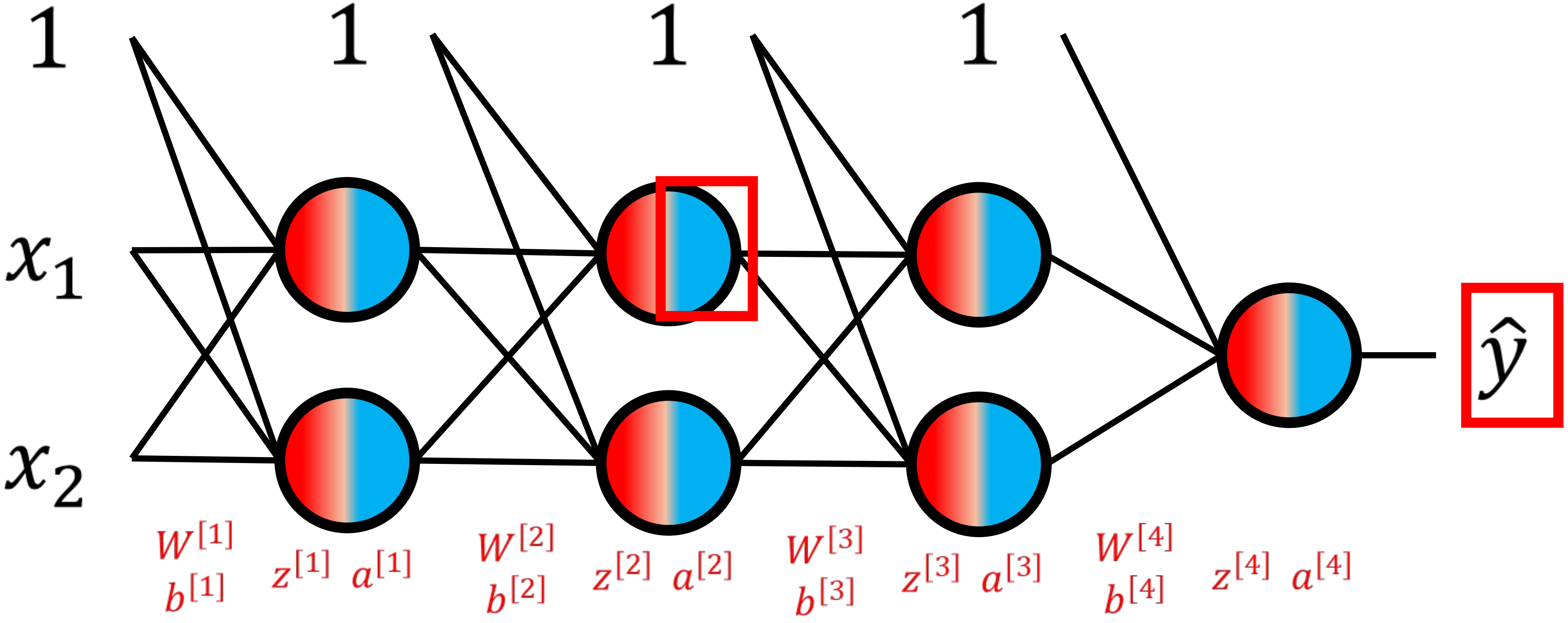
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



