Copyright Notice

Optimization in Neural Networks and NewtonÕs Method

Regression with a perceptron

Regression Problem Motivation

Regression Problem Motivation

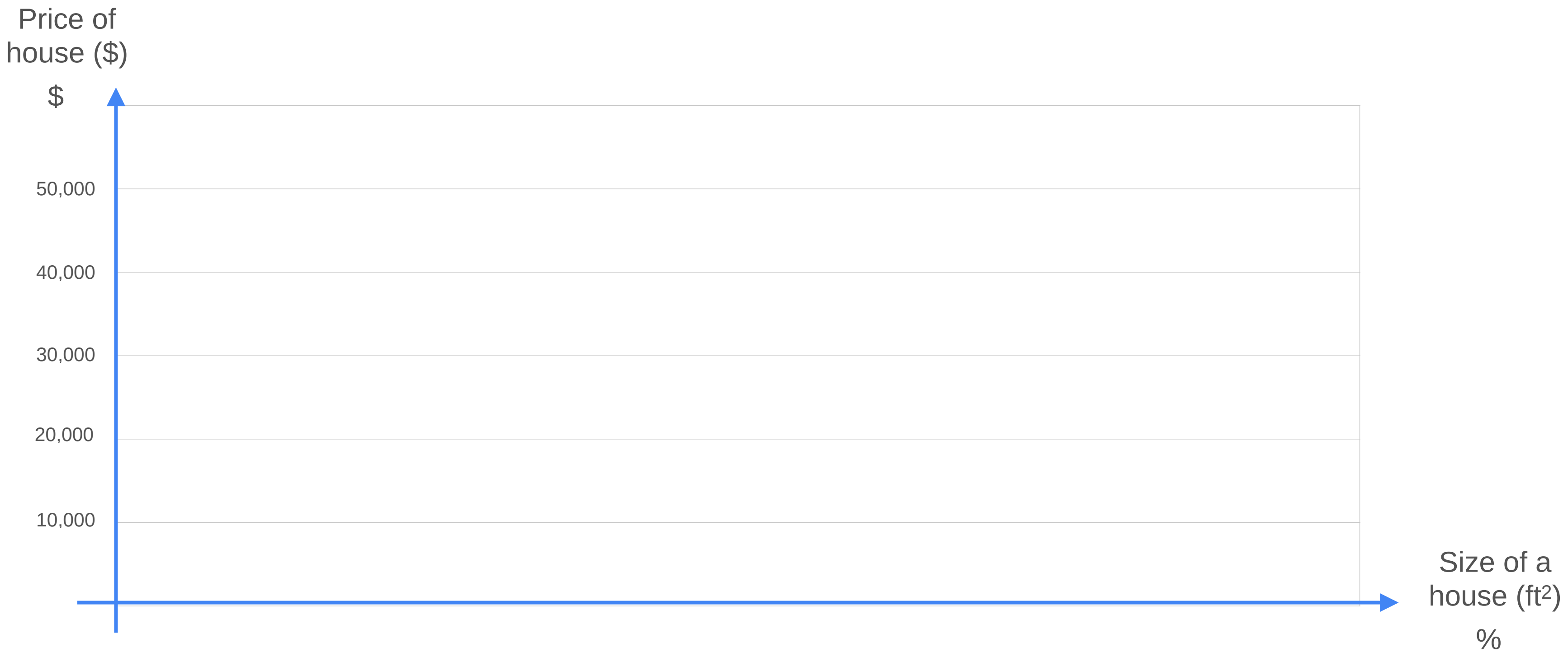
Predicting

the price of a house

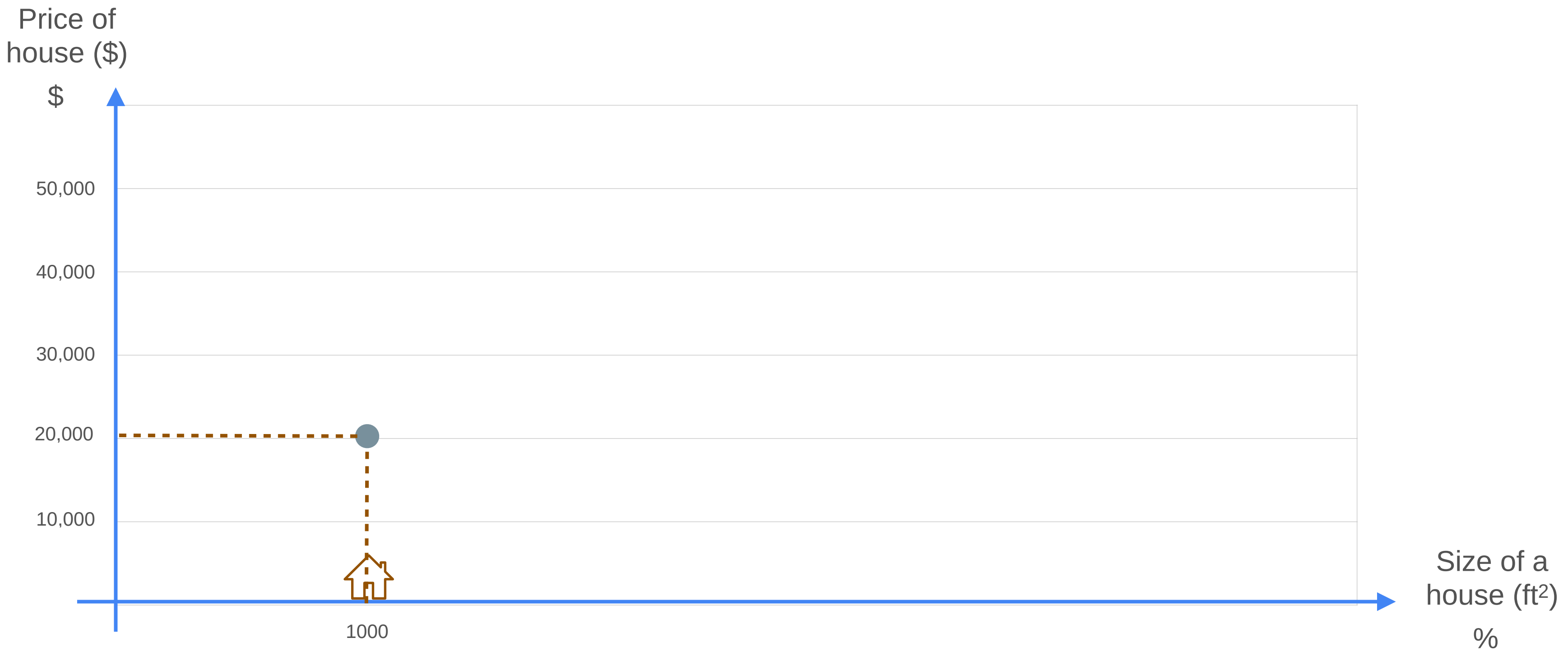
from

the size of the house

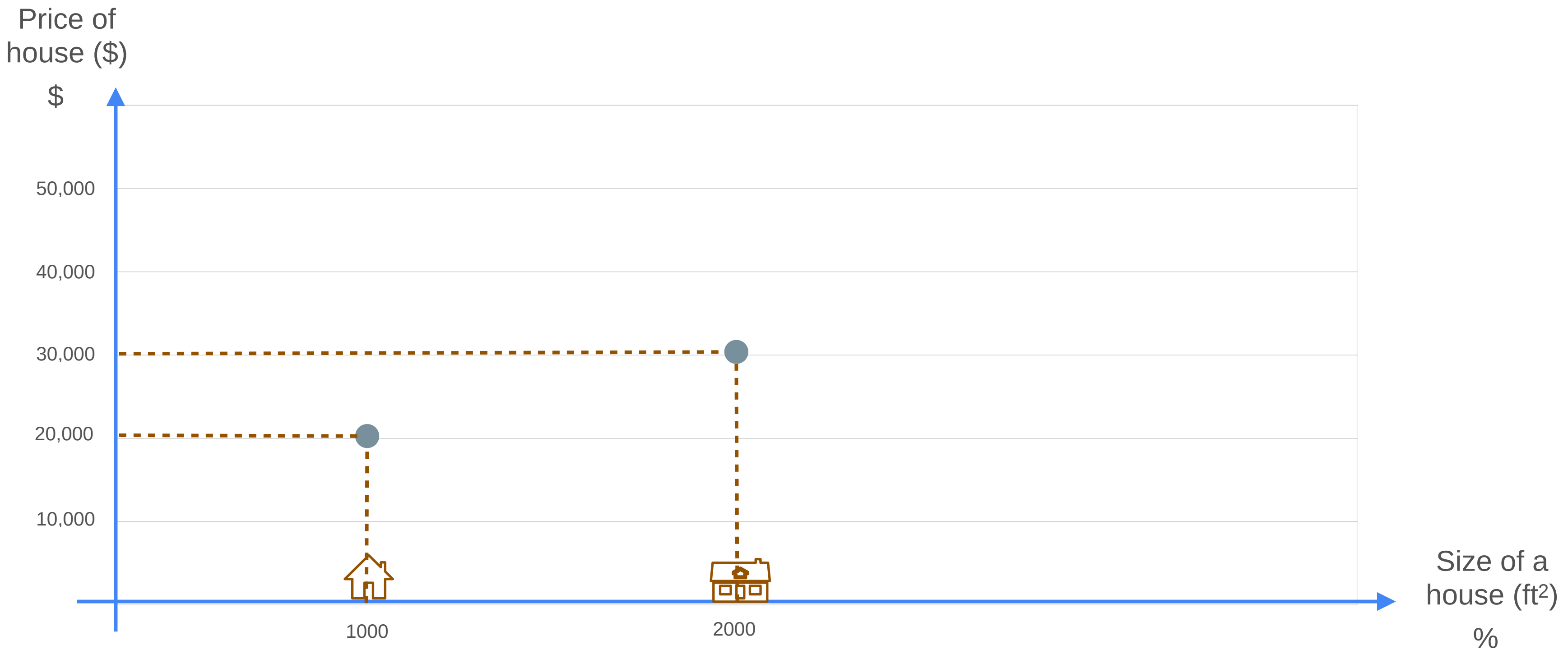
Regression Problem Motivation



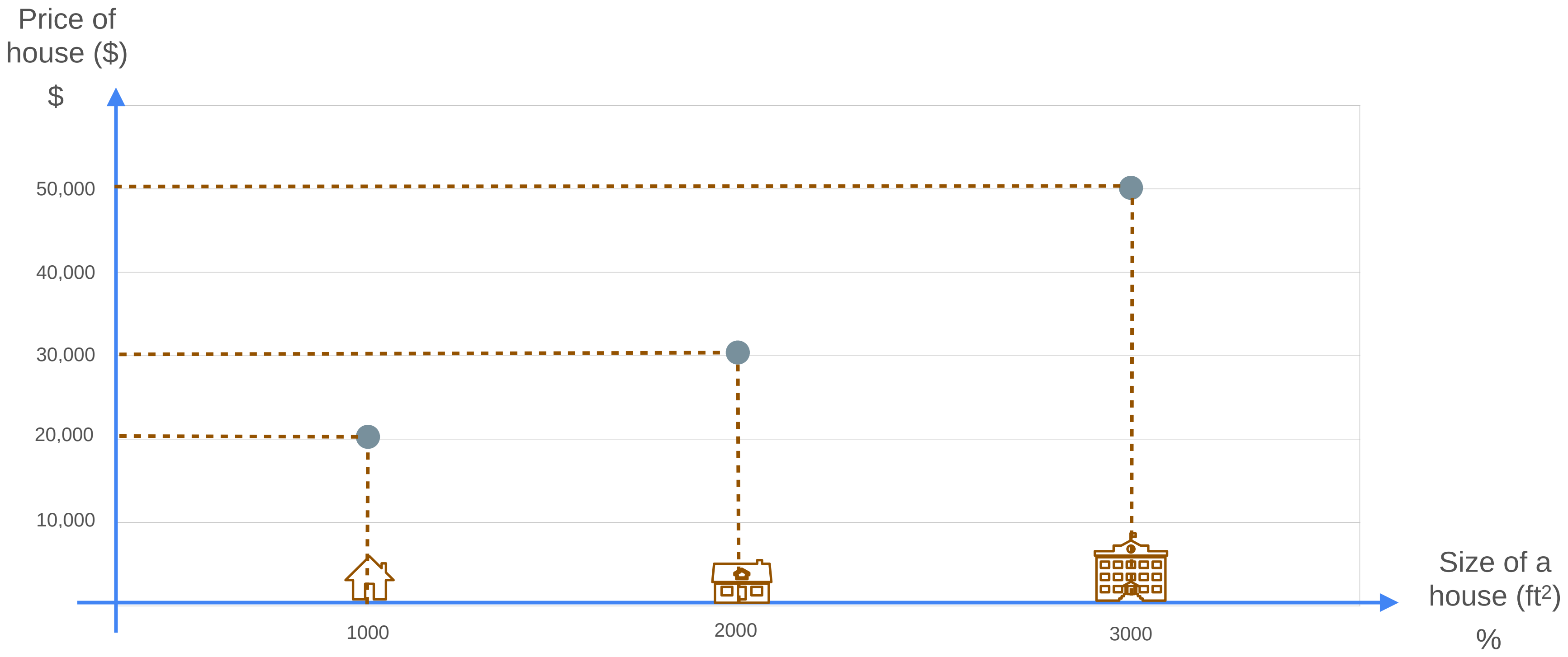
Regression Problem Motivation



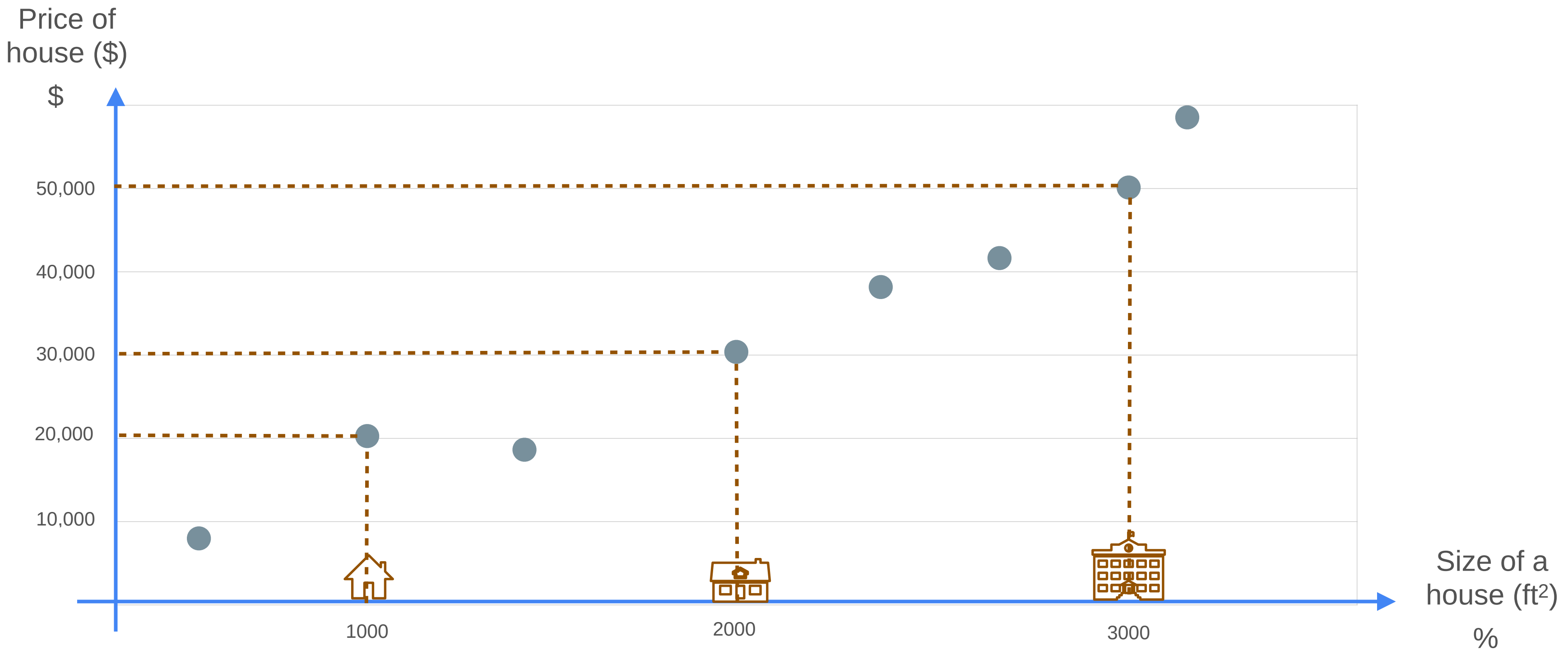
Regression Problem Motivation



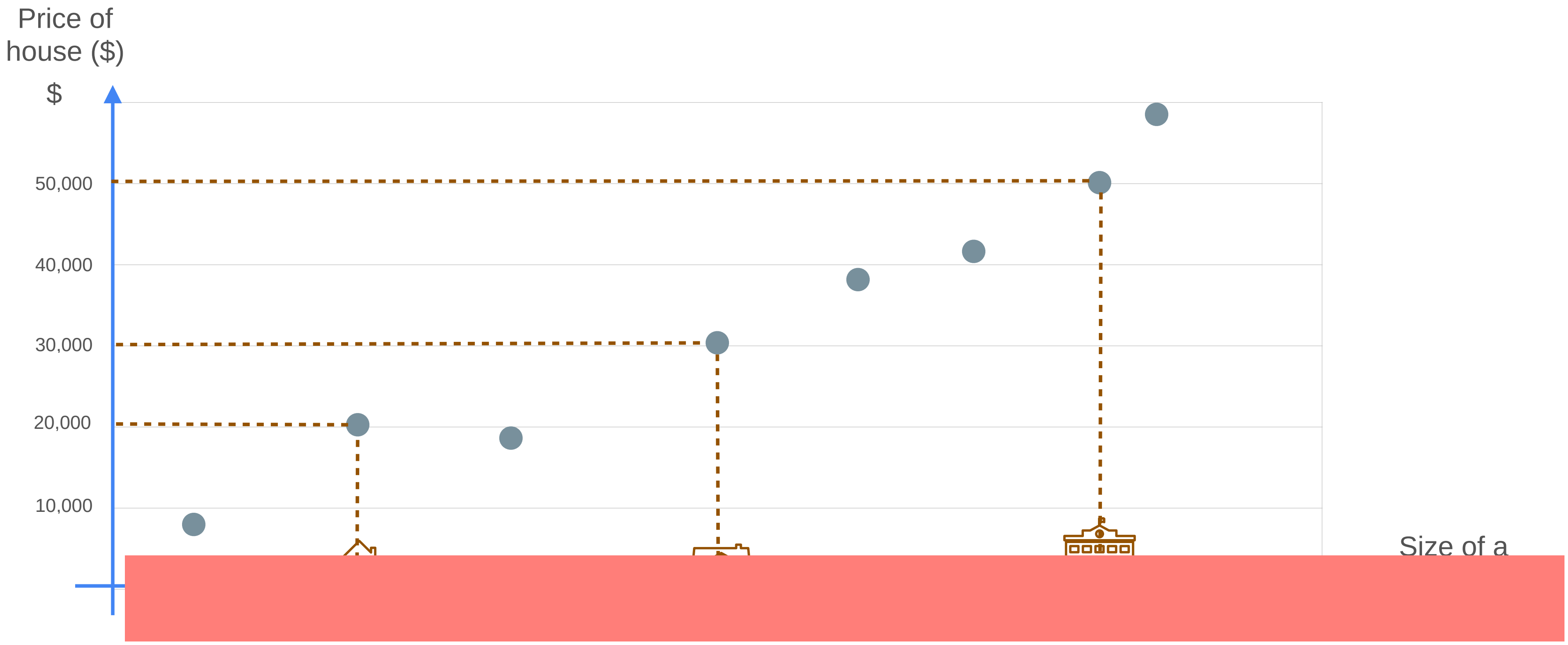
Regression Problem Motivation



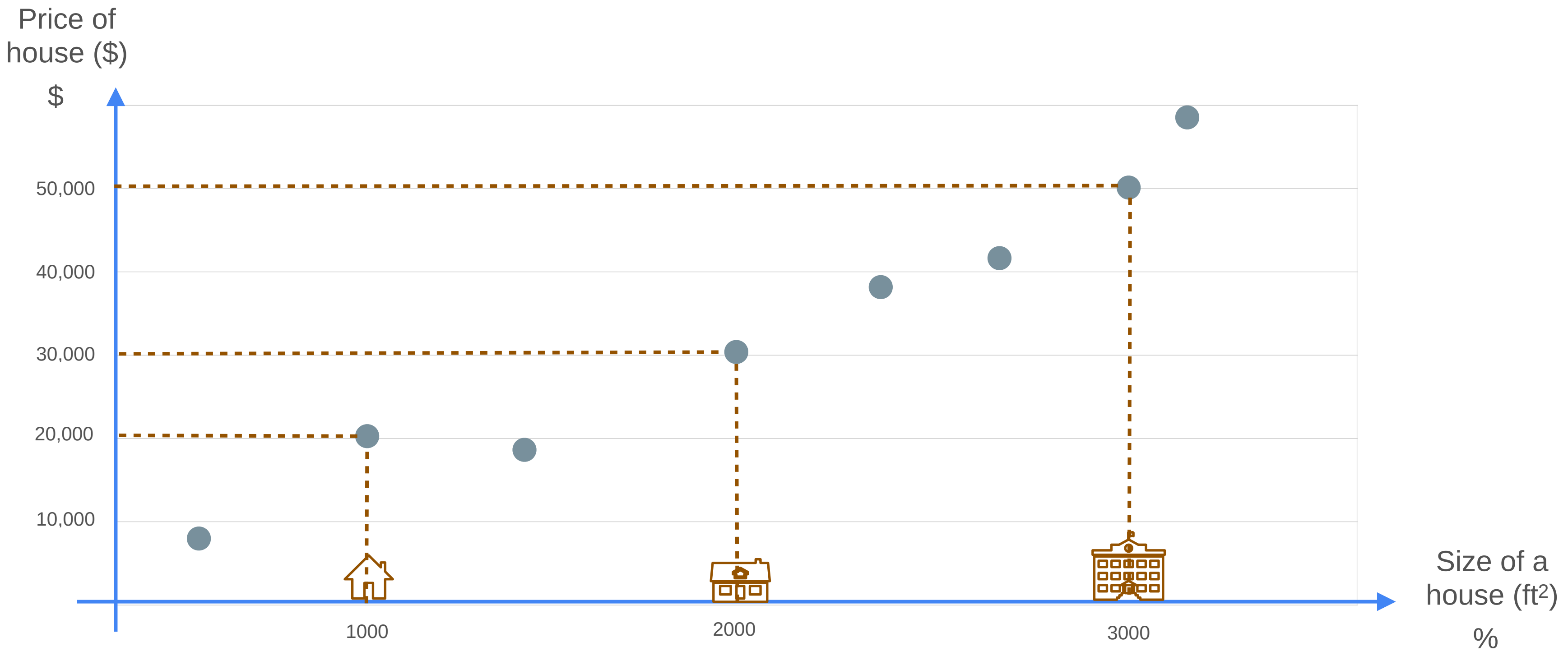
Regression Problem Motivation



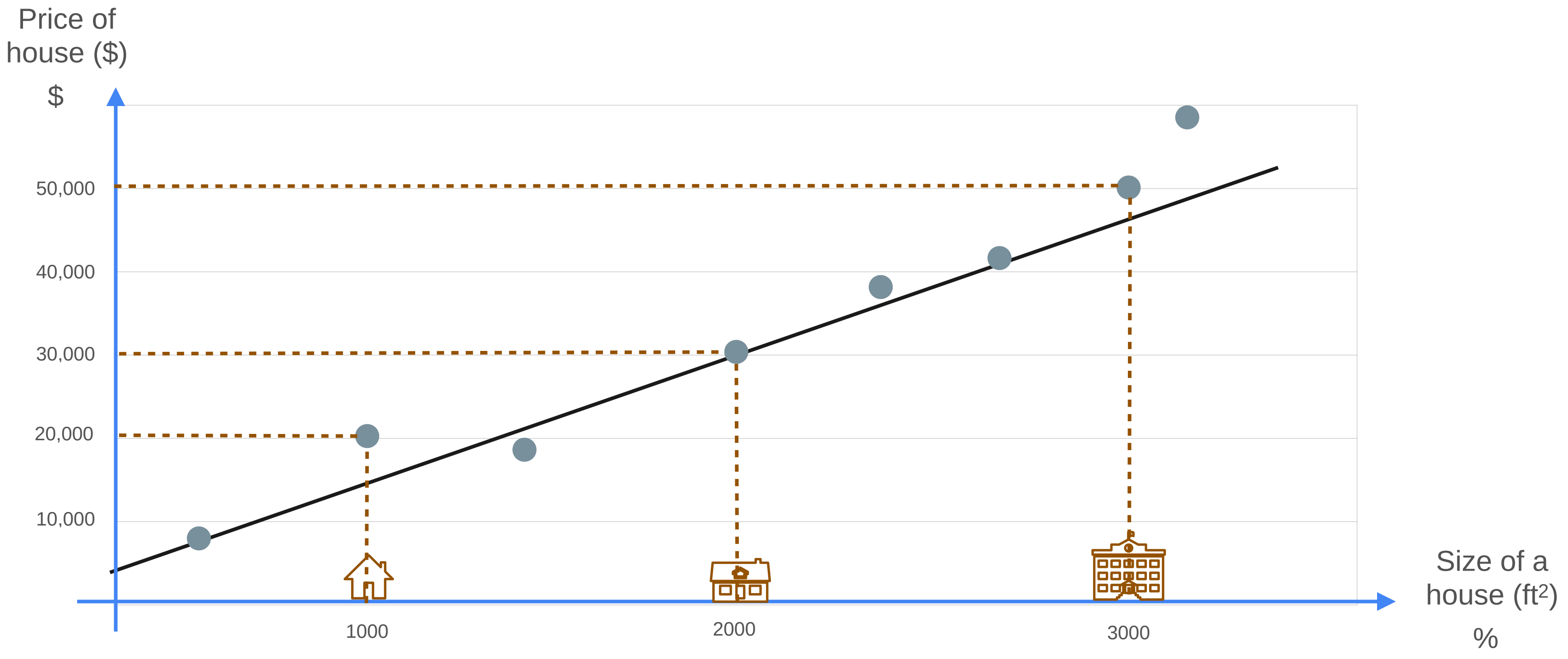
Regression Problem Motivation



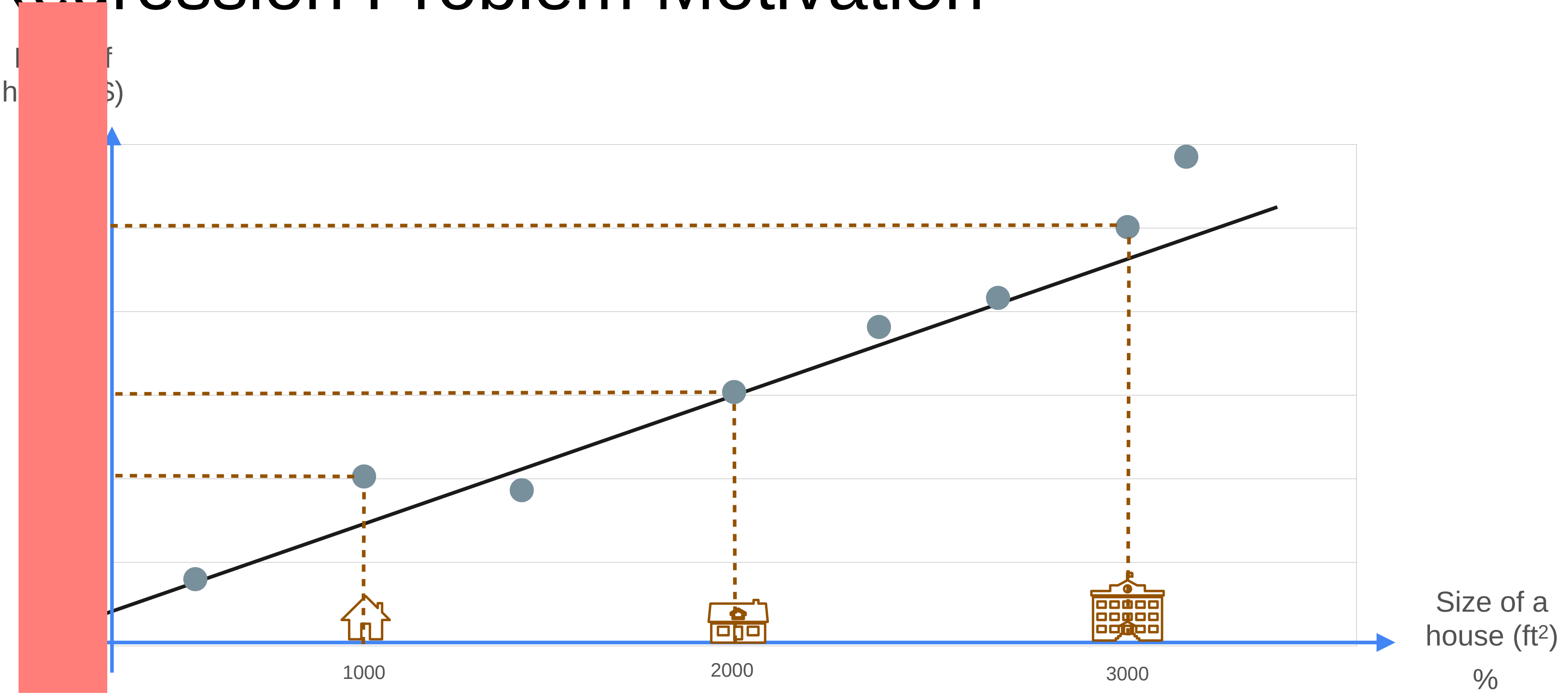
Regression Problem Motivation



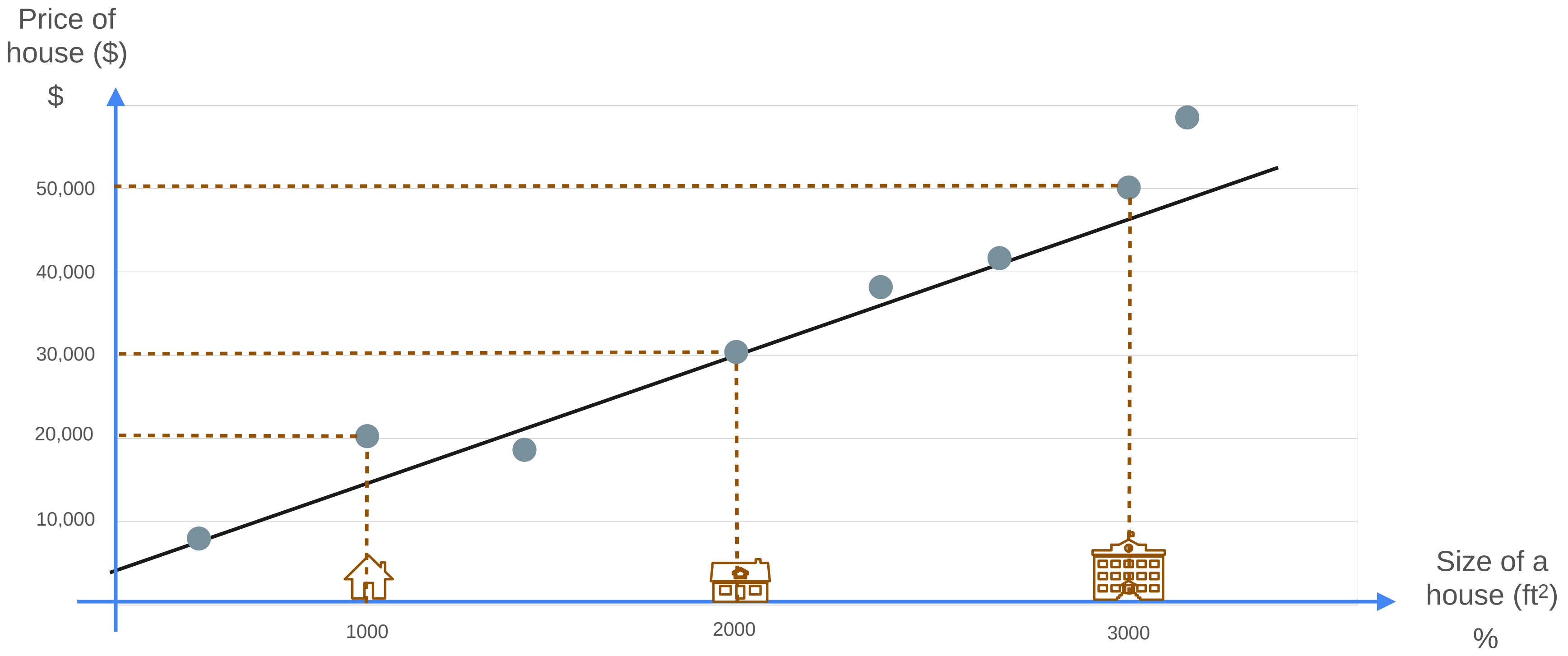
Regression Problem Motivation



Regression Problem Motivation



Regression Problem Motivation



"Regression With a Perceptron

	Size of a house (ft^2) %		Price of house (\$) \$
			
			
			

"Regression With a Perceptron

	Size of a house (ft ^{2}) %		Price of house ($) $
	1000ft^2		$20,000
	2000ft^2		$30,000
	3000ft^2		$50,000

"Regression With a Perceptron

	Size of a house (ft^2) %	Number of rooms &	Price of house ($) $
	$1000ft^2$	2	$20,000
	$2000ft^2$	4	$30,000
	$3000ft^2$	7	$50,000

"Regression With a Perceptron

Inputs

Size of a house (ft^2)

%

Number of rooms

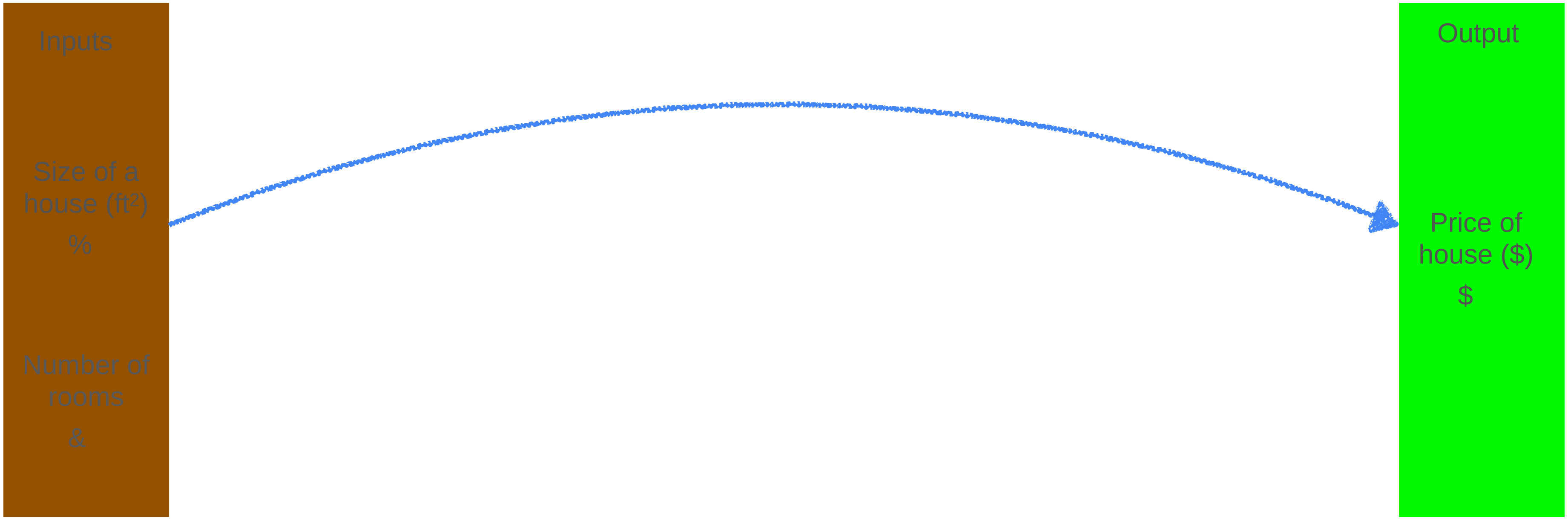
&

Output

Price of house ($)

$

"Regression With a Perceptron



"Regression With a Perceptron

Single Layer Neural Network Perceptron

Inputs

Size of a house (ft^2)

%

Number of rooms

&

Output

Price of house ($)