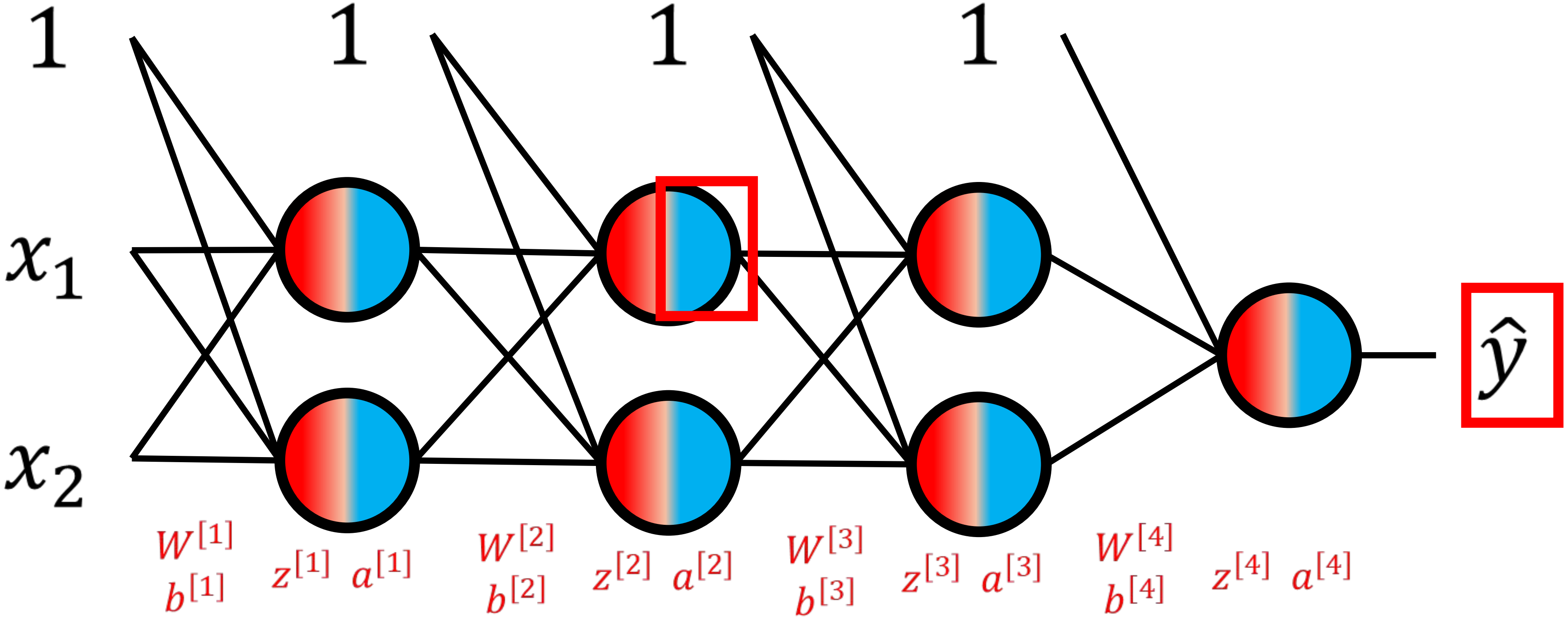
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial a_1^{[2]}} = \left(\frac{\partial \hat{y}}{\partial z_1^{[3]}} \times \frac{\partial z_1^{[3]}}{\partial a_1^{[2]}}\right) + \left(\frac{\partial \hat{y}}{\partial z_2^{[3]}} \times \frac{\partial z_2^{[3]}}{\partial a_1^{[2]}}\right)$$

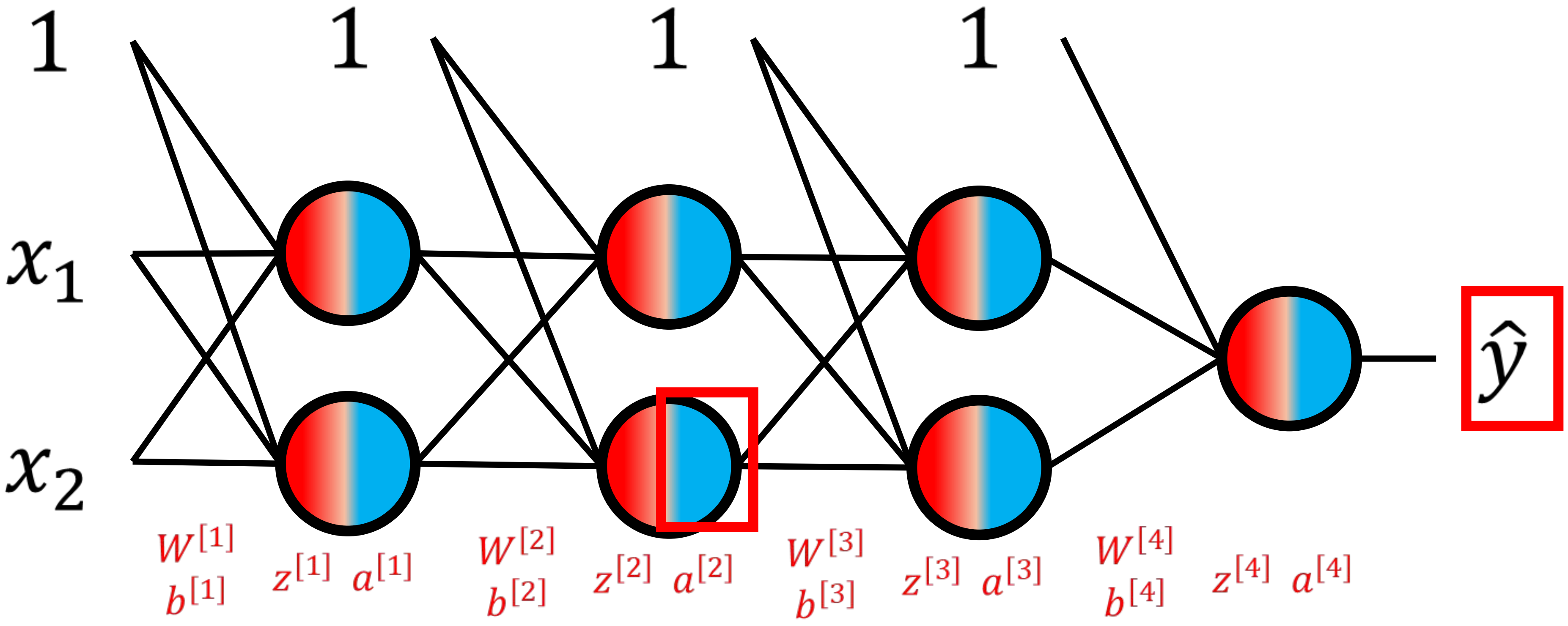
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