$

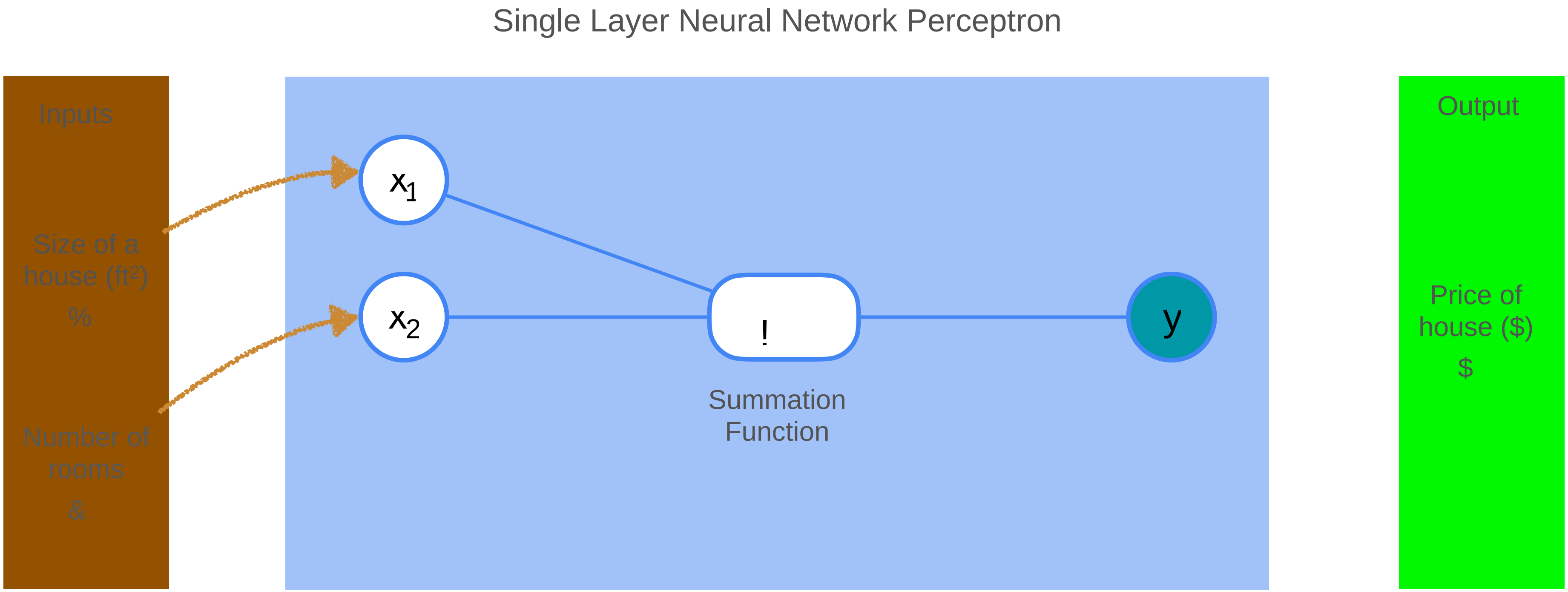
"Regression With a Perceptron

Single Layer Neural Network Perceptron

"Regression With a Perceptron

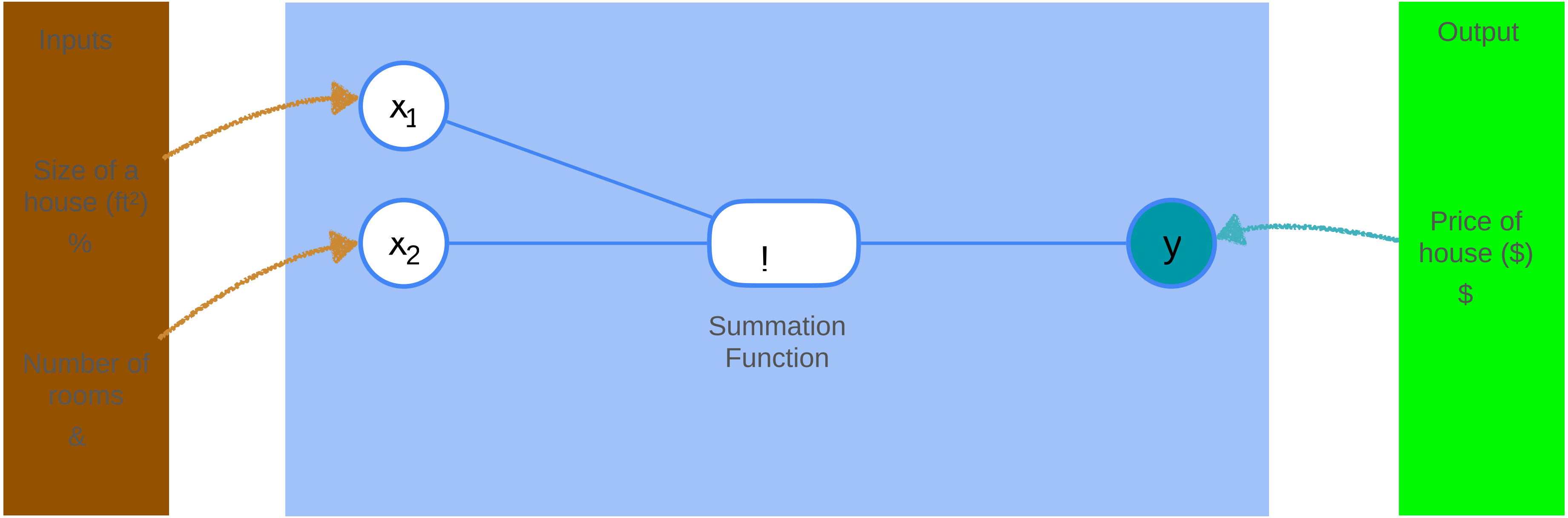
Single Layer Neural Network Perceptron

"Regression With a Perceptron



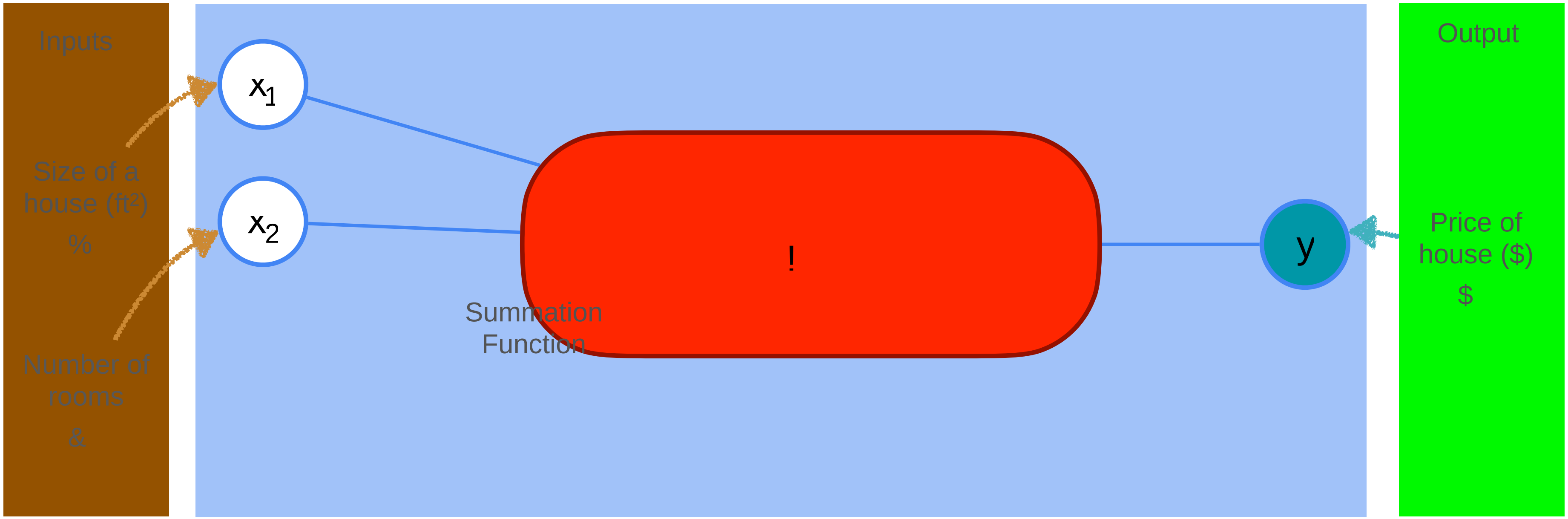
"Regression With a Perceptron

Single Layer Neural Network Perceptron

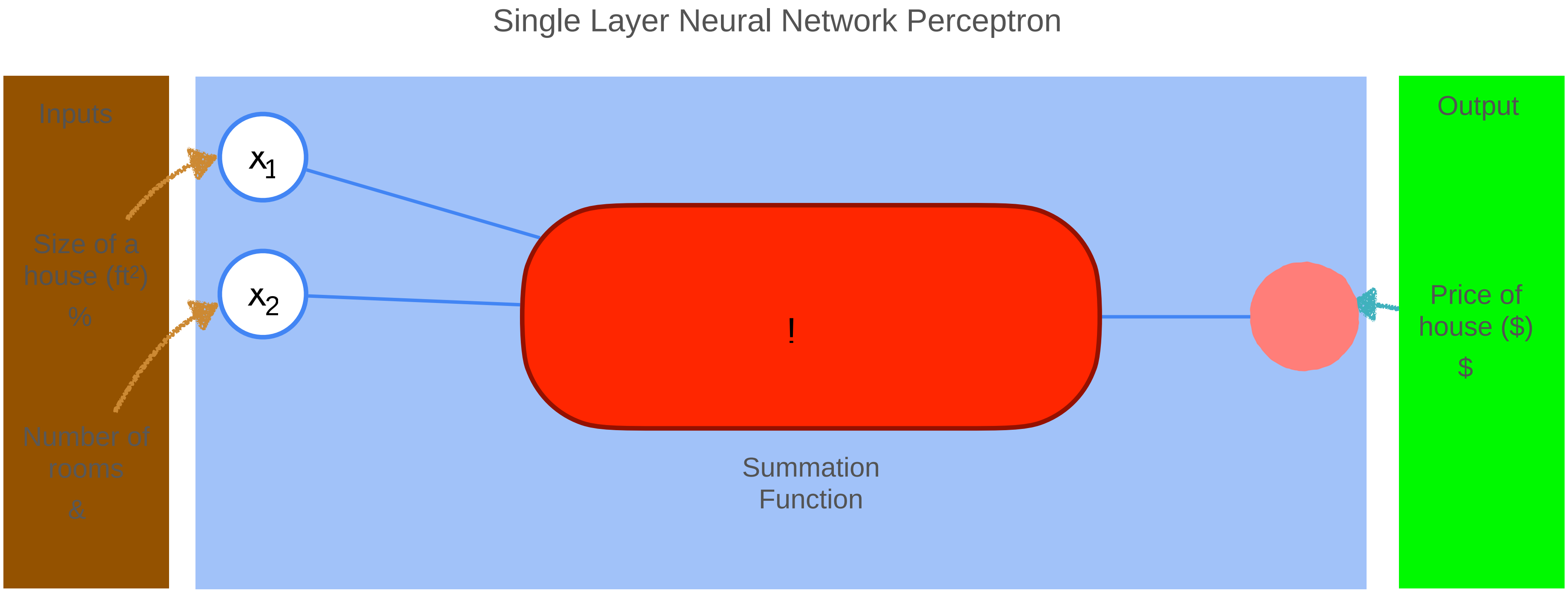


"Regression With a Perceptron

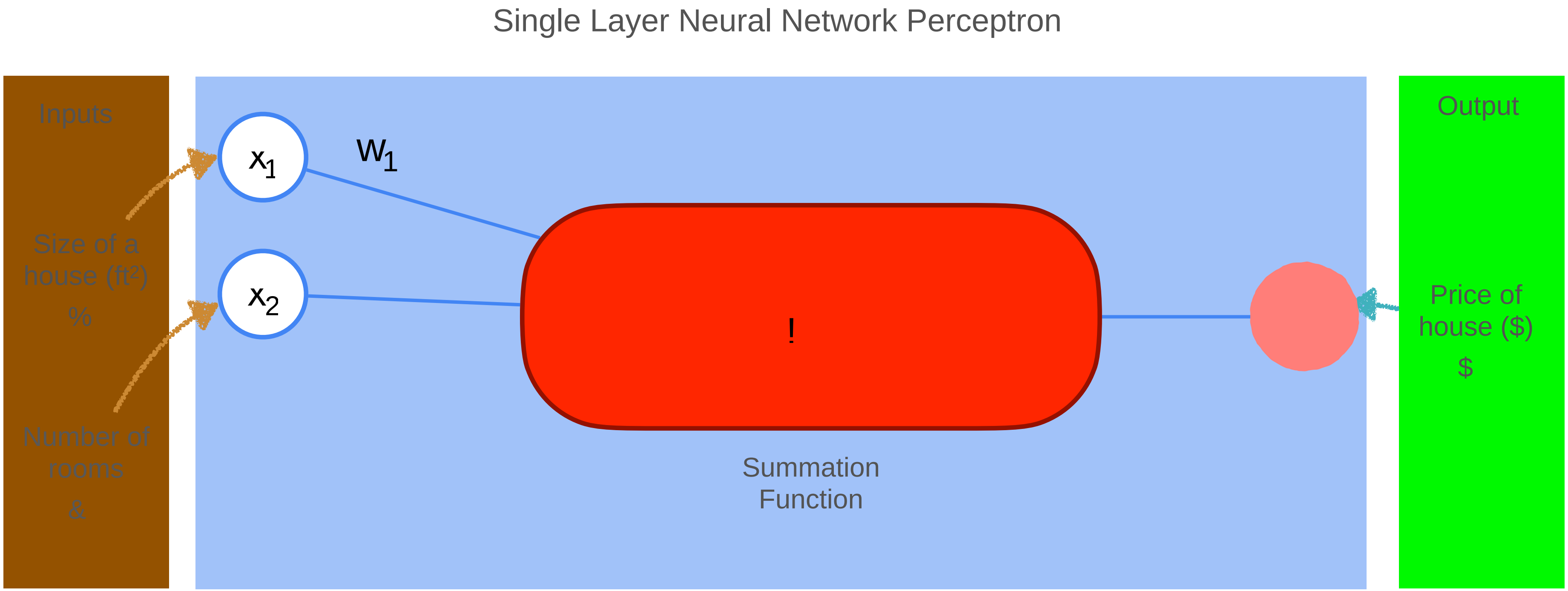
Single Layer Neural Network Perceptron



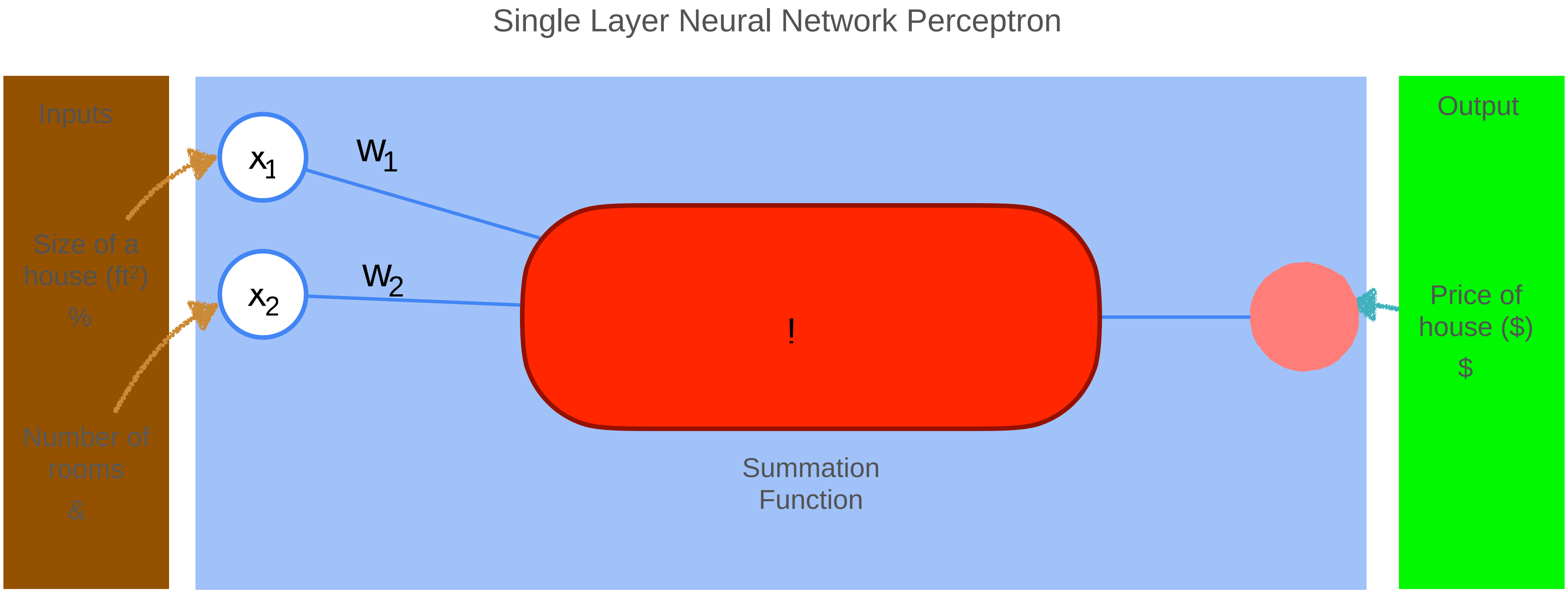
"Regression With a Perceptron



"Regression With a Perceptron

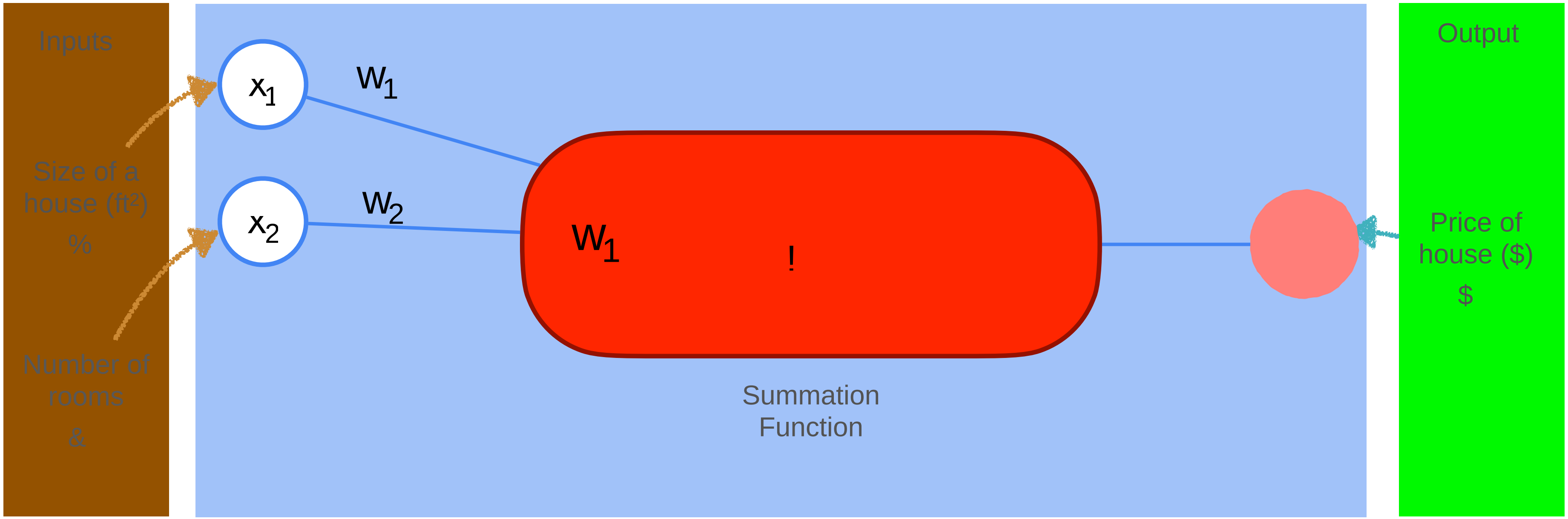


"Regression With a Perceptron

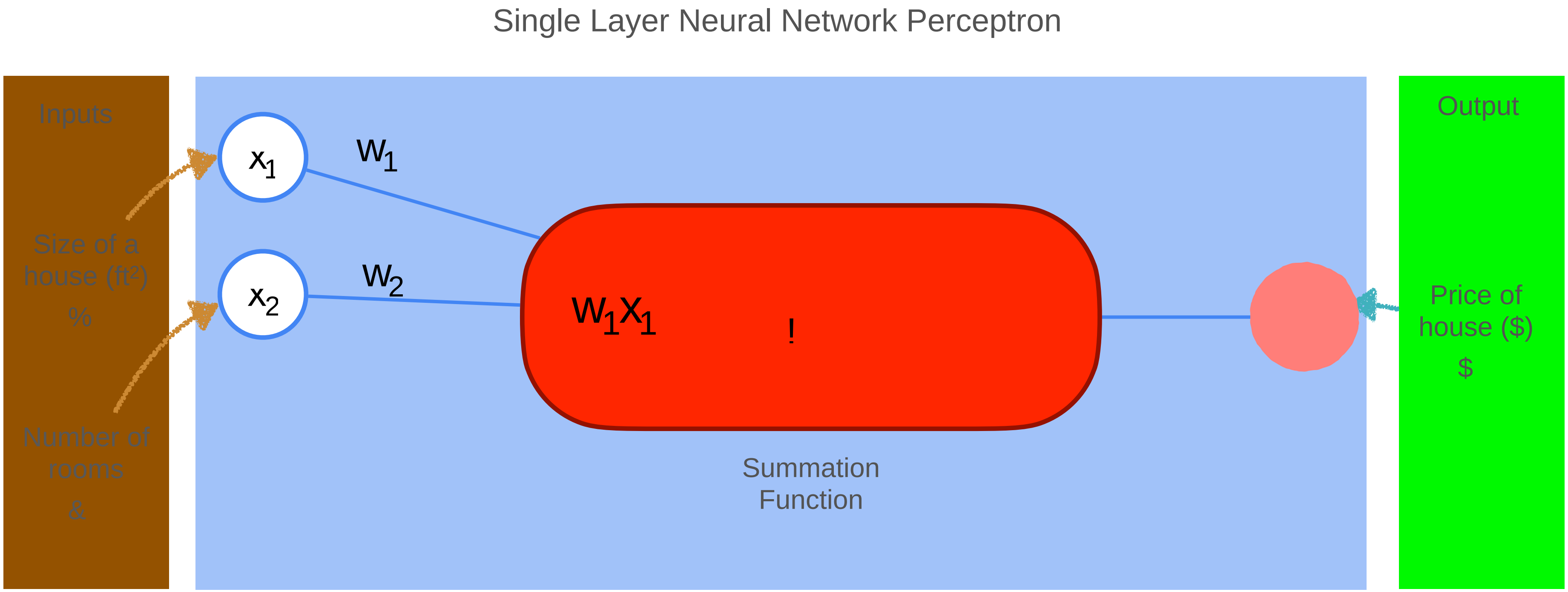


"Regression With a Perceptron

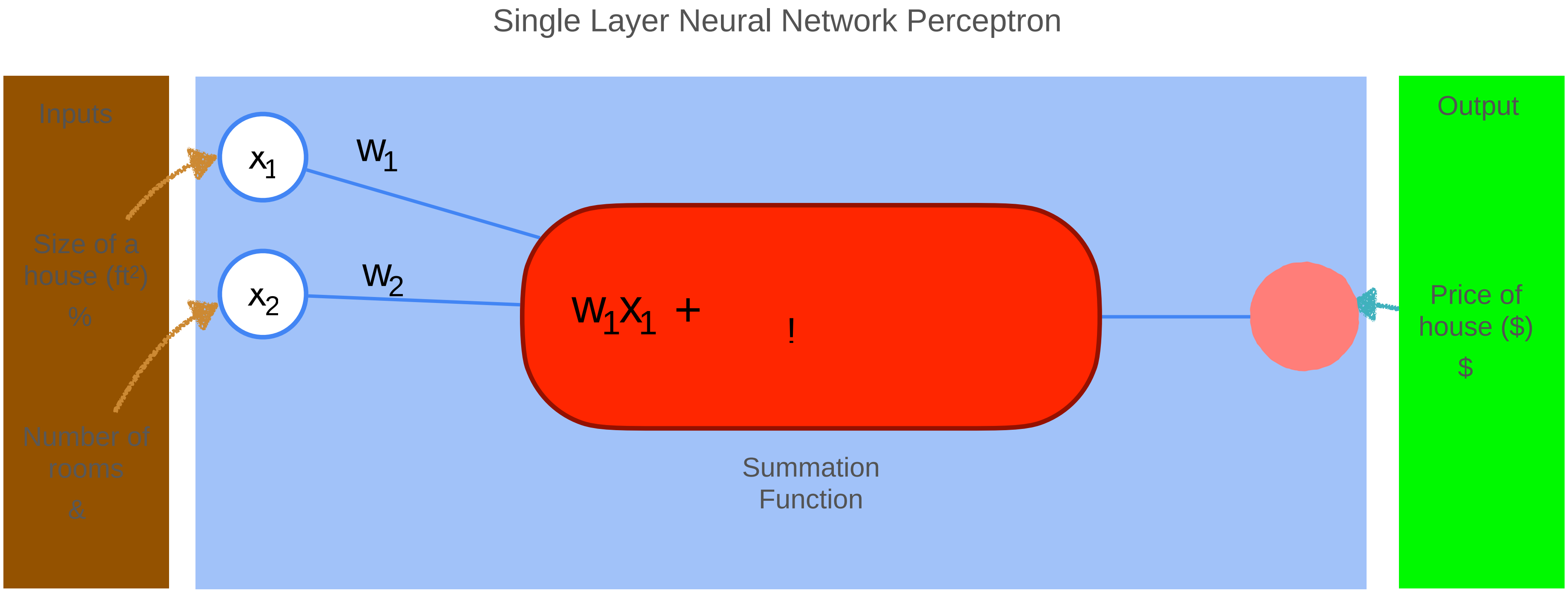
Single Layer Neural Network Perceptron



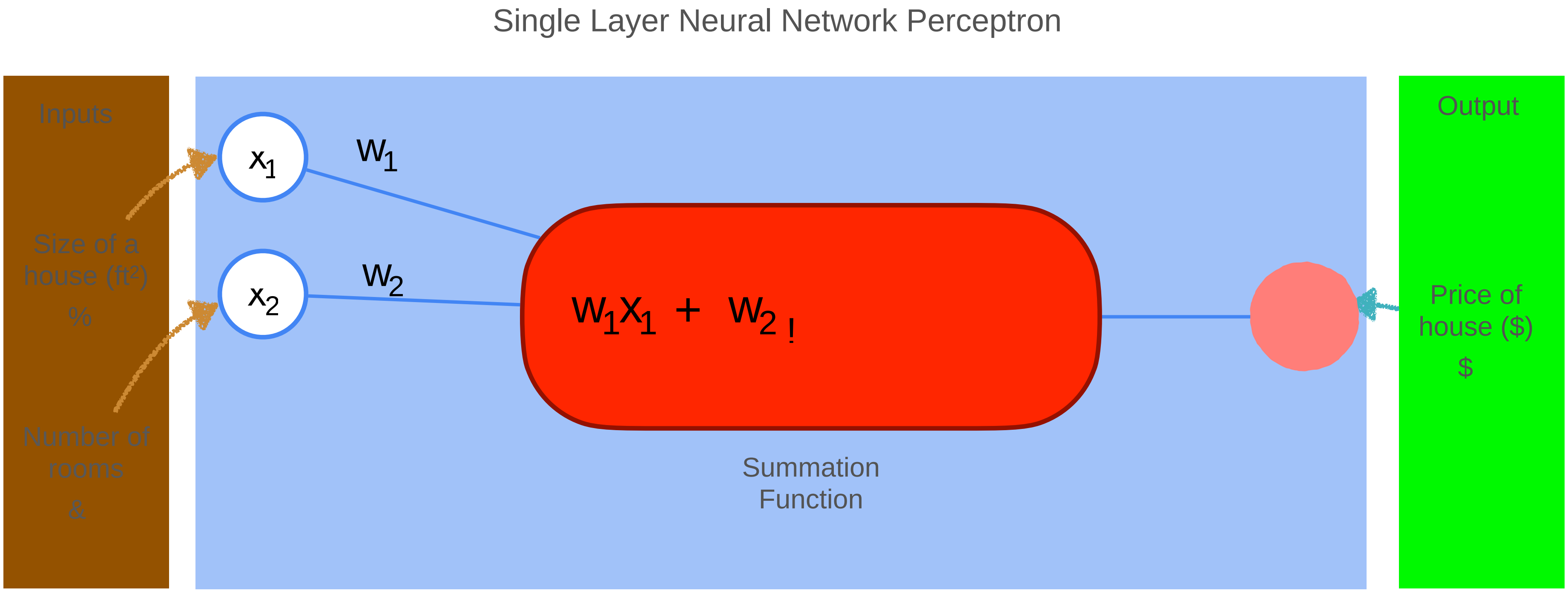
"Regression With a Perceptron



"Regression With a Perceptron

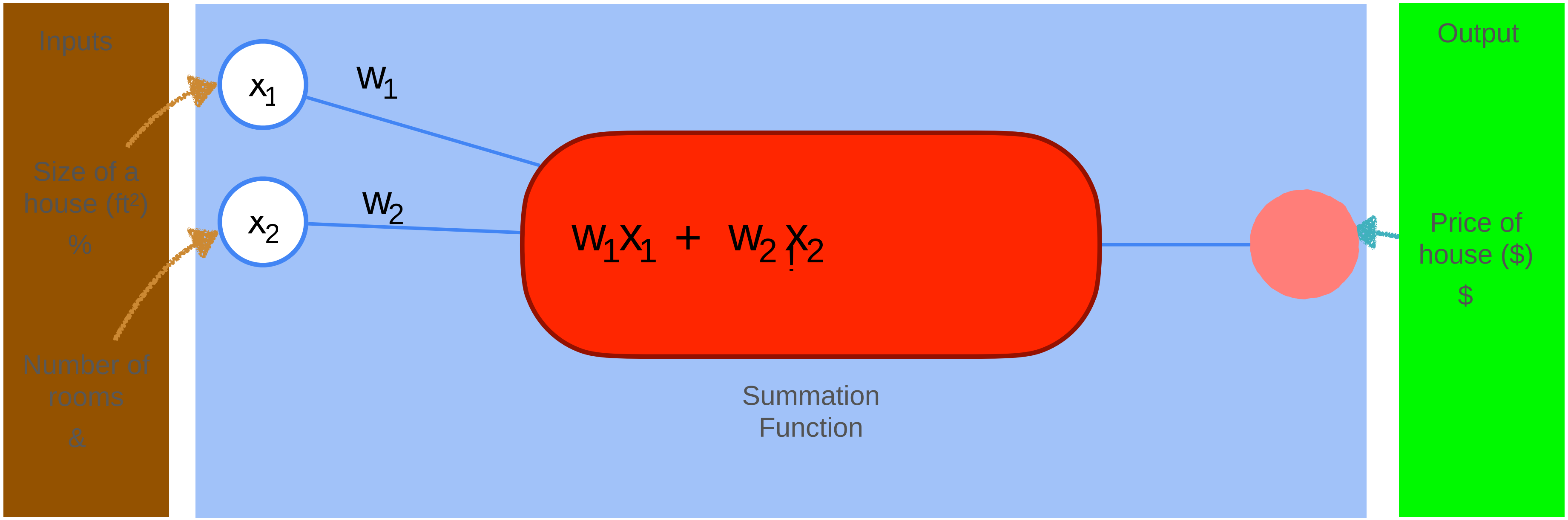


"Regression With a Perceptron



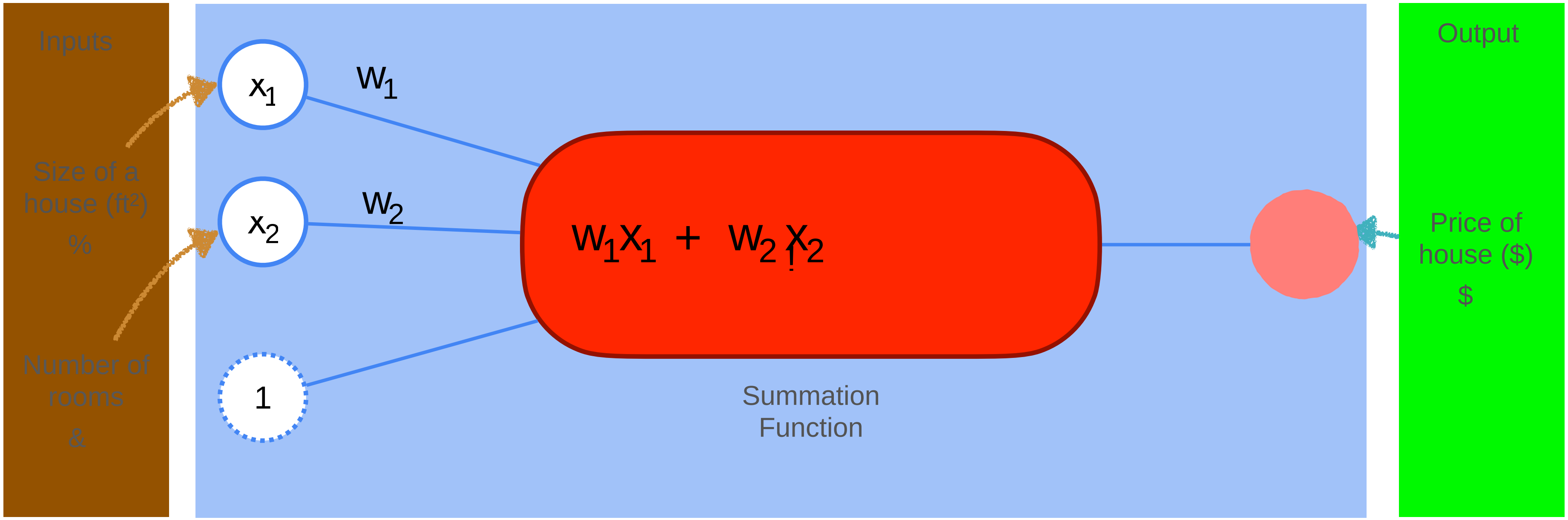
"Regression With a Perceptron

Single Layer Neural Network Perceptron



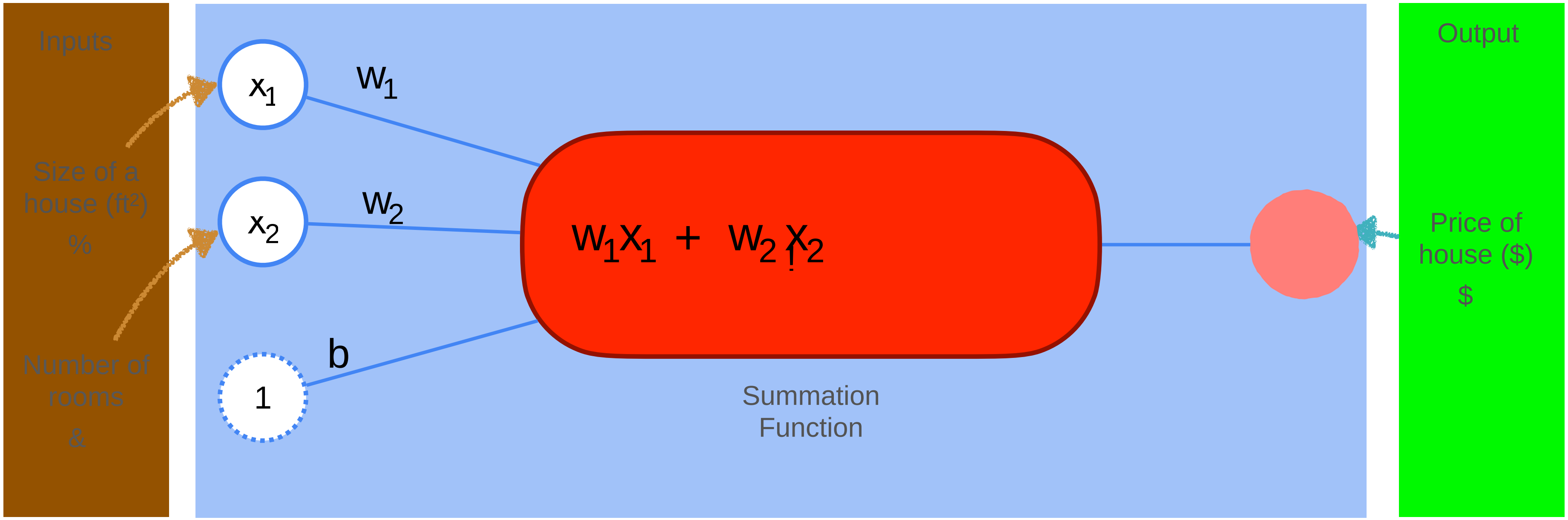
"Regression With a Perceptron

Single Layer Neural Network Perceptron



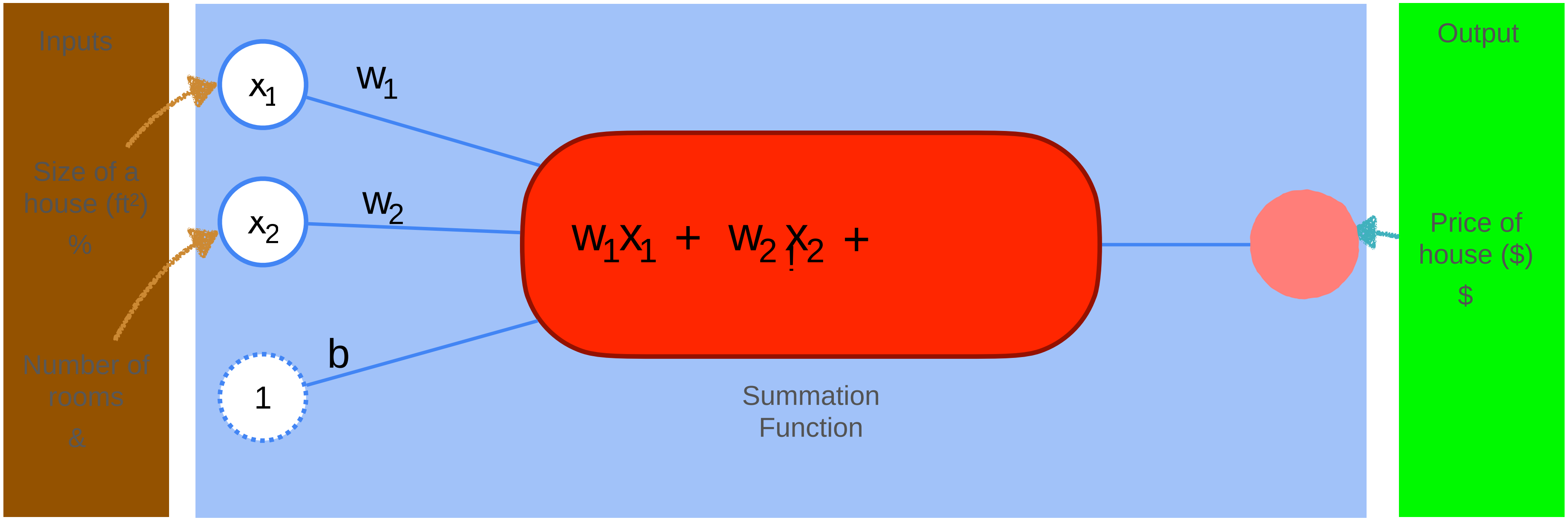
"Regression With a Perceptron

Single Layer Neural Network Perceptron



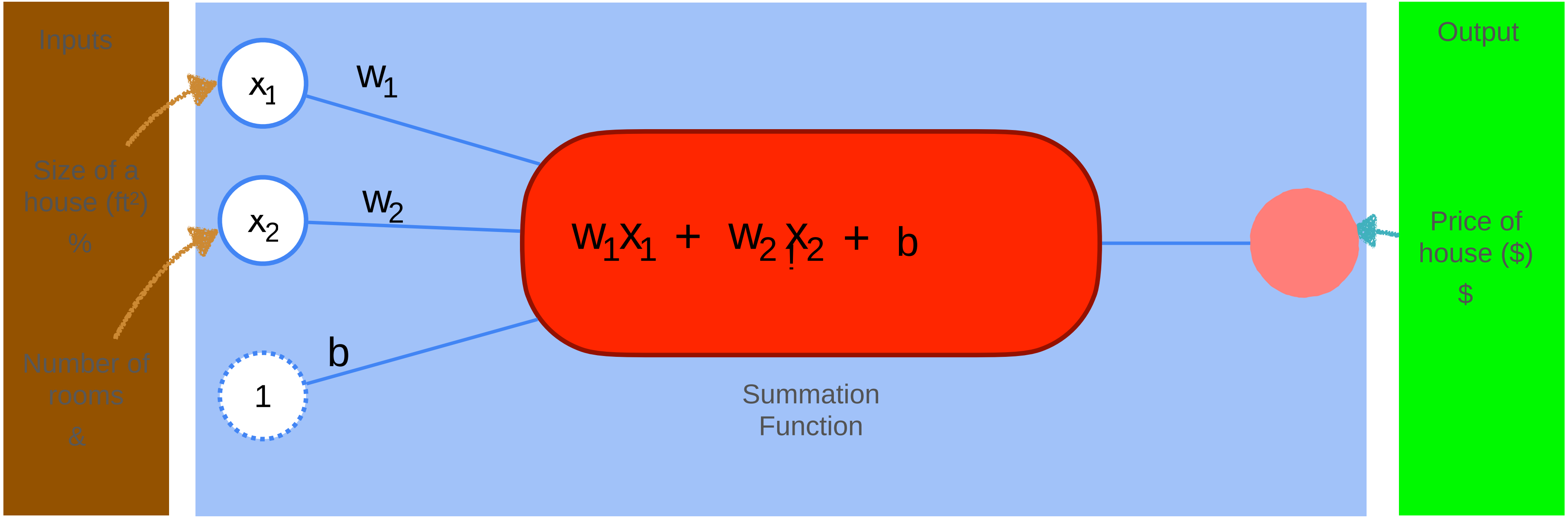
"Regression With a Perceptron

Single Layer Neural Network Perceptron



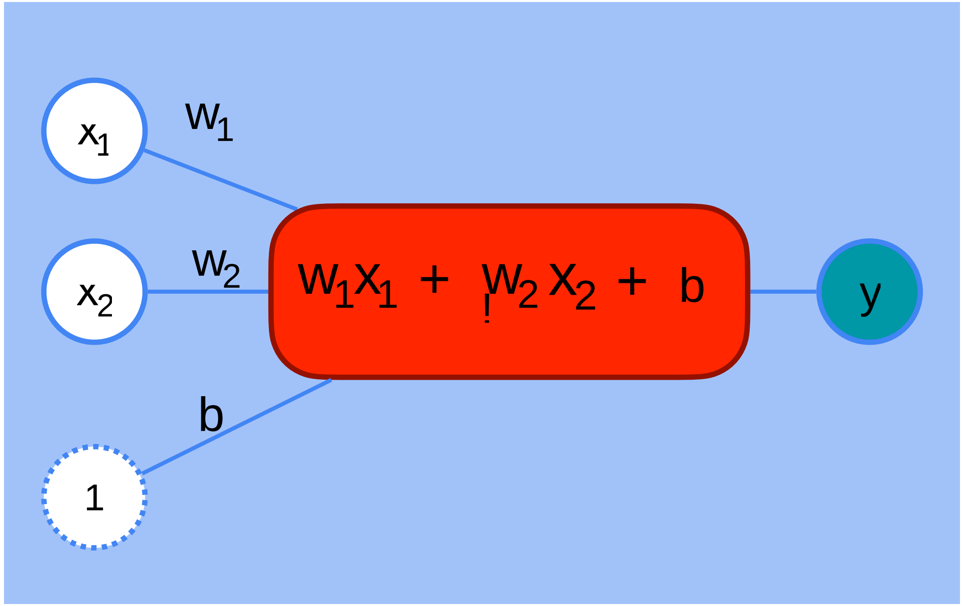
"Regression With a Perceptron

Single Layer Neural Network Perceptron



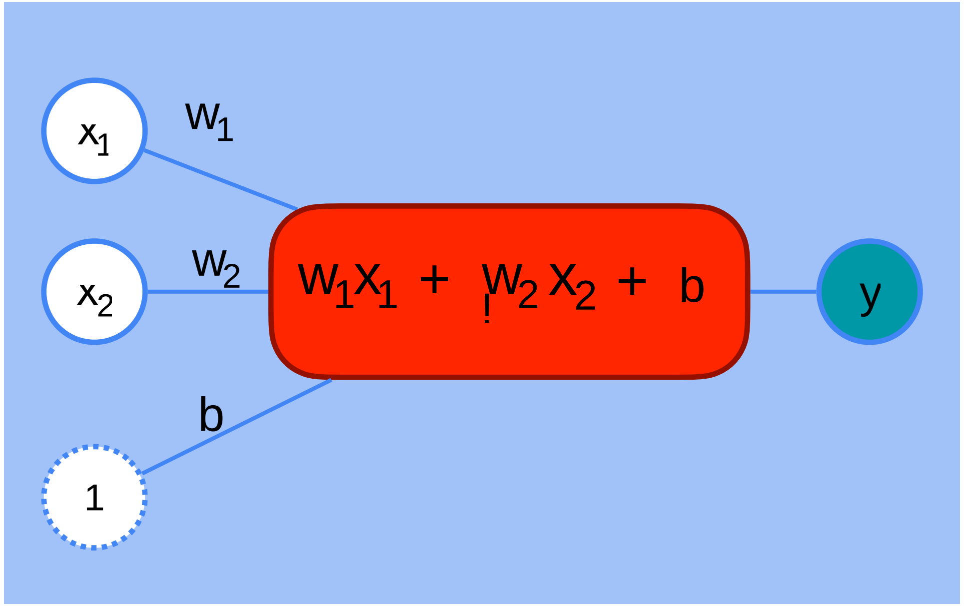
"Regression With a Perceptron

Single Layer Neural Network Perceptron



"Regression With a Perceptron

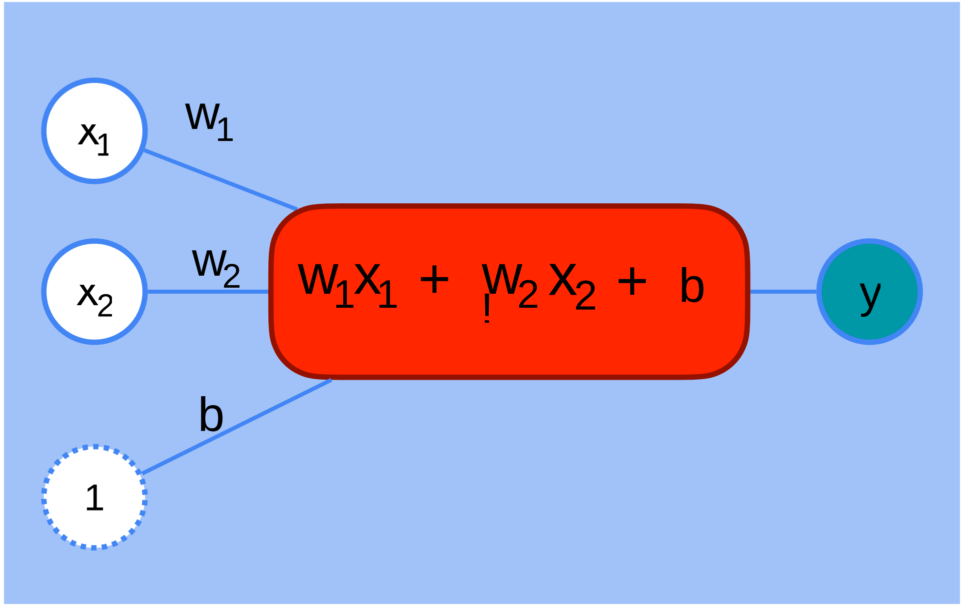
Single Layer Neural Network Perceptron



y1

"Regression With a Perceptron

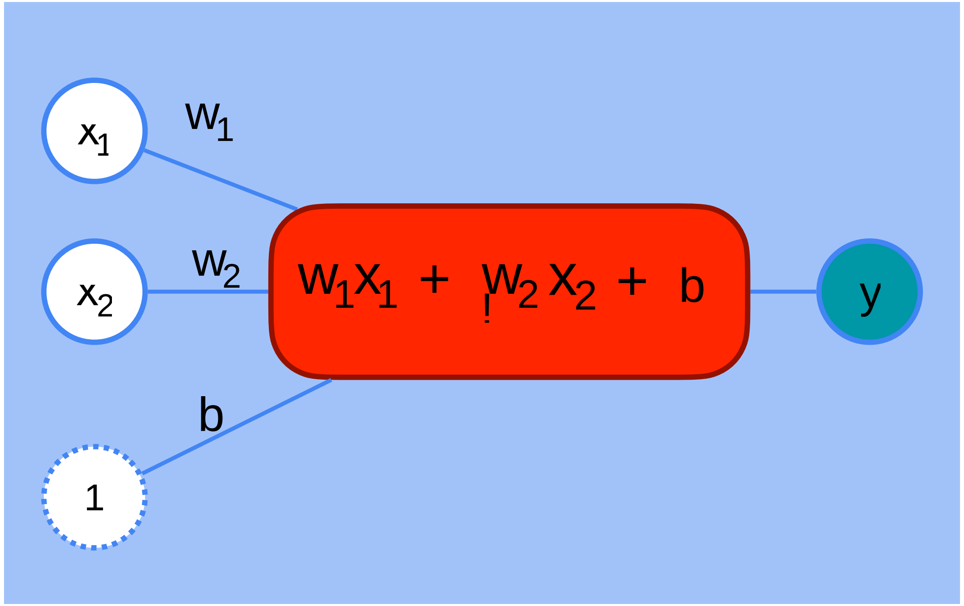
Single Layer Neural Network Perceptron



$$y1= w_1x_1 + w_2x_2 + b$$

"Regression With a Perceptron

Single Layer Neural Network Perceptron

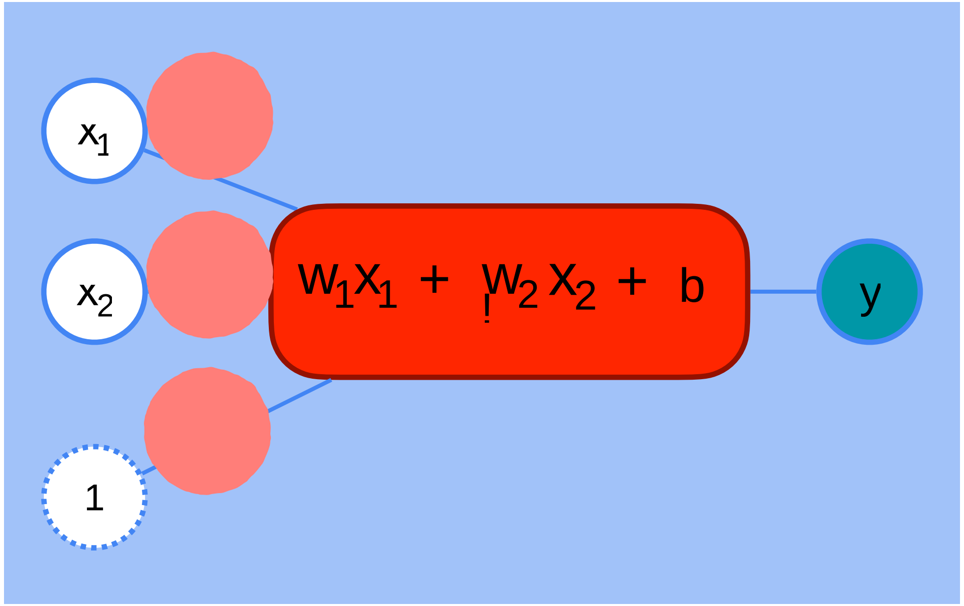


$$y1= w_1x_1 + w_2x_2 + b$$

Main Goal:

"Regression With a Perceptron

Single Layer Neural Network Perceptron

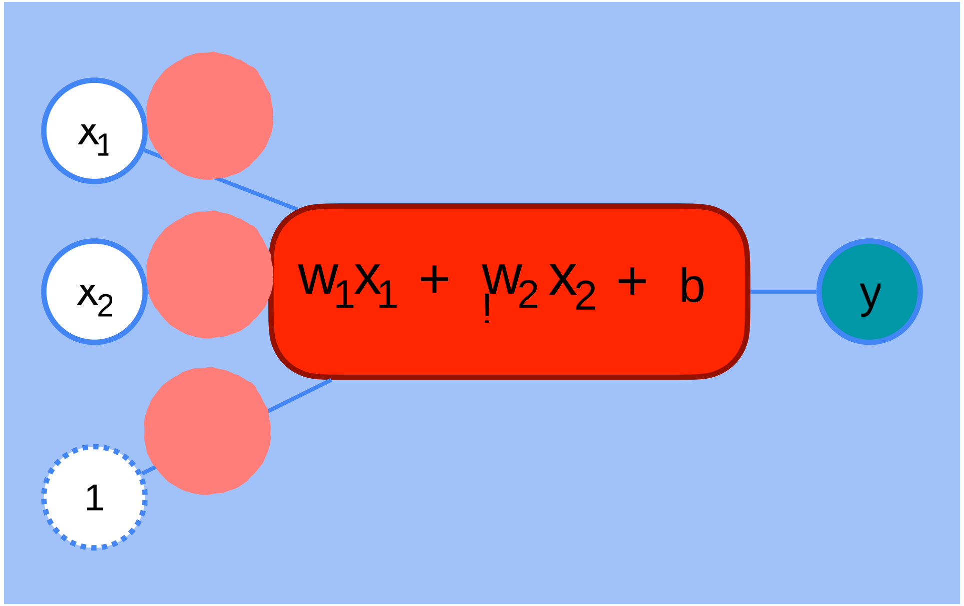


$$y1 = \quad x_1 \ + \quad x_2 \ + $$

Main Goal:

"Regression With a Perceptron

Single Layer Neural Network Perceptron



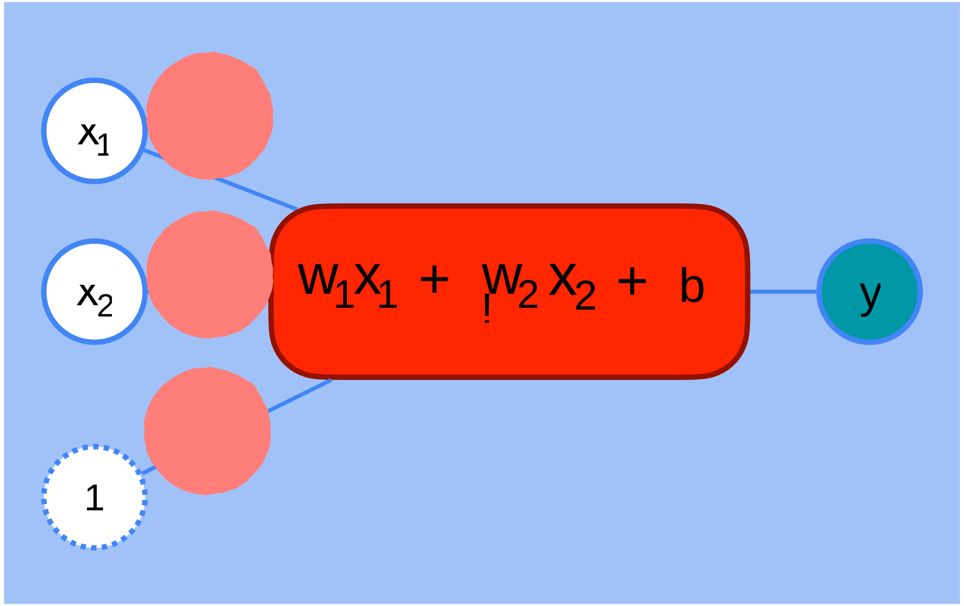
$y1 = \quad x_1 + \quad x_2 +$

Main Goal:

Find weights and bias that will optimise the predictions.

"Regression With a Perceptron

Single Layer Neural Network Perceptron



$$y1 = \quad x_1 + \quad x_2 +$$

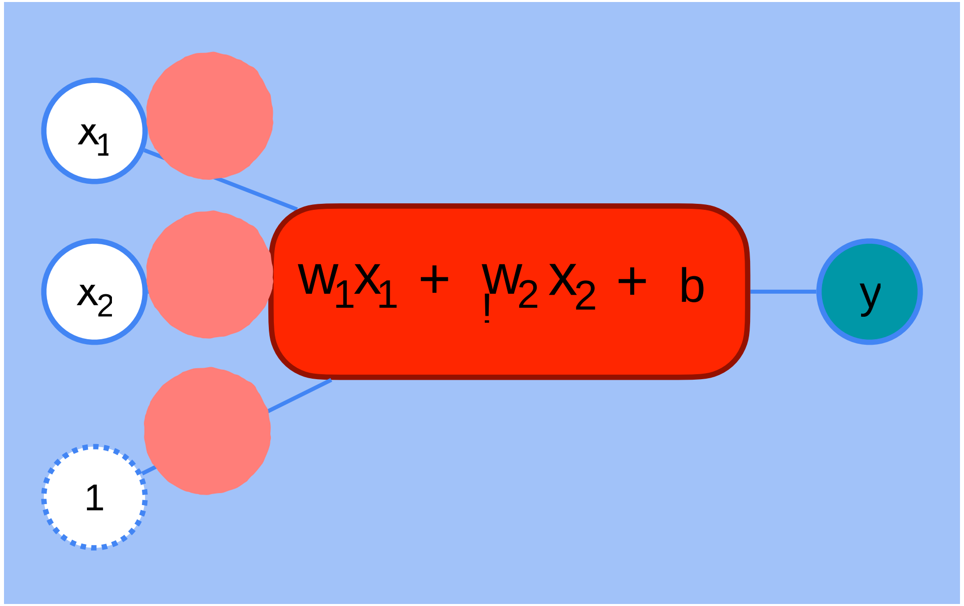
Main Goal:

Find weights and bias that will optimise the predictions.

ie. Reduce the errors in the predictions

"Regression With a Perceptron

Single Layer Neural Network Perceptron



$y1 = \quad x_1 + \quad x_2 + $

Main Goal:

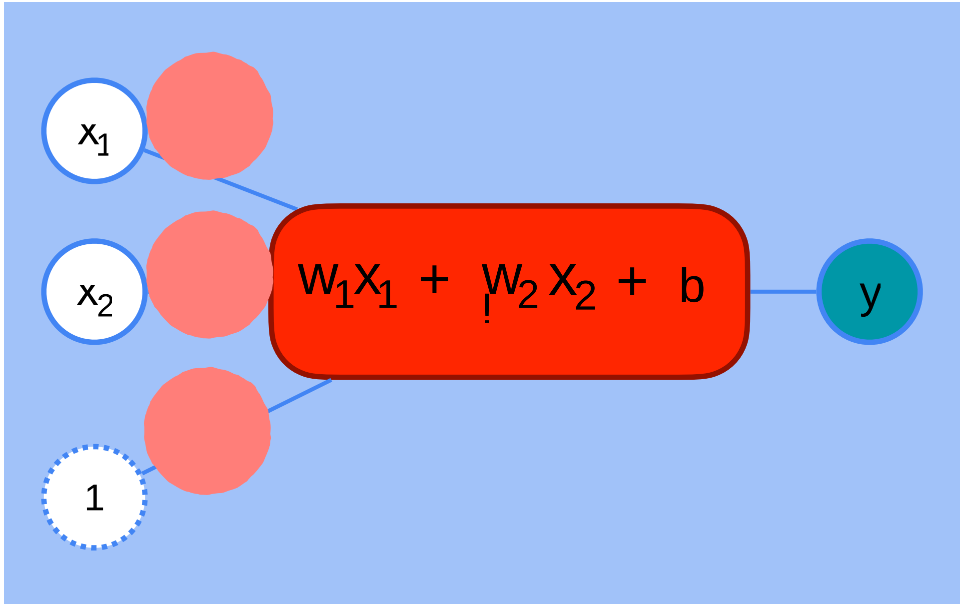
Find weights and bias that will optimise the predictions.

ie. Reduce the errors in the predictions

'

"Regression With a Perceptron

Single Layer Neural Network Perceptron



$y1 = \;\; x_1 \; + \;\; x_2 \; +$

Main Goal:

Find weights and bias that will optimise the predictions.

ie. Reduce the errors in the predictions

'

The Loss Function

Optimization in Neural Networks and NewtonÕs Method

Regression with a perceptron: Loss function

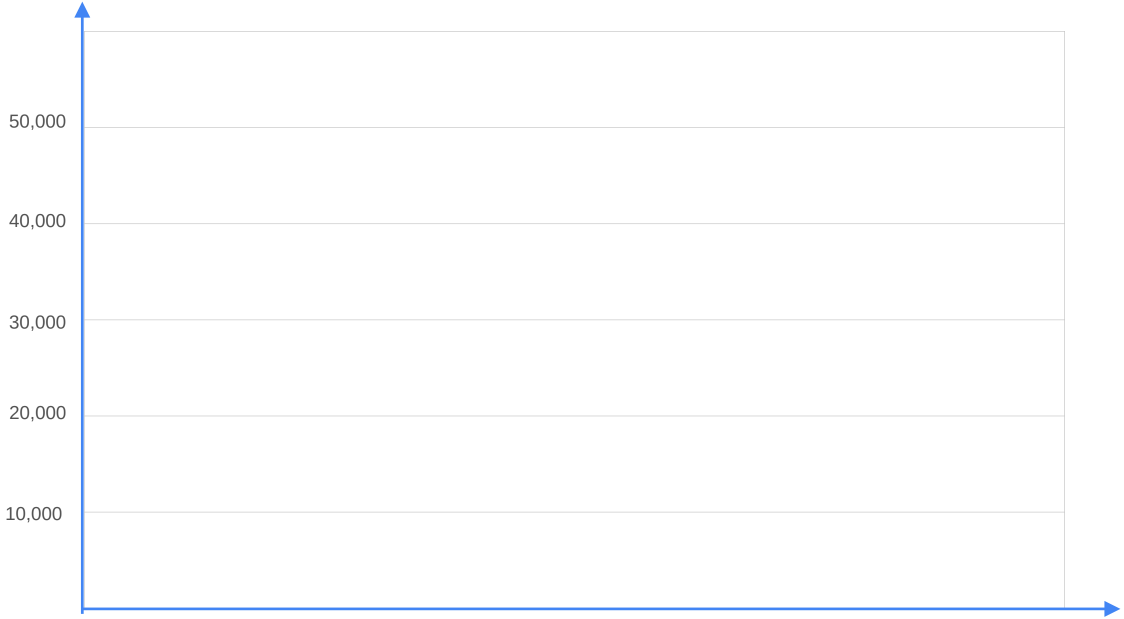
Mean Squared Error

Mean Squared Error

	y		
	\$20,000		
	\$30,000		
	\$50,000		

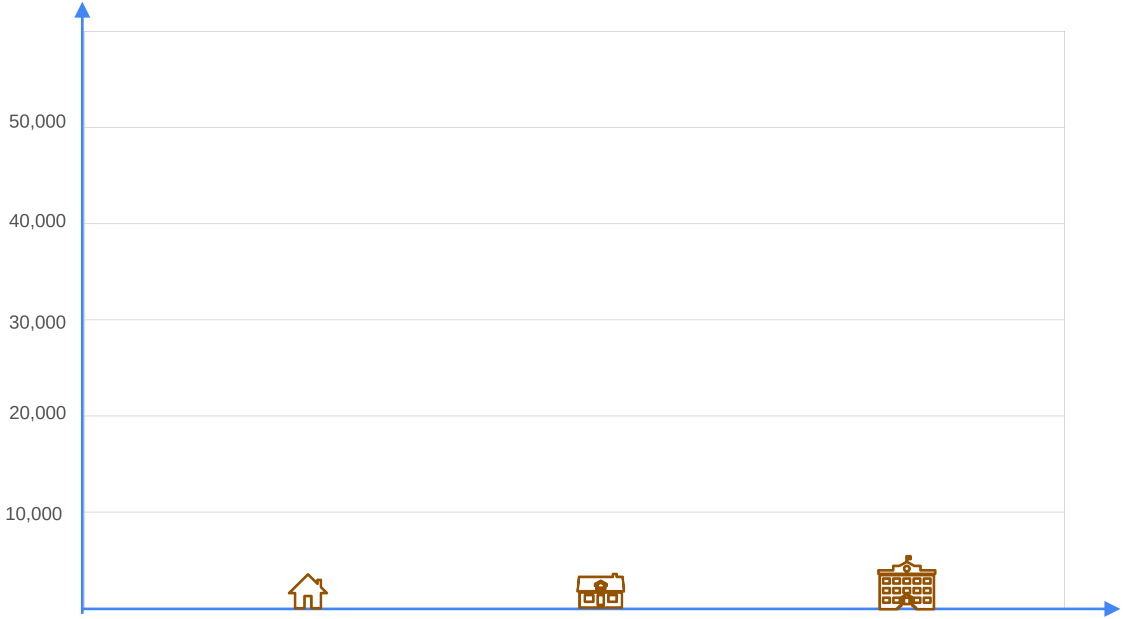
Mean Squared Error

	y		
	$20,000		
	$30,000		
	$50,000		



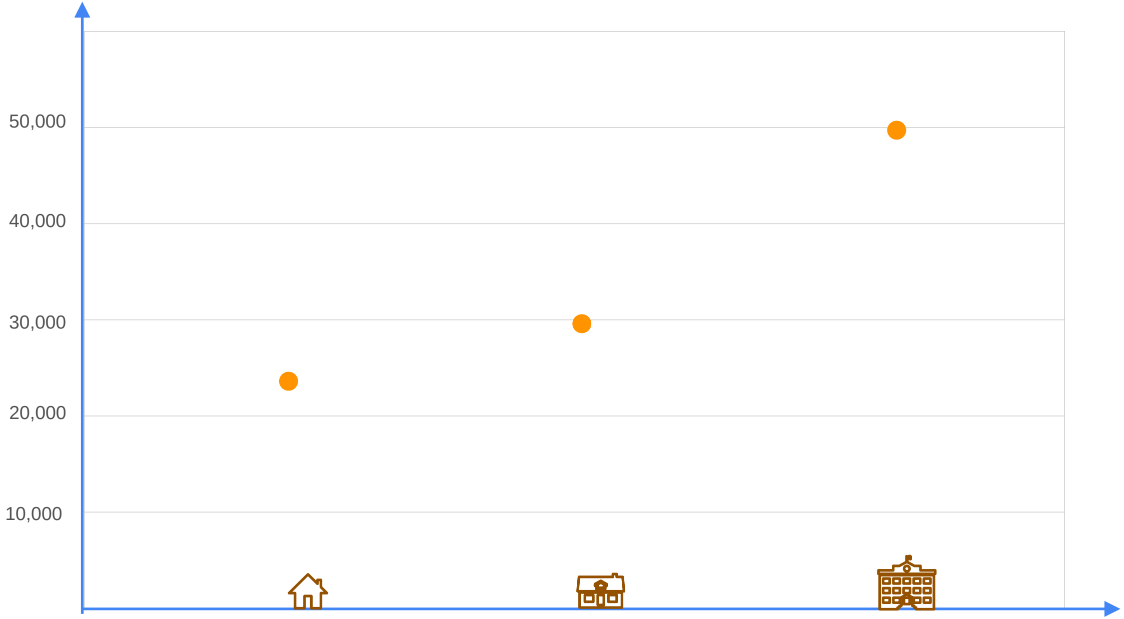
Mean Squared Error

	y		
	$20,000		
	$30,000		
	$50,000		



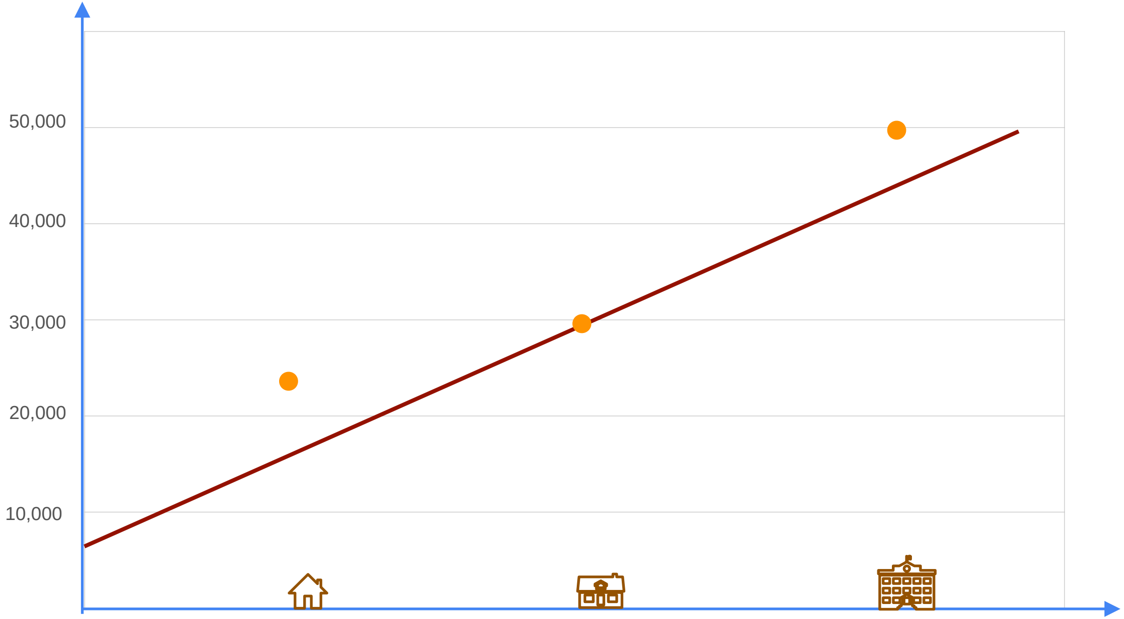
Mean Squared Error

	y		
	$20,000		
	$30,000		
	$50,000		



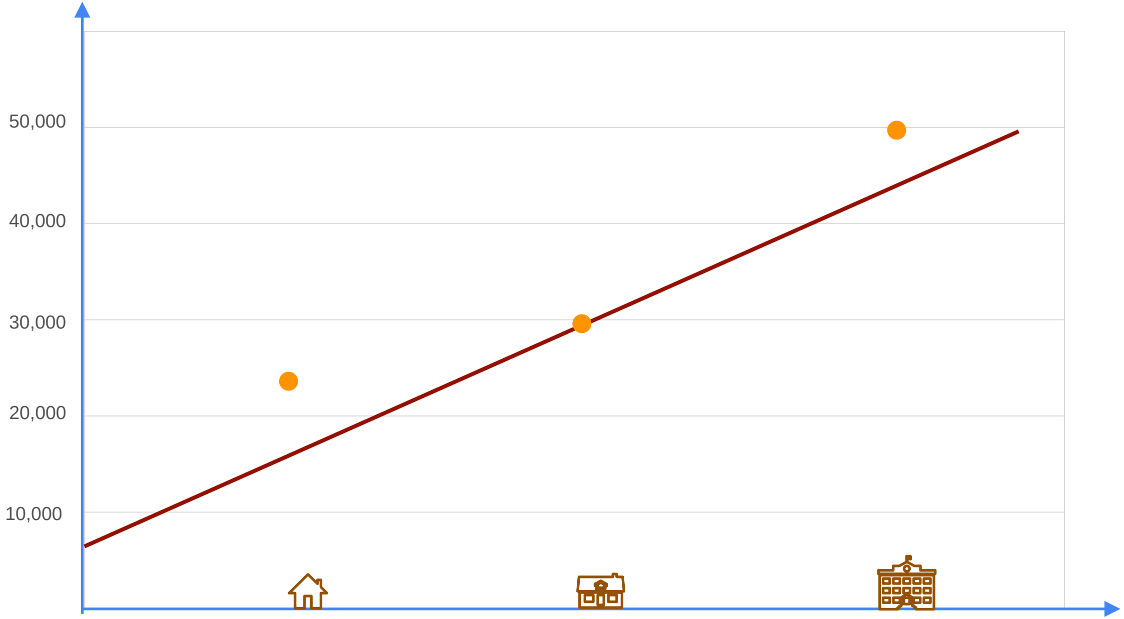
Mean Squared Error

	y		
	$20,000		
	$30,000		
	$50,000		



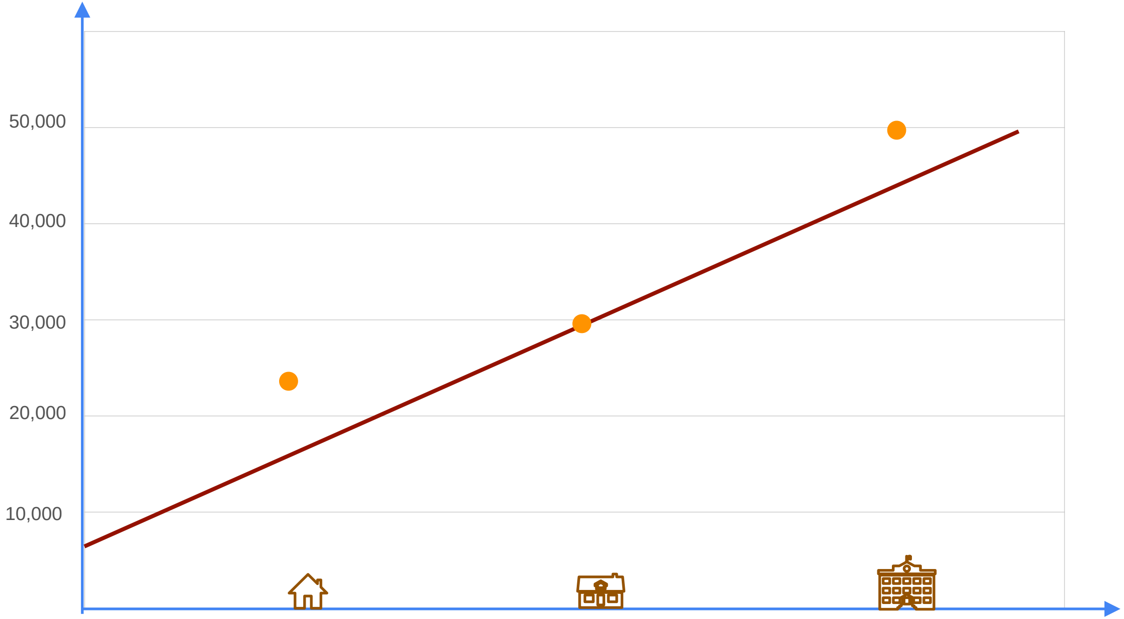
Mean Squared Error

	y	y	
	$20,000	$15,000	
	$30,000	$30,000	
	$50,000	$45,000	



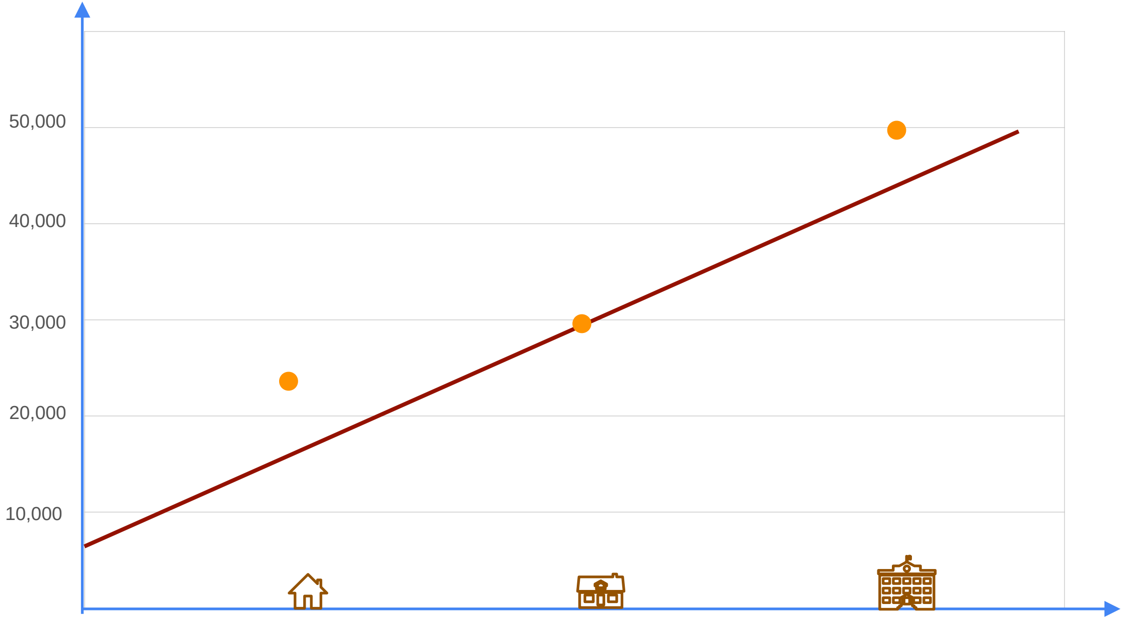
Mean Squared Error

	y	y	y" y
	$20,000	$15,000	
	$30,000	$30,000	
	$50,000	$45,000	



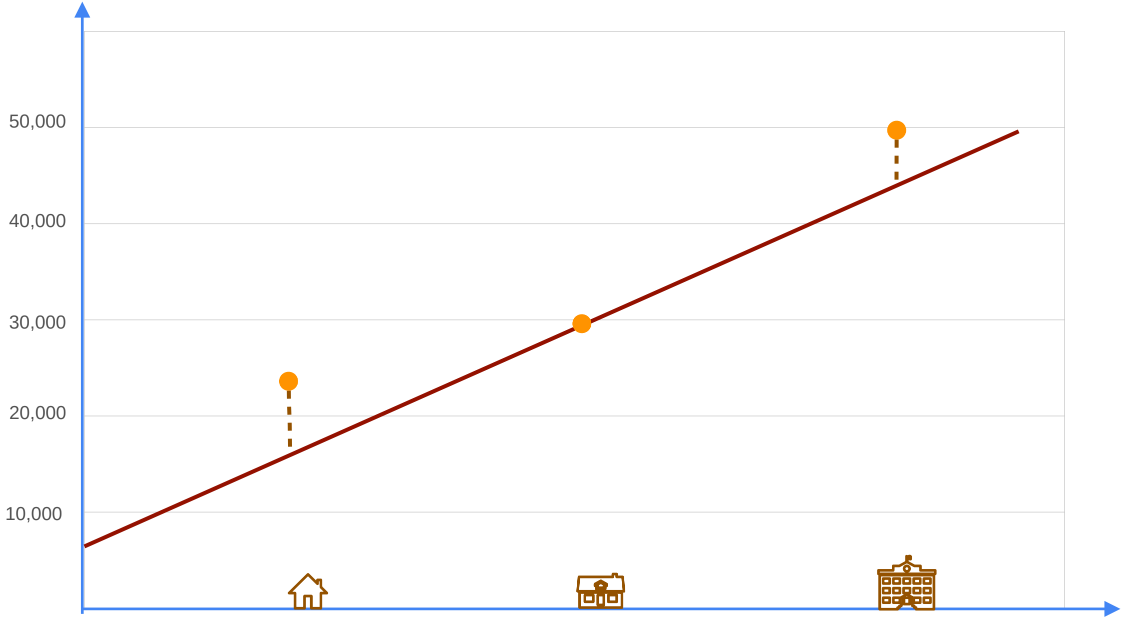
Mean Squared Error

	y	y	y " y
	$20,000	$15,000	Error
	$30,000	$30,000	Error
	$50,000	$45,000	Error



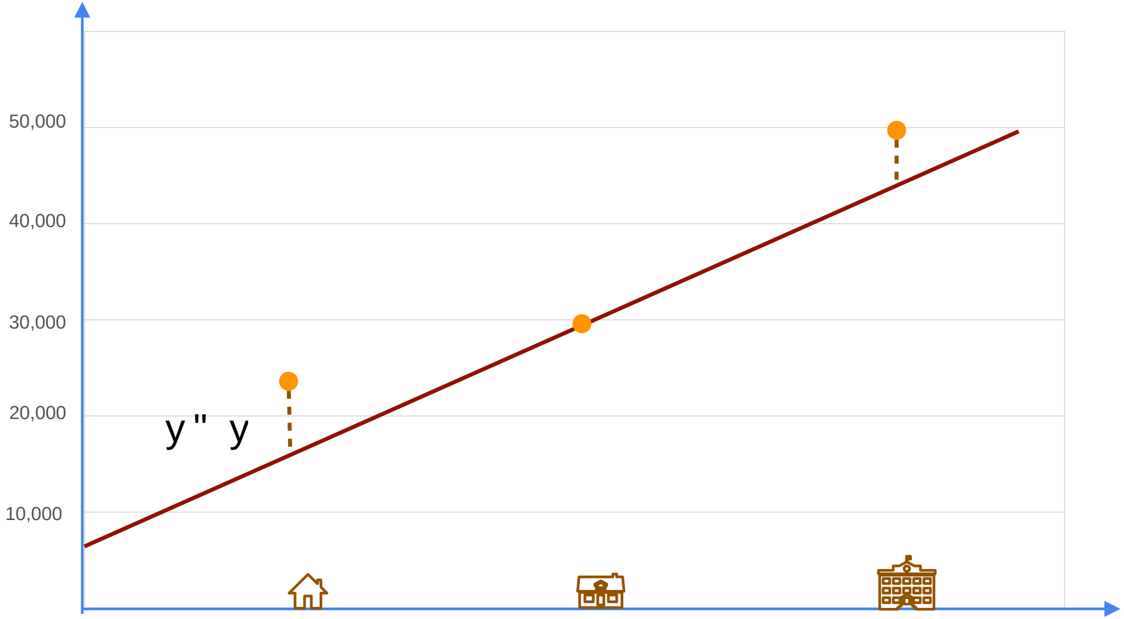
Mean Squared Error

	y	y	y " y
	$20,000	$15,000	Error
	$30,000	$30,000	Error
	$50,000	$45,000	Error



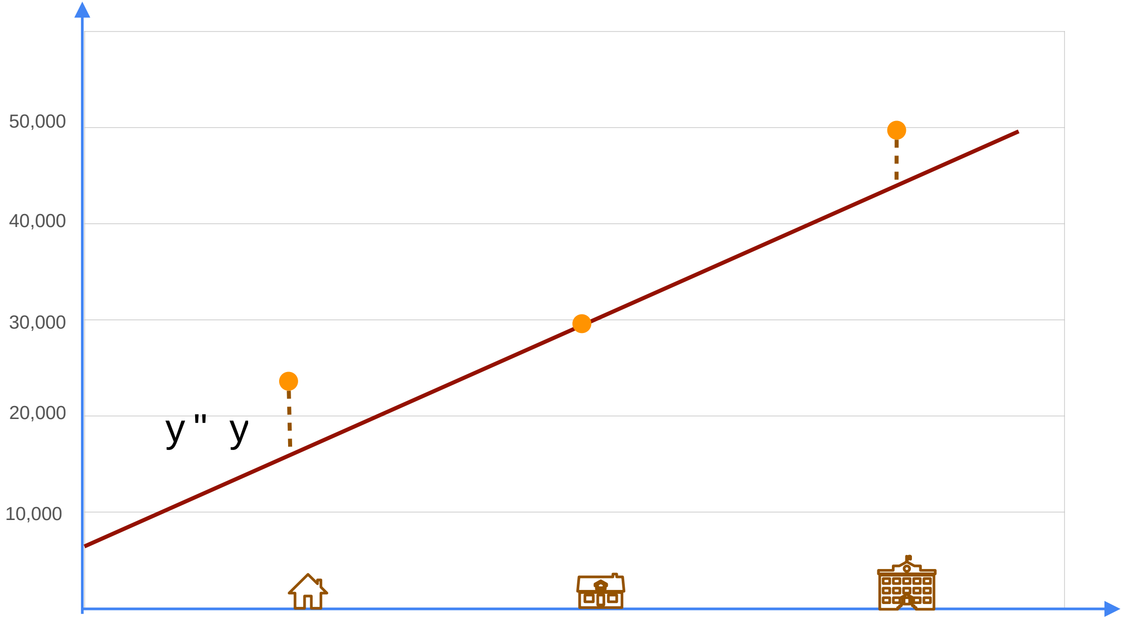
Mean Squared Error

	y	y	y " y
	$20,000	$15,000	Error
	$30,000	$30,000	Error
	$50,000	$45,000	Error



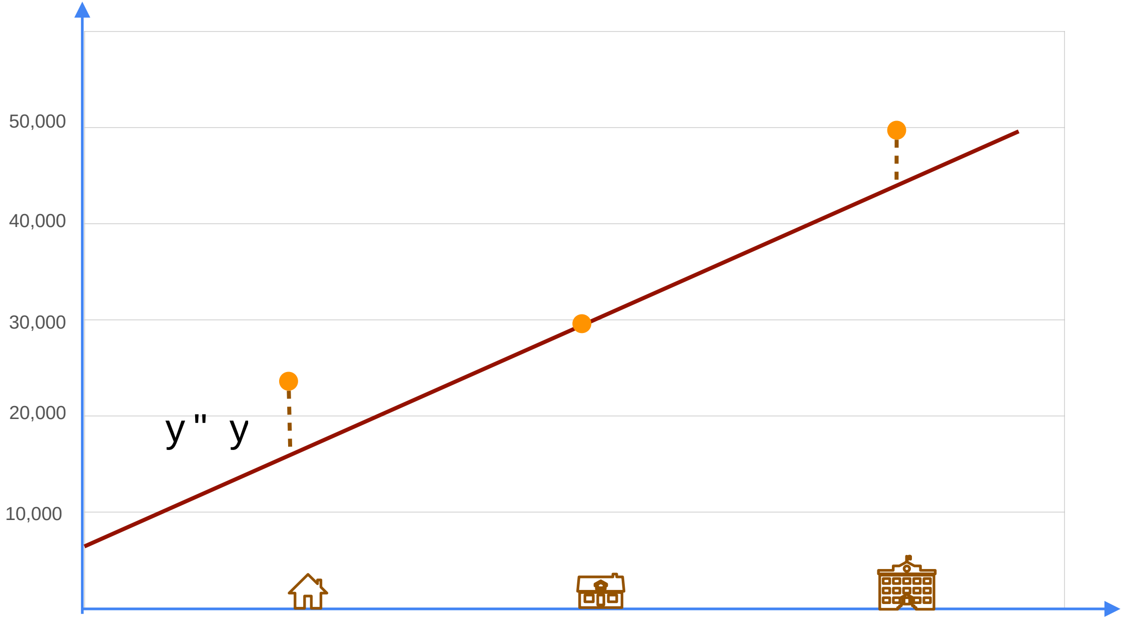
Mean Squared Error

	y	y	(y" y)²
	$20,000	$15,000	Error
	$30,000	$30,000	Error
	$50,000	$45,000	Error



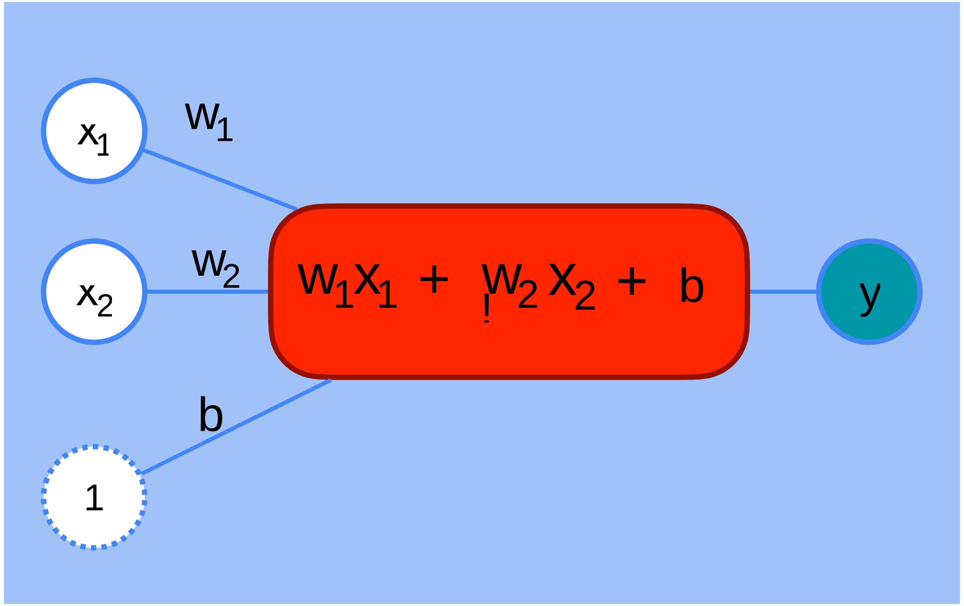
Mean Squared Error

	y	y	$\frac{1}{2}(y"\ y)^2$
	$20,000	$15,000	Error
	$30,000	$30,000	Error
	$50,000	$45,000	Error



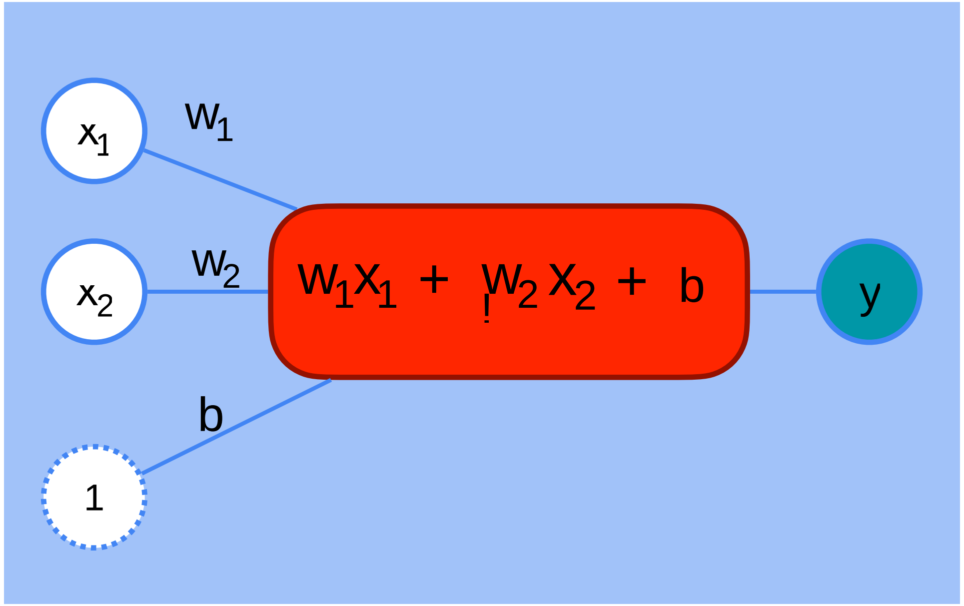
"Regression With a Perceptron

Single Layer Neural Network Perceptron



"Regression With a Perceptron

Single Layer Neural Network Perceptron

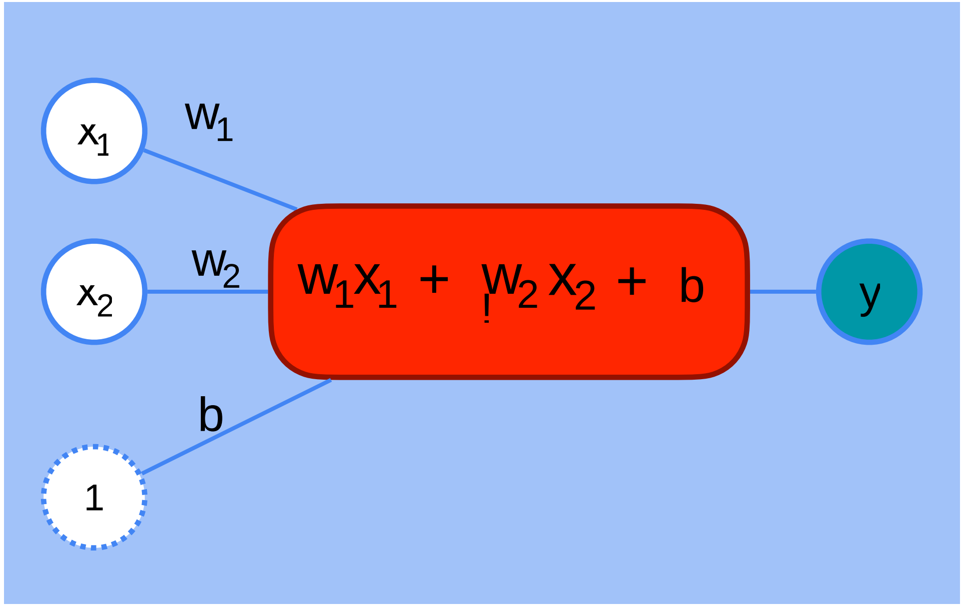


Prediction Function:

y1

"Regression With a Perceptron

Single Layer Neural Network Perceptron

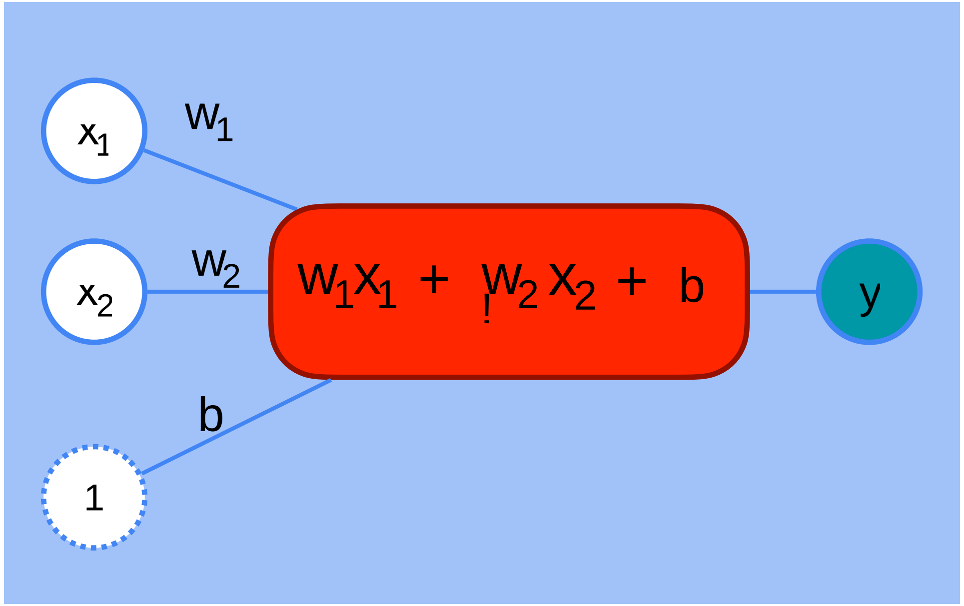


Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

"Regression With a Perceptron

Single Layer Neural Network Perceptron



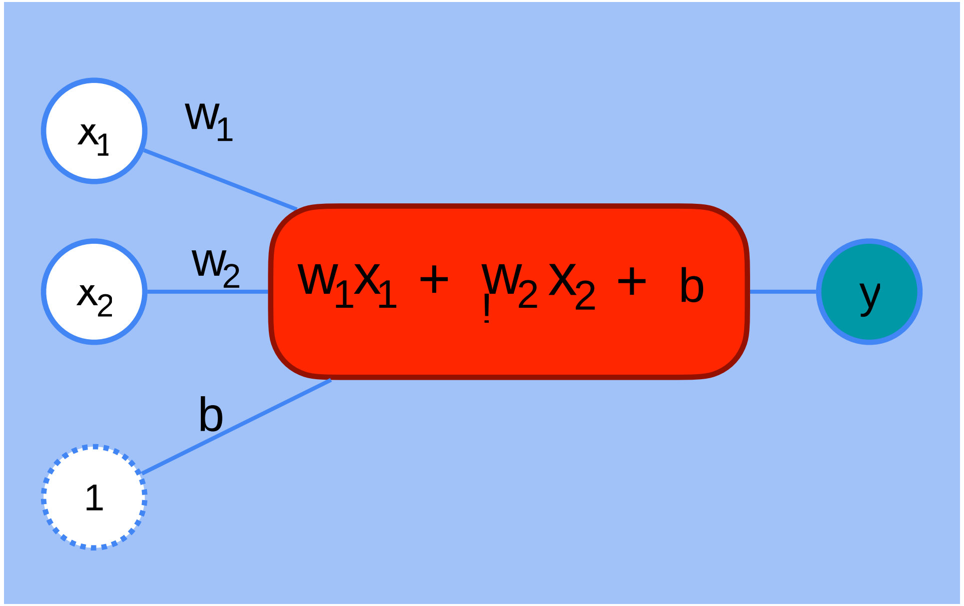
Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

"Regression With a Perceptron

Single Layer Neural Network Perceptron



Prediction Function:

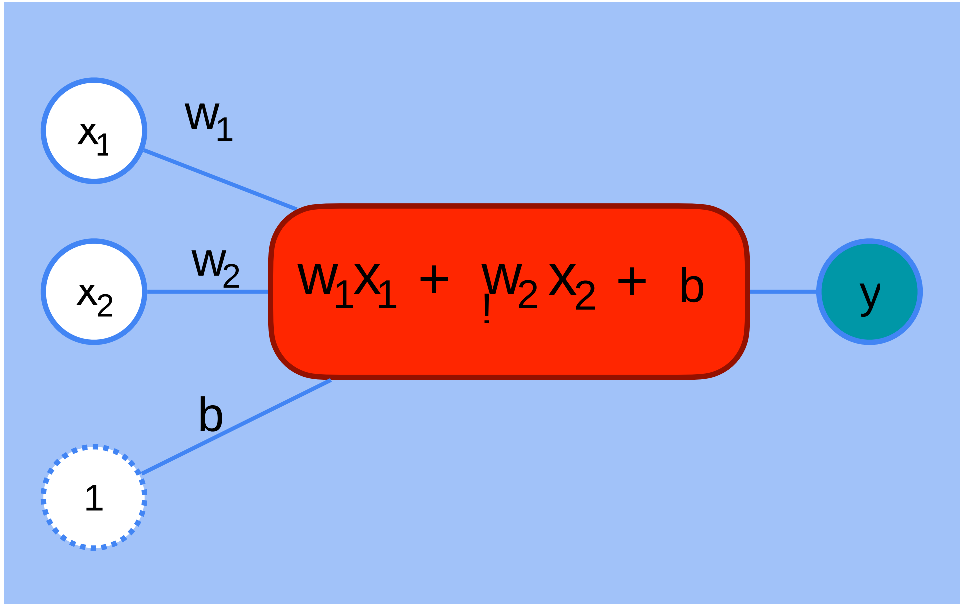
$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$= \frac{1}{2}(y\text{"}\ y1)^2$$

"Regression With a Perceptron

Single Layer Neural Network Perceptron



Prediction Function:

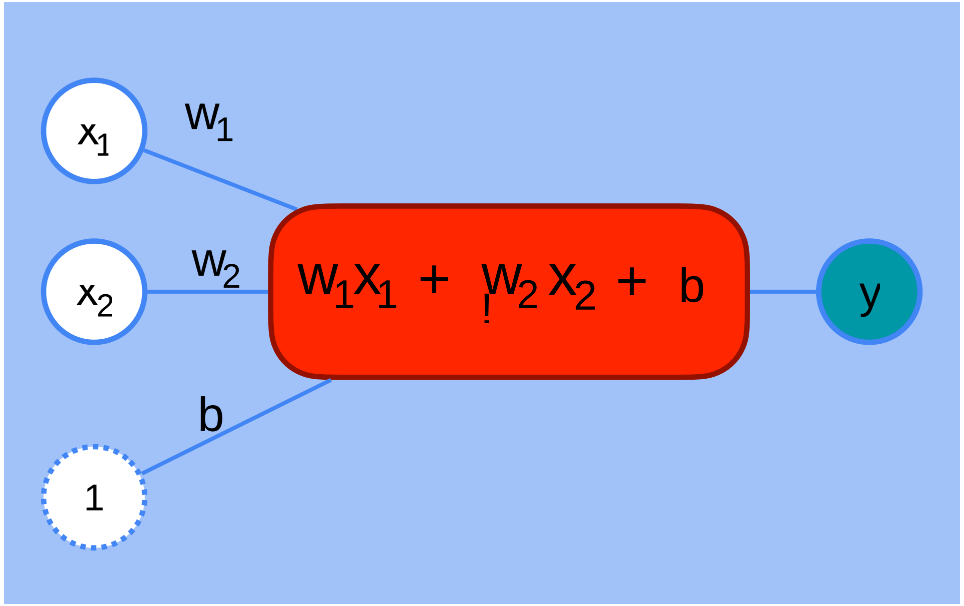
$$\hat{y} = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

"Regression With a Perceptron

Single Layer Neural Network Perceptron



Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

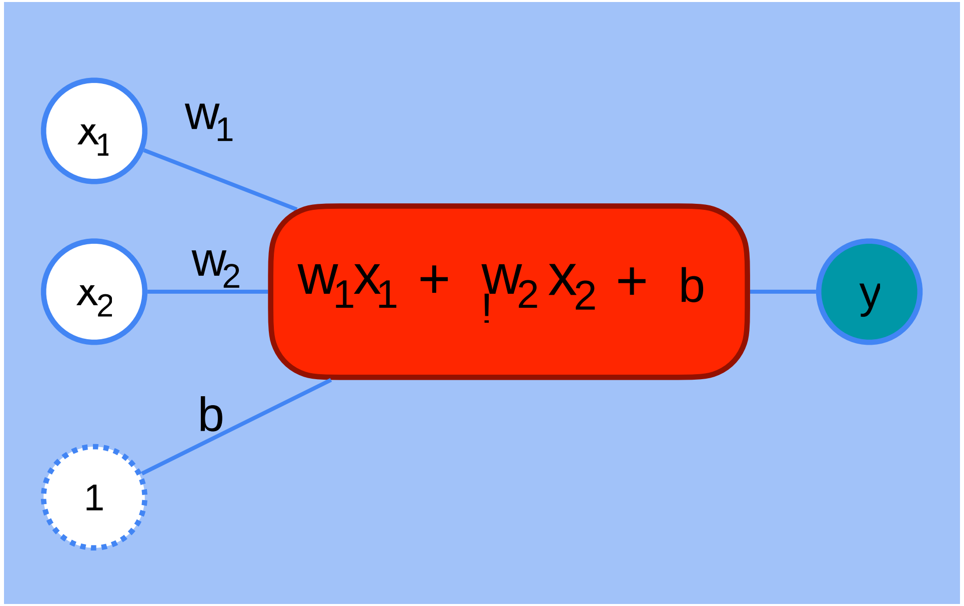
Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"} \; y1)^2$$

Main Goal:

"Regression With a Perceptron

Single Layer Neural Network Perceptron



Prediction Function:

$$\hat{y} = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

Main Goal:

Find w_1 , w_2 , b that give y with the least error

Optimization in Neural Networks and NewtonÕs Method

Regression with a perceptron: Gradient Descent

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

Main Goal:

Find w_1 , w_2 , b that give y with the least error

"Regression With a Perceptron

Prediction Function:

$$\hat{y} = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

Main Goal:

Find w_1 , w_2 , b that give $\hat{y}$ with the least error

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \,"\, y1)^2$$

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:

"Regression With a Perceptron

Prediction Function:

$$\hat{y} = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

Main Goal:

Find w_1 , w_2 , b that give $\hat{y}$ with the least error

To find optimal values for:

w_1 , w_2 , b

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \,\text{"}\, y1)^2$$

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:
w_1 , w_2 , b

You need gradient descent

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"} \; y1)^2$$

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$$w_1 \; \$ \; w_1 \text{"} \; \# \frac{, L}{, w_1}$$

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$$w_1 \ \$ \ w_1 \text{ " } \# \frac{, L}{, w_1}$$

$$w_2 \ \$ \ w_2 \text{ " } \# \frac{, L}{, w_2}$$

"Regression With a Perceptron

Prediction Function:

$$\hat{y} = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

Main Goal:

Find w_1 , w_2 , b that give $\hat{y}$ with the least error

To find optimal values for:
w_1 , w_2 , b

You need gradient descent

$$w_1 \leftarrow w_1 - \alpha \frac{\partial L}{\partial w_1}$$

$$w_2 \leftarrow w_2 - \alpha \frac{\partial L}{\partial w_2}$$

$$b \leftarrow b - \alpha \frac{\partial L}{\partial b}$$

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\, ''\, y1)^2$$

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$$w_1 \ \$ \quad ' \ \# \frac{,L}{,w_1}$$

$$w_2 \ \$ \quad ' \ \#' \frac{L}{,w_2}$$

$$b \ \$ \quad \#' \frac{L}{,b}$$

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \;" \; y1)^2$$

Some initial starting values

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:
w_1 , w_2 , b

You need gradient descent

$$w_1 \;\$ \quad ' \;\# \frac{,L}{,w_1}$$

$$w_2 \;\$ \quad ' \;\# \frac{,L}{,w_2}$$

$$b \;\$ \quad \# \frac{,L}{,b}$$

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\;"\;\; y1)^2$$

Some initial starting values

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:
w_1 , w_2 , b

You need gradient descent

$$w_1 \;\$ \quad '\;\#$$

$$w_2 \;\$ \quad '\;\#\frac{,L}{,w_2}$$

$$b \;\$ \quad \#\frac{,L}{,b}$$

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

Some initial starting values

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:

w_1 , w_2 , b

You need gradient descent

w_1 \$ ' #

w_2 \$ ' #

b \$ $\# \frac{,L}{,b}$

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y " y1)^2$$

Some initial starting values

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:
w_1 , w_2 , b

You need gradient descent

w_1 $ ' #

w_2 $ ' #

b $ #

"Regression With a Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

Some initial starting values

Main Goal:

Find w_1 , w_2 , b that give y with the least error

To find optimal values for:

w_1 , w_2 , b

You need gradient descent

w_1 $ ' #

w_2 $ ' #

b $ #

SUB-TASK

Find the following partial derivatives

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y " y1)^2$$

$$\frac{, L}{, b}$$

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

Using chain rule:

$$\frac{,L}{,b}$$

$$\frac{,L}{,w_1}$$

$$\frac{,L}{,w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"} \; y1)^2$$

Using chain rule:

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y" \ y1)^2$$

Using chain rule:

=

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y" \; y1)^2$$

Using chain rule:

$$=$$

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y " \; y1)^2$$

Using chain rule:

$$= \; , L$$

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$(y, y)1 = \frac{1}{2}(y\text{''}\ 1^2$$

Using chain rule:

$$= \ , L$$

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y\,'\,1)^2$$

Using chain rule:

$$= \frac{,\,L}{,\,y1}$$

$$\frac{,\,L}{,\,w_1}$$

$$\frac{,\,L}{,\,w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y\ '\ 1)^2$$

Using chain rule:

$$= \frac{,L}{,y1} \cdot$$

$$\frac{,L}{,w_1}$$

$$\frac{,L}{,w_2}$$

"Regression With Perceptron

Prediction Function:

$$1 = w_1x_1 + w_2x_2 +$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y"\ y1)^2$$

Using chain rule:

$$= \frac{, L}{, y1} \cdot$$

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$1 = w_1 x_1 + w_2 x_2 +$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

Using chain rule:

$$= \frac{, L}{, y1} \cdot \ , y1$$

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$1 = w_1x_1 + w_2x_2 +$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y" \;\; y1)^2$$

Using chain rule:

$$= \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$1 = w_1 x_1 + w_2 x_2 +$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

Using chain rule:

$$\frac{,\ L}{,\ b} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ b}$$

$$\frac{,\ L}{,\ w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y \,''\, 1)^2$$

Using chain rule:

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y " 1)^2$$

Using chain rule:

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$=$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1= w_1x_1 + w_2x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y " 1)^2$$

Using chain rule:

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$= \frac{, L}{, y1}$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y'' 1^2$$

Using chain rule:

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$= \frac{, L}{, y1} \cdot$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$1 = {}_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

Using chain rule:

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$= \frac{, L}{, y1} \cdot$$

$$\frac{, L}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$1 = \quad {}_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"} \; y1)^2$$

Using chain rule:

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$= \frac{,L}{,y1} \cdot \; ,y1$$

$$\frac{,L}{,w_2}$$

"Regression With Perceptron

Prediction Function:

$$1 = \;_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\; y1)^2$$

Using chain rule:

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$= \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

$$\frac{,L}{,w_2}$$

"Regression With Perceptron

Prediction Function:

$$1 = 1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

Using chain rule:

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_1} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

"Regression With Perceptron

Prediction Function:

$$1 = \;{}_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

Using chain rule:

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_1} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

$$=$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y '' 1)^2$$

Using chain rule:

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$=$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y " \; 1)^2$$

Using chain rule:

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$= \frac{, L}{, y1}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$(y, y1) = \frac{1}{2}(y \text{"} 1)^2$$

Using chain rule:

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$= \frac{, L}{, y1} \cdot$$

"Regression With Perceptron

Prediction Function:

$$1 = w_1 x_1 + \quad x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"} \; y1)^2$$

Using chain rule:

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$= \frac{, L}{, y1} \cdot$$

"Regression With Perceptron

Prediction Function:

$$1 = w_1x_1 + \quad x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

Using chain rule:

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_1} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

$$= \frac{,L}{,y1} \cdot \ ,y1$$

"Regression With Perceptron

Prediction Function:

$$1 = w_1x_1 + \quad x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

Using chain rule:

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_1} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

$$= \frac{,L}{,y1} \cdot \frac{,y1}{,w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

Using chain rule:

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_1} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

$$\frac{,L}{,w_2} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

Using chain rule:

$$\frac{, L}{, b} = \cdot$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"} \; y1)^2$$

$\frac{, L}{, y1}$	
$\frac{, y1}{, b}$	
$\frac{, y1}{, w_1}$	
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"} \; y1)^2$$

$\frac{,\, y1}{,\, b}$	
$\frac{,\, y1}{,\, w_1}$	
$\frac{,\, y1}{,\, w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) =$$

$$\frac{, y1}{, b}$$

$$\frac{, y1}{, w_1}$$

$$\frac{, y1}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) =$$

	$= (y" \; y1)$
$\frac{, y1}{, b}$	
$\frac{, y1}{, w_1}$	
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(\quad y1)^2$$

	$= (y \text{"} \ y1)$
$\frac{, y1}{, b}$	
$\frac{, y1}{, w_1}$	
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(\quad y1)^2$$

	$= -(y \text{"} \; y1)$
$\frac{, y1}{, b}$	
$\frac{, y1}{, w_1}$	
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1x_1 + w_2x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

	$= - (y \text{ " } y1)$
$\frac{, y1}{, b}$	
$\frac{, y1}{, w_1}$	
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$\hat{y} = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

$$\frac{\partial L}{\partial \hat{y}} = -(y - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial \hat{y}}{\partial w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = \quad + \ b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

$$\frac{, L}{, y1} = -(y \text{ " } y1)$$

$$\frac{, y1}{, w_1}$$

$$\frac{, y1}{, w_2}$$

"Regression With Perceptron

Prediction Function:

$$y1 = \qquad + \ b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

$\frac{, L}{, y1}$	$= -(y \text{ " } y1)$
	$= 1$
$\frac{, y1}{, w_1}$	
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{ " } y1)^2$$

$\frac{, L}{, y1}$	$= - (y \text{ " } y1)$
	$= 1$
$\frac{, y1}{, w_1}$	
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y " y1)^2$$

$\frac{, L}{, y1}$	$= -(y " y1)$
$\frac{, y1}{, b}$	$= 1$
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$\hat{y} = w_1 x_1 +$$

Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 +$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\text{"}\ y1)^2$$

$\frac{, L}{, y1}$	$= -(y\text{"}\ y1)$
$\frac{, y1}{, b}$	$= 1$
	$= x_1$
$\frac{, y1}{, w_2}$	