How does $a_2^{[2]}$ affect $\hat{y}$?

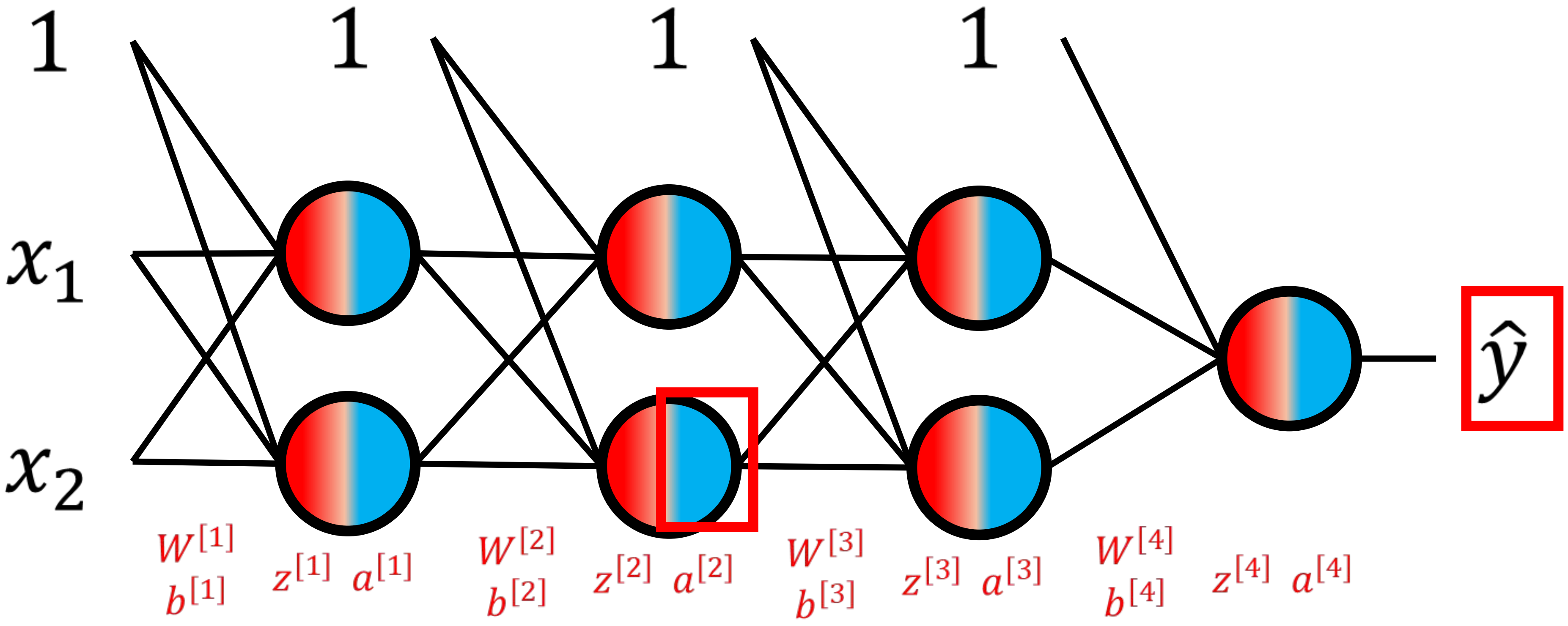
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial a_2^{[2]}}$$