"Regression With Perceptron

Prediction Function:

$$\hat{y} = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
	$= x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y " \ y1)^2$$

$\frac{, L}{, y1}$	$= - (y " \ y1)$
$\frac{, y1}{, b}$	$= 1$
$\frac{, y1}{, w_1}$	$= x_1$

"Regression With Perceptron

Prediction Function:

$$y1 = \quad + \quad w_2 x_2 \quad +$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"} \; y1)^2$$

$$\frac{,L}{,y1} = -(y \text{"} \; y1)$$

$$\frac{,y1}{,b} = 1$$

$$\frac{,y1}{,w_1} = x_1$$

"Regression With Perceptron

Prediction Function:

$$y1 = \qquad + \quad w_2 x_2 \qquad$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y\ "\ y1)^2$$

$\frac{, L}{, y1}$	$= - (y\ "\ y1)$
$\frac{, y1}{, b}$	$= 1$
$\frac{, y1}{, w_1}$	$= x_1$
	$= x_2$

"Regression With Perceptron

Prediction Function:

$$y1 = w_1 x_1 + w_2 x_2 + b$$

Loss Function:

$$L(y, y1) = \frac{1}{2}(y \text{"}\ y1)^2$$

$\frac{, L}{, y1}$	$= -(y \text{"}\ y1)$
$\frac{, y1}{, b}$	$= 1$
$\frac{, y1}{, w_1}$	$= x_1$
	$= x_2$

"Regression With Perceptron

$$\frac{, L}{, y1} = -(y \,"\, y1)$$

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{,L}{,y1} = -(y\," \;\; y1)$$

$$\frac{,y1}{,b} = 1$$

$$\frac{,y1}{,w_1} = x_1$$

$$\frac{,y1}{,w_2} = x_2$$

$$= \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_1} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

$$\frac{,L}{,w_2} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_2}$$

"Regression With Perceptron

$$\frac{,\, y1}{,\, b} = 1$$

$$\frac{,\, y1}{,\, w_1} = x_1$$

$$\frac{,\, y1}{,\, w_2} = x_2$$

$$= \; L \cdot \frac{,\, y1}{,\, b}$$

$$\frac{,\, L}{,\, w_1} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_1}$$

$$\frac{,\, L}{,\, w_2} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$= \; L \cdot \frac{, y1}{, b}$$

$$= -(y \text{ " } y1$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{,\, y1}{,\, b} = 1$$

$$\frac{,\, y1}{,\, w_1} = x_1$$

$$\frac{,\, y1}{,\, w_2} = x_2$$

$$= \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, b}$$

$$= -(y\text{"}\ y1)$$

$$\frac{,\, L}{,\, w_1} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_1}$$

$$\frac{,\, L}{,\, w_2} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$= \frac{, L}{, y1} \cdot \; L$$

$$= -(y\,"\; y)1$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$= \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$= -(y" \ y1)$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

= -(y" y)1

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{,\ y1}{,\ b} = 1$$

$$\frac{,\ y1}{,\ w_1} = x_1$$

$$\frac{,\ y1}{,\ w_2} = x_2$$

$$\frac{,\ L}{,\ b} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ b}$$

= -(y " y1)

$$= \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ w_1}$$

$$\frac{,\ L}{,\ w_2} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

= -(y" y)1

$$= \quad \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$= -(y\text{"}\ y)1$

$$= \cdot \frac{, y1}{, w_1}$$

$= -(y\text{"}\ y)1$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$= -(y \text{"}\ y)1$$

$$= \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$= -(y \text{"}\ y)1$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$\frac{, y1}{, b} = 1$

$\frac{, y1}{, w_2} = x_2$

$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$ = -(y" y)1

$= \frac{, L}{, y1} \cdot$ 1 = -(y" y)1

$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$= -(y" \ y)1$

$$= \frac{, L}{, y1} \cdot \ldots 1$$

$= -(y" \ y)1 x_1$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b} = -(y \text{ " } y)1$$

$$= \frac{, L}{, y1} \cdot \frac{, y1}{, w_1} = -(y \text{ " } y)1 x_1$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{,\,y1}{,\,b} = 1$$

$$\frac{,\,y1}{,\,w_1} = x_1$$

$$\frac{,\,y1}{,\,w_2} = x_2$$

$$\frac{,\,L}{,\,b} = \frac{,\,L}{,\,y1} \cdot \frac{,\,y1}{,\,b} \qquad = -(y\,"\ \,y)1$$

$$\frac{,\,L}{,\,w_1} = \frac{,\,L}{,\,y1} \cdot \frac{,\,y1}{,\,w_1} \qquad = -(y\,"\ \,y)1x_1$$

$$\frac{,\,L}{,\,w_2} = \frac{,\,L}{,\,y1} \cdot \frac{,\,y1}{,\,w_2}$$

"Regression With Perceptron

$$\frac{,\, y1}{,\, b} = 1$$

$$\frac{,\, y1}{,\, w_1} = x_1$$

$$\frac{,\, y1}{,\, w_2} = x_2$$

$$\frac{,\, L}{,\, b} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, b} = -(y \text{ " } y)1$$

$$\frac{,\, L}{,\, w_1} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_1} = -(y \text{ " } y)1 x_1$$

$$= \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$= -(y\,"\ y1)$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$= -(y\,"\ y1)x_1$

$$= \quad L \cdot \frac{, y1}{, w_2}$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$= -(y\,''\ y)1$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$= -(y\,''\ y)1 x_1$$

$$= \quad L \cdot \frac{, y1}{, w_2}$$

$$= -(y\,''\ y)1$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b} = -(y" \ y1)$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1} = -(y" \ y1)x_1$$

$$= \frac{, L}{, y1} \cdot \frac{, y1}{, w_2} = -(y" \ y1)$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

= -(y" y)1

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

= -(y" y)1 x_1

$$= \frac{, L}{, y1} \cdot$$ L

= -(y" y)1

"Regression With Perceptron

$$\frac{,\, y1}{,\, b} = 1$$

$$\frac{,\, y1}{,\, w_1} = x_1$$

$$\frac{,\, L}{,\, b} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, b}$$

= -(y" y)1

$$\frac{,\, L}{,\, w_1} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_1}$$

= -(y" y)1x_1

$$= \frac{,\, L}{,\, y1} \cdot$$ L

= -(y" y)1x_2

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b} = -(y \text{"} \ y1)$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1} = -(y \text{"} \ y1)x_1$$

$$= \frac{, L}{, y1} \cdot \frac{, y1}{, w_2} = -(y \text{"} \ y1)x_2$$

"Regression With Perceptron

$$\frac{, y1}{, b} = 1$$

$$\frac{, y1}{, w_1} = x_1$$

$$\frac{, y1}{, w_2} = x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b} = -(y\,"\ y)1$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1} = -(y\,"\ y)1x_1$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2} = -(y\,"\ y)1x_2$$

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:

w_1 , w_2 , b

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 \text{"} \# \frac{, L}{, w_1}$$

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$w_1 = w_1$ " #

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

w_1 = w_1 " #(-x_1(y " y))

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 \text{ " } \# (-x_1 (y \text{ " } y))$$

$$w_2 = w_2 \text{ " } \# \frac{, L}{, w_2}$$

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 \text{ " } \# (-x_1(y \text{ " } y))$$

$$w_2 = w_2 \text{ " } \#$$

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 \text{ " } \#(-x_1(y \text{ " } \hat{y}))$$

$$w_2 = w_2 \text{ " } \#(-x_2(y \text{ " } \hat{y}))$$

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 \,\text{"}\, \#(-x_1(y \,\text{"}\, \hat{y}))$$

$$w_2 = w_2 \,\text{"}\, \#(-x_2(y \,\text{"}\, \hat{y}))$$

$$b = b \,\text{"}\, \# \frac{,L}{,b}$$

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 " \#(-x_1(y " y))$$

$$w_2 = w_2 " \#(-x_2(y " y))$$

$$b = b " \#$$

"Regression With a Perceptron

Main Goal:

Find w_1 , w_2 , b that give y with the least error

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 \text{"} \; \# (-x_1(y \text{"} \; y))$$

$$w_2 = w_2 \text{"} \; \# (-x_2(y \text{"} \; y))$$

$$b = b \text{"} \; \# (-(y \text{"} \; y))$$

Optimization in Neural Networks and NewtonÕs Method

Classification with a perceptron

Classification Problem Motivation

Classification Problem Motivation

Classification Problem Motivation

Sentence			

Classification Problem Motivation

Sentence			
Aack aack aack!			

Classification Problem Motivation

Sentence			
Aack aack aack!			
Beep beep!			

Classification Problem Motivation

Sentence			
Aack aack aack!			
Beep beep!			
Aack beep beep beep!			

Classification Problem Motivation

Sentence			
Aack aack aack!			
Beep beep!			
Aack beep beep beep!			
Aack beep aack!			

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			
Beep beep!			
Aack beep beep beep!			
Aack beep aack!			

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			!
Beep beep!			
Aack beep beep beep!			
Aack beep aack!			

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			Happy !
Beep beep!			
Aack beep beep beep!			
Aack beep aack!			

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			Happy !
Beep beep!			"
Aack beep beep beep!			
Aack beep aack!			

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			Happy !
Beep beep!			Sad "
Aack beep beep beep!			
Aack beep aack!			

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			Happy !
Beep beep!			Sad "
Aack beep beep beep!			"
Aack beep aack!			

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			Happy !
Beep beep!			Sad "
Aack beep beep beep!			Sad "
Aack beep aack!			

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			Happy !
Beep beep!			Sad "
Aack beep beep beep!			Sad "
Aack beep aack!			!

Classification Problem Motivation

Sentence			Mood
Aack aack aack!			Happy !
Beep beep!			Sad "
Aack beep beep beep!			Sad "
Aack beep aack!			Happy !

Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!			Happy !
Beep beep!			Sad "
Aack beep beep beep!			Sad "
Aack beep aack!			Happy !

Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3		Happy !
Beep beep!			Sad "
Aack beep beep beep!			Sad "
Aack beep aack!			Happy !

Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!			Sad "
Aack beep beep beep!			Sad "
Aack beep aack!			Happy !

Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0		Sad "
Aack beep beep beep!			Sad "
Aack beep aack!			Happy !

Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0	2	Sad "
Aack beep beep beep!			Sad "
Aack beep aack!			Happy !

Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0	2	Sad "
Aack beep beep beep!	1		Sad "
Aack beep aack!			Happy !

Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0	2	Sad "
Aack beep beep beep!	1	3	Sad "
Aack beep aack!			Happy !

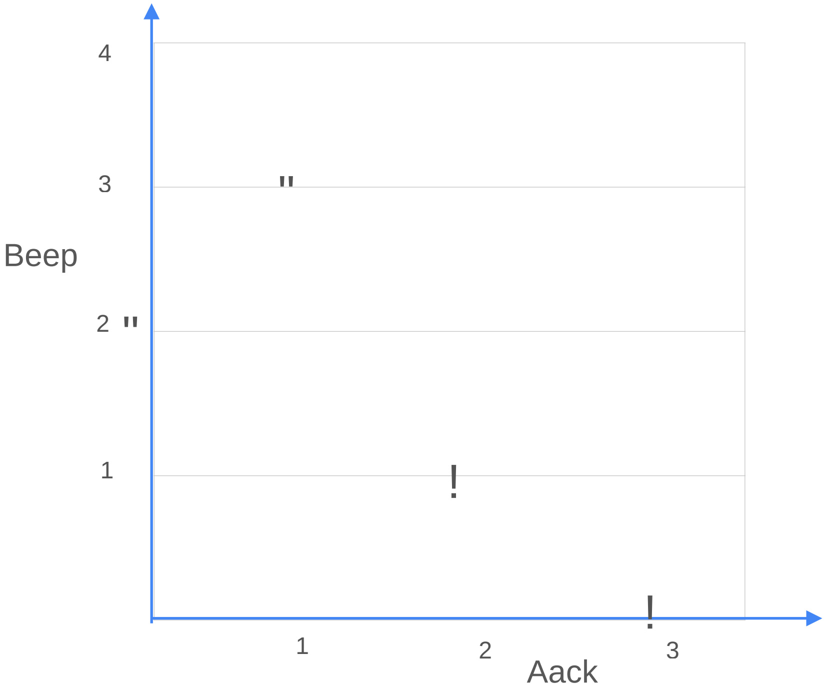
Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0	2	Sad "
Aack beep beep beep!	1	3	Sad "
Aack beep aack!	2		Happy !

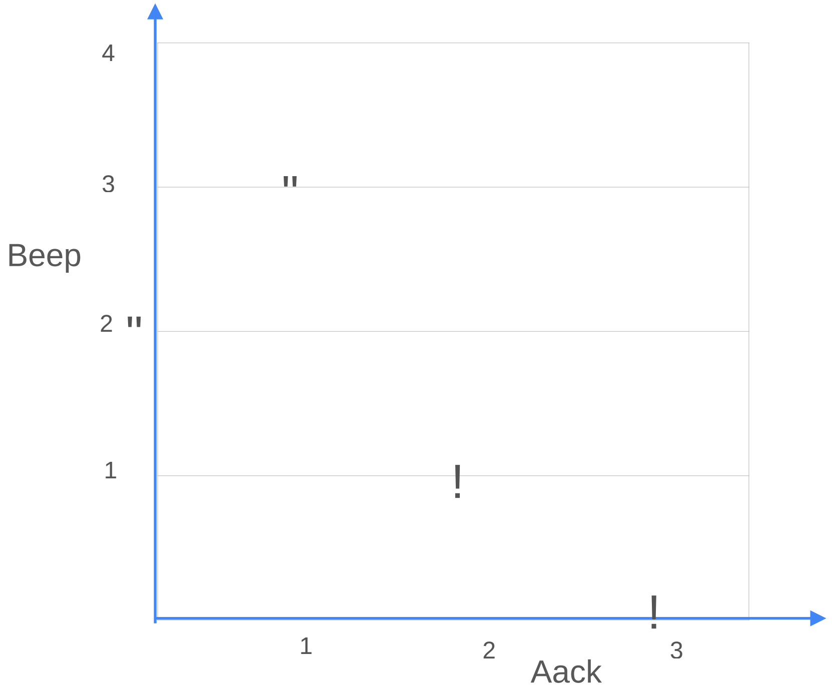
Classification Problem Motivation

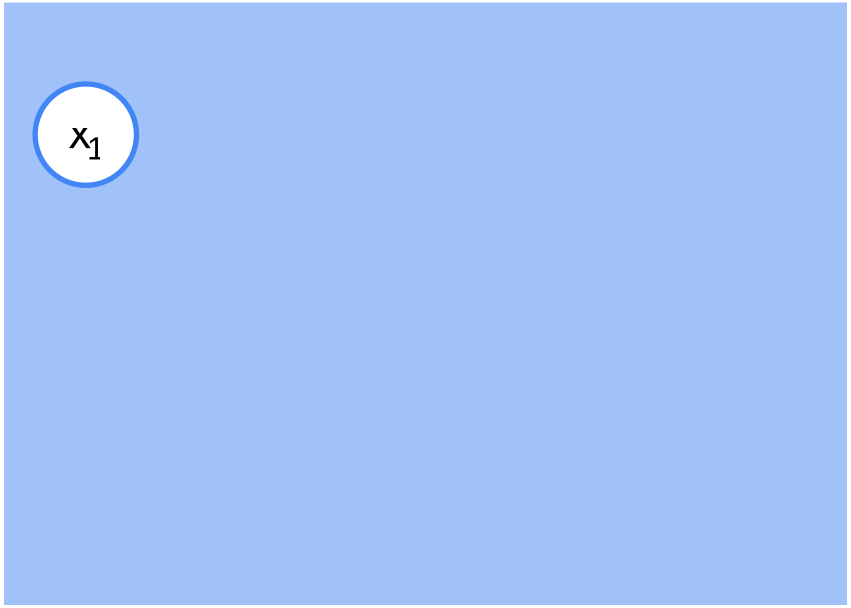
Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0	2	Sad "
Aack beep beep beep!	1	3	Sad "
Aack beep aack!	2	1	Happy !

Classification Problem Motivation

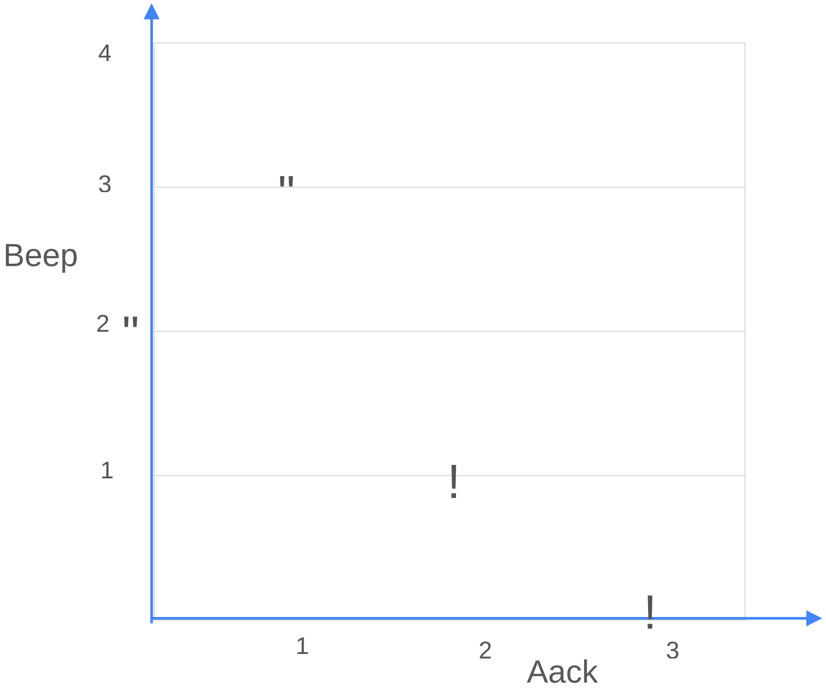


Classification Problem Motivation

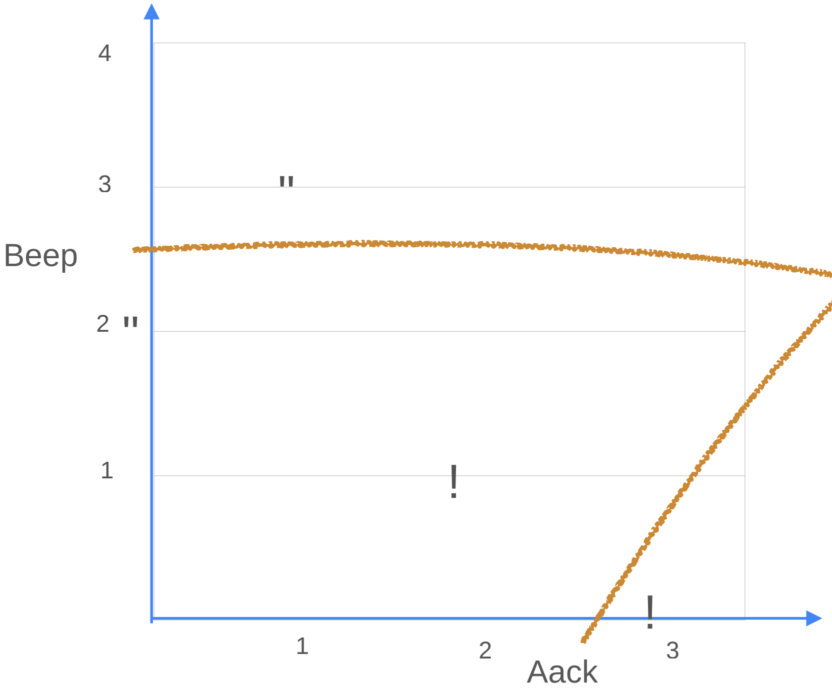




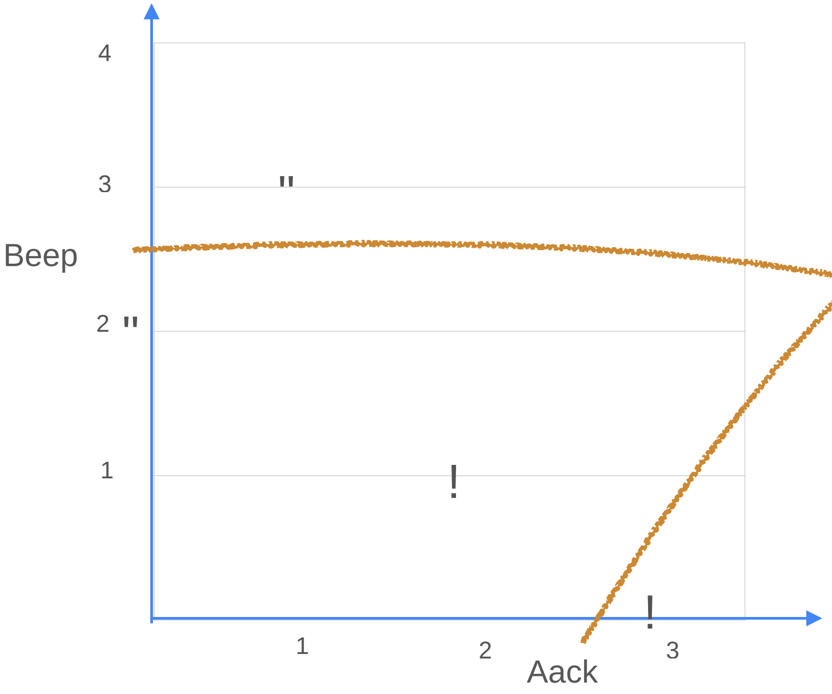
Classification Problem Motivation

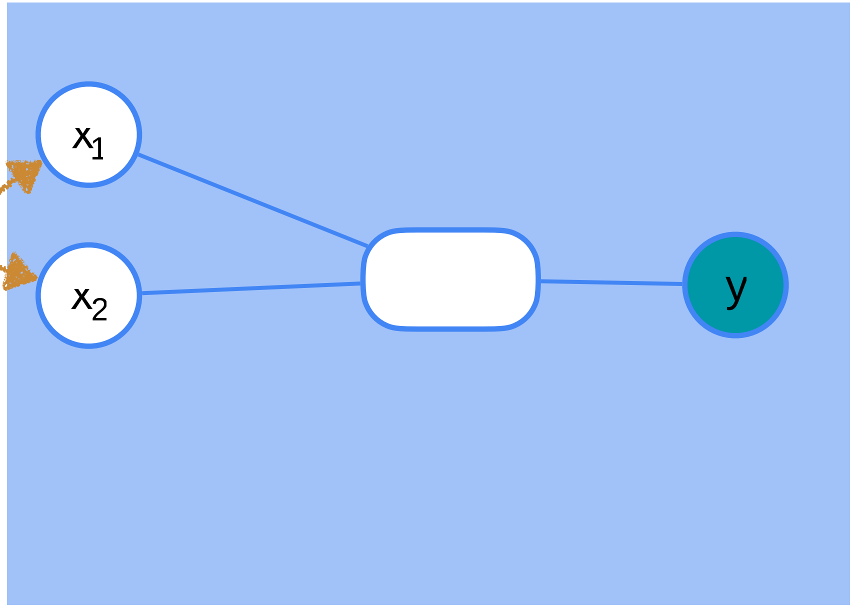


Classification Problem Motivation

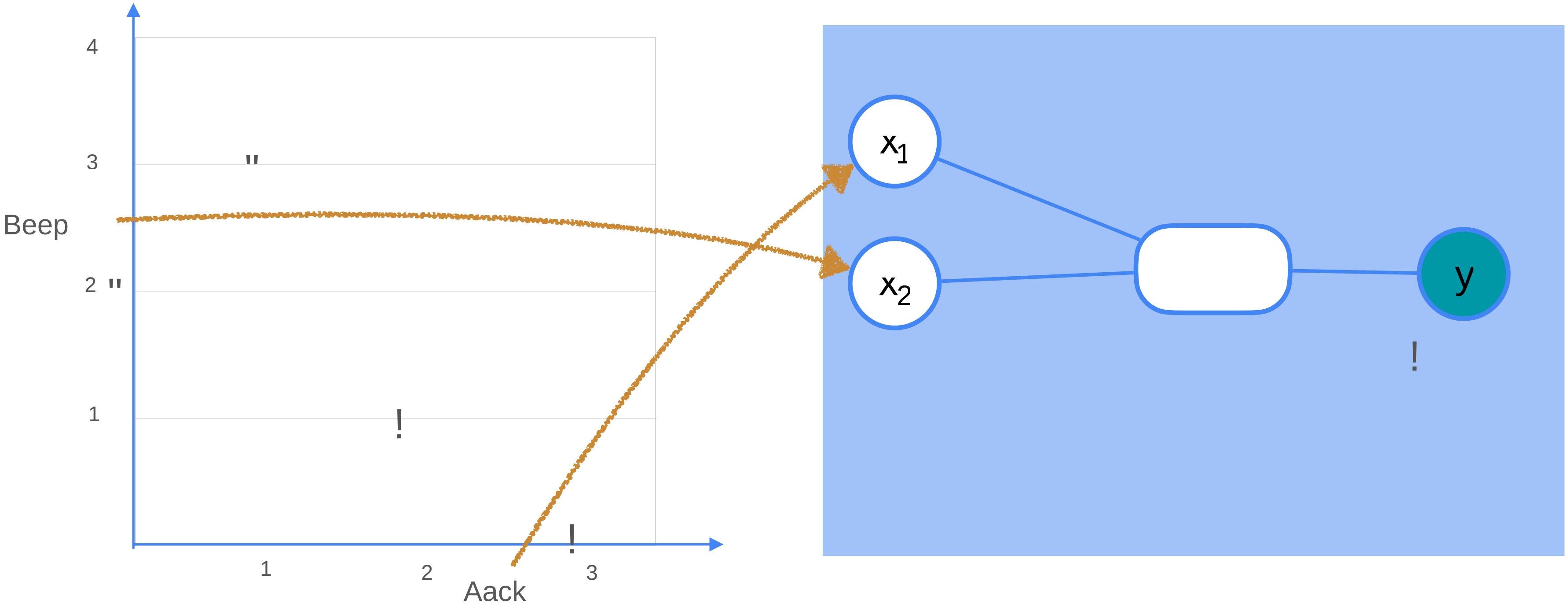


Classification Problem Motivation

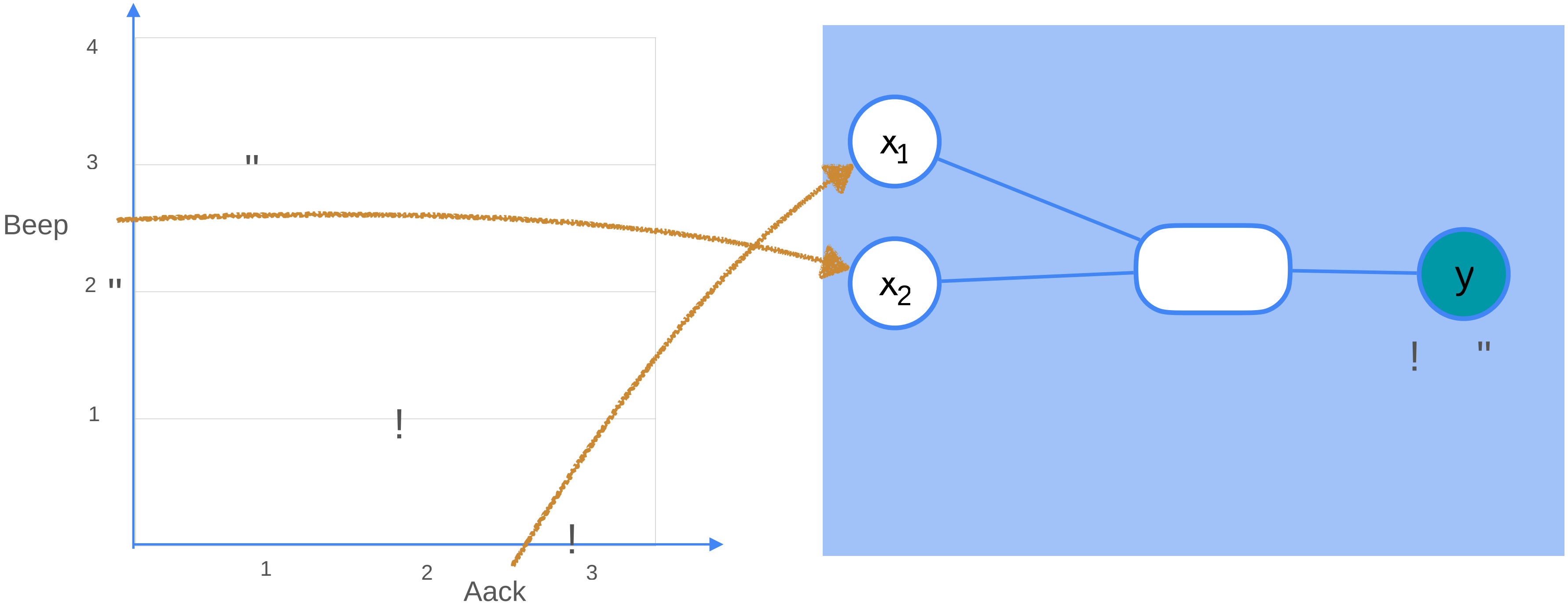




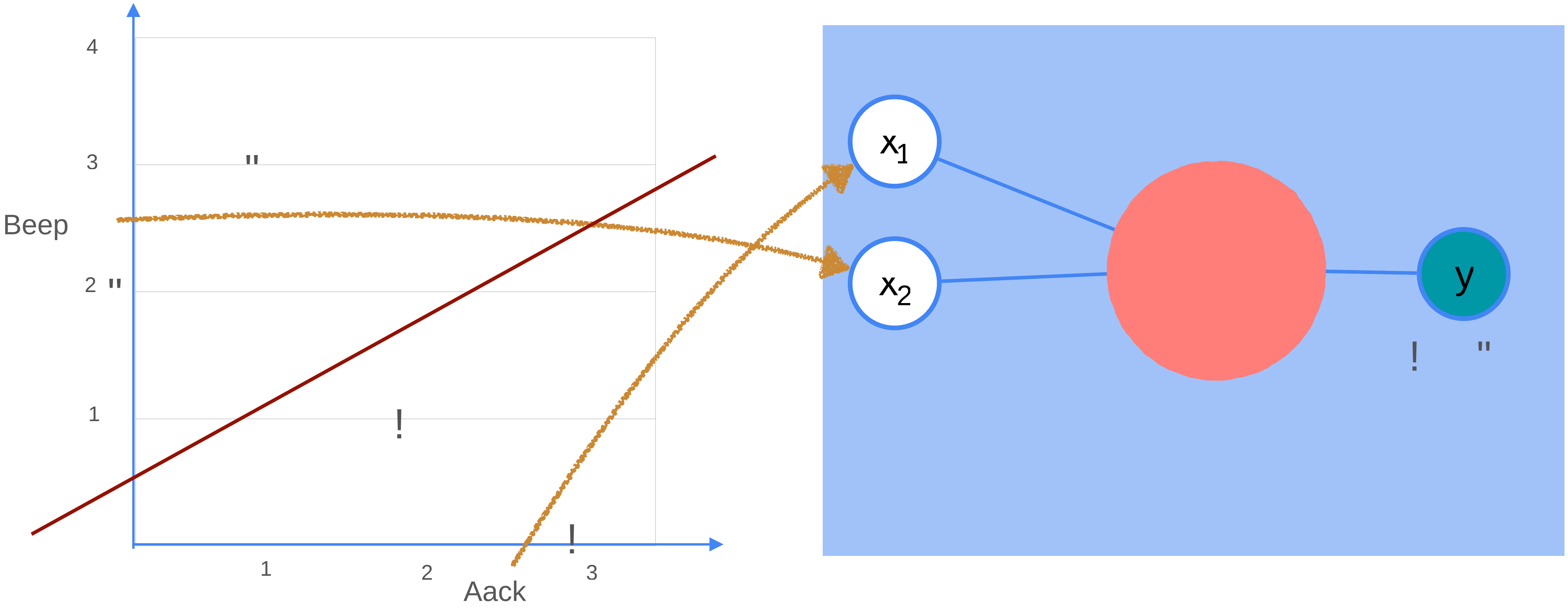
Classification Problem Motivation



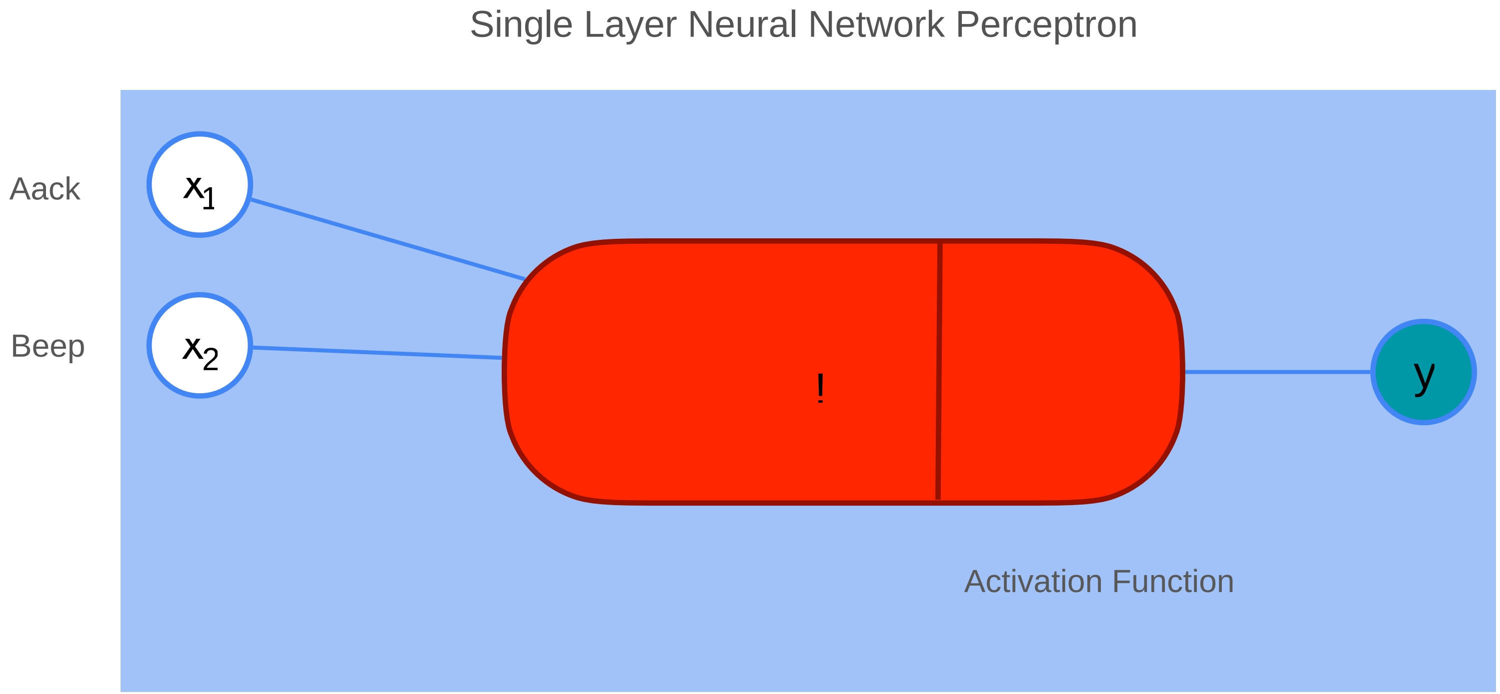
Classification Problem Motivation



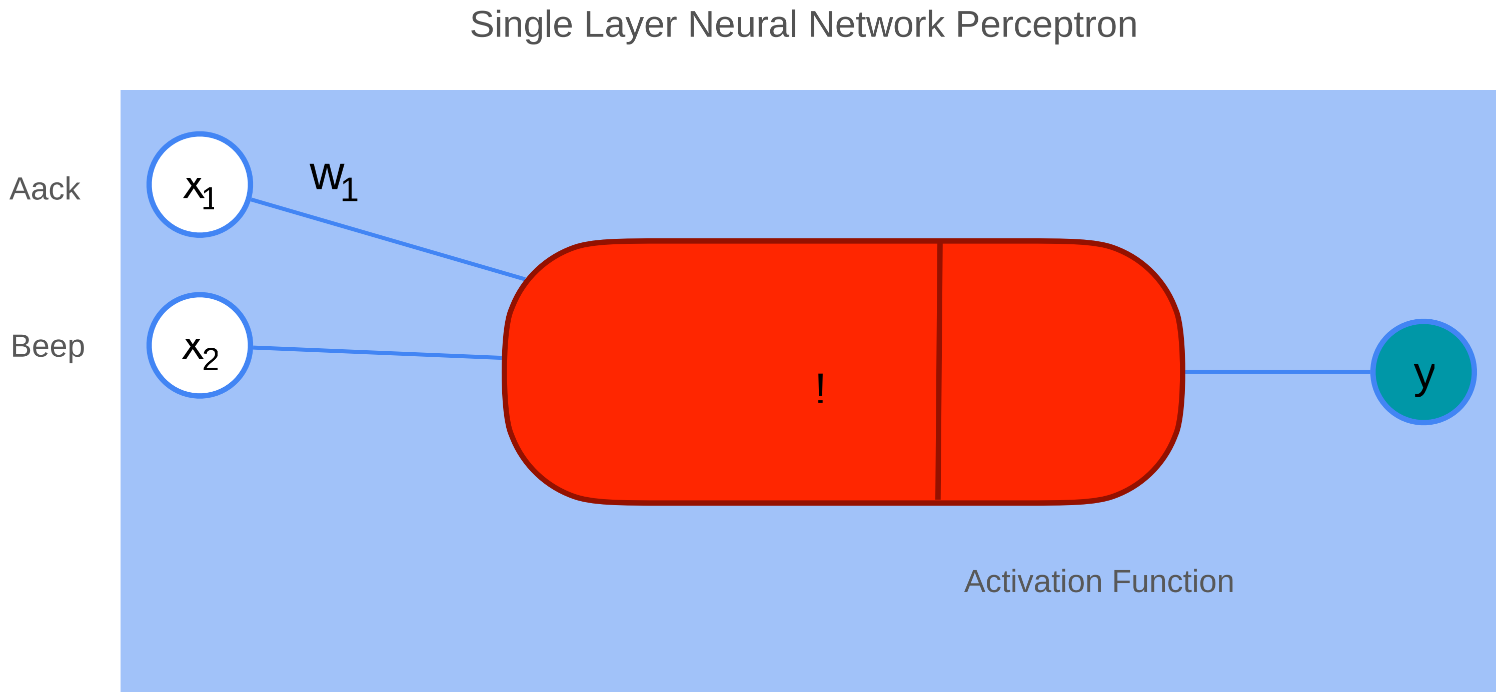
Classification Problem Motivation



"Classification With a Perceptron

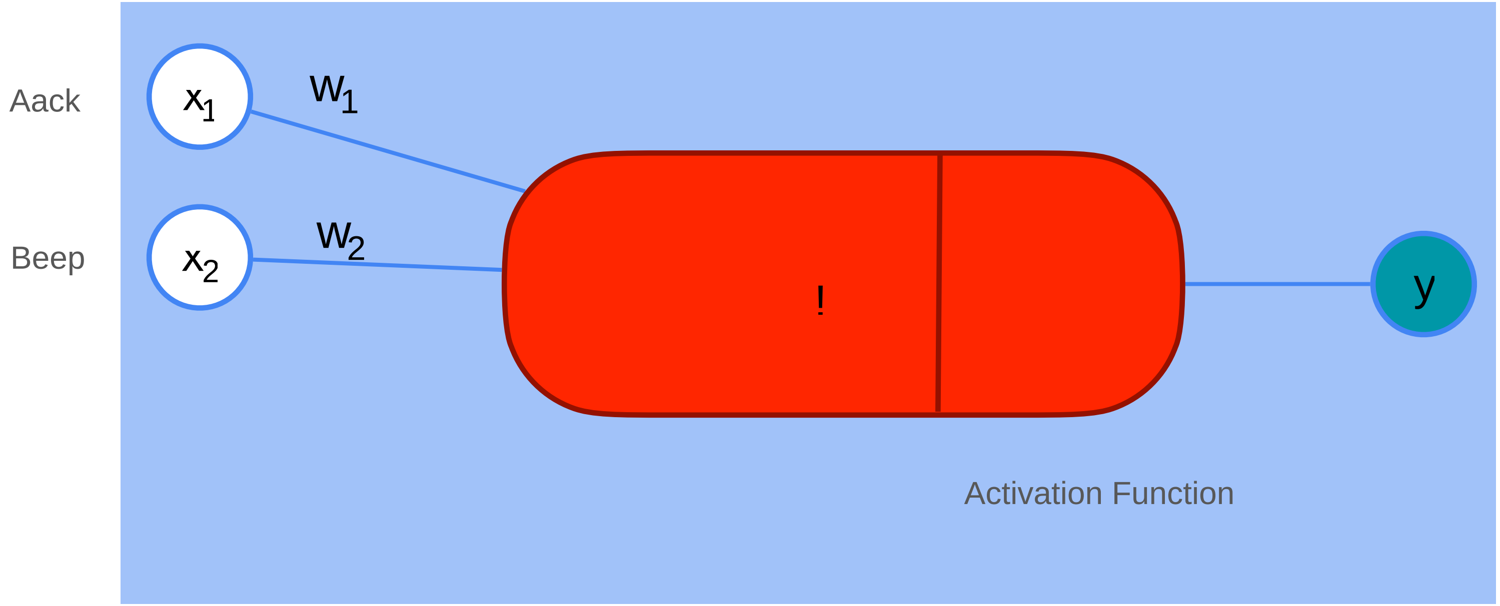


"Classification With a Perceptron



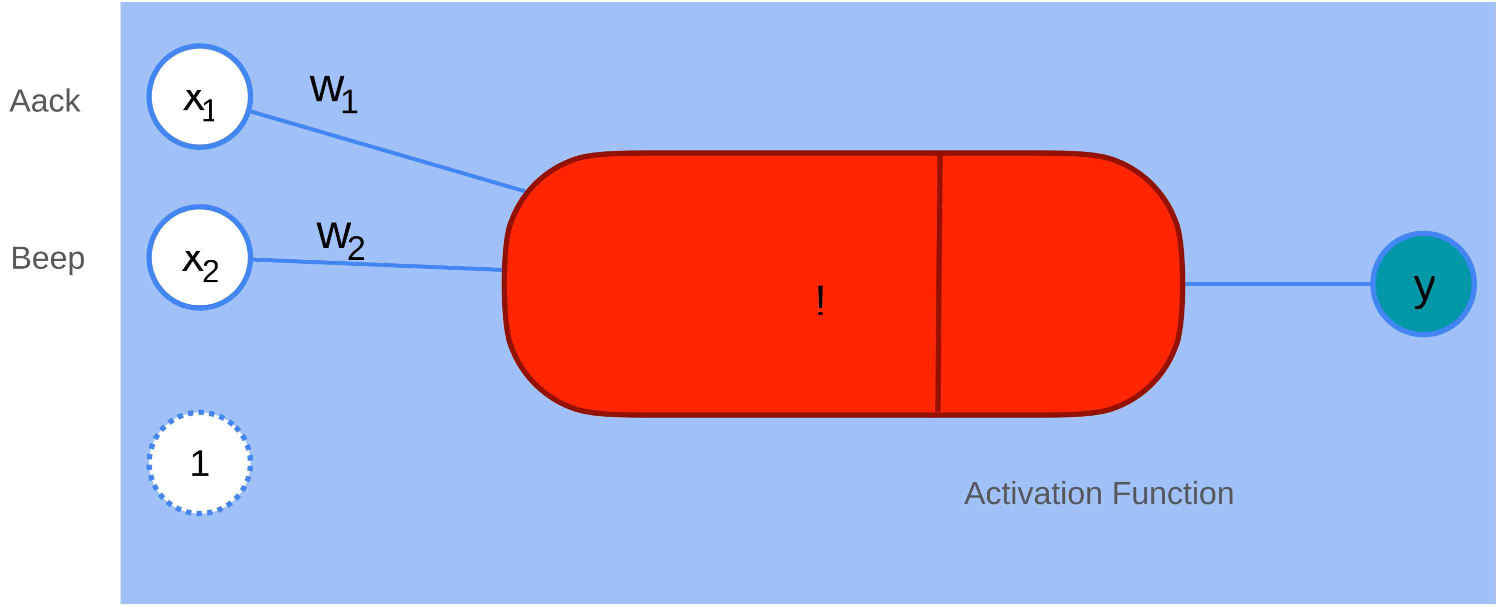
"Classification With a Perceptron

Single Layer Neural Network Perceptron

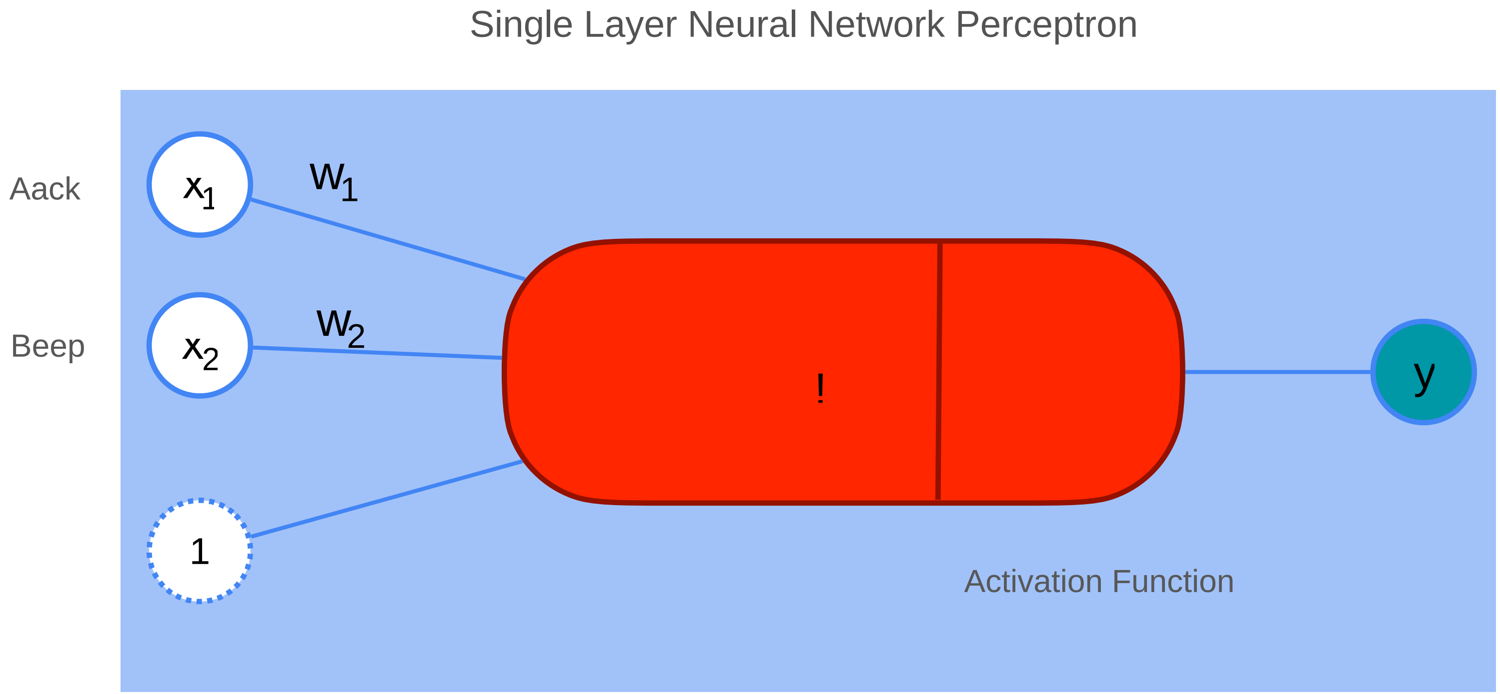


"Classification With a Perceptron

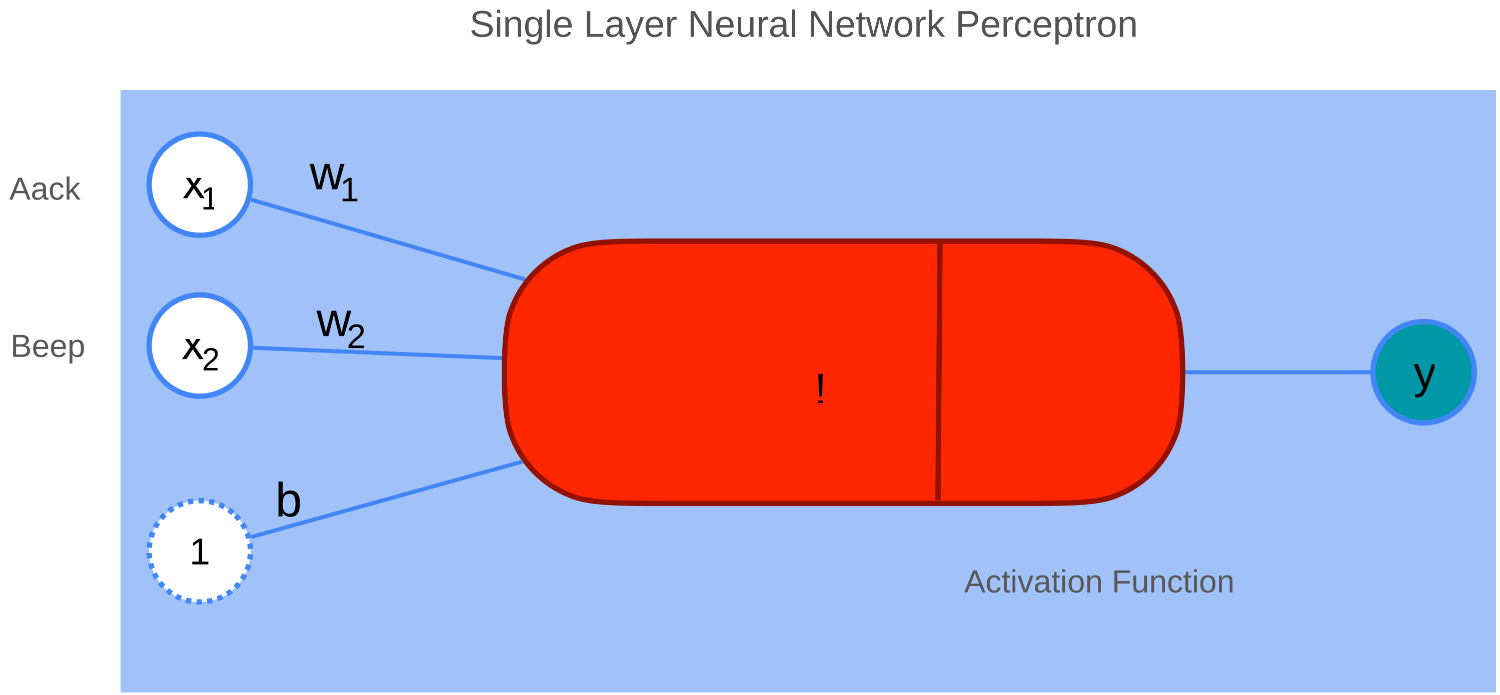
Single Layer Neural Network Perceptron



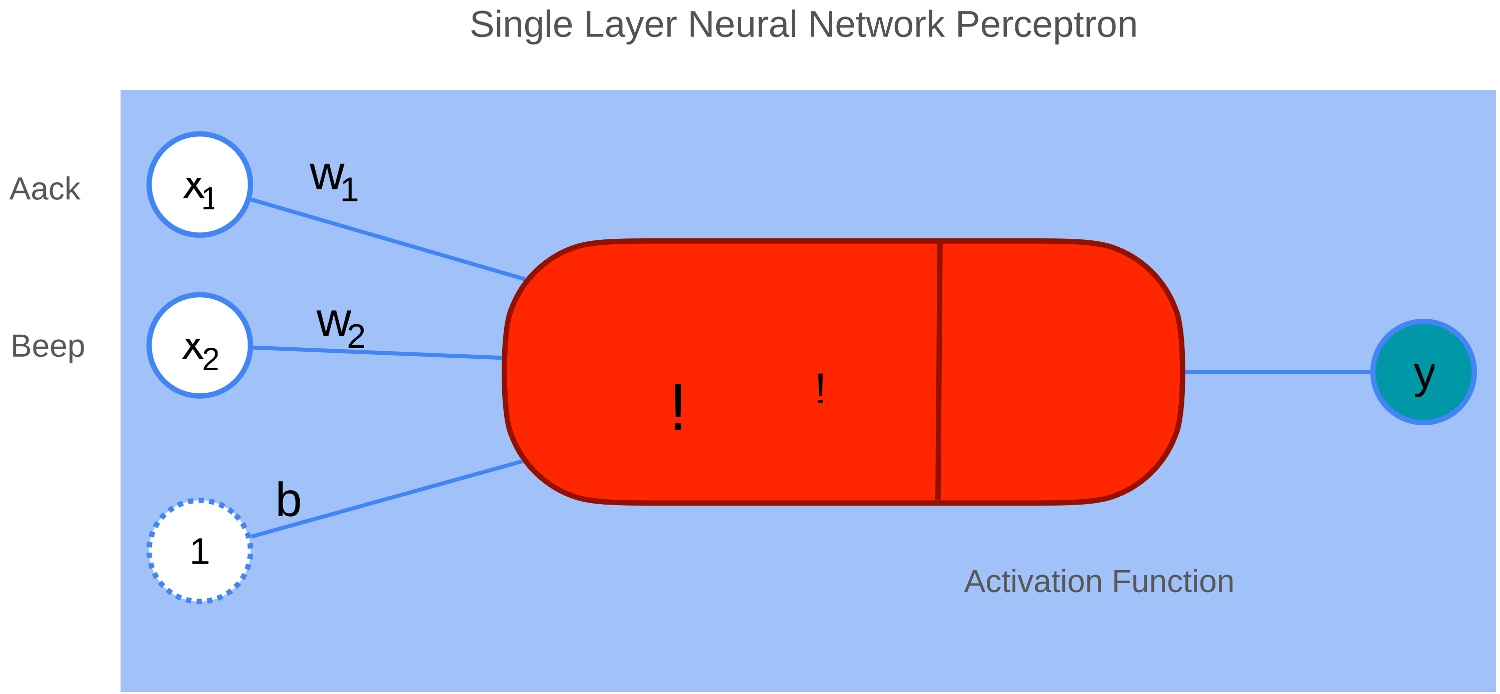
"Classification With a Perceptron



"Classification With a Perceptron

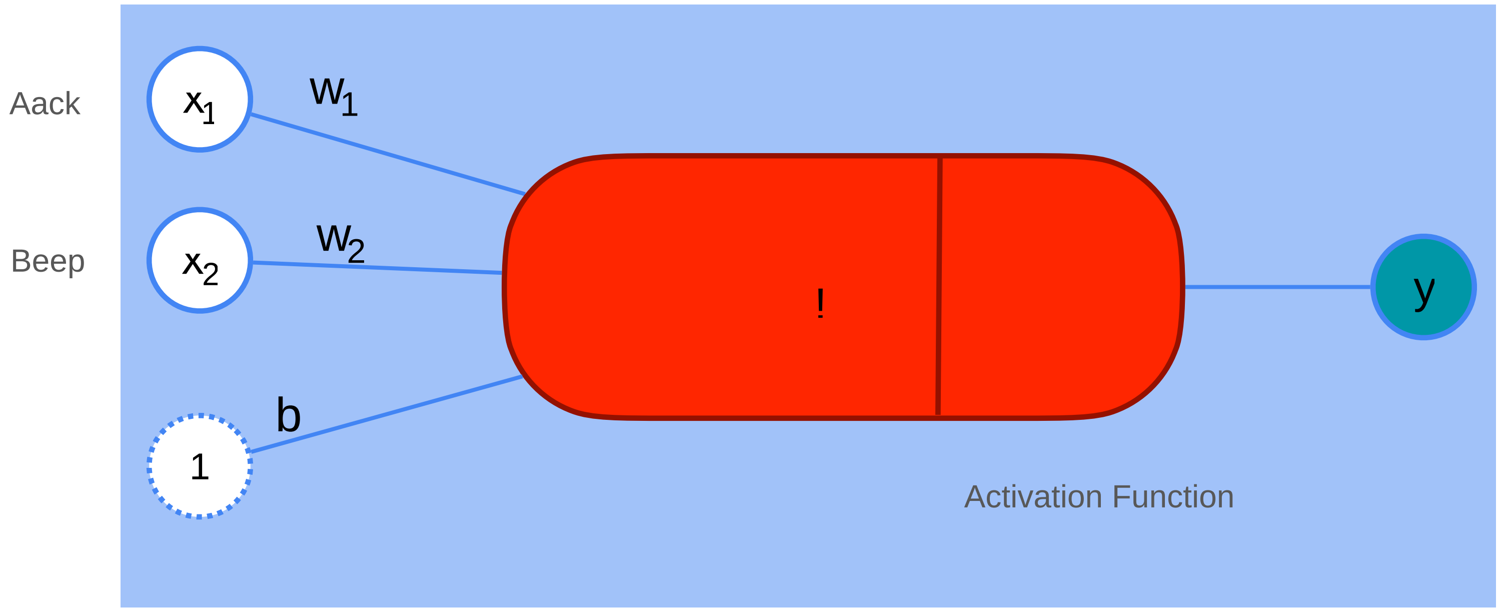


"Classification With a Perceptron



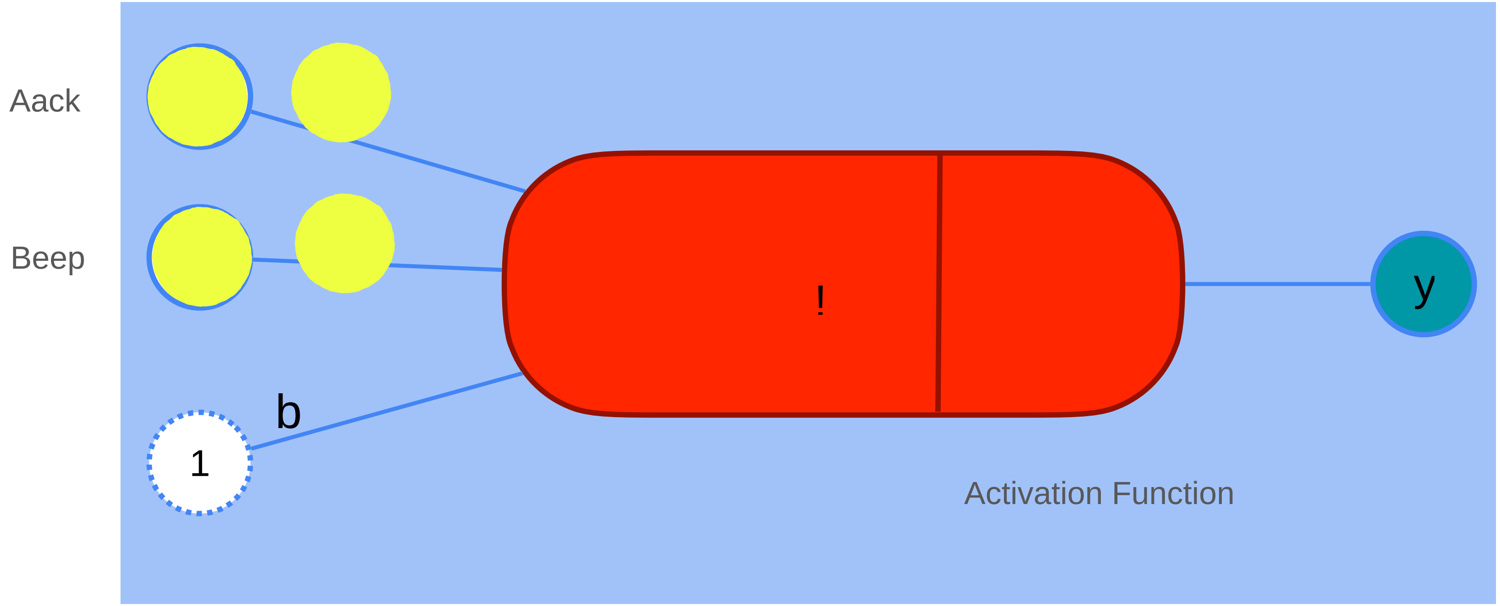
"Classification With a Perceptron

Single Layer Neural Network Perceptron



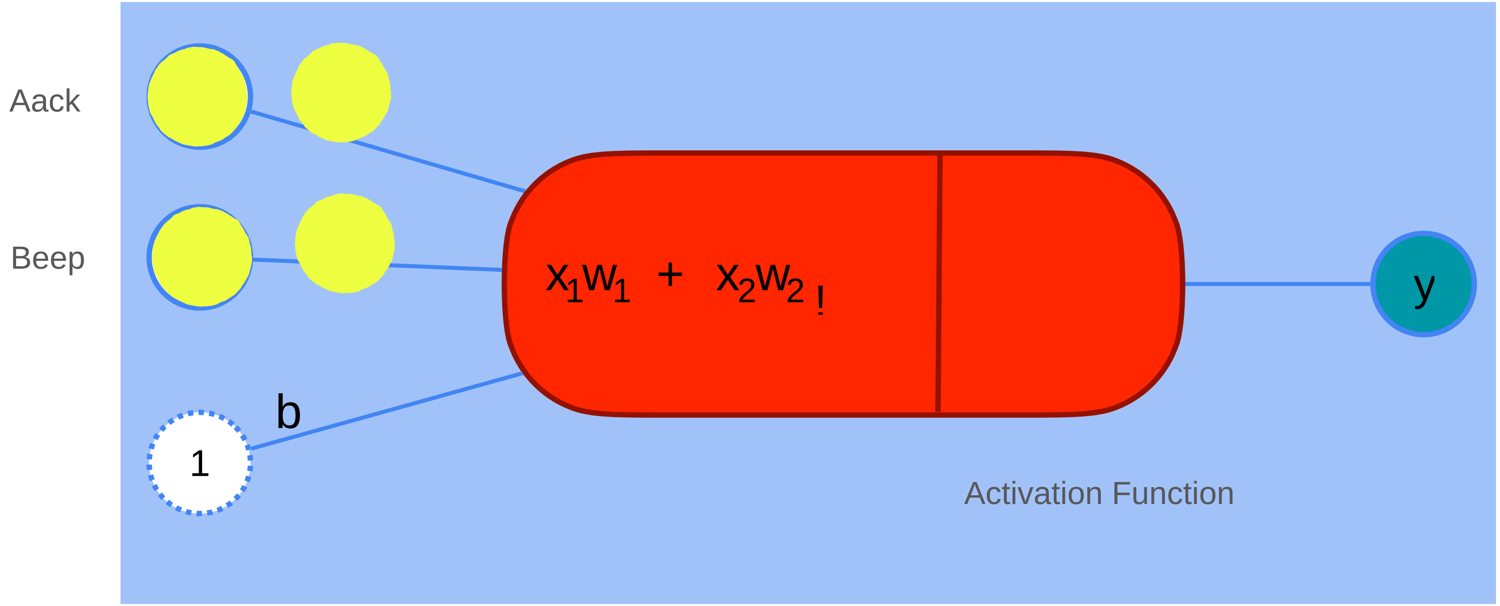
"Classification With a Perceptron

Single Layer Neural Network Perceptron

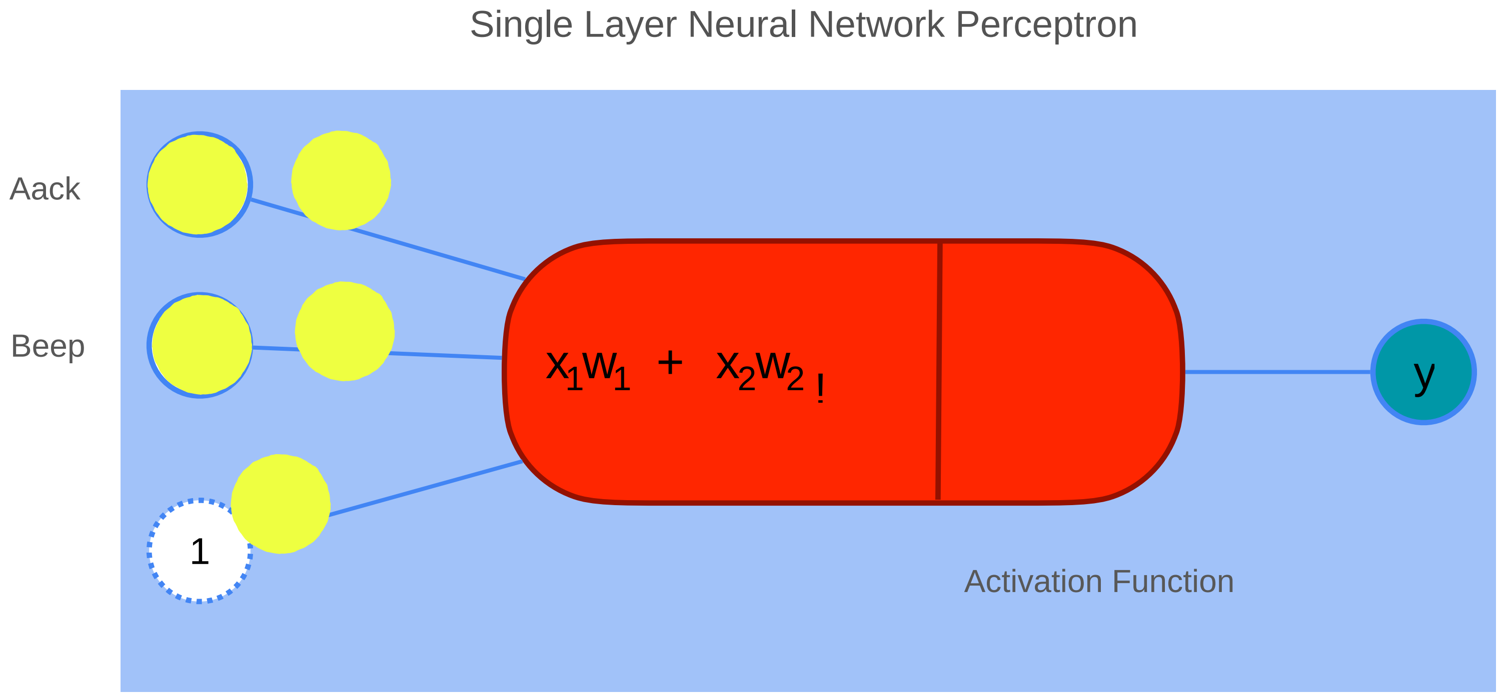


"Classification With a Perceptron

Single Layer Neural Network Perceptron

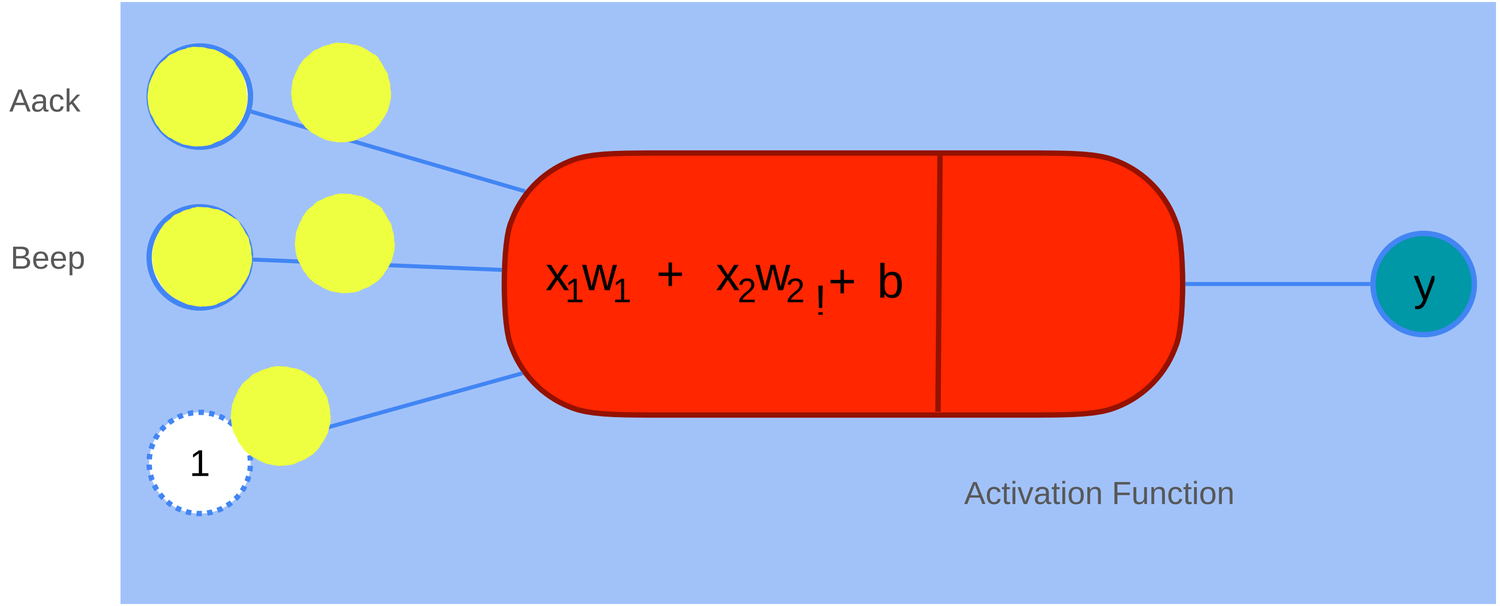


"Classification With a Perceptron

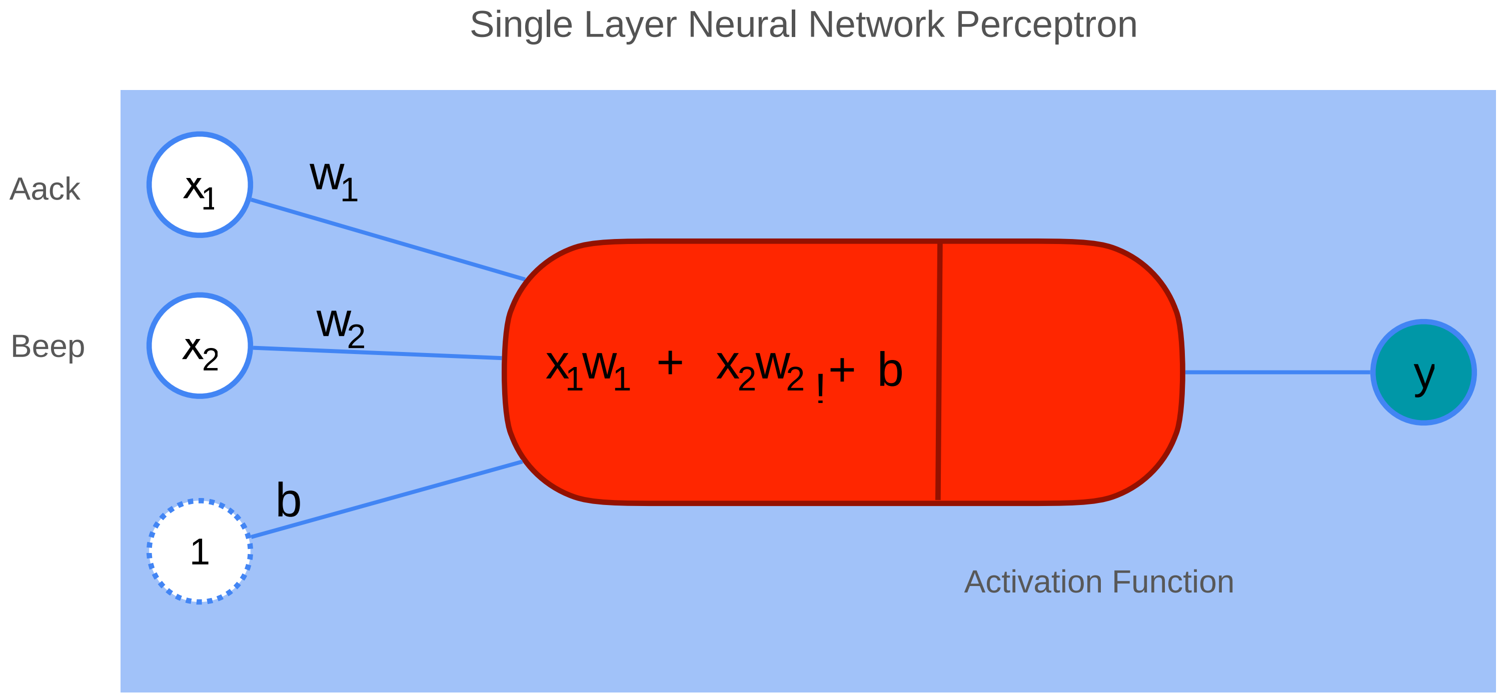


"Classification With a Perceptron

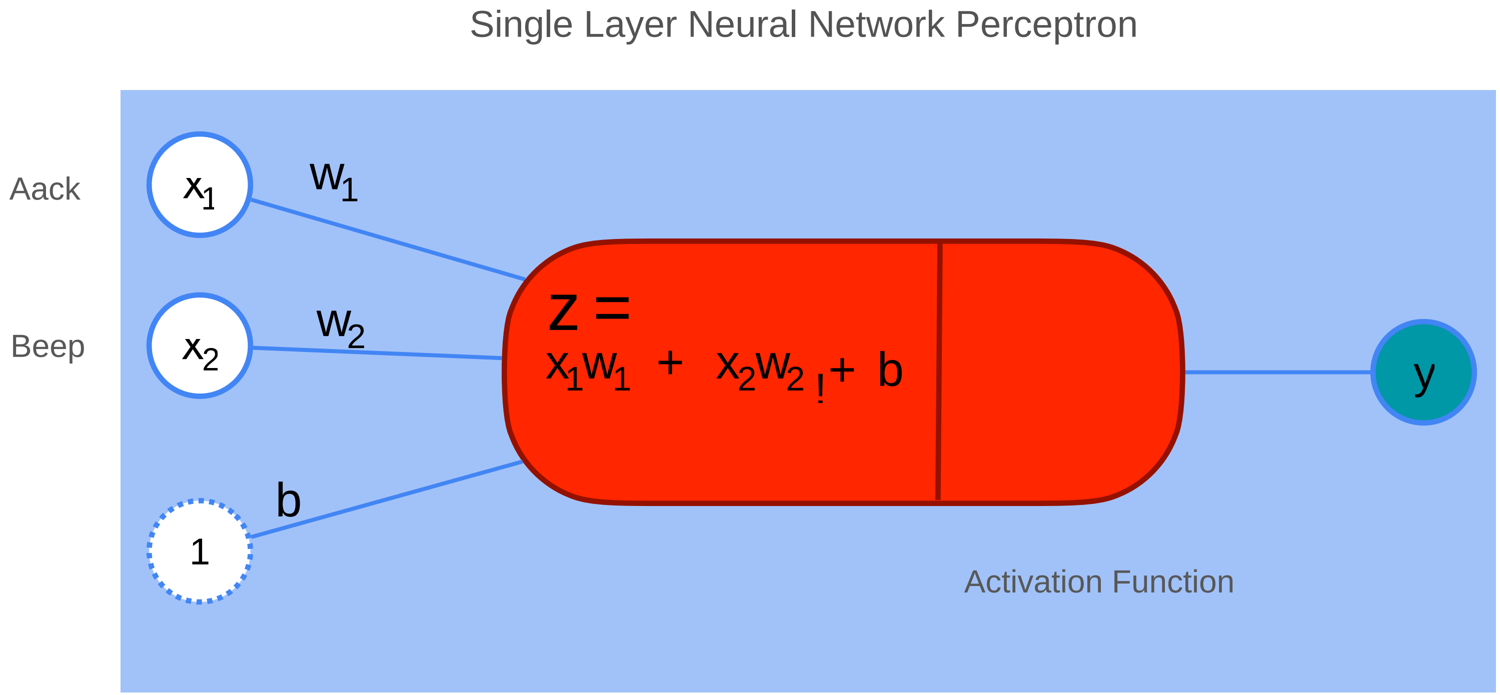
Single Layer Neural Network Perceptron



"Classification With a Perceptron

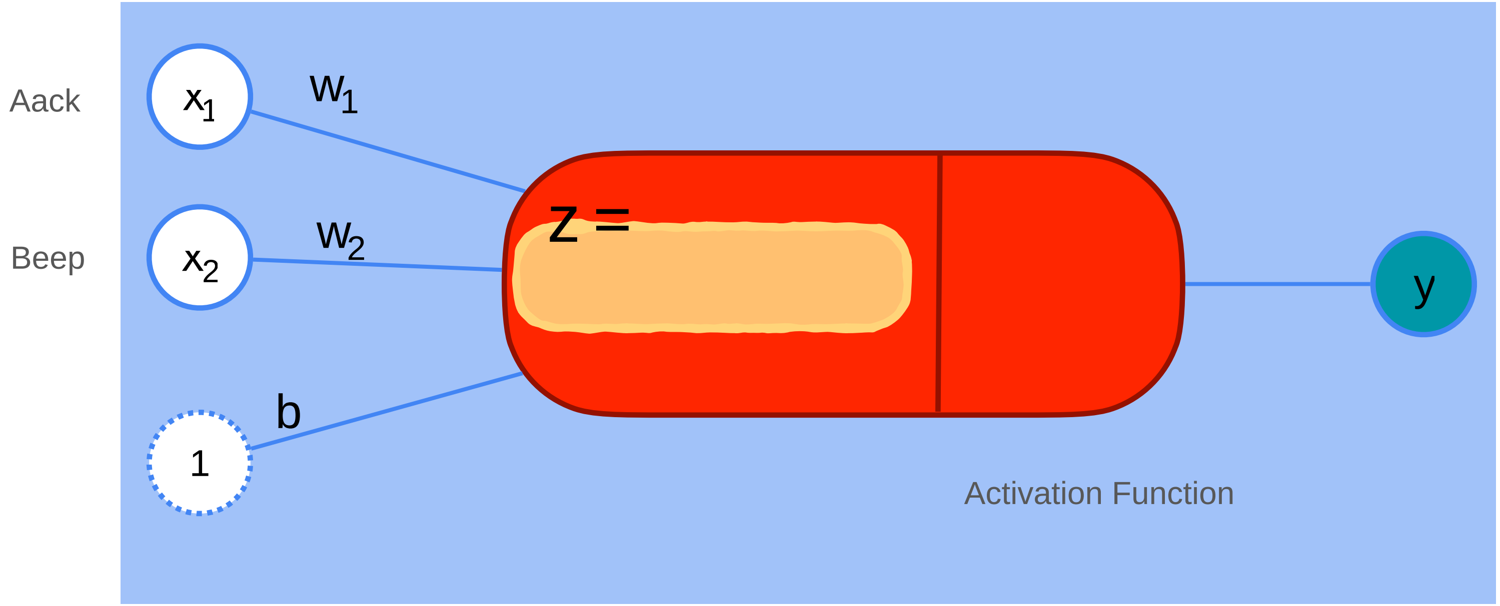


"Classification With a Perceptron



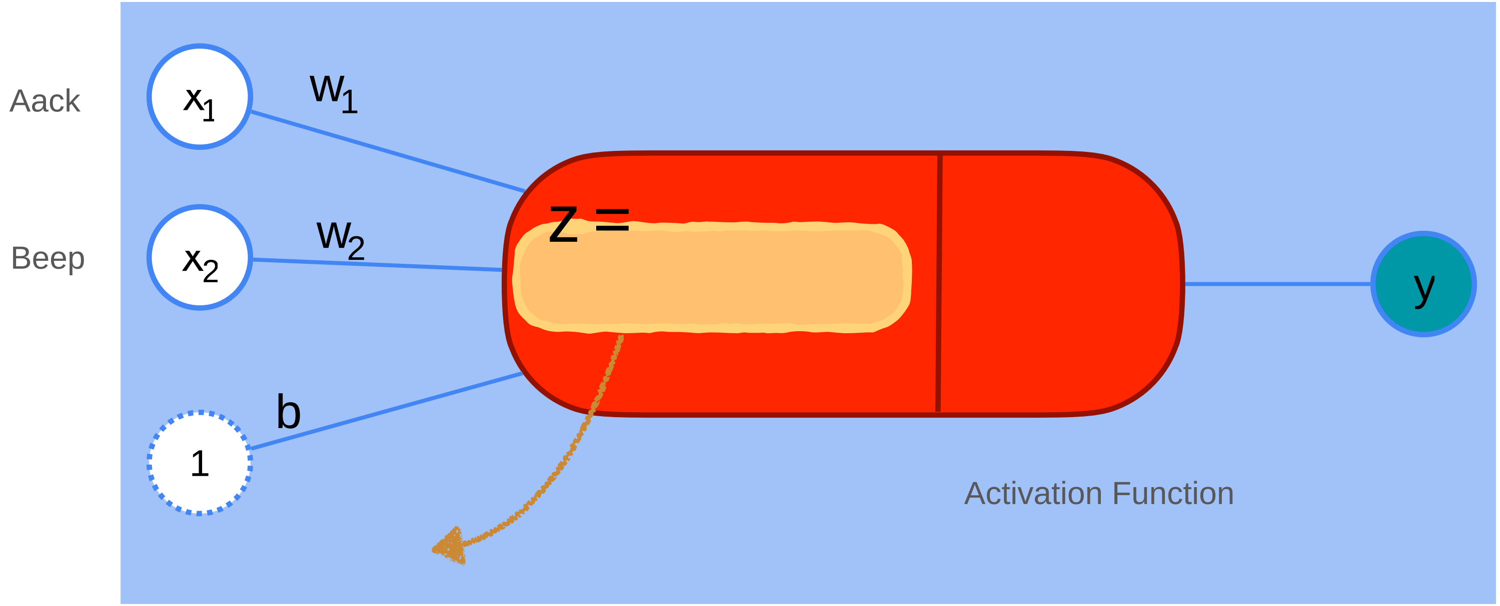
"Classification With a Perceptron

Single Layer Neural Network Perceptron



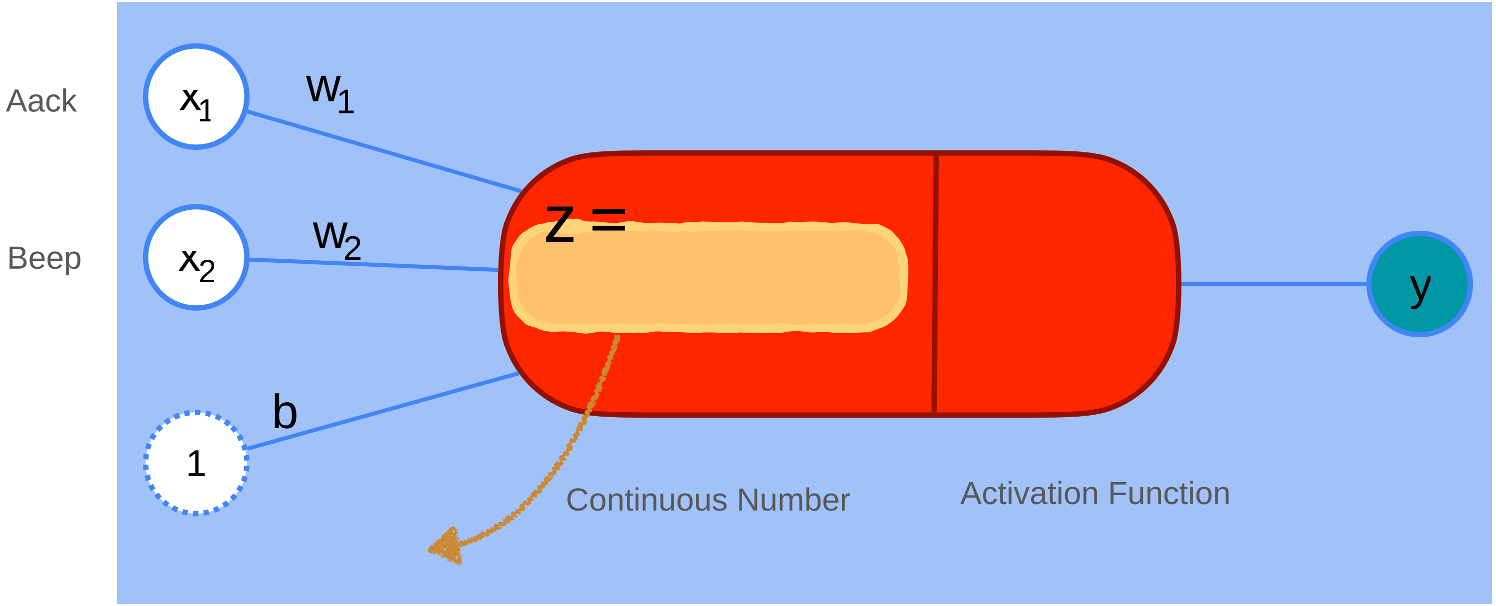
"Classification With a Perceptron

Single Layer Neural Network Perceptron



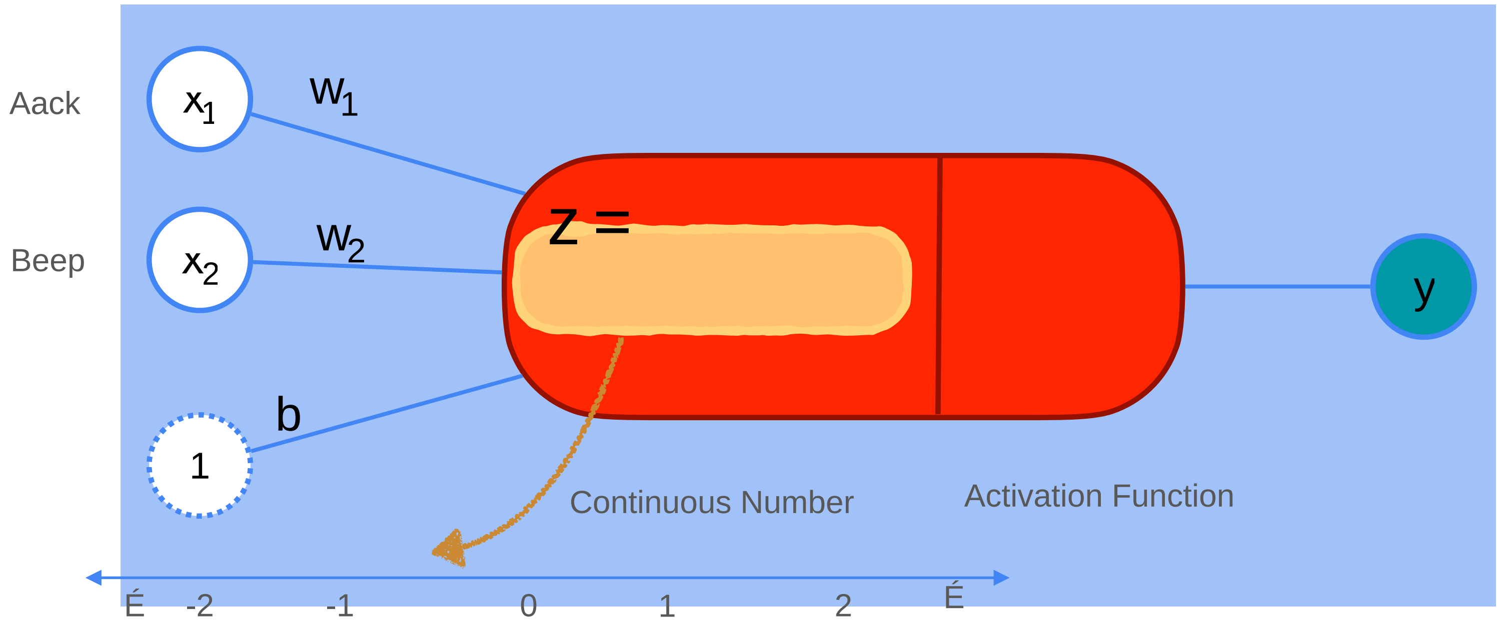
"Classification With a Perceptron

Single Layer Neural Network Perceptron



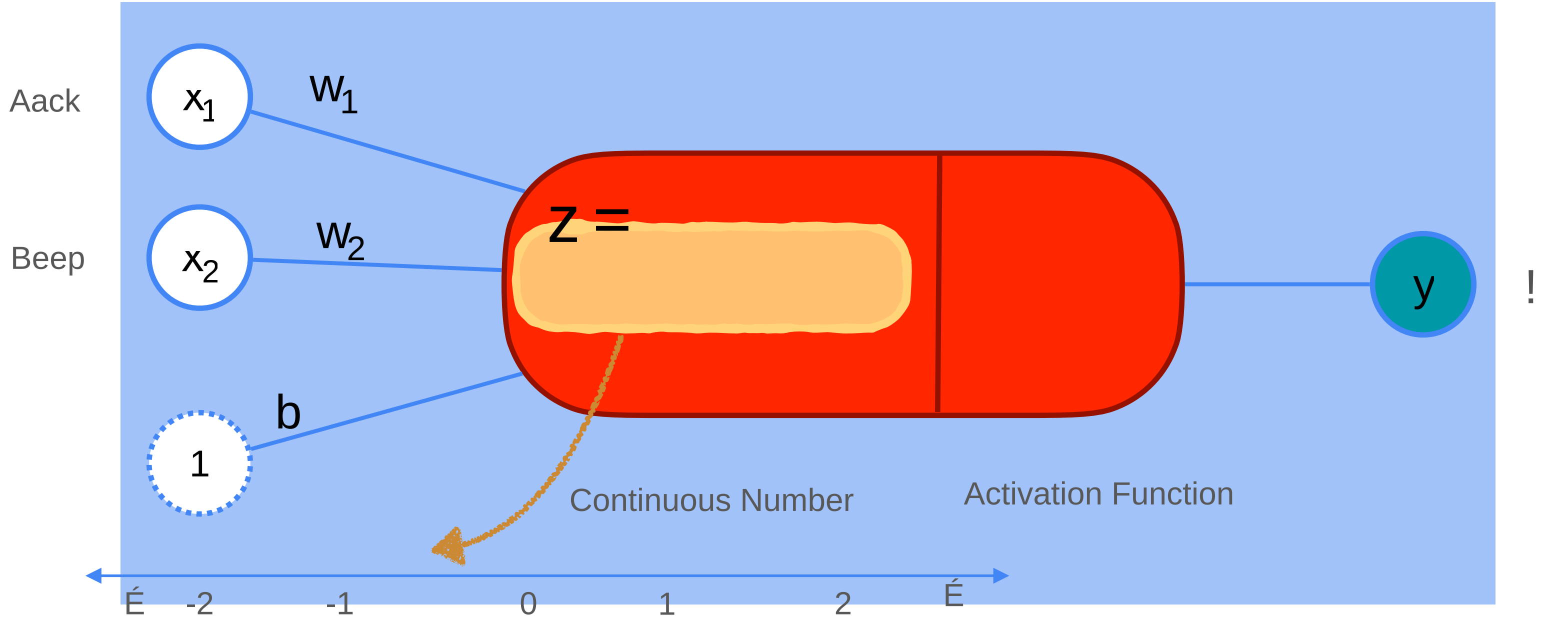
"Classification With a Perceptron

Single Layer Neural Network Perceptron

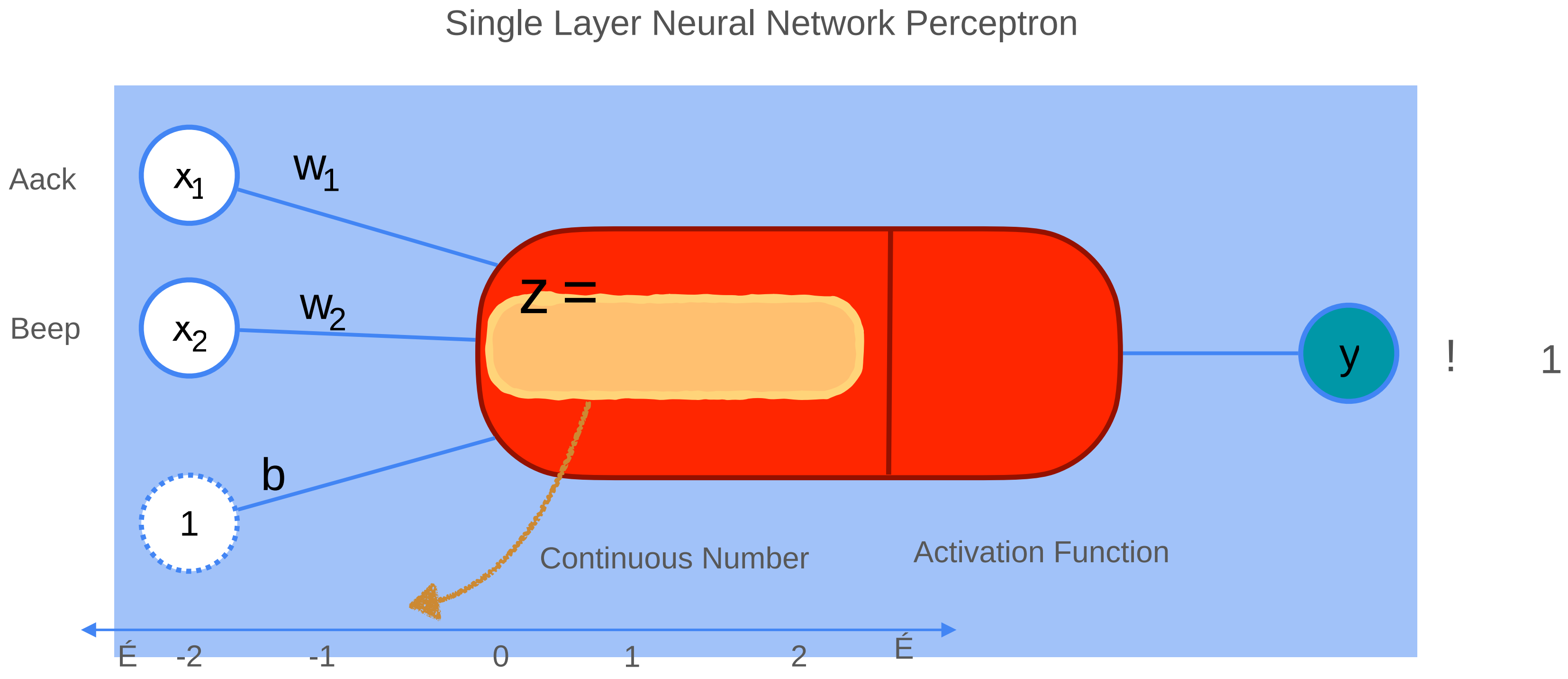


"Classification With a Perceptron

Single Layer Neural Network Perceptron

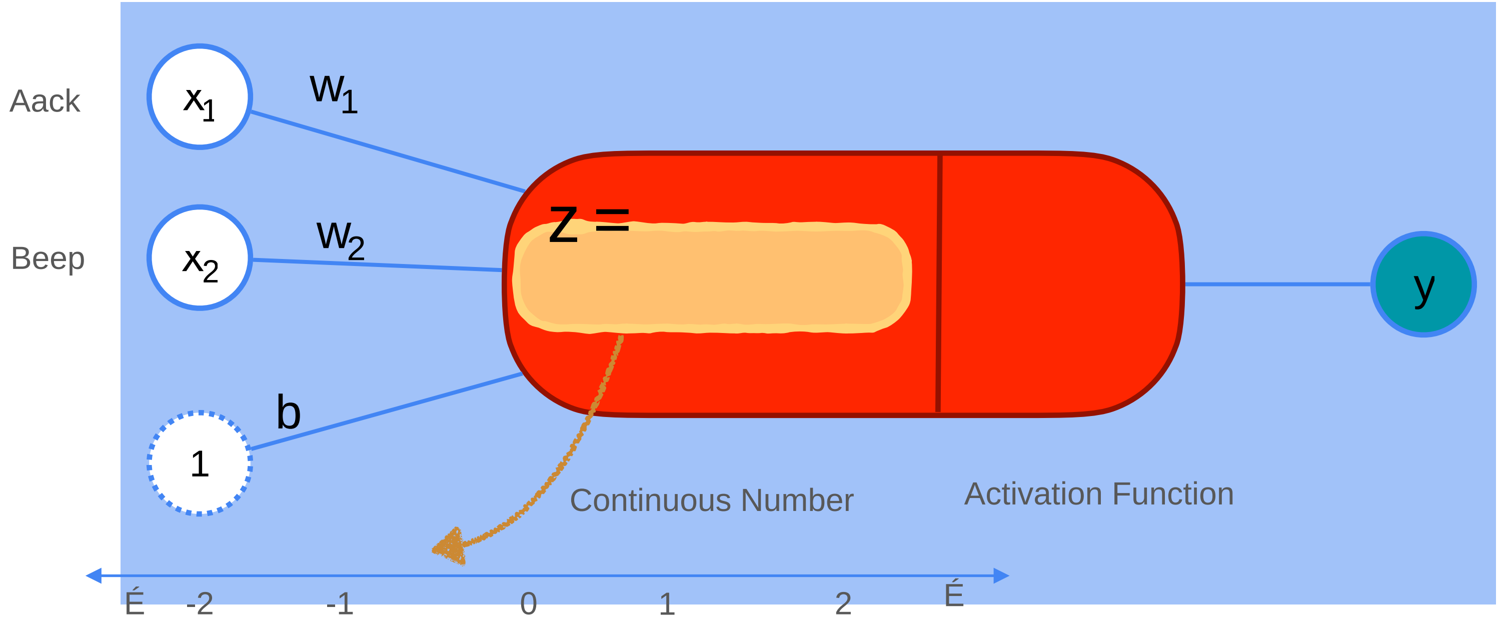


"Classification With a Perceptron



"Classification With a Perceptron

Single Layer Neural Network Perceptron

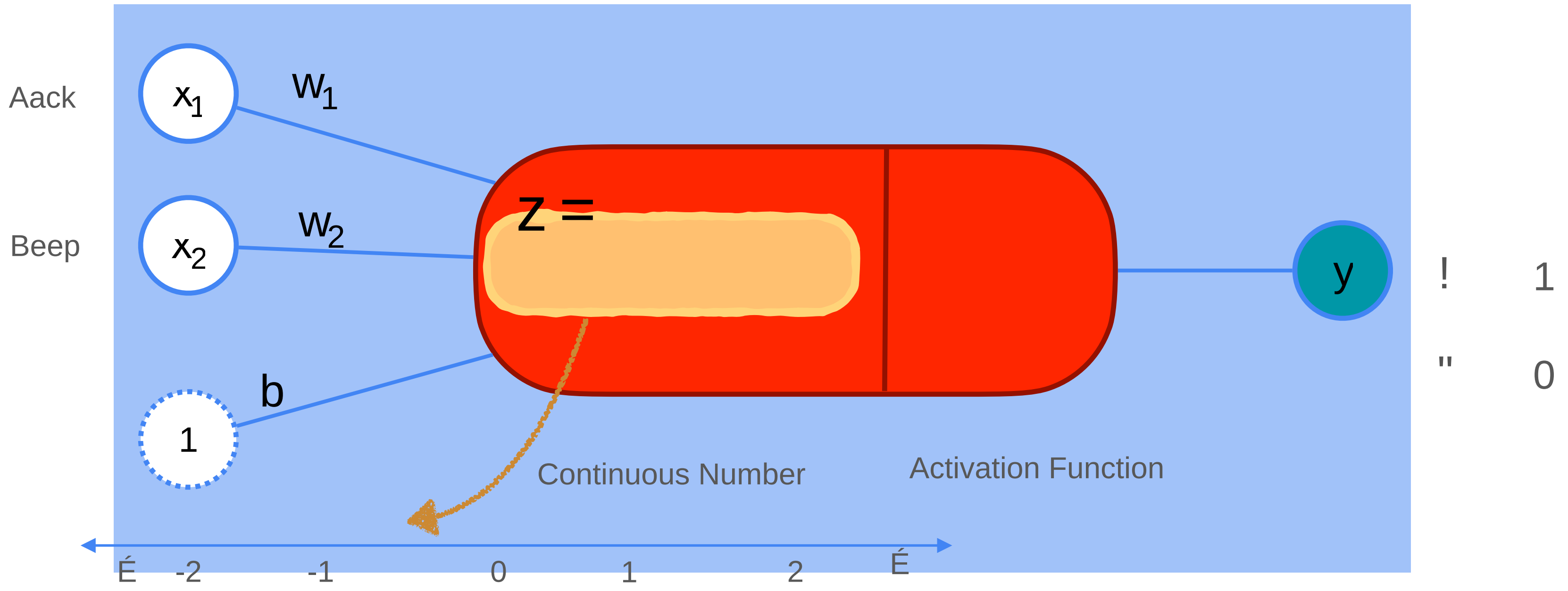


! 1

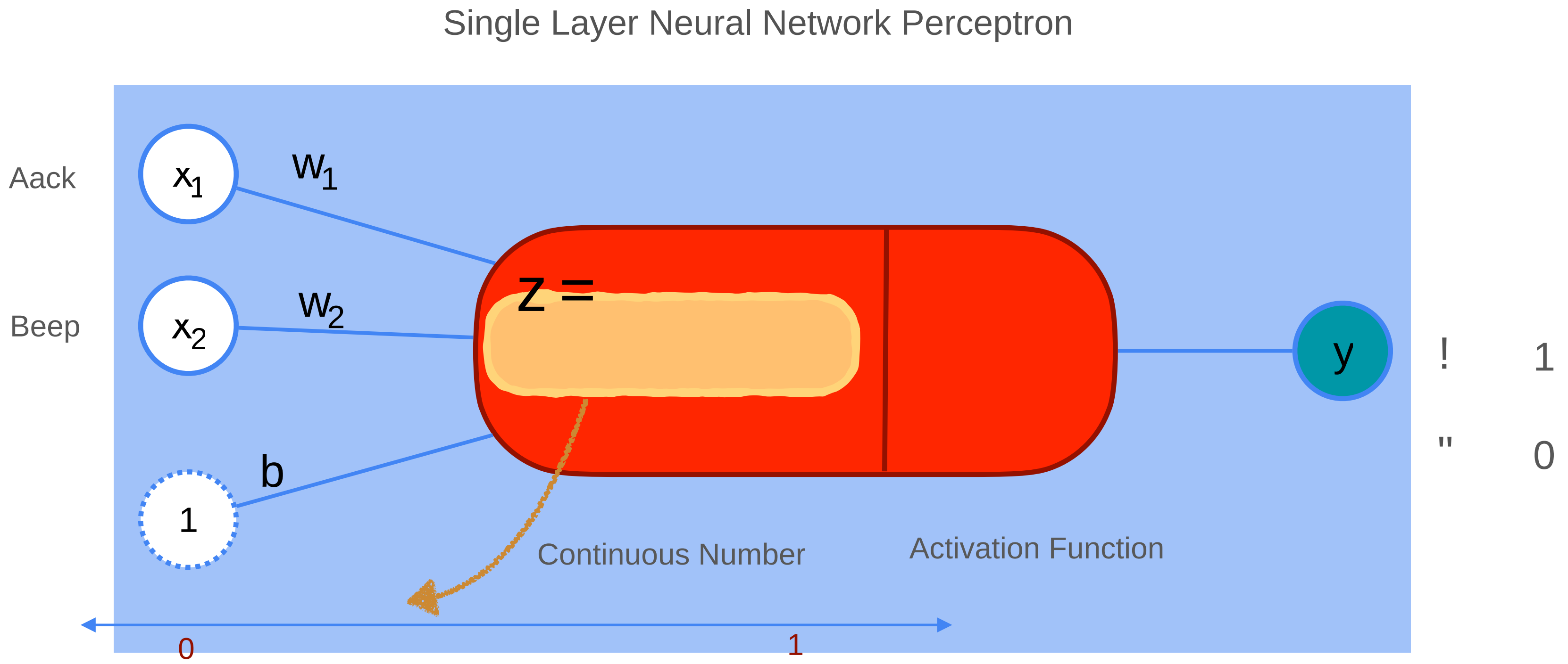
"