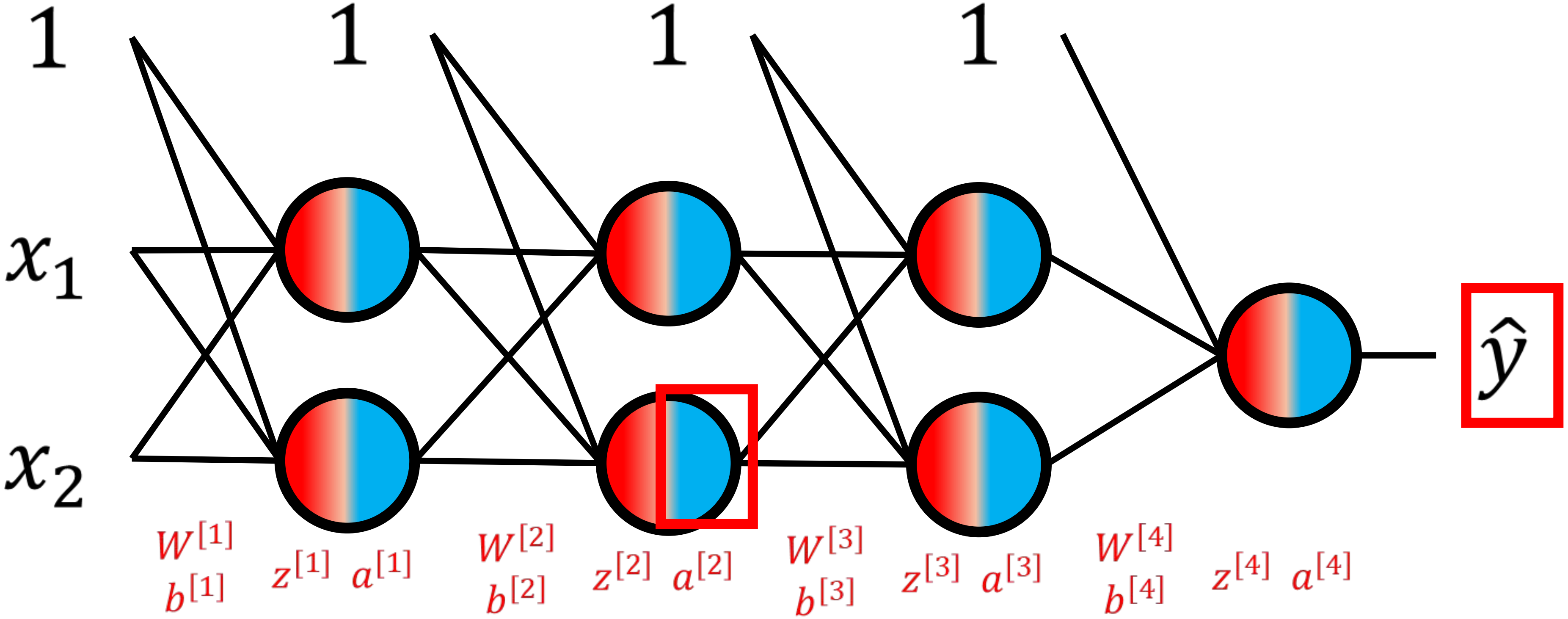
$$\frac{\partial \hat{y}}{\partial a^{[4]}} = 1$$

$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = \sigma(z^{[i]})\left(1 - \sigma(z^{[i]})\right)$$

$$\frac{\partial z^{[i]}}{\partial W^{[i]}} = a^{[i-1]}$$

$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = W^{[i]}$$

$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$



$$\frac{\partial \hat{y}}{\partial a_2^{[2]}} = \left(\frac{\partial \hat{y}}{\partial z_1^{[3]}} \times \frac{\partial z_1^{[3]}}{\partial a_2^{[2]}}\right) + \left(\frac{\partial \hat{y}}{\partial z_2^{[3]}} \times \frac{\partial z_2^{[3]}}{\partial a_2^{[2]}}\right)$$

Remember:

- In these slides, we only computed how each weight affects the **prediction**.

- But in reality, what we really need to know is how the prediction affects the **loss**.

- So, you still need to compute the **derivative of the loss with respect to the prediction**, and multiply it at the end of the chain rule (you can try out this step by yourself).