"Classification With a Perceptron

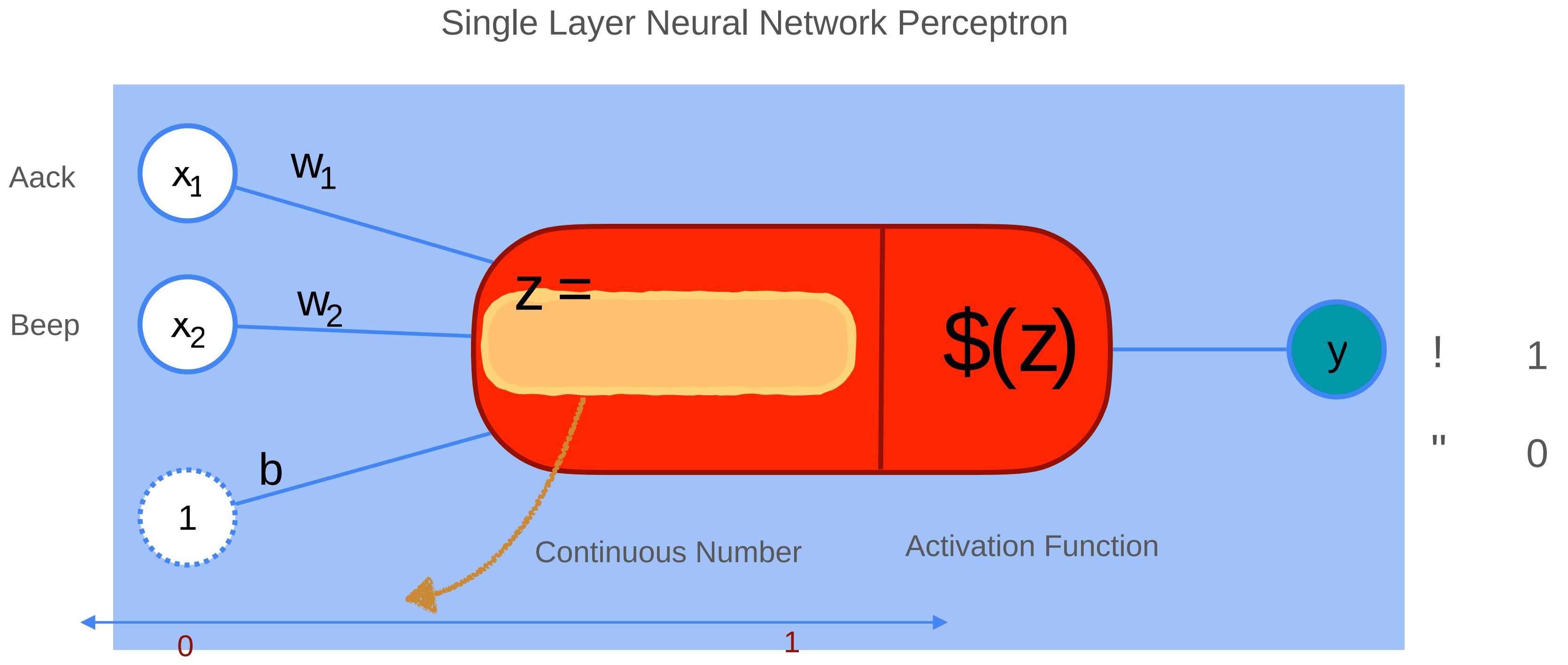
Single Layer Neural Network Perceptron



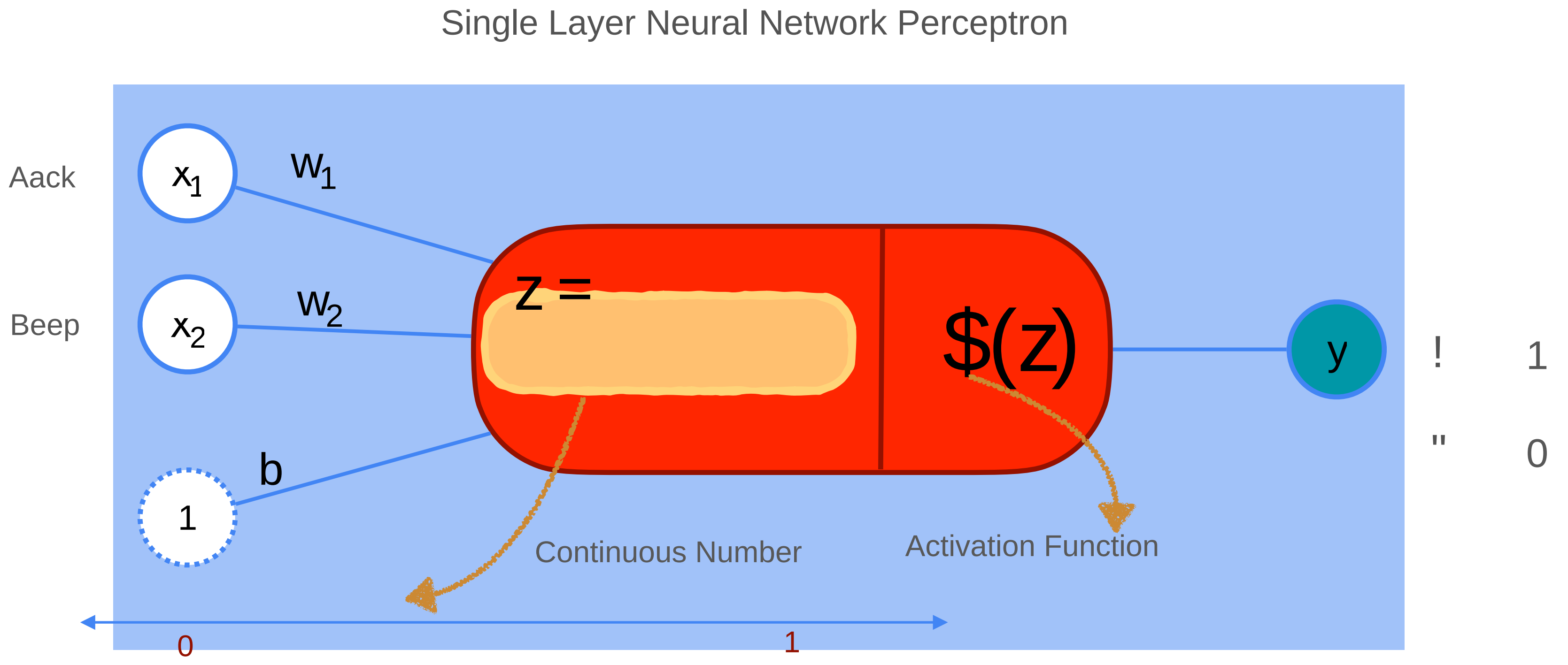
"Classification With a Perceptron



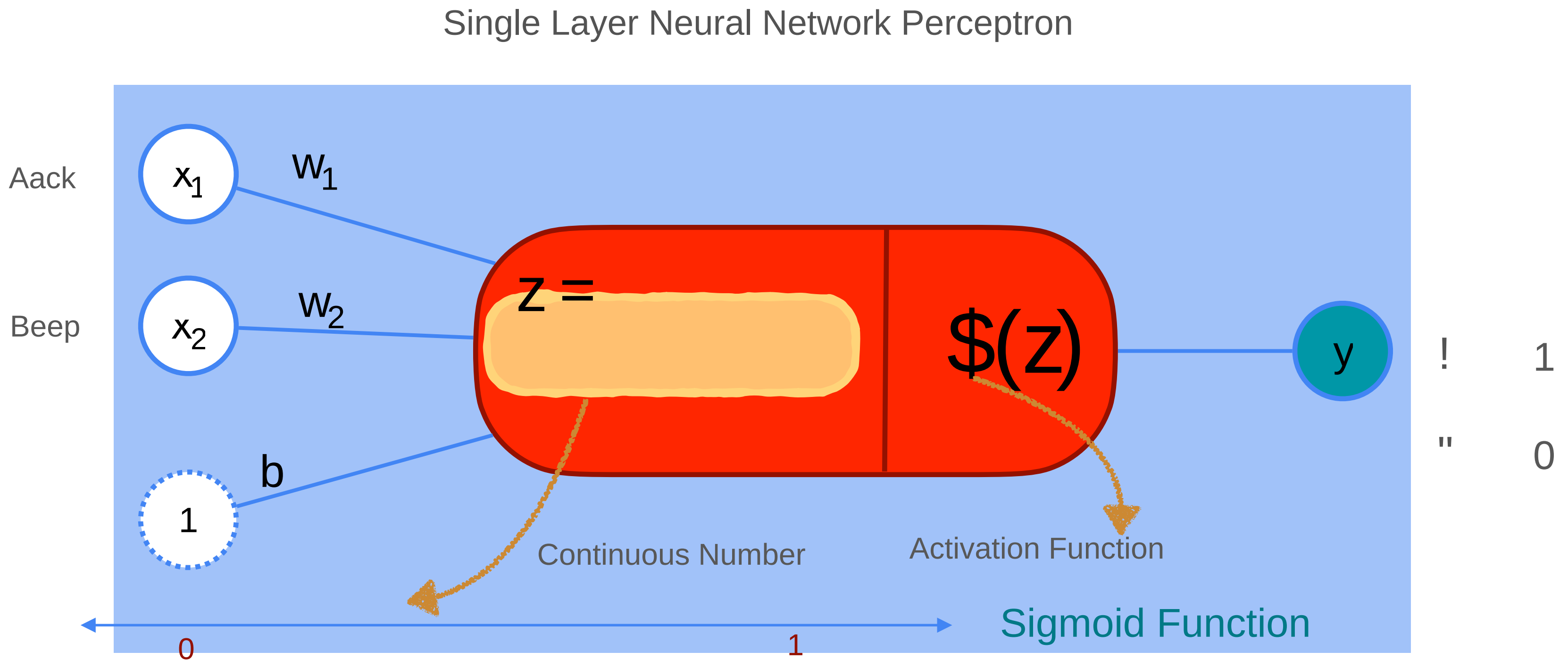
"Classification With a Perceptron



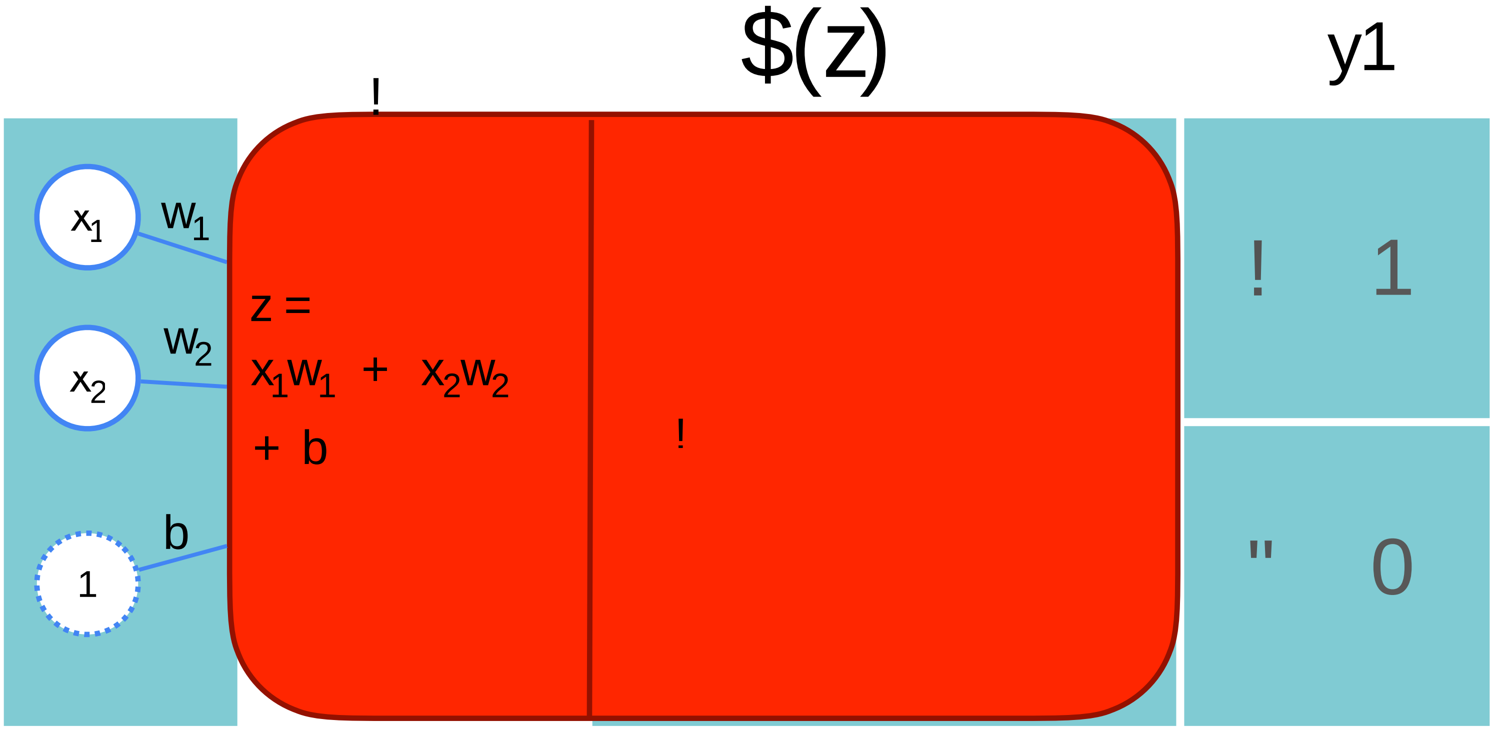
"Classification With a Perceptron



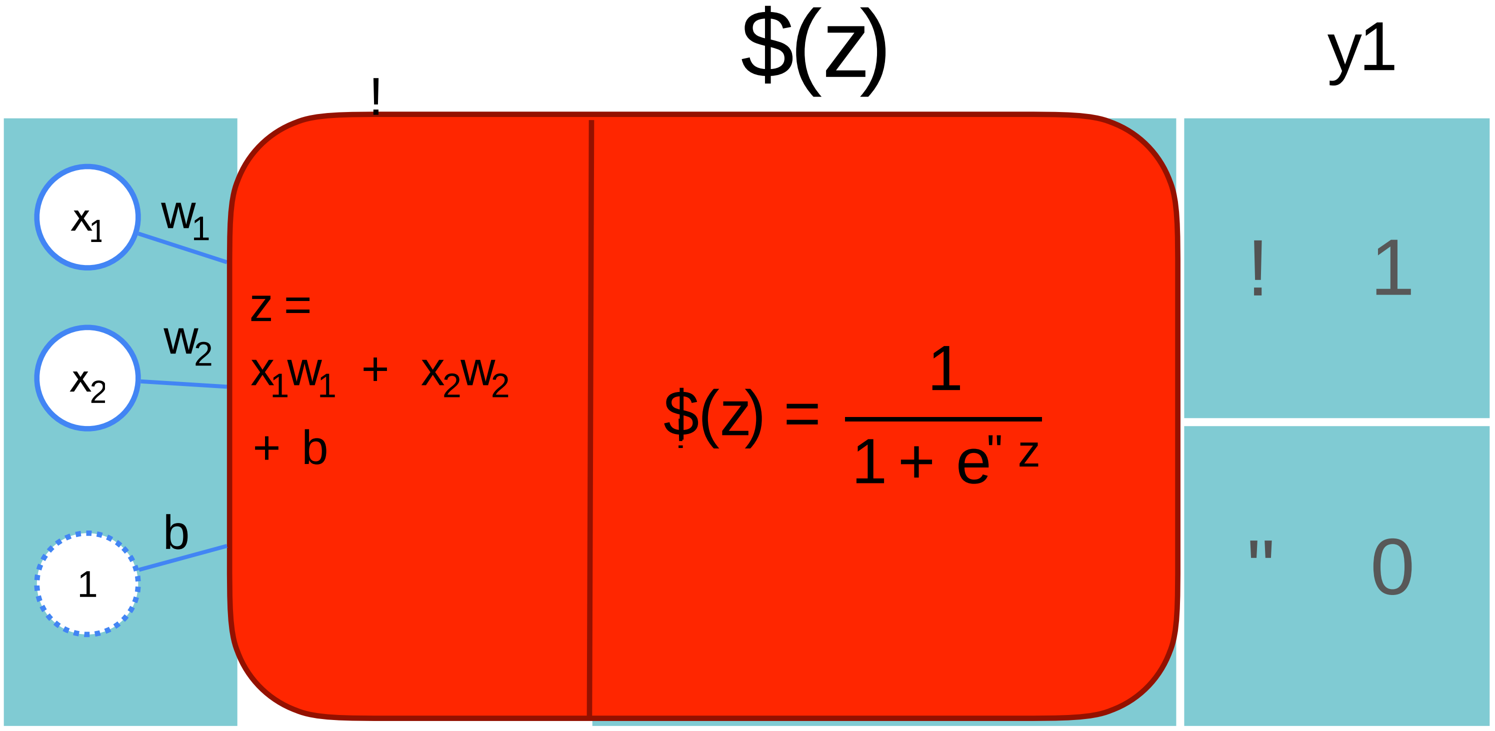
"Classification With a Perceptron



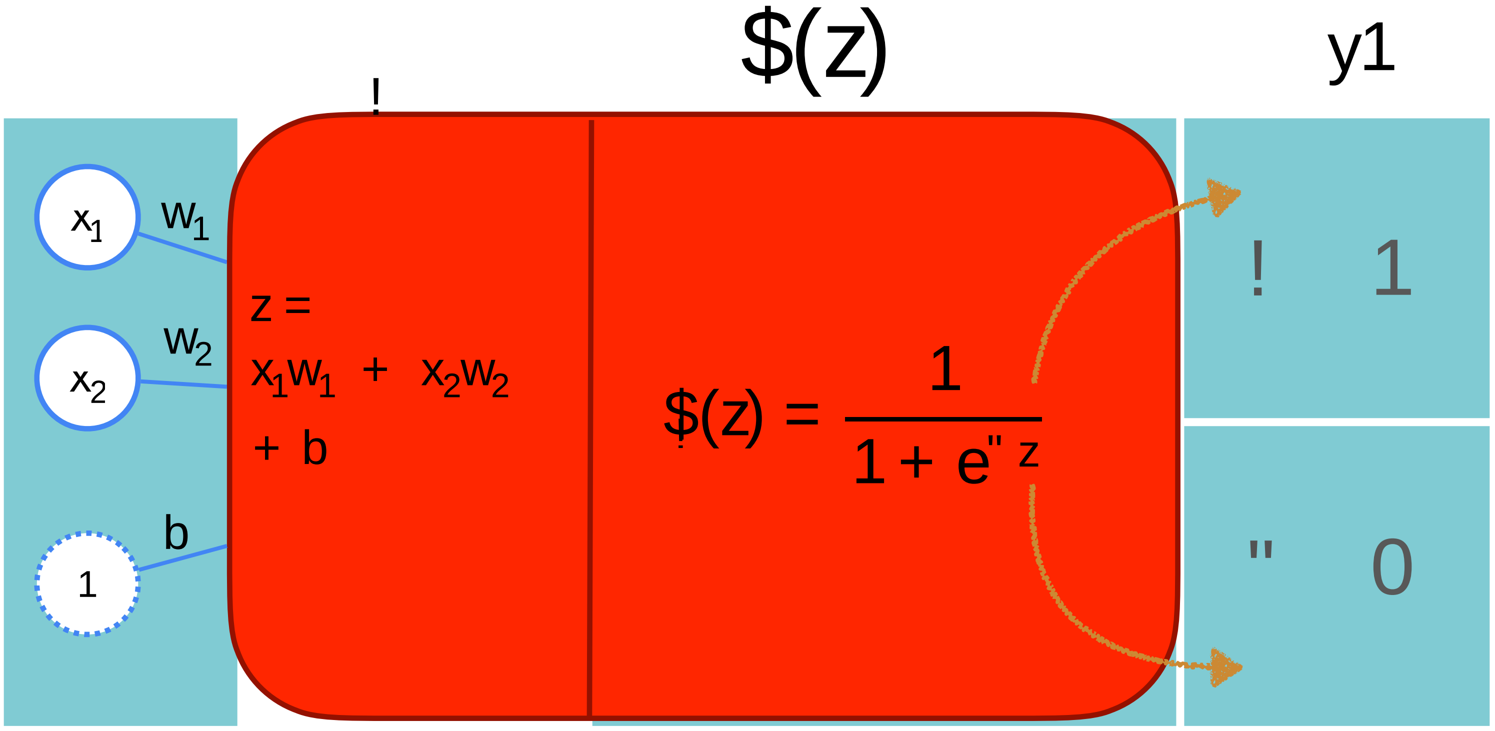
Sigmoid Function



Sigmoid Function



Sigmoid Function



Optimization in Neural Networks and NewtonÕs Method

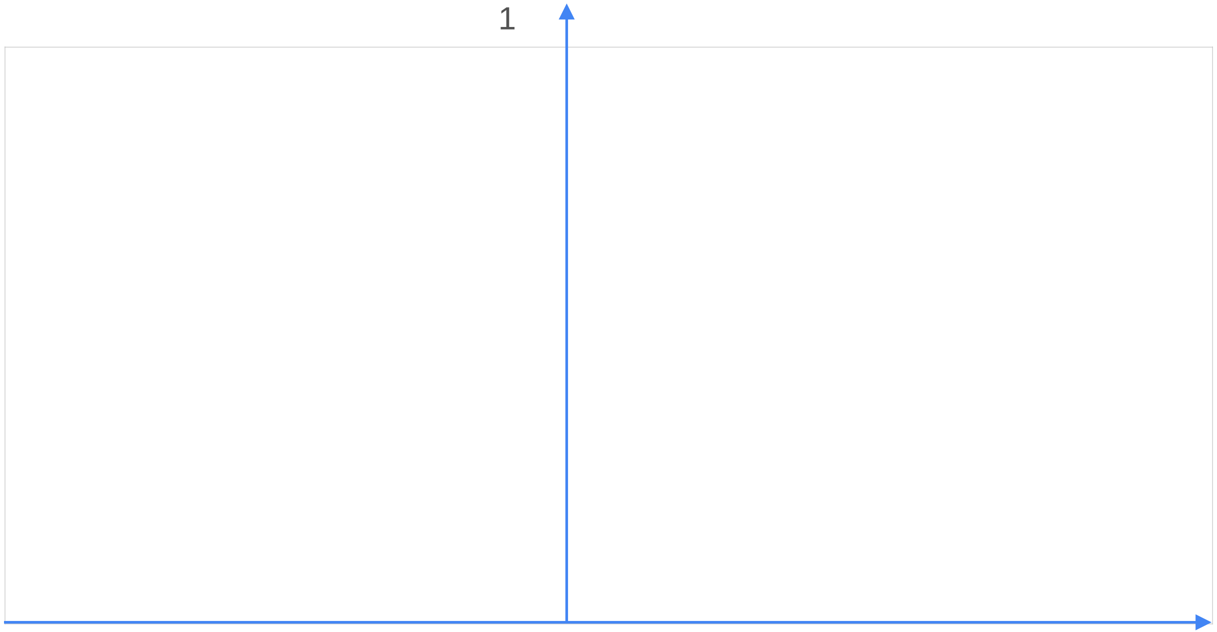
Classification with a perceptron:
The sigmoid function

Sigmoid Function

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

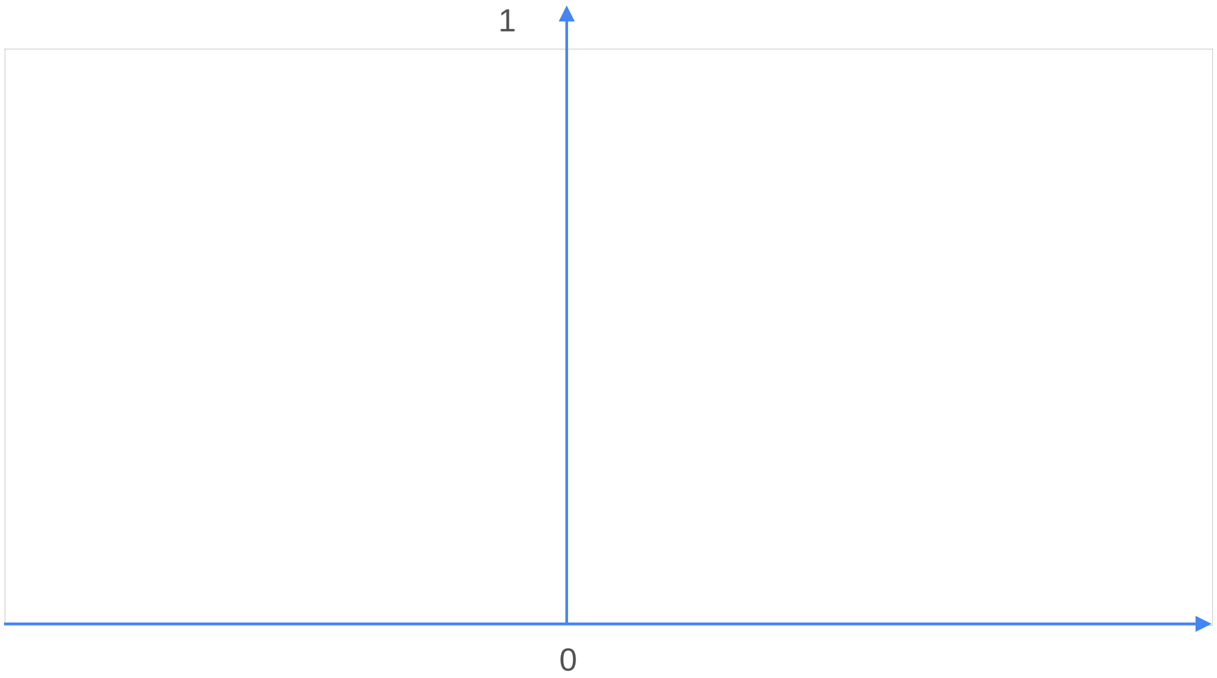
Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{"\, z}}$$



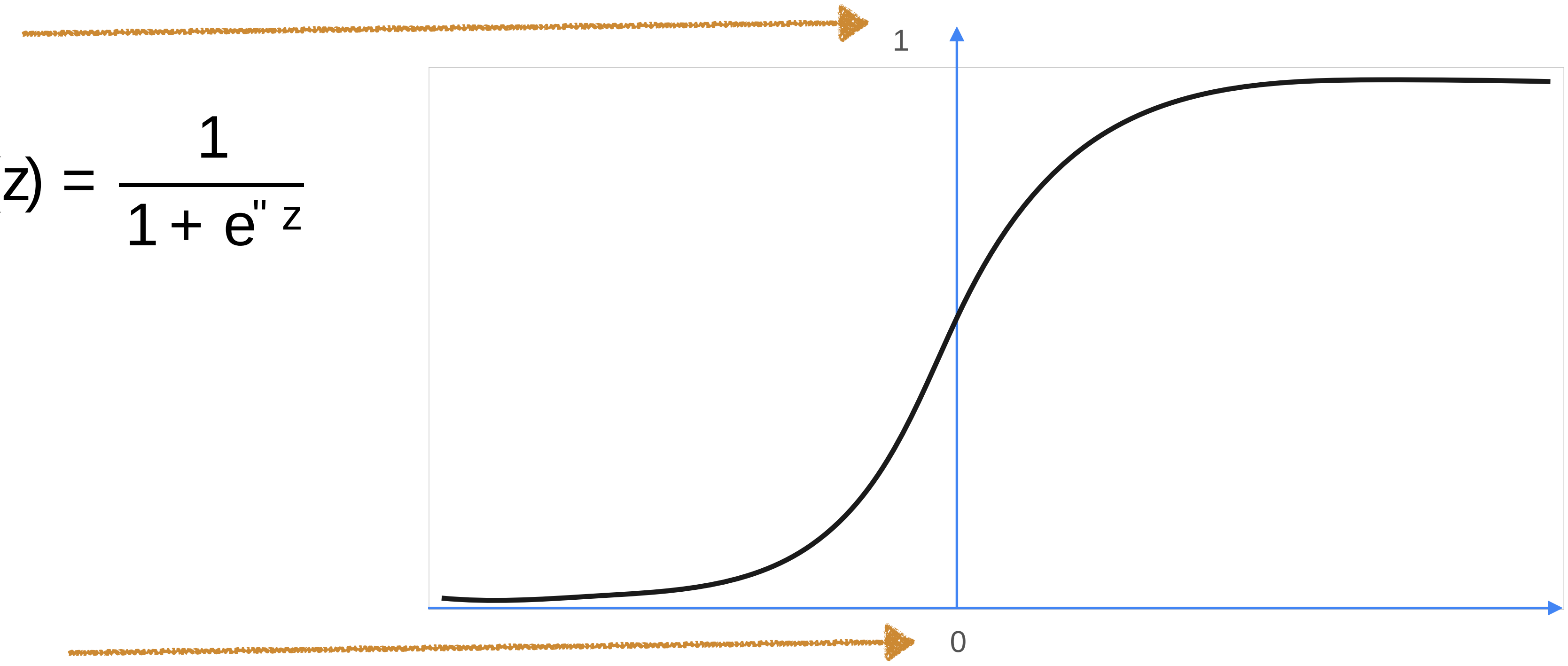
Sigmoid Function

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$



Sigmoid Function

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$



"Derivative of a Sigmoid Function

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{\text{"}\, z}}$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{'' z}}$$

$$\$(z) = (1 + e^{'' z})^{'' 1}$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{'' z}}$$

$$\$(z) = (1 + e^{'' z})^{'' 1}$$

$$\frac{d}{dz}\$(z)$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{"z}}$$

$$\$(z) = (1 + e^{"z})^{"1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{"z})^{"1}$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z)$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = " 1$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{\text{"}\, z}}$$

$$\$(z) = (1 + e^{\text{"}\, z})^{\text{"}\, 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{\text{"}\, z})^{\text{"}\, 1}$$

$$\frac{d}{dz}\$(z) = \text{"}\; 1\,(1 + e^{\text{"}\, z})^{\text{"}\, 1\text{"}\, 1}$$

"Derivative of a Sigmoid Function

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

$$\sigma(z) = (1 + e^{-z})^{-1}$$

$$\frac{d}{dz}\sigma(z) = \frac{d}{dz}(1 + e^{-z})^{-1}$$

$$\frac{d}{dz}\sigma(z) = -1\,(1 + e^{-z})^{-1-1}\left(\frac{d}{dz}(1 + e^{-z})\right)$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{''\,z}}$$

$$\$(z) = (1 + e^{''\,z})^{''\,1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{''\,z})^{''\,1}$$

$$\frac{d}{dz}\$(z) = ''\ 1\ (1 + e^{''\,z})^{''\,1''\,1}\ (\frac{d}{dz}(1 + e^{''\,z}))$$

$$= ''\ 1$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1+e^{"\, z}}$$

$$\$(z) = (1+e^{"\, z})^{"\, 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1+e^{"\, z})^{"\, 1}$$

$$\frac{d}{dz}\$(z) = "\ 1\ (1+e^{"\, z})^{"\, 1"\, 1}\ (\frac{d}{dz}(1+e^{"\, z}))$$

$$= "\ 1\ (1+e^{"\, z})^{"\, 2}$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = " \ 1 \ (1 + e^{" z})^{" 1" 1} \ (\frac{d}{dz}(1 + e^{" z}))$$

$$= " \ 1 \ (1 + e^{" z})^{" 2} \ (\frac{d}{dz}(1)$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{'' z}}$$

$$\$(z) = (1 + e^{'' z})^{'' 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{'' z})^{'' 1}$$

$$\frac{d}{dz}\$(z) = '' \; 1 \; (1 + e^{'' z})^{'' 1'' 1} \; (\frac{d}{dz}(1 + e^{'' z}))$$

$$= '' \; 1 \; (1 + e^{'' z})^{'' 2} \; (\frac{d}{dz}(1) + \frac{d}{dz}(e^{'' z}))$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = " 1 (1 + e^{" z})^{" 1" 1} (\frac{d}{dz}(1 + e^{" z}))$$

$$= " 1 (1 + e^{" z})^{" 2} (\frac{d}{dz}(1) + \frac{d}{dz}(e^{" z}))$$

$$= " 1$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = " 1 (1 + e^{" z})^{" 1" 1} (\frac{d}{dz}(1 + e^{" z}))$$

$$= " 1 (1 + e^{" z})^{" 2} (\frac{d}{dz}(1) + \frac{d}{dz}(e^{" z}))$$

$$= " 1 (1 + e^{" z})^{" 2}$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = " \ 1 \ (1 + e^{" z})^{" 1" 1} \ (\frac{d}{dz}(1 + e^{" z}))$$

$$= " \ 1 \ (1 + e^{" z})^{" 2} \ (\frac{d}{dz}(1) + \frac{d}{dz}(e^{" z}))$$

$$= " \ 1 \ (1 + e^{" z})^{" 2} \ (0$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1+e^{" z}}$$

$$\$(z) = (1+e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1+e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = "\ 1\ (1+e^{" z})^{" 1" 1}\ (\frac{d}{dz}(1+e^{" z}))$$

$$= "\ 1\ (1+e^{" z})^{" 2}\ (\frac{d}{dz}(1) + \frac{d}{dz}(e^{" z}))$$

$$= "\ 1\ (1+e^{" z})^{" 2}\ (0 + e^{" z}(\frac{d}{dz}(" z)))$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{\text{"}\,z}}$$

$$\$(z) = (1 + e^{\text{"}\,z})^{\text{"}\,1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{\text{"}\,z})^{\text{"}\,1}$$

$$\frac{d}{dz}\$(z) = \text{"}\; 1\ (1 + e^{\text{"}\,z})^{\text{"}\,1\text{"}\,1}\ (\frac{d}{dz}(1 + e^{\text{"}\,z}))$$

$$= \text{"}\; 1\ (1 + e^{\text{"}\,z})^{\text{"}\,2}\ (\frac{d}{dz}(1) + \frac{d}{dz}(e^{\text{"}\,z}))$$

$$= \text{"}\; 1\ (1 + e^{\text{"}\,z})^{\text{"}\,2}\ (0 + e^{\text{"}\,z}(\frac{d}{dz}(\text{"}\; z)))$$

$$= \text{"}\; 1$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{'' z}}$$

$$\$(z) = (1 + e^{'' z})^{'' 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{'' z})^{'' 1}$$

$$\frac{d}{dz}\$(z) = '' \ 1 \ (1 + e^{'' z})^{'' 1 '' 1} \ (\frac{d}{dz}(1 + e^{'' z}))$$

$$= '' \ 1 \ (1 + e^{'' z})^{'' 2} \ (\frac{d}{dz}(1) + \frac{d}{dz}(e^{'' z}))$$

$$= '' \ 1 \ (1 + e^{'' z})^{'' 2} \ (0 + e^{'' z}(\frac{d}{dz}('' z)))$$

$$= '' \ 1 \ (1 + e^{'' z})^{'' 2}$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1+e^{\text{"} z}}$$

$$\$(z) = (1+e^{\text{"} z})^{\text{"} 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1+e^{\text{"} z})^{\text{"} 1}$$

$$\frac{d}{dz}\$(z) = \text{"}\ 1\ (1+e^{\text{"} z})^{\text{"} 1 \text{"} 1}\ (\frac{d}{dz}(1+e^{\text{"} z}))$$

$$= \text{"}\ 1\ (1+e^{\text{"} z})^{\text{"} 2}\ (\frac{d}{dz}(1) + \frac{d}{dz}(e^{\text{"} z}))$$

$$= \text{"}\ 1\ (1+e^{\text{"} z})^{\text{"} 2}\ (0 + e^{\text{"} z}(\frac{d}{dz}(\text{"}\ z)))$$

$$= \text{"}\ 1\ (1+e^{\text{"} z})^{\text{"} 2}\ (e^{\text{"} z})$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = " 1 (1 + e^{" z})^{" 1" 1} (\frac{d}{dz}(1 + e^{" z}))$$

$$= " 1 (1 + e^{" z})^{" 2} (\frac{d}{dz}(1) + \frac{d}{dz}(e^{" z}))$$

$$= " 1 (1 + e^{" z})^{" 2} (0 + e^{" z}(\frac{d}{dz}(" z)))$$

$$= " 1 (1 + e^{" z})^{" 2} (e^{" z}) (" 1)$$

"Derivative of a Sigmoid Function

$$\$(z) = \frac{1}{1 + e^{" z}}$$

$$\$(z) = (1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = \frac{d}{dz}(1 + e^{" z})^{" 1}$$

$$\frac{d}{dz}\$(z) = " 1 (1 + e^{" z})^{" 1" 1} (\frac{d}{dz}(1 + e^{" z}))$$

$$= " 1 (1 + e^{" z})^{" 2} (\frac{d}{dz}(1) + \frac{d}{dz}(e^{" z}))$$

$$= " 1 (1 + e^{" z})^{" 2} (0 + e^{" z}(\frac{d}{dz}(" z)))$$

$$= " 1 (1 + e^{" z})^{" 2} (e^{" z}) (" 1)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \text{"}\ 1\ (1 + e^{\text{"}\ z})^{\text{"}\ 2}\ (e^{\text{"}\ z})\ (\text{"}\ 1)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \cancel{"\ 1}\ (1 + e^{"z})^{"2}\ (e^{"z})\ ("\ 1)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\sigma(z) = \cancel{-1}\,(1+e^{-z})^{-2}\,(e^{-z})\,\cancel{(-1)}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\sigma(z) = \cancel{-1}\,(1+e^{-z})^{-2}\,(e^{-z})\,\cancel{(-1)}$$

$$= (1+e^{-z})^{-2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \cancel{"1}\,(1+e^{"z})^{"2}\,(e^{"z})\,\cancel{("1)}$$

$$= (1+e^{"z})^{"2}\,(e^{"z})$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\sigma(z) = \cancel{-1}\,(1+e^{-z})^{-2}\,(e^{-z})\,\cancel{(-1)}$$

$$= (1+e^{-z})^{-2}\,(e^{-z})$$

$$= \frac{1}{(1+e^{-z})^{2}}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \cancel{"\ 1}\,(1+e^{"\ z})^{"\ 2}\ (e^{"\ z})\ \cancel{("\ 1)}$$

$$= (1+e^{"\ z})^{"\ 2}\ (e^{"\ z})$$

$$= \frac{1}{(1+e^{"\ z})^{2}}\ (e^{"\ z})$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\sigma(z) = \cancel{-1}\,(1+e^{-z})^{-2}\,(e^{-z})\,\cancel{(-1)}$$

$$= (1+e^{-z})^{-2}\,(e^{-z})$$

$$= \frac{1}{(1+e^{-z})^{2}}\,(e^{-z})$$

$$= \frac{e^{-z}}{(1+e^{-z})^{2}}$$

"Derivative of a Sigmoid Function

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{''\,z}}{(1 + e^{''\,z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{" z} + 1 " \; 1}{(1 + e^{" z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{"z} + 1 " 1}{(1 + e^{"z})^2}$$

$$= \frac{1 + e^{"z} " 1}{(1 + e^{"z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{\text{"}\,z} + 1 \text{ "}\ 1}{(1 + e^{\text{"}\,z})^2}$$

$$= \frac{1 + e^{\text{"}\,z} \text{ "}\ 1}{(1 + e^{\text{"}\,z})^2}$$

$$= \frac{1 + e^{\text{"}\,z}}{(1 + e^{\text{"}\,z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\sigma(z) = \frac{e^{-z} + 1 - 1}{(1 + e^{-z})^2}$$

$$= \frac{1 + e^{-z} - 1}{(1 + e^{-z})^2}$$

$$= \frac{1 + e^{-z}}{(1 + e^{-z})^2} - \frac{1}{(1 + e^{-z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{" z} + 1 " 1}{(1 + e^{" z})^2}$$

$$= \frac{1 + e^{" z} " 1}{(1 + e^{" z})^2}$$

$$= \frac{\cancel{1 + e^{" z}}}{(1 + e^{" z})^2} " \frac{1}{(1 + e^{" z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{\text{"}\, z} + 1 \text{"}\ 1}{(1 + e^{\text{"}\, z})^2}$$

$$= \frac{1 + e^{\text{"}\, z} \text{"}\ 1}{(1 + e^{\text{"}\, z})^2}$$

$$= \frac{\cancel{1 + e^{\text{"}\, z}}}{(\cancel{1 + e^{\text{"}\, z}})^2} \ \text{"}\ \frac{1}{(1 + e^{\text{"}\, z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{" z} + 1 " 1}{(1 + e^{" z})^2}$$

$$= \frac{1 + e^{" z} " 1}{(1 + e^{" z})^2}$$

$$= \frac{\cancel{1 + e^{" z}}}{(\cancel{1 + e^{" z}})^2} \;"\; \frac{1}{(1 + e^{" z})^2}$$

$$= \frac{1}{(1 + e^{" z})}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{''\,z} + 1 \,''\, 1}{(1+e^{''\,z})^2}$$

$$= \frac{1+e^{''\,z} \,''\, 1}{(1+e^{''\,z})^2}$$

$$= \frac{\cancel{1+e^{''\,z}}}{(\cancel{1+e^{''\,z}})^2} \;''\; \frac{1}{(1+e^{''\,z})^2}$$

$$= \frac{1}{(1+e^{''\,z})} \;''\; \frac{1}{(1+e^{''\,z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\sigma(z) = \frac{e^{-z} + 1 - 1}{(1+e^{-z})^2}$$

$$= \frac{1 + e^{-z} - 1}{(1+e^{-z})^2}$$

$$= \frac{\cancel{1+e^{-z}}}{(1+e^{-z})^{\cancel{2}}} - \frac{1}{(1+e^{-z})^2}$$

$$= \frac{1}{(1+e^{-z})} - \frac{1}{(1+e^{-z})^2}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{\text{"}\,z} + 1 \text{ " } 1}{(1 + e^{\text{"}\,z})^2}$$

$$= \frac{1 + e^{\text{"}\,z} \text{ " } 1}{(1 + e^{\text{"}\,z})^2}$$

$$= \frac{\cancel{1 + e^{\text{"}\,z}}}{(\cancel{1 + e^{\text{"}\,z}})^2} \text{ " } \frac{1}{(1 + e^{\text{"}\,z})^2}$$

$$= \frac{1}{(1 + e^{\text{"}\,z})} \text{ " } \frac{1}{(1 + e^{\text{"}\,z})^2}$$

$$\frac{d}{dz}\$(z)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\sigma(z) = \frac{e^{-z} + 1 - 1}{(1+e^{-z})^2}$$

$$= \frac{1 + e^{-z} - 1}{(1+e^{-z})^2}$$

$$= \frac{\cancel{1 + e^{-z}}}{(\cancel{1+e^{-z}})^2} - \frac{1}{(1+e^{-z})^2}$$

$$= \frac{1}{(1+e^{-z})} - \frac{1}{(1+e^{-z})^2}$$

$$\frac{d}{dz}\sigma(z) = \frac{1}{(1+e^{-z})}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{''z} + 1 '' 1}{(1+e^{''z})^2}$$

$$= \frac{1+e^{''z} '' 1}{(1+e^{''z})^2}$$

$$= \frac{\cancel{1+e^{''z}}}{(\cancel{1+e^{''z}})^2} \; '' \; \frac{1}{(1+e^{''z})^2}$$

$$= \frac{1}{(1+e^{''z})} \; '' \; \frac{1}{(1+e^{''z})^2}$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1+e^{''z})} \; ''$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{" z} + 1 " 1}{(1 + e^{" z})^2}$$

$$= \frac{1 + e^{" z} " 1}{(1 + e^{" z})^2}$$

$$= \frac{\cancel{1 + e^{" z}}}{(\cancel{1 + e^{" z}})^2} \;" \; \frac{1}{(1 + e^{" z})^2}$$

$$= \frac{1}{(1 + e^{" z})} \;"$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1 + e^{" z})} \;"$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{\text{"}z} + 1 \text{"} 1}{(1 + e^{\text{"}z})^2}$$

$$= \frac{1 + e^{\text{"}z} \text{"} 1}{(1 + e^{\text{"}z})^2}$$

$$= \frac{\cancel{1 + e^{\text{"}z}}}{(\cancel{1 + e^{\text{"}z}})^2} \text{"} \frac{1}{(1 + e^{\text{"}z})^2}$$

$$= \frac{1}{(1 + e^{\text{"}z})} \text{"}$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1 + e^{\text{"}z})} \text{"} \left(\frac{1}{(1 + e^{\text{"}z})} \right)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\sigma(z) = \frac{e^{-z}+1-1}{(1+e^{-z})^2}$$

$$= \frac{1+e^{-z}-1}{(1+e^{-z})^2}$$

$$= \frac{\cancel{1+e^{-z}}}{(\cancel{1+e^{-z}})^2} - \frac{1}{(1+e^{-z})^2}$$

$$= \frac{1}{(1+e^{-z})} -$$

$$\frac{d}{dz}\sigma(z) = \frac{1}{(1+e^{-z})} - \left(\frac{1}{(1+e^{-z})}\right)\left(\frac{1}{(1+e^{-z})}\right)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{\text{"}\, z} + 1 \text{"}\; 1}{(1+e^{\text{"}\, z})^2}$$

$$= \frac{1+e^{\text{"}\, z} \text{"}\; 1}{(1+e^{\text{"}\, z})^2}$$

$$= \frac{\cancel{1+e^{\text{"}\, z}}}{(\cancel{1+e^{\text{"}\, z}})^2} \;\text{"}\; \frac{1}{(1+e^{\text{"}\, z})^2}$$

$$= \frac{1}{(1+e^{\text{"}\, z})} \;\text{"}$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1+e^{\text{"}\, z})} \;\text{"}\; \left(\frac{1}{(1+e^{\text{"}\, z})}\right)\left(\frac{1}{(1+e^{\text{"}\, z})}\right)$$

$$= \frac{1}{(1+e^{\text{"}\, z})}$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{" z} + 1 " \; 1}{(1 + e^{" z})^2}$$

$$= \frac{1 + e^{" z} " \; 1}{(1 + e^{" z})^2}$$

$$= \frac{\cancel{1 + e^{" z}}}{(\cancel{1 + e^{" z}})^2} \; " \; \frac{1}{(1 + e^{" z})^2}$$

$$= \frac{1}{(1 + e^{" z})} \; "$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1 + e^{" z})} \; " \; \left(\frac{1}{(1 + e^{" z})} \right) \left(\frac{1}{(1 + e^{" z})} \right)$$

$$= \frac{1}{(1 + e^{" z})} \left(1 \; " \; \frac{1}{(1 + e^{" z})} \right)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{\text{"}\, z} + 1 \text{"}\ 1}{(1+e^{\text{"}\, z})^2}$$

$$= \frac{1+e^{\text{"}\, z} \text{"}\ 1}{(1+e^{\text{"}\, z})^2}$$

$$= \frac{\cancel{1+e^{\text{"}\, z}}}{(\cancel{1+e^{\text{"}\, z}})^2} \text{"}\ \frac{1}{(1+e^{\text{"}\, z})^2}$$

$$= \frac{1}{(1+e^{\text{"}\, z})} \text{"}$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1+e^{\text{"}\, z})} \text{"} \left(\frac{1}{(1+e^{\text{"}\, z})}\right)\left(\frac{1}{(1+e^{\text{"}\, z})}\right)$$

$$= \frac{1}{(1+e^{\text{"}\, z})}\left(1 \text{"} \frac{1}{(1+e^{\text{"}\, z})}\right)$$

Recall that:

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{''z} + 1 '' 1}{(1+e^{''z})^2}$$

$$= \frac{1+e^{''z}''1}{(1+e^{''z})^2}$$

$$= \frac{\cancel{1+e^{''z}}}{(\cancel{1+e^{''z}})^2} \;''\; \frac{1}{(1+e^{''z})^2}$$

$$= \frac{1}{(1+e^{''z})} \;''$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1+e^{''z})} \;''\; \left(\frac{1}{(1+e^{''z})}\right)\left(\frac{1}{(1+e^{''z})}\right)$$

$$= \frac{1}{(1+e^{''z})}\left(1 '' \frac{1}{(1+e^{''z})}\right)$$

Recall that: $\$(z) = \frac{1}{1+e^{''z}}$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{"z} + 1 " 1}{(1 + e^{"z})^2}$$

$$= \frac{1 + e^{"z} " 1}{(1 + e^{"z})^2}$$

$$= \frac{\cancel{1 + e^{"z}}}{(\cancel{1 + e^{"z}})^2} " \frac{1}{(1 + e^{"z})^2}$$

$$= \frac{1}{(1 + e^{"z})} "$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1 + e^{"z})} " \left(\frac{1}{(1 + e^{"z})}\right)\left(\frac{1}{(1 + e^{"z})}\right)$$

$$= \frac{1}{(1 + e^{"z})} \left(1 " \frac{1}{(1 + e^{"z})}\right)$$

Recall that: $\$(z) = \frac{1}{1 + e^{"z}}$

$$\frac{d}{dz}\$(z)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{\text{"} z} + 1 \text{"} \; 1}{(1+e^{\text{"} z})^2}$$

$$= \frac{1+e^{\text{"} z} \text{"} \; 1}{(1+e^{\text{"} z})^2}$$

$$= \frac{1\cancel{+e^{\text{"} z}}}{(1\cancel{+e^{\text{"} z}})^2} \; \text{"} \; \frac{1}{(1+e^{\text{"} z})^2}$$

$$= \frac{1}{(1+e^{\text{"} z})} \; \text{"}$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1+e^{\text{"} z})} \; \text{"} \; \left(\frac{1}{(1+e^{\text{"} z})}\right)\left(\frac{1}{(1+e^{\text{"} z})}\right)$$

$$= \frac{1}{(1+e^{\text{"} z})}\left(1 \text{"} \frac{1}{(1+e^{\text{"} z})}\right)$$

Recall that: $\$(z) = \frac{1}{1+e^{\text{"} z}}$

$$\frac{d}{dz}\$(z) = \$(z)$$

"Derivative of a Sigmoid Function

$$\frac{d}{dz}\$(z) = \frac{e^{"z} + 1 " 1}{(1+e^{"z})^2}$$

$$= \frac{1+e^{"z} " 1}{(1+e^{"z})^2}$$

$$= \frac{\cancel{1+e^{"z}}}{(\cancel{1+e^{"z}})^2} " \frac{1}{(1+e^{"z})^2}$$

$$= \frac{1}{(1+e^{"z})} "$$

$$\frac{d}{dz}\$(z) = \frac{1}{(1+e^{"z})} " \left(\frac{1}{(1+e^{"z})}\right)\left(\frac{1}{(1+e^{"z})}\right)$$

$$= \frac{1}{(1+e^{"z})} \left(1 " \frac{1}{(1+e^{"z})}\right)$$

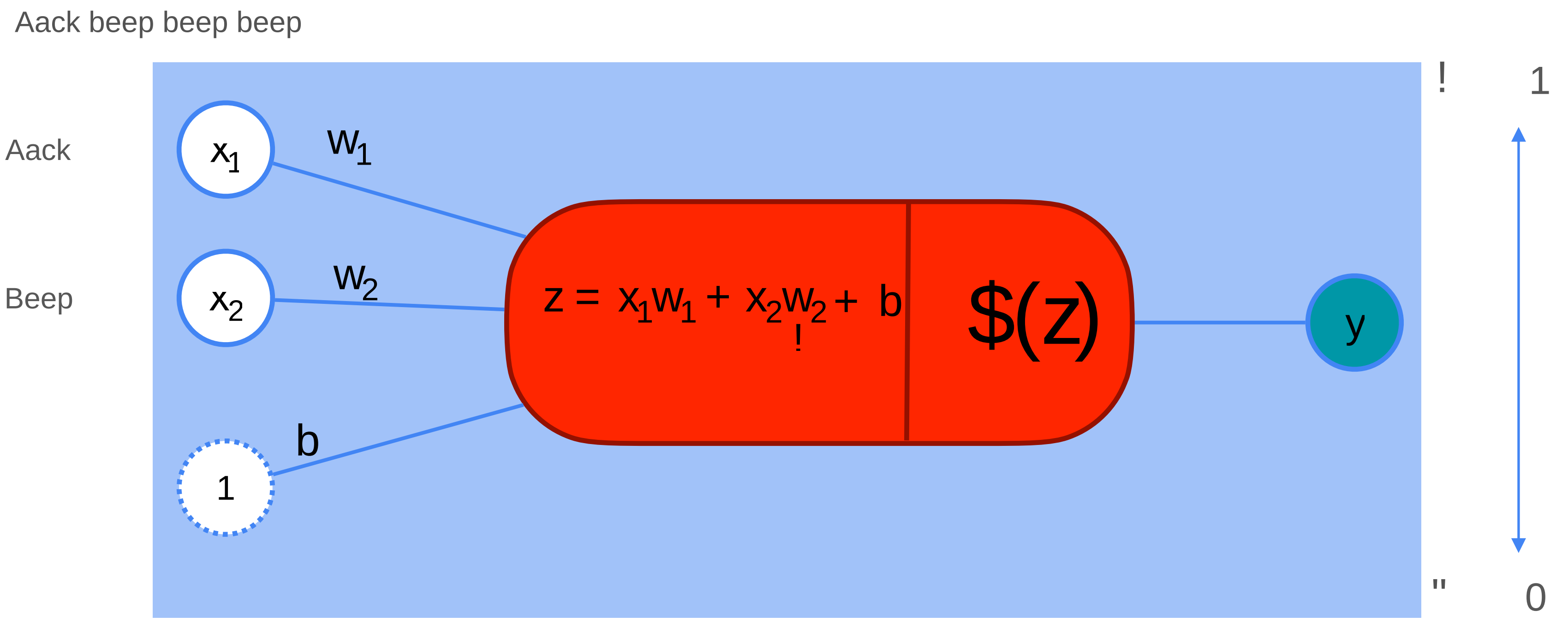
Recall that: $\$(z) = \frac{1}{1+e^{"z}}$

$$\frac{d}{dz}\$(z) = \$(z)\,(1 " \$(z))$$

Optimization in Neural Networks and NewtonÕs Method

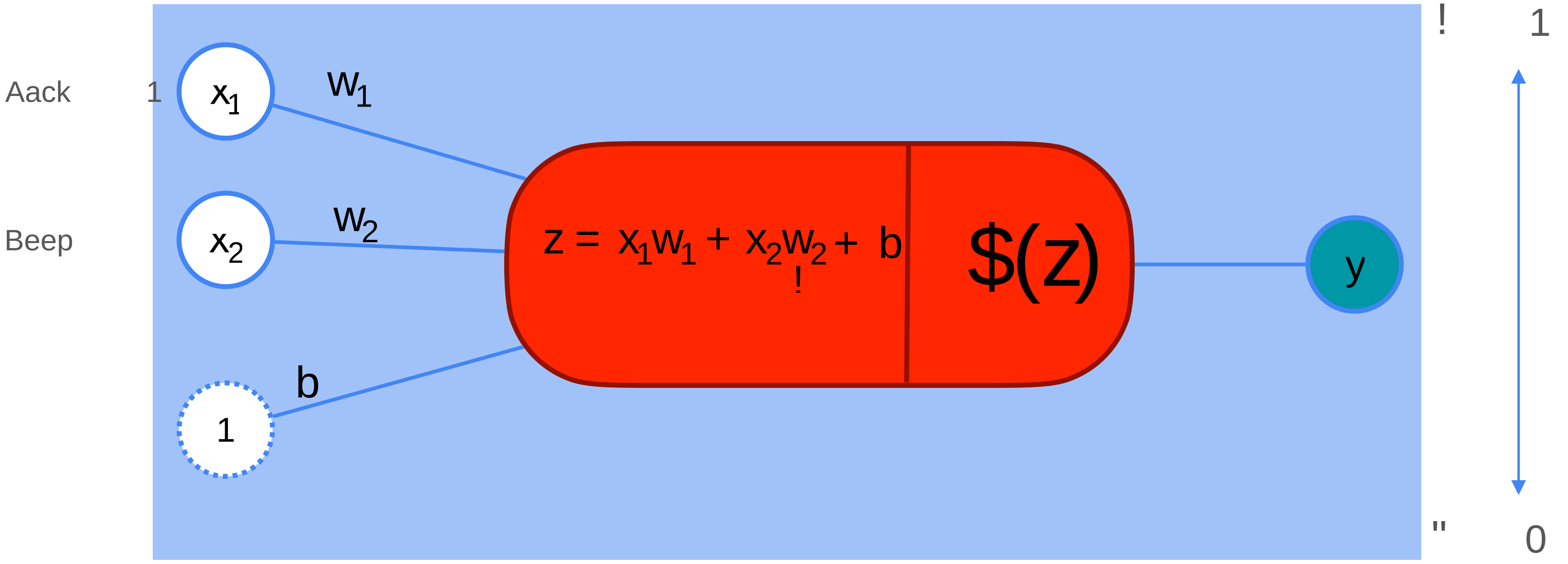
Classification with a perceptron: Gradient Descent

"Classification With a Perceptron



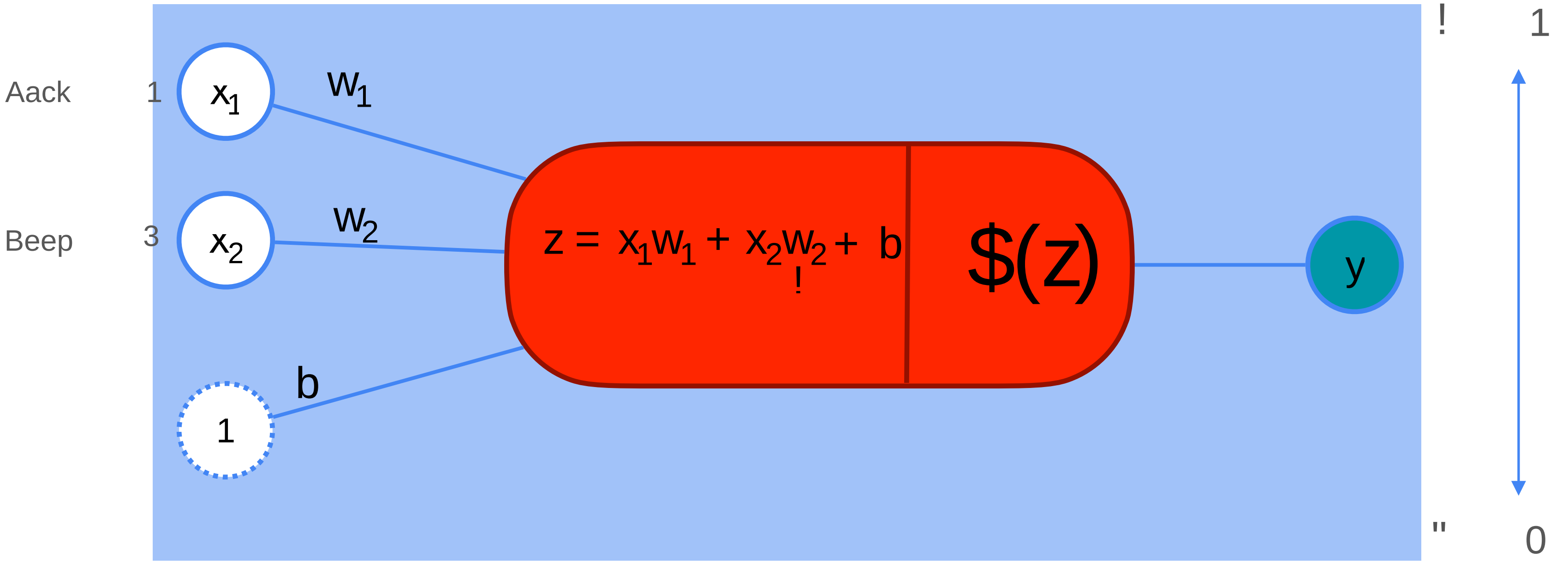
"Classification With a Perceptron

Aack beep beep beep



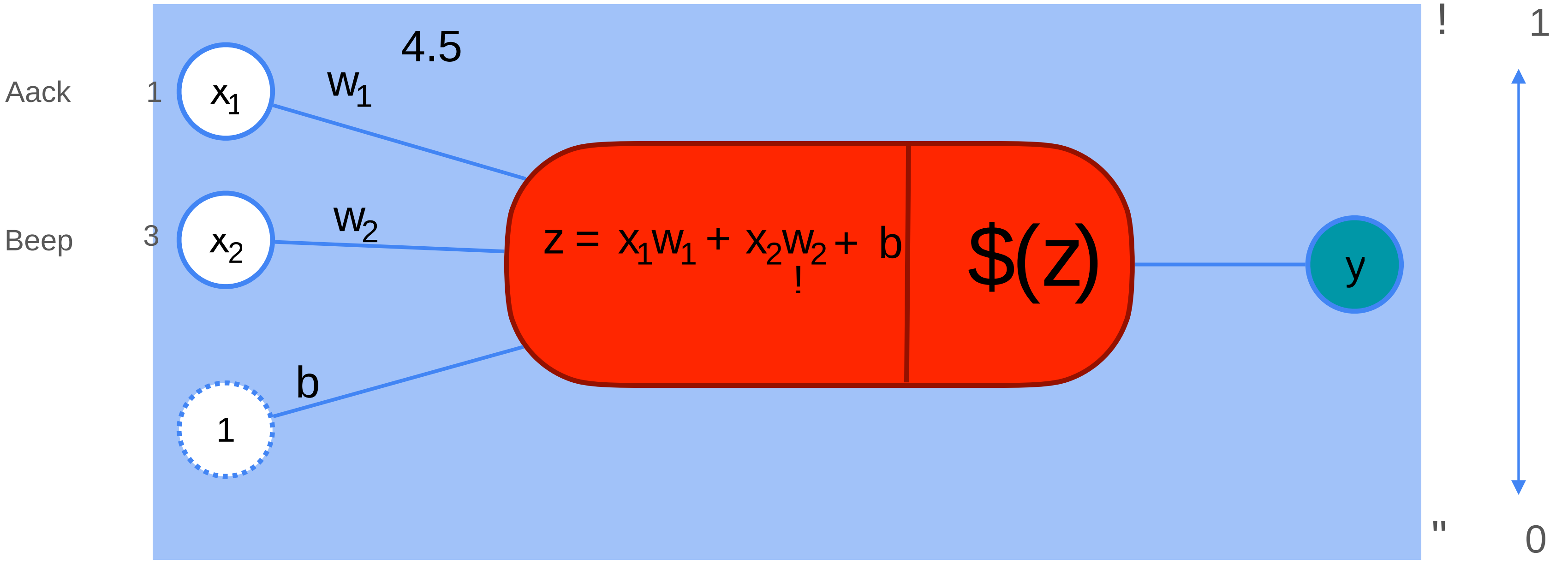
"Classification With a Perceptron

Aack beep beep beep



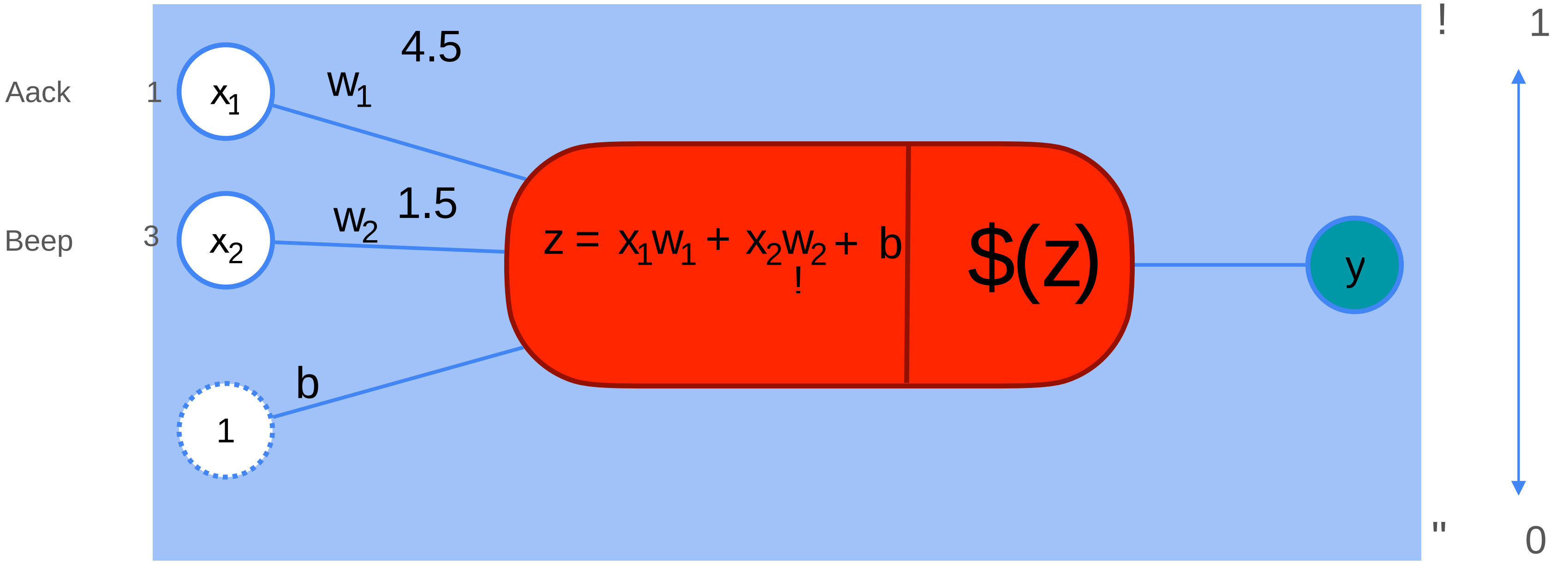
"Classification With a Perceptron

Aack beep beep beep



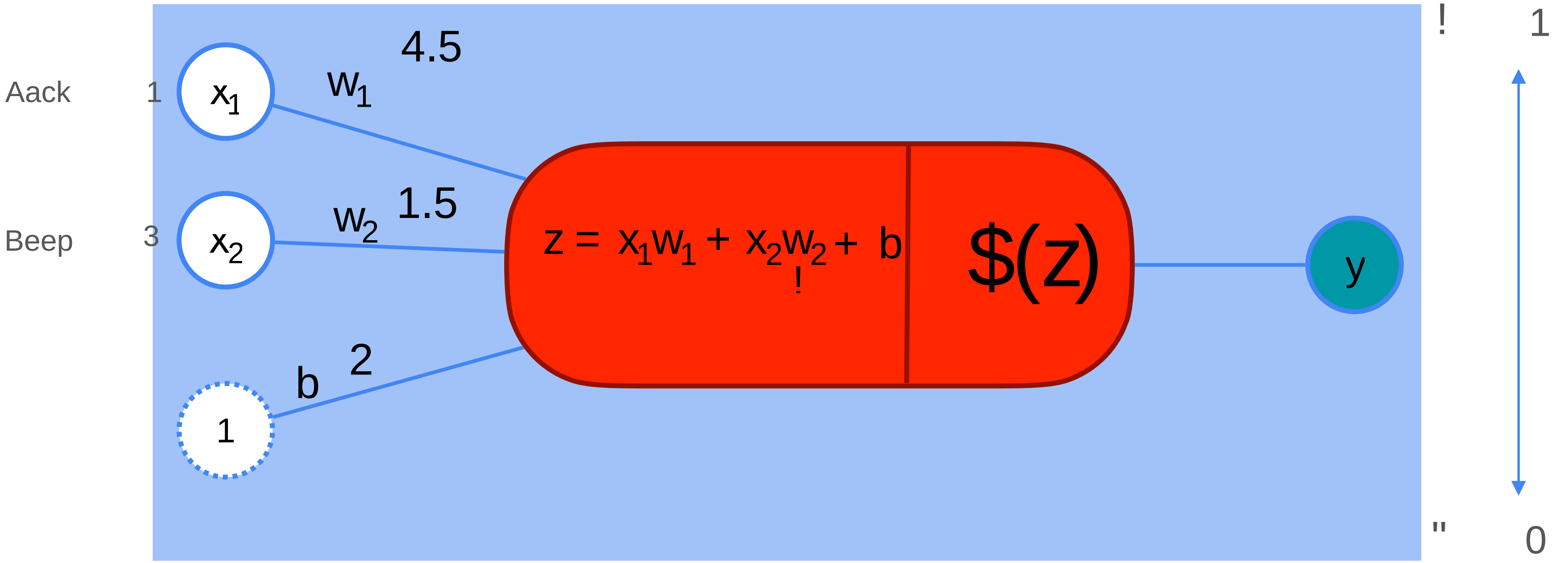
"Classification With a Perceptron

Aack beep beep beep



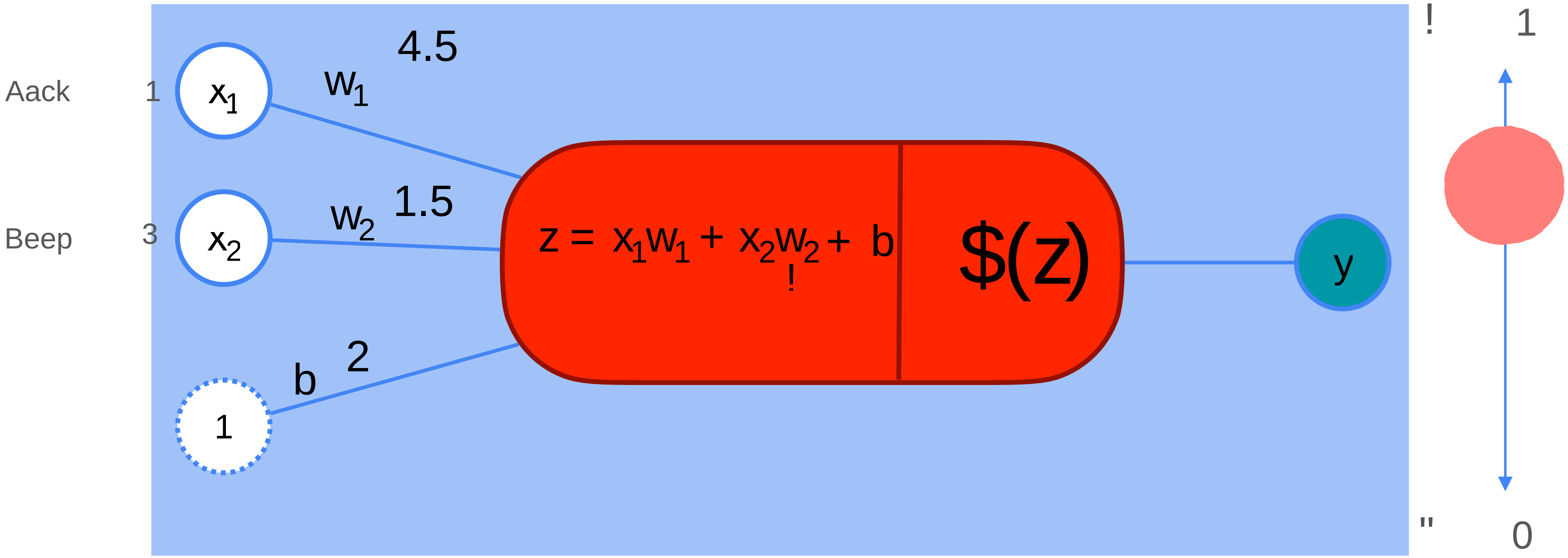
"Classification With a Perceptron

Aack beep beep beep



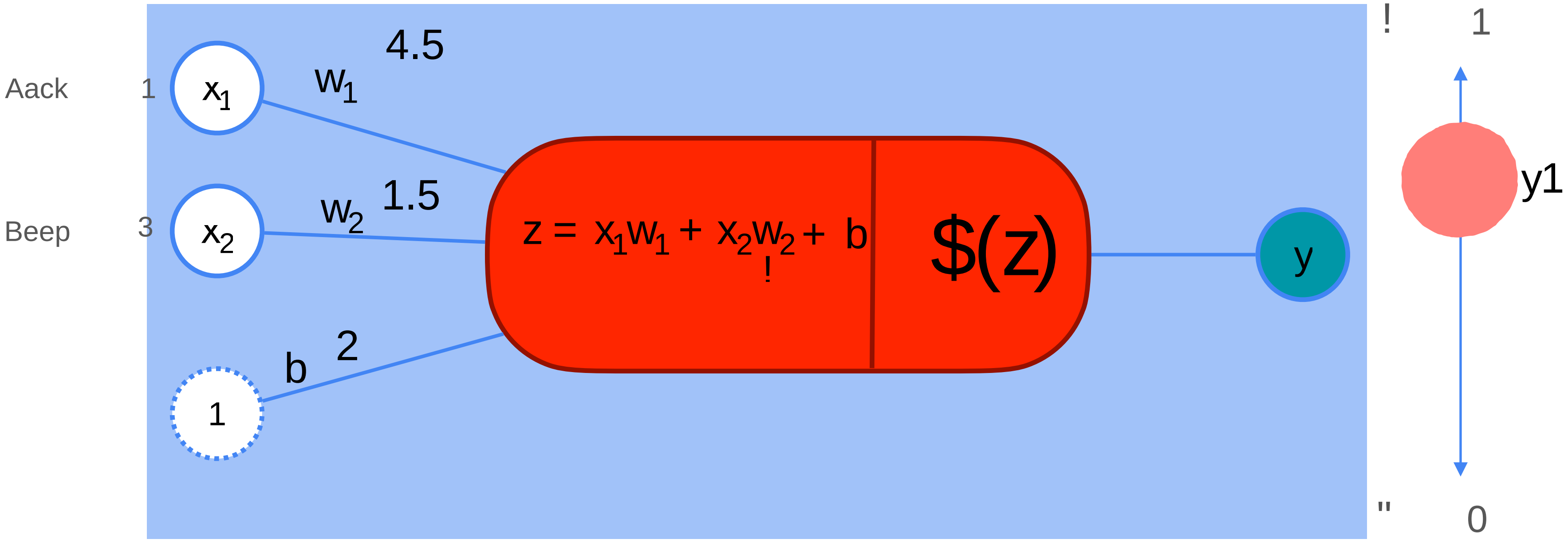
"Classification With a Perceptron

Aack beep beep beep

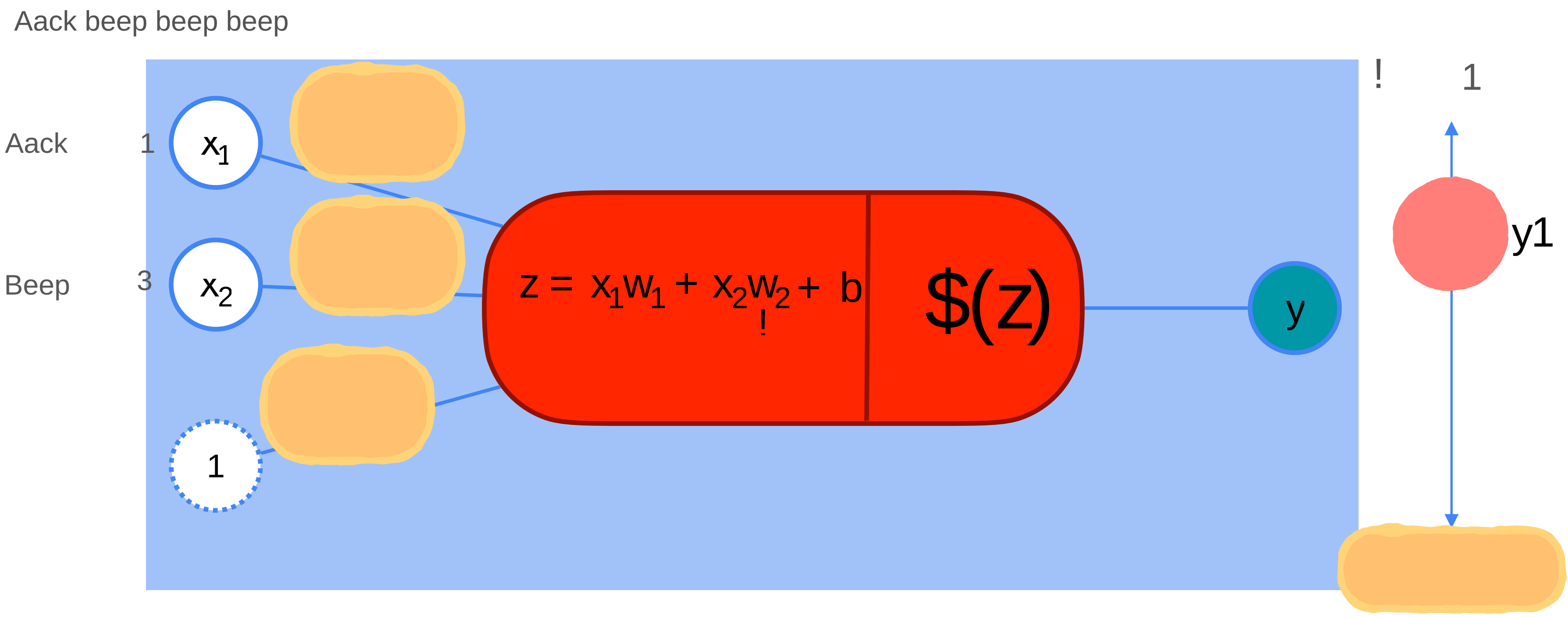


"Classification With a Perceptron

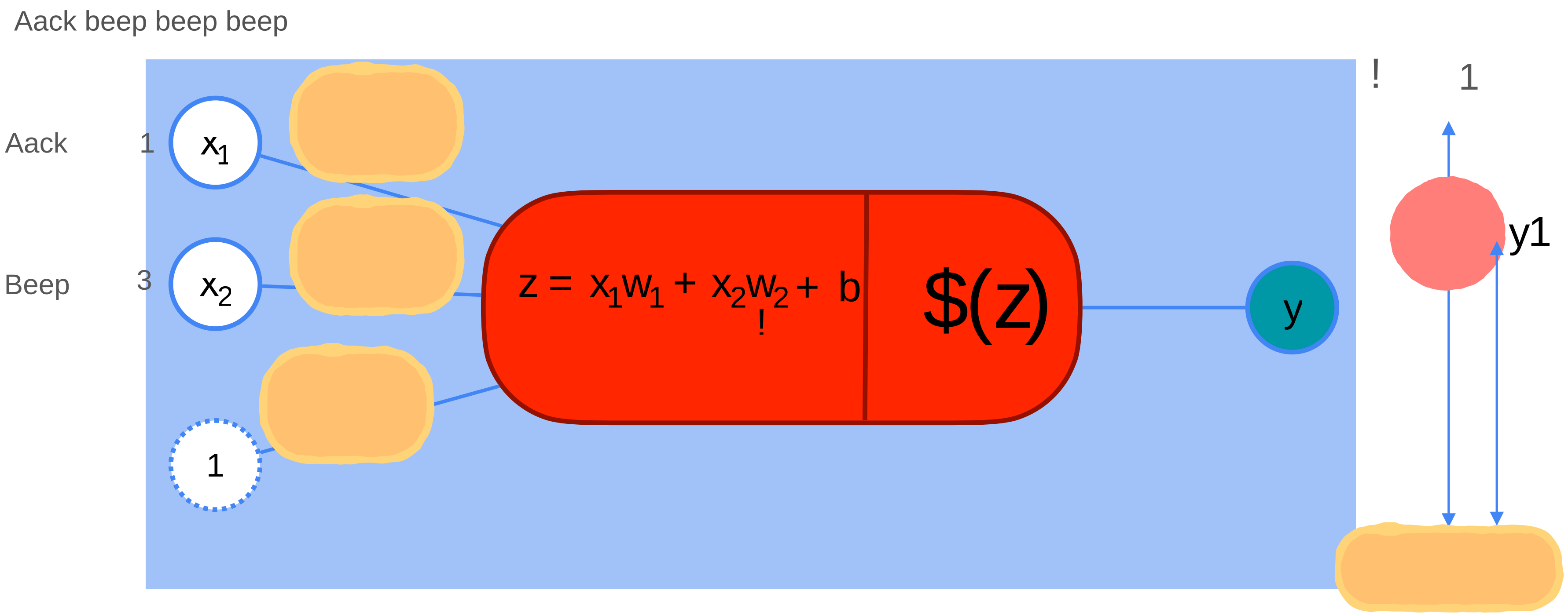
Aack beep beep beep



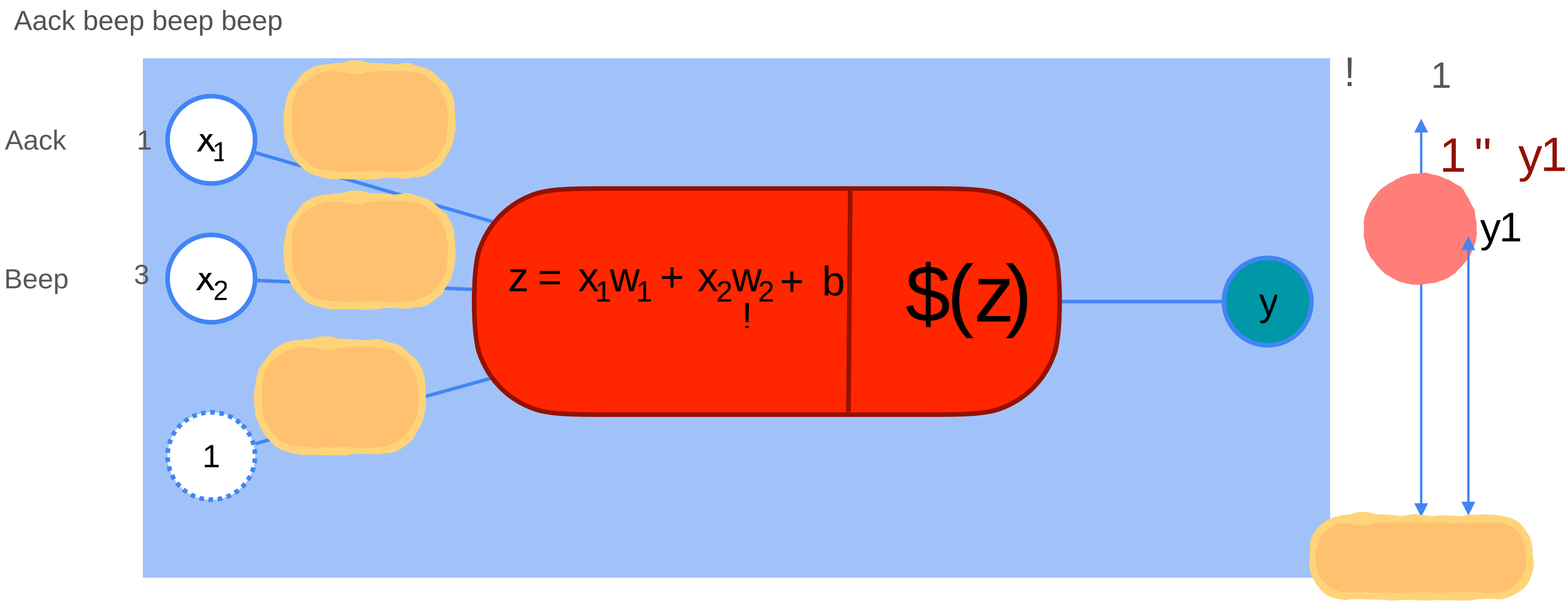
"Classification With a Perceptron



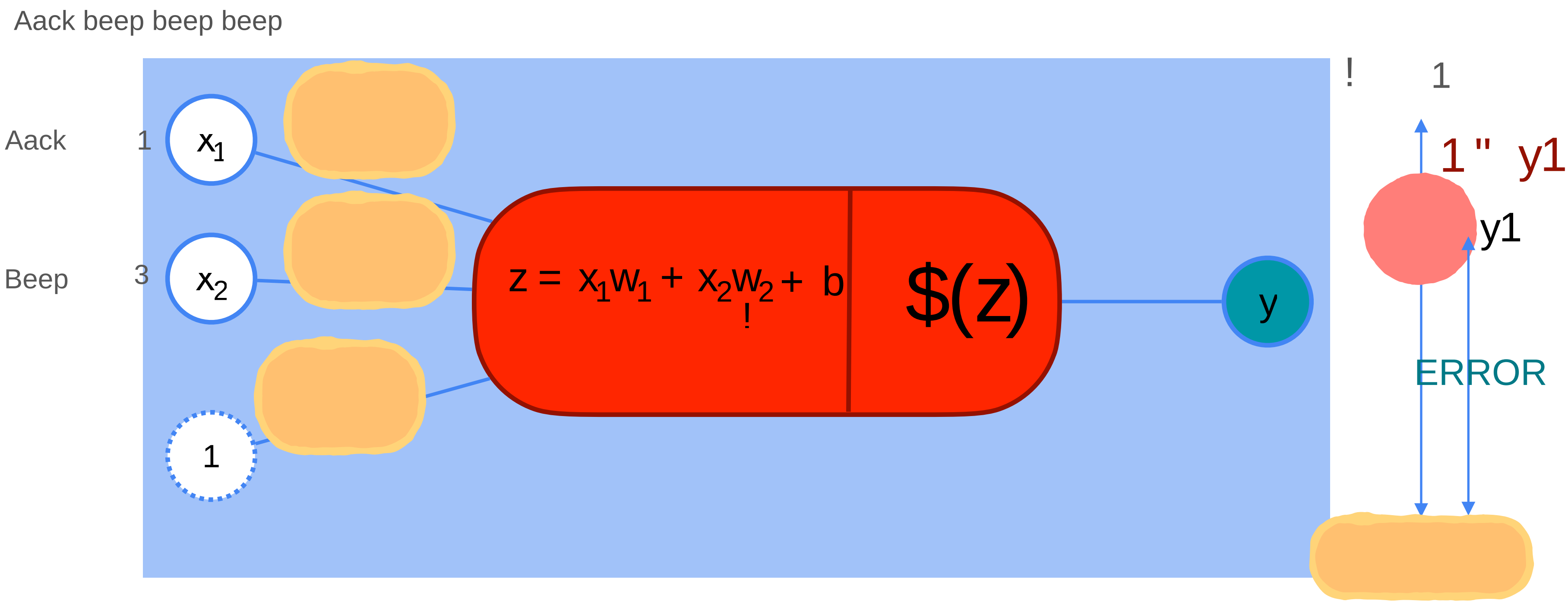
"Classification With a Perceptron



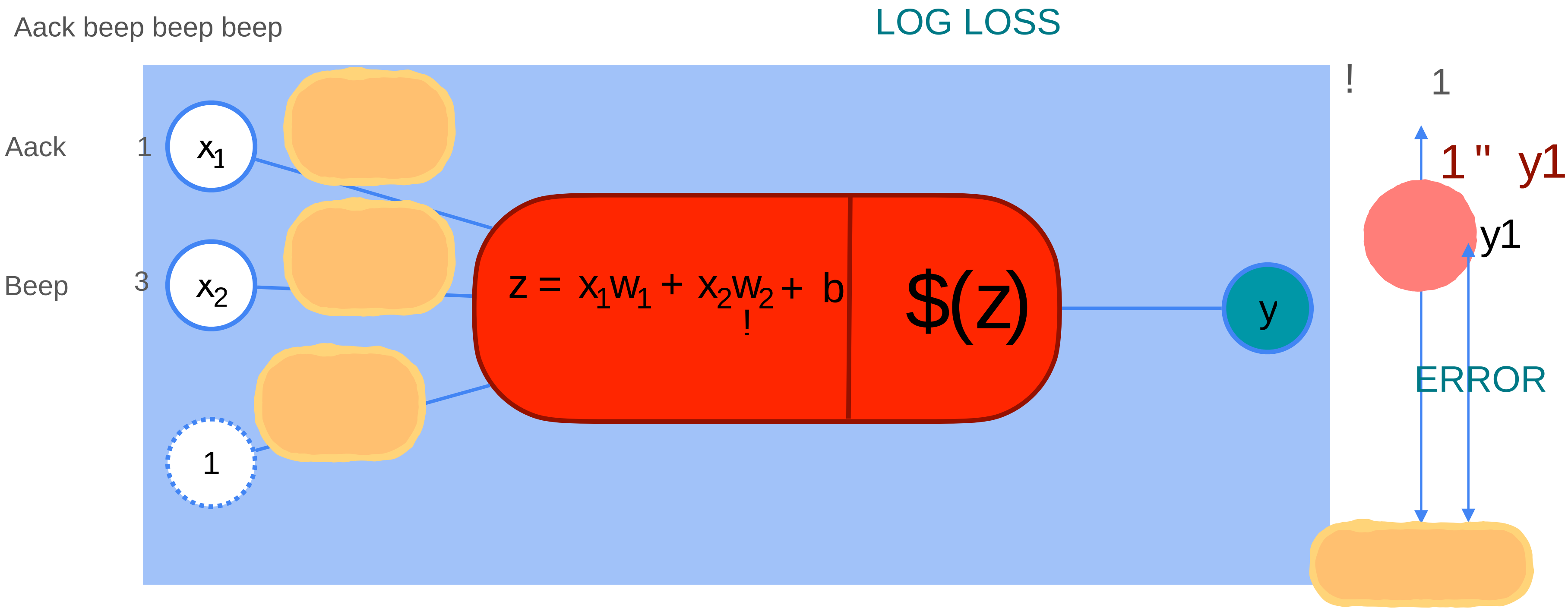
"Classification With a Perceptron



"Classification With a Perceptron

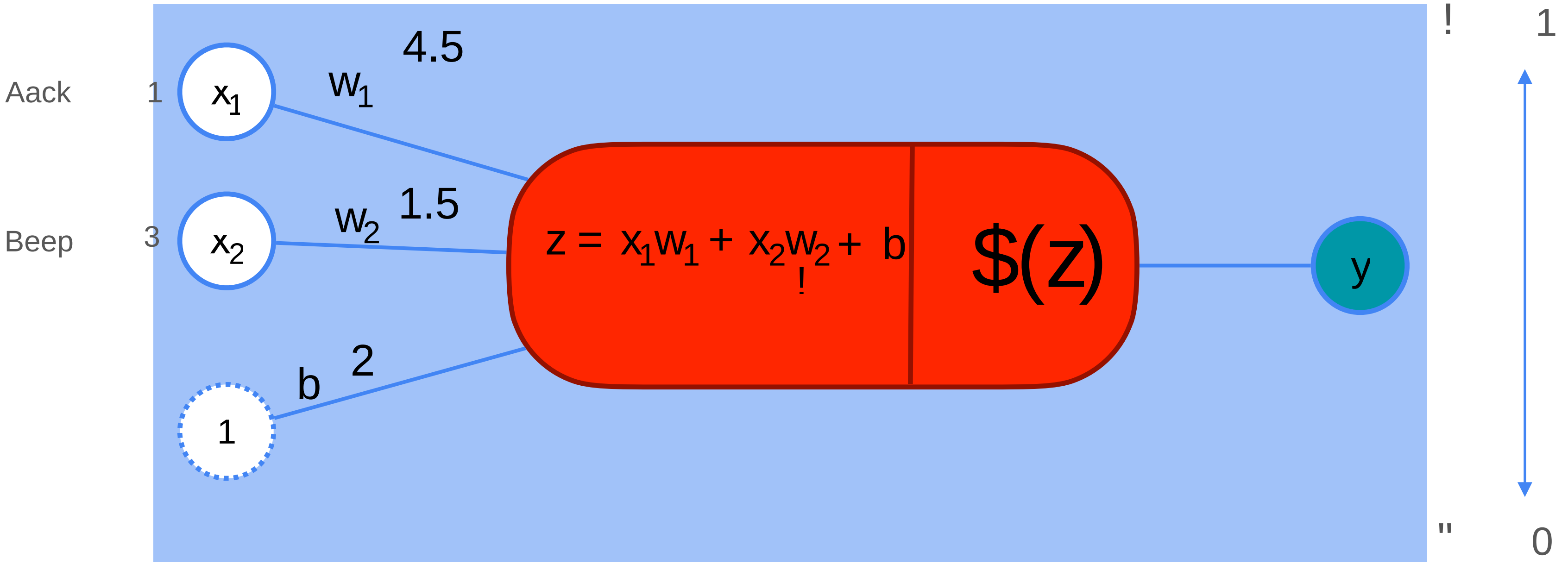


"Classification With a Perceptron



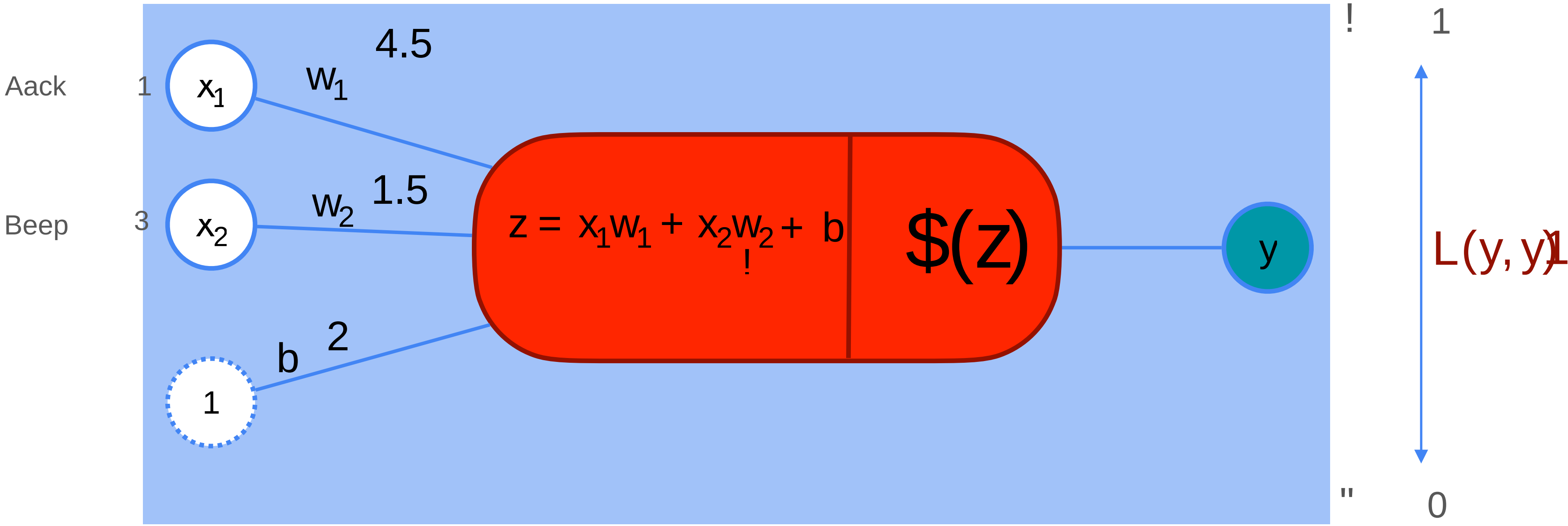
"Classification With a Perceptron

Aack beep beep beep

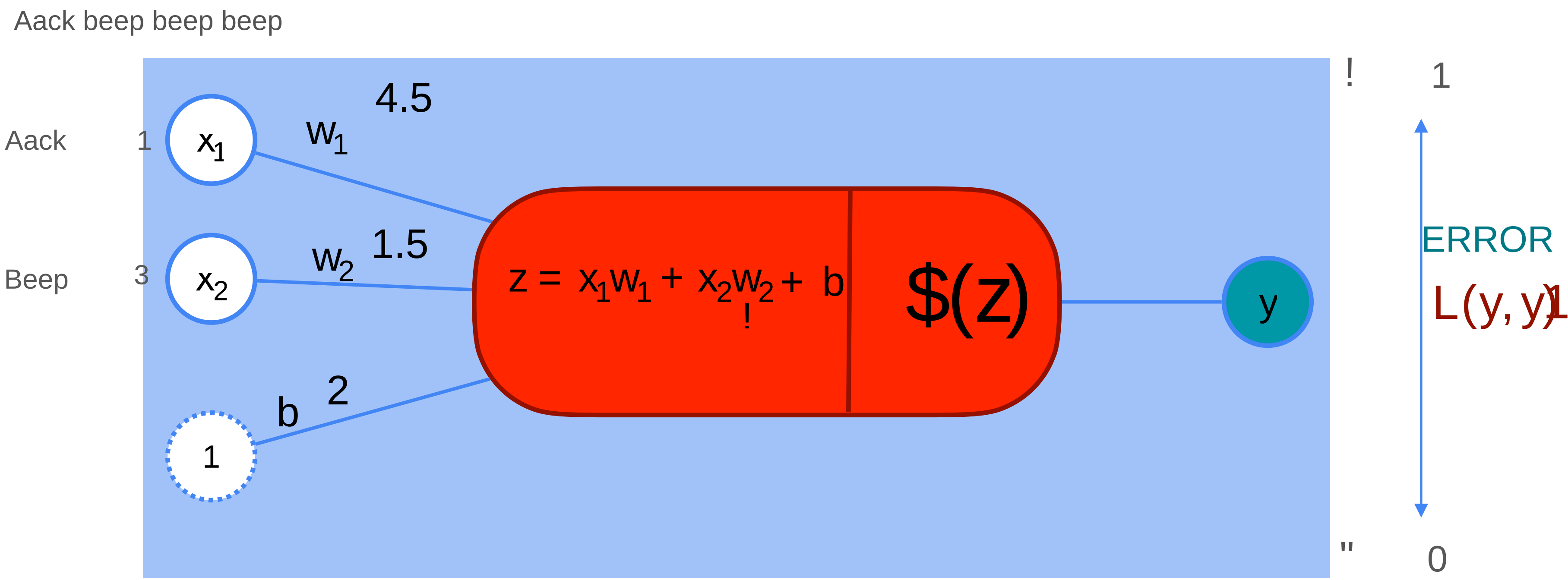


"Classification With a Perceptron

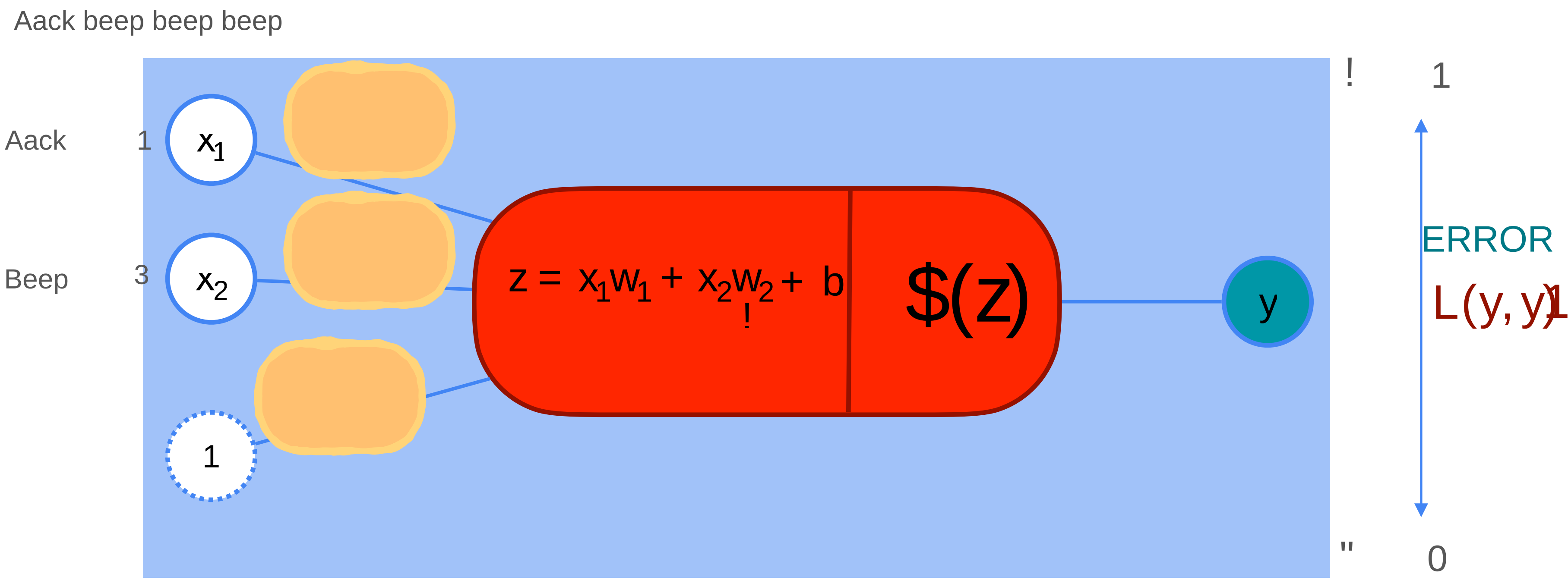
Aack beep beep beep



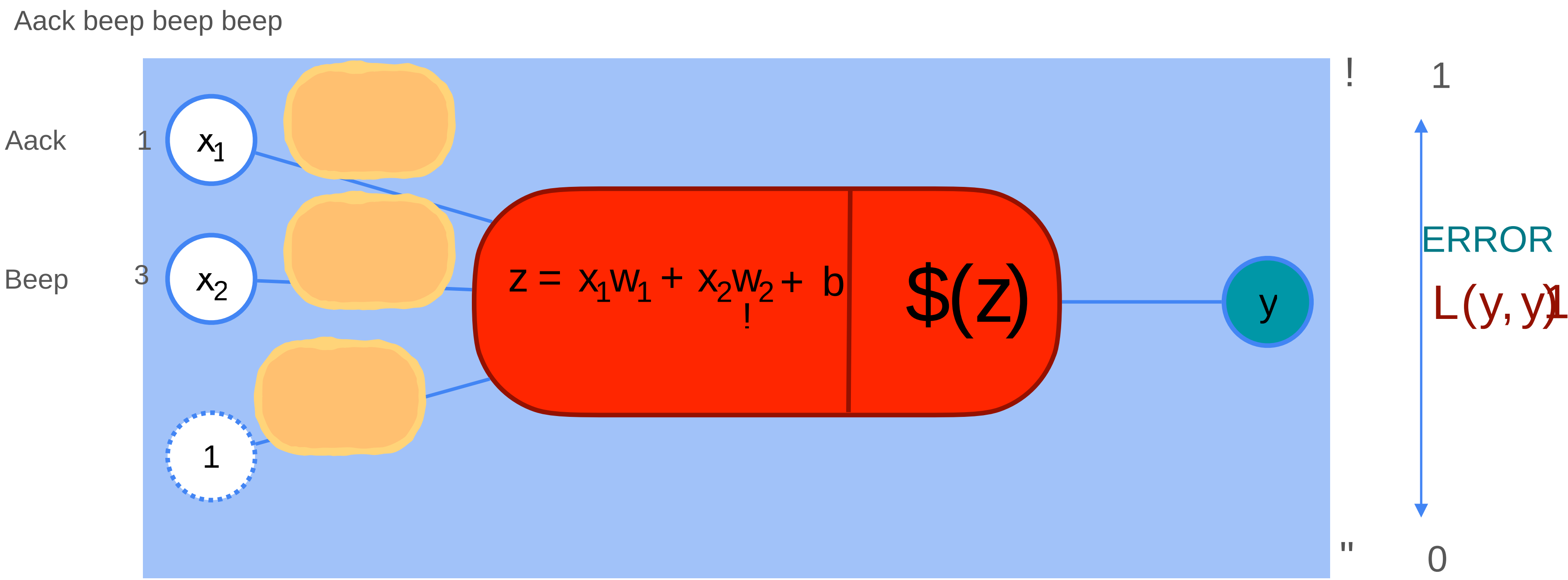
"Classification With a Perceptron



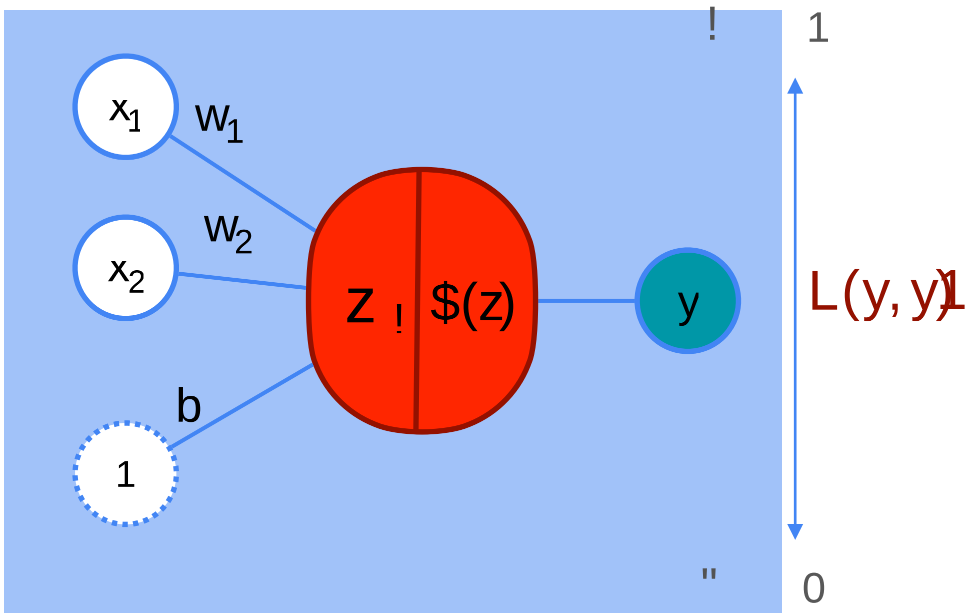
"Classification With a Perceptron



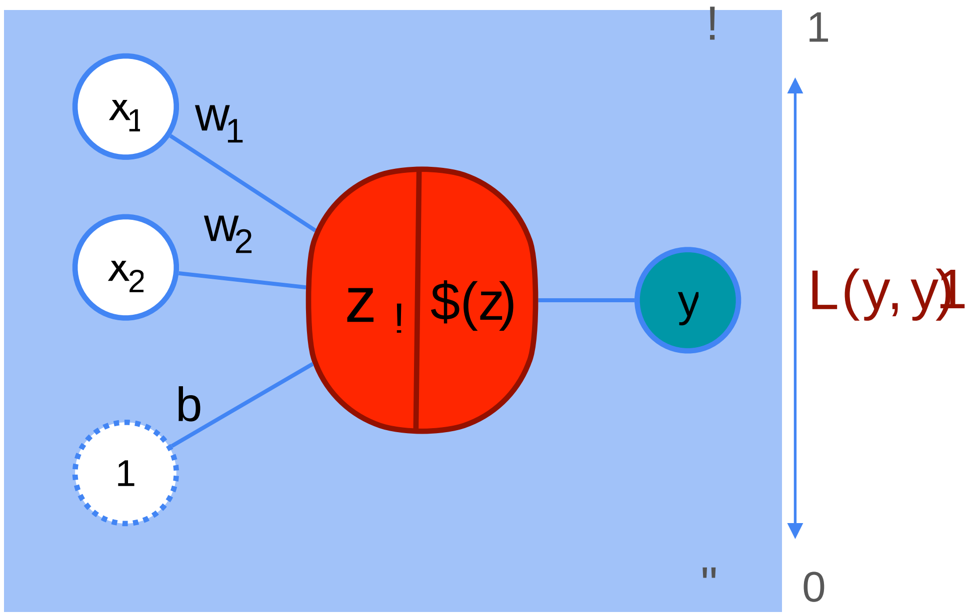
"Classification With a Perceptron



"Classification With a Perceptron



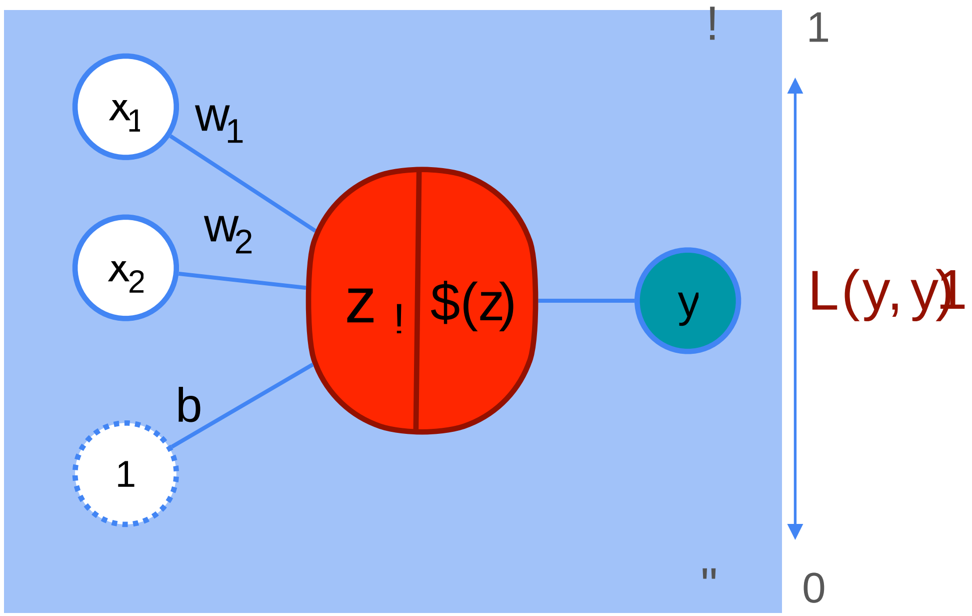
"Classification With a Perceptron



Prediction Function:

y1

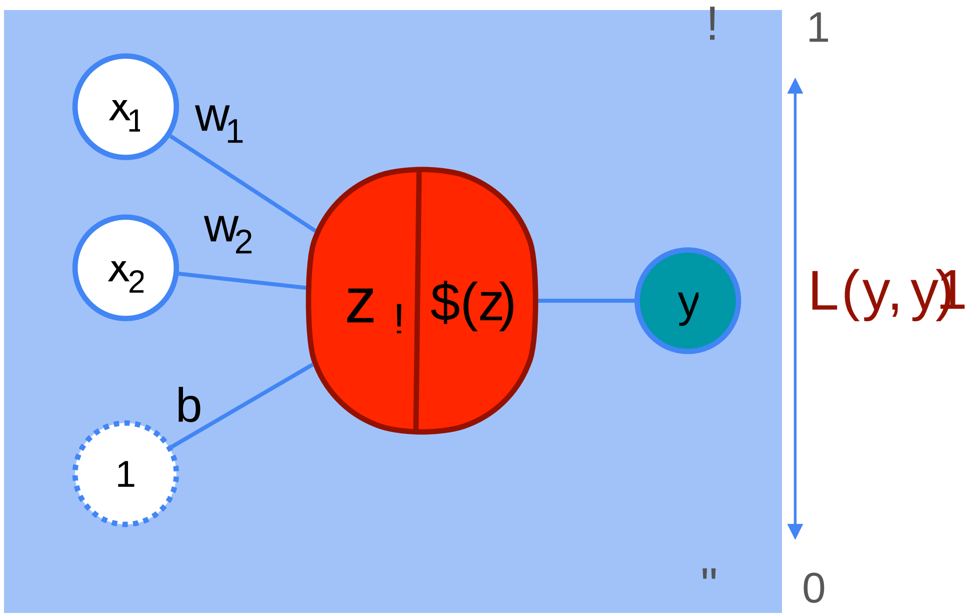
"Classification With a Perceptron



Prediction Function:

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

"Classification With a Perceptron

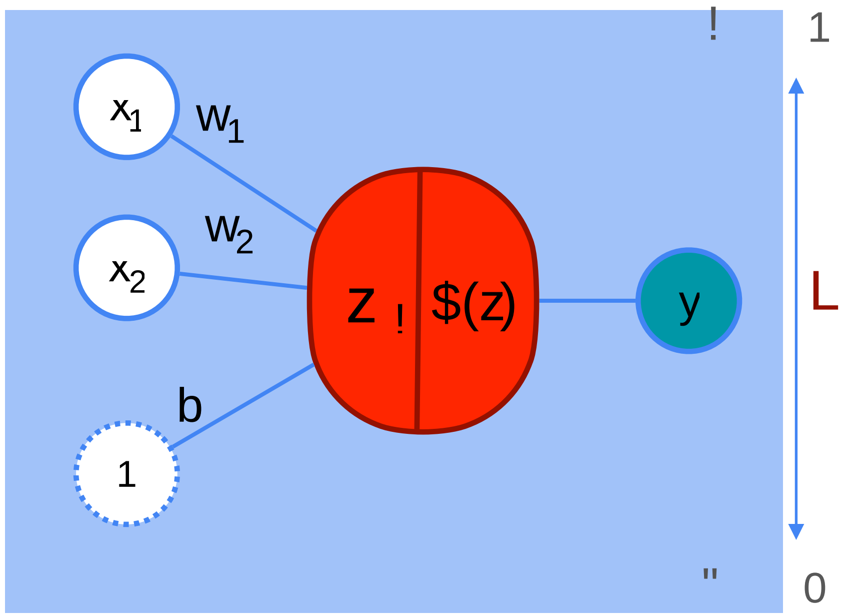


Prediction Function:

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

Loss Function:

Classification With a Perceptron



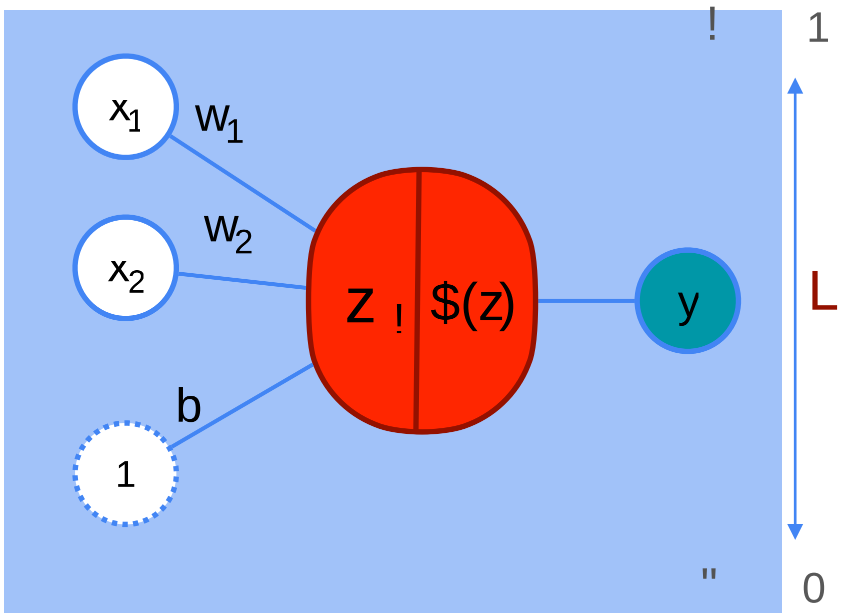
Prediction Function:

$$\hat{y} = \sigma(w_1 x_1 + w_2 x_2 + b)$$

Loss Function:

$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1 - y)\ln(1 - \hat{y})$$

"Classification With a Perceptron



Prediction Function:

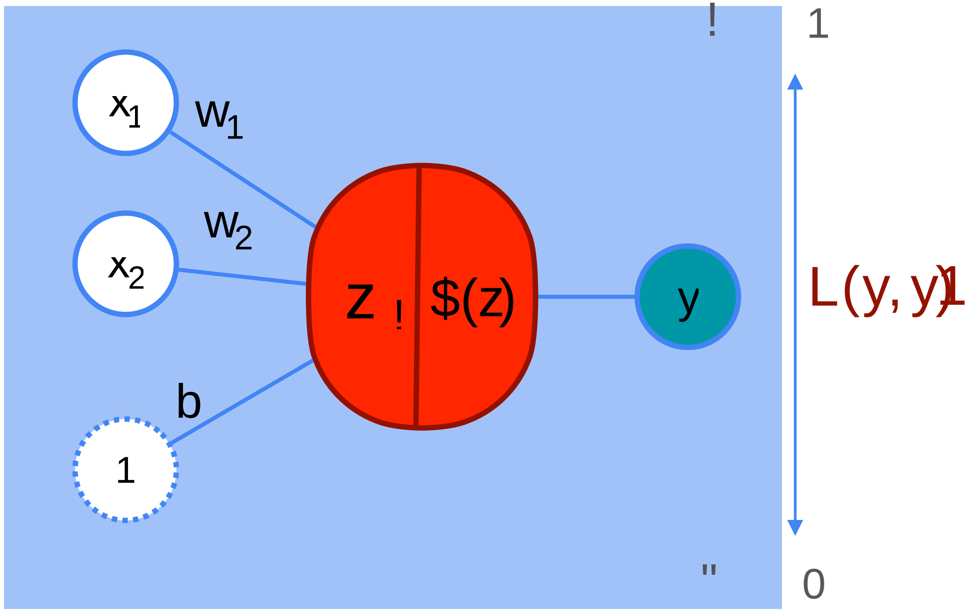
$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

Loss Function:

L(y, y1 L(y, y1 = -y ln(y1 " (1 " y)ln(1 " y1

Main Goal:

"Classification With a Perceptron



Prediction Function:

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

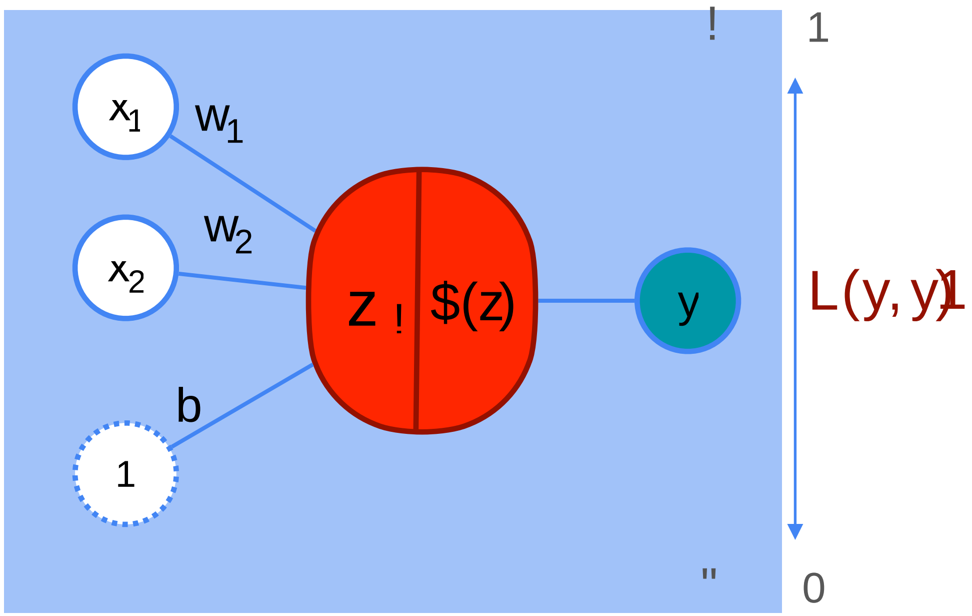
Loss Function:

$$L(y, y1 = -y\ln(y1 \text{ " } (1 \text{ " } y)\ln(1 \text{ " } y1$$

Main Goal:

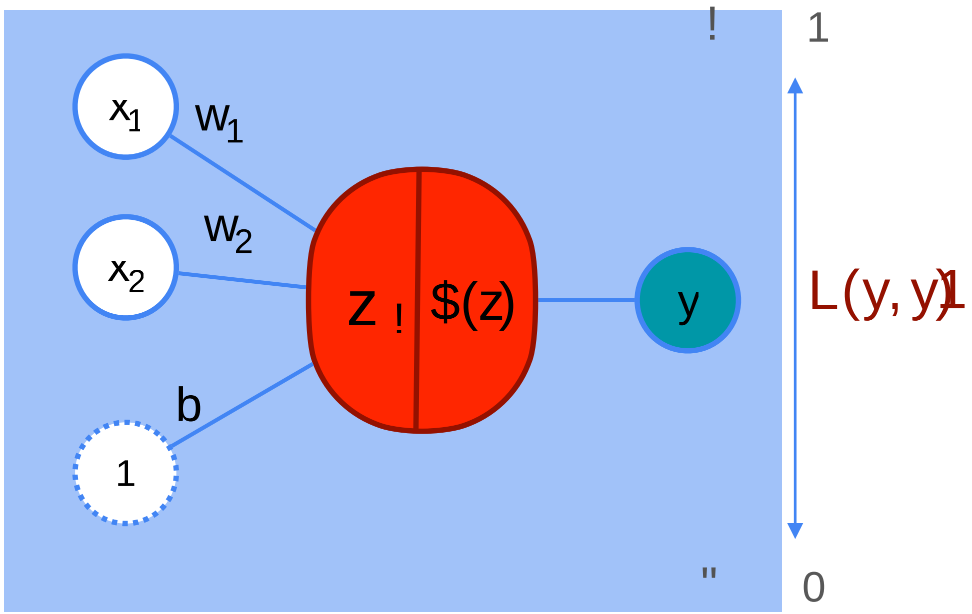
Find w_1 , w_2 , b that give y with the least error

"Classification With a Perceptron

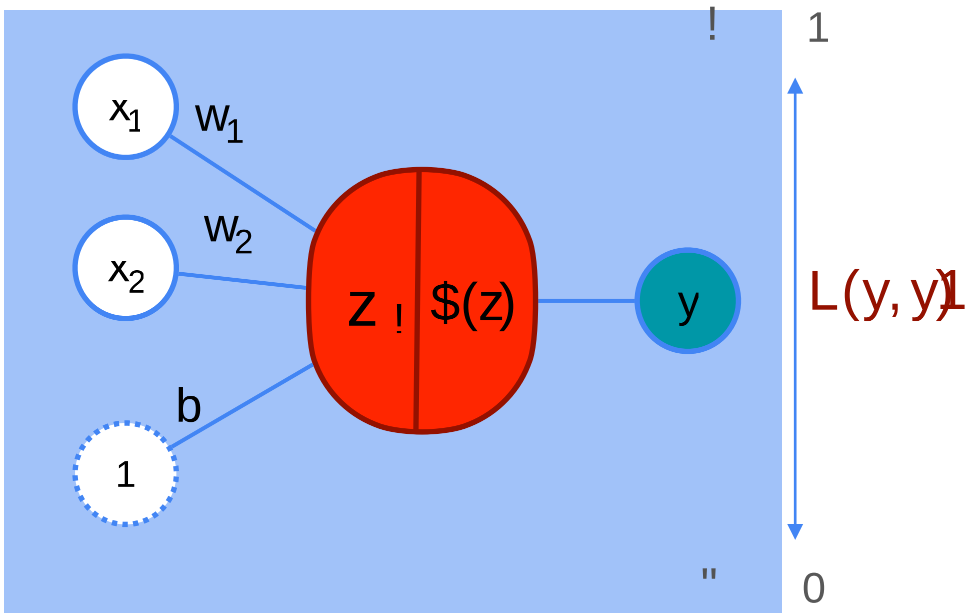


"Classification With a Perceptron

To find optimal values for:



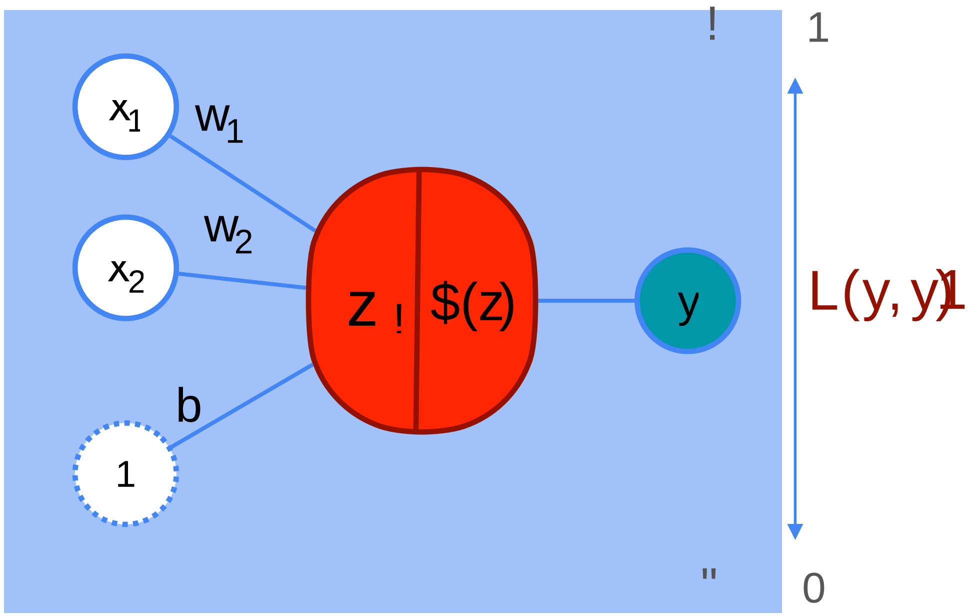
"Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

"Classification With a Perceptron

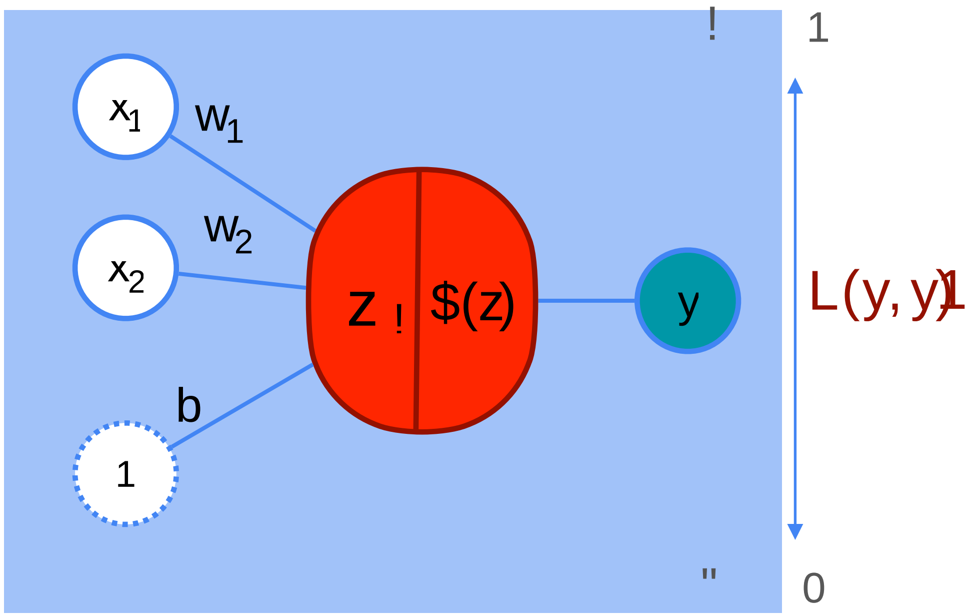


To find optimal values for:

w_1 , w_2 , b

You need gradient descent

"Classification With a Perceptron



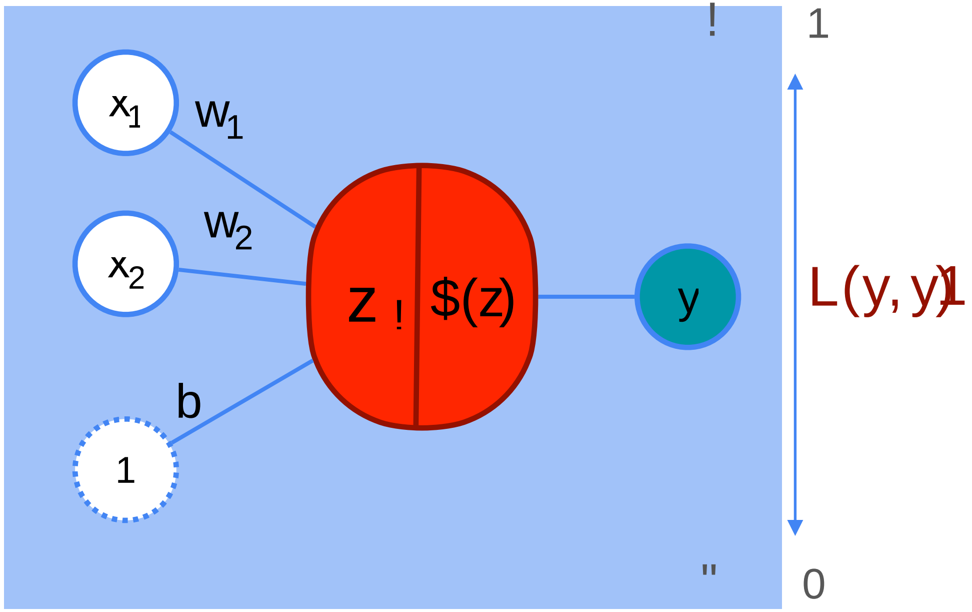
To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$$w_1 \; \$ \; w_1 \, " \; \# \frac{, L}{, w_1}$$

"Classification With a Perceptron



To find optimal values for:

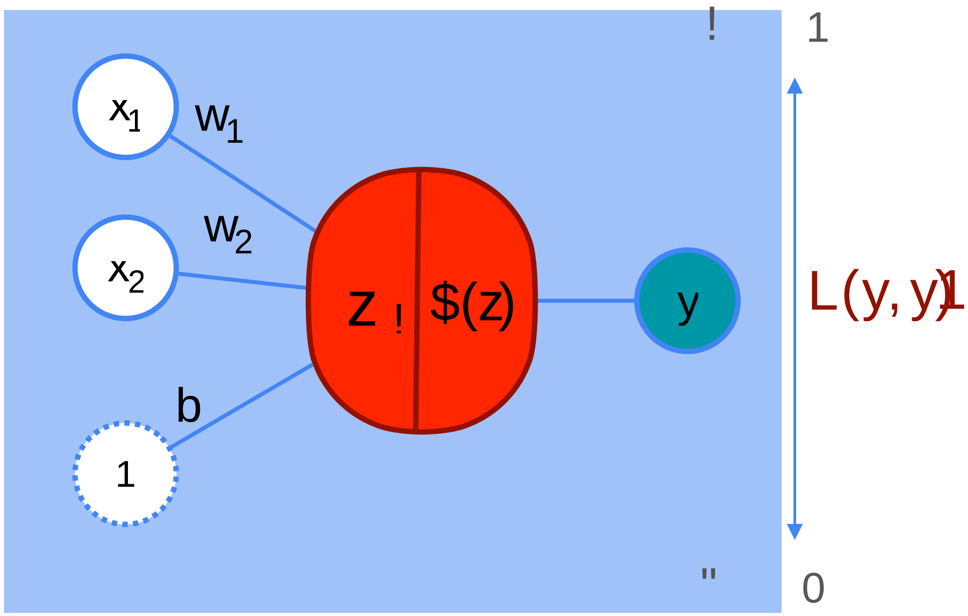
w_1 , w_2 , b

You need gradient descent

$$w_1 \; \$ \; w_1 \, " \; \# \frac{, L}{, w_1}$$

$$w_2 \; \$ \; w_2 \, " \; \# \frac{, L}{, w_2}$$

"Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

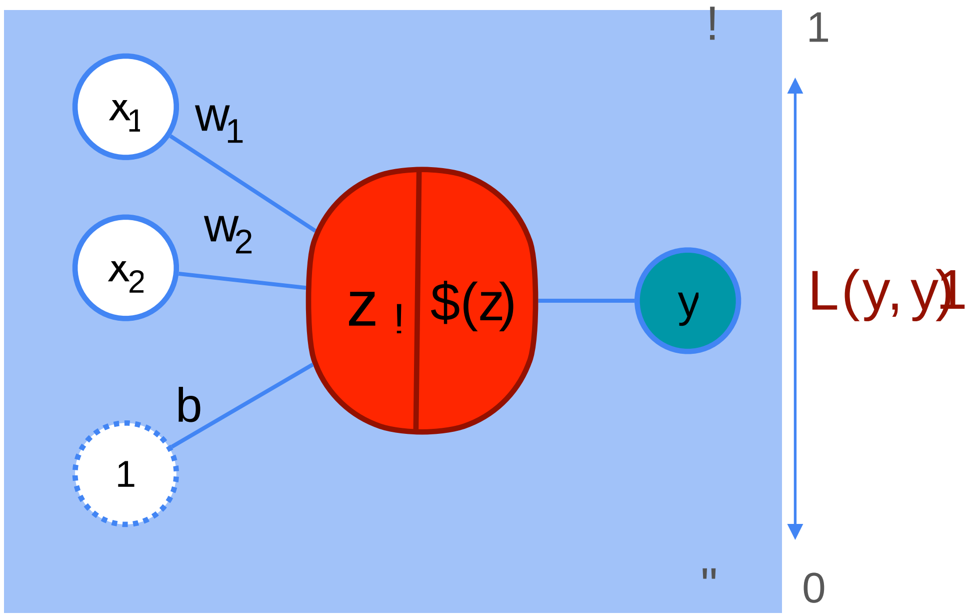
You need gradient descent

$$w_1 \ \$ \ \ w_1 \text{"} \ \# \frac{,L}{,w_1}$$

$$w_2 \ \$ \ \ w_2 \text{"} \ \# \frac{,L}{,w_2}$$

$$b \ \$ \ \ b \text{"} \ \# \frac{,L}{,b}$$

"Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

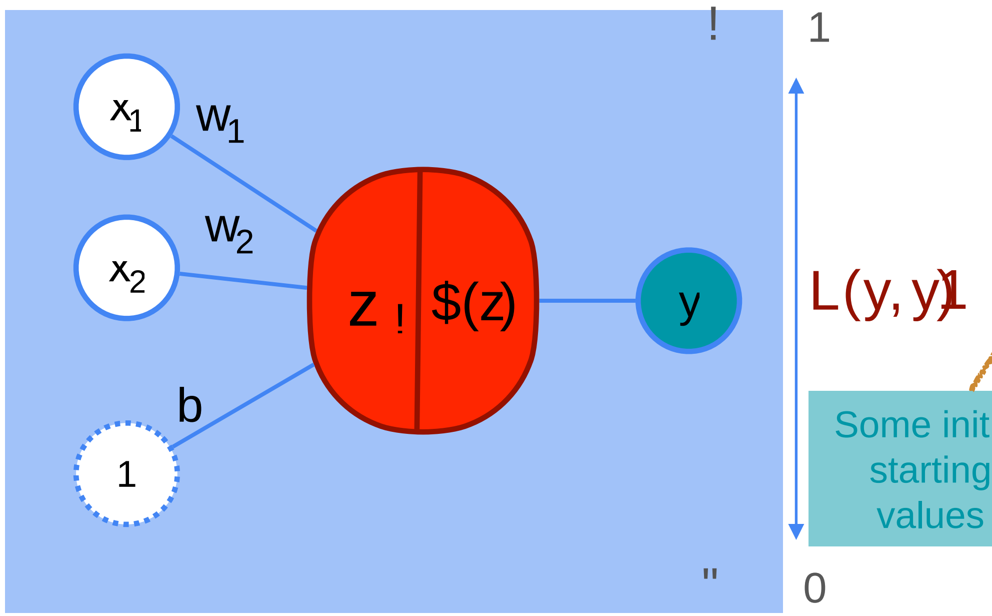
You need gradient descent

w_1 \$ " $\#\frac{,L}{,w_1}$

w_2 \$ " $\#\frac{,L}{,w_2}$

b \$ $\#\frac{,L}{,b}$

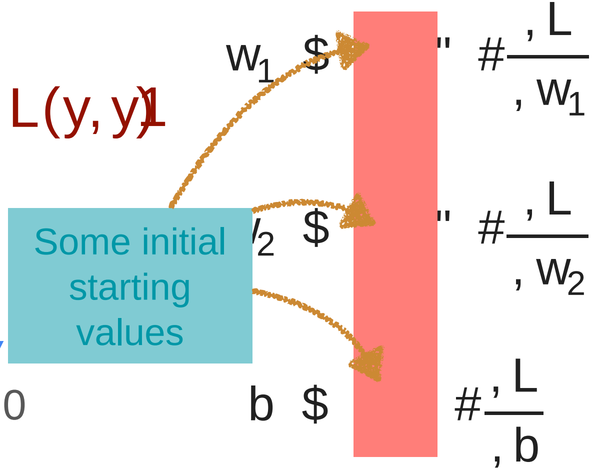
"Classification With a Perceptron



To find optimal values for:

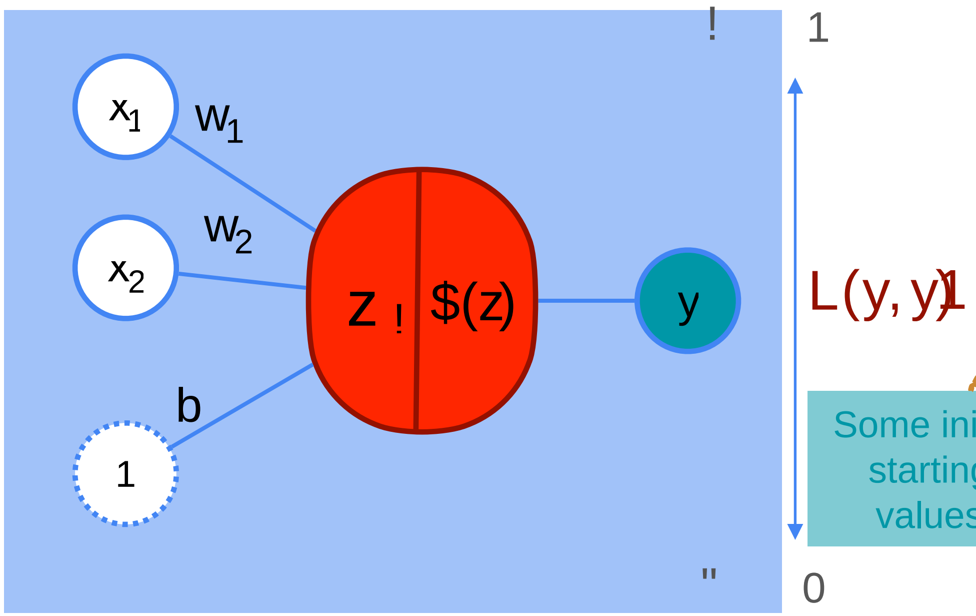
w_1 , w_2 , b

You need gradient descent

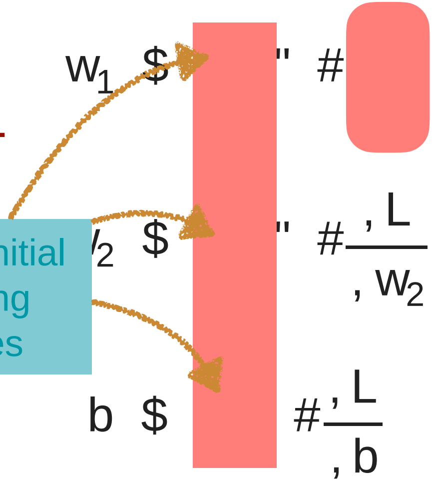


"Classification With a Perceptron

To find optimal values for:
w_1 , w_2 , b



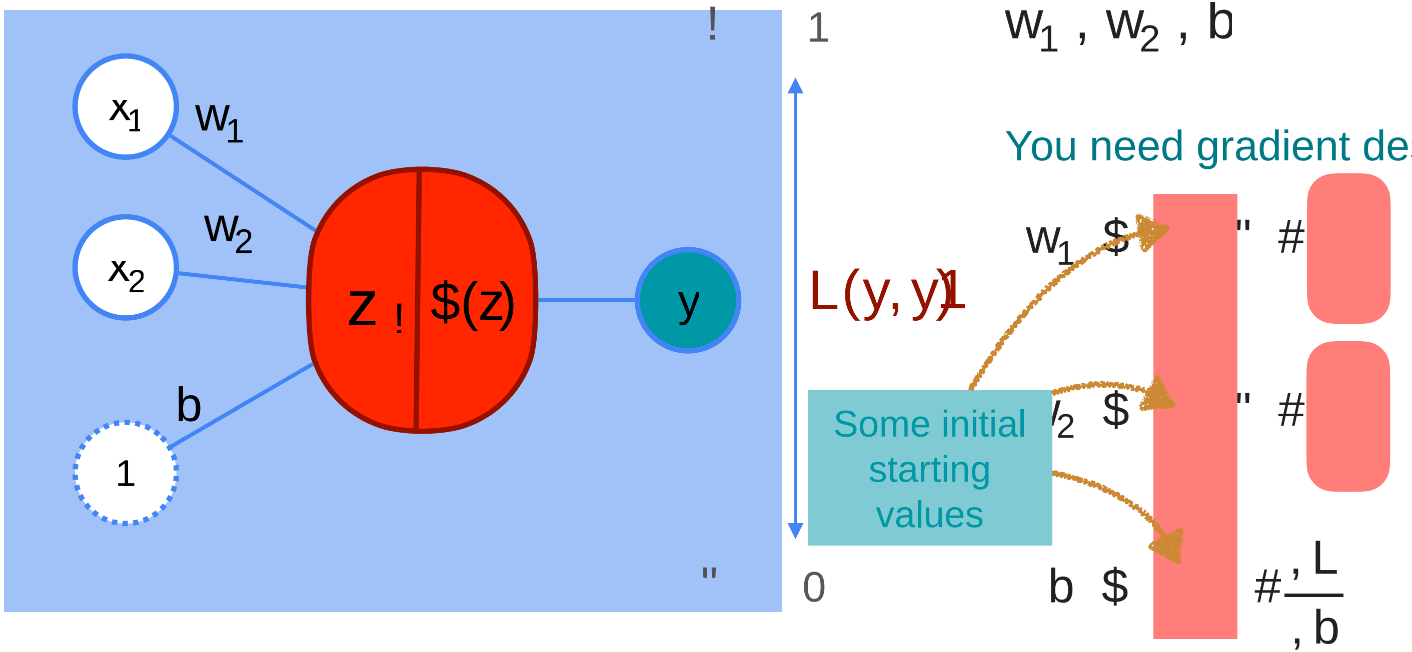
You need gradient descent



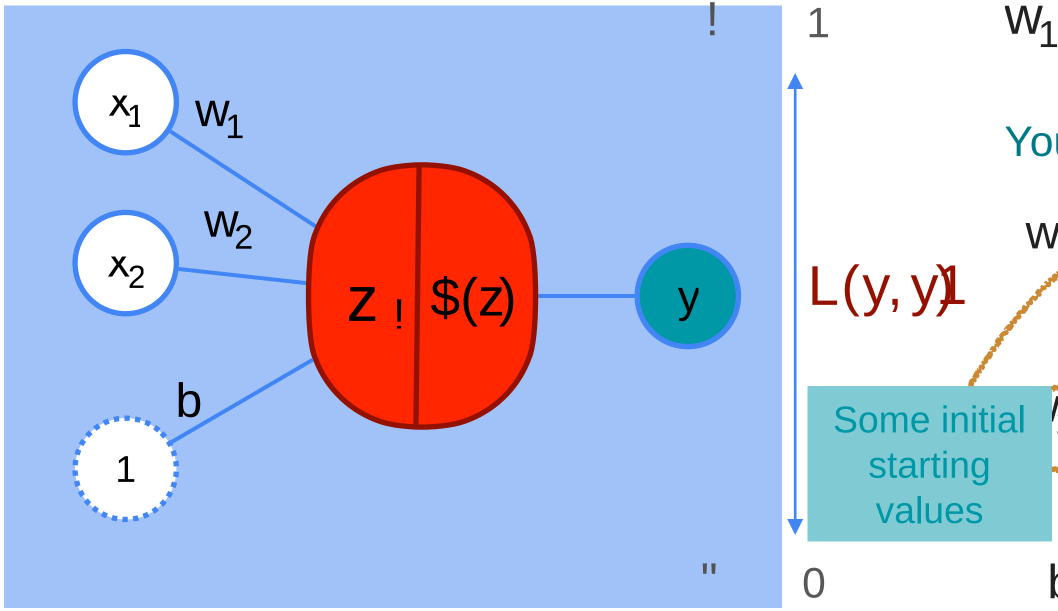
"Classification With a Perceptron

To find optimal values for:
w_1 , w_2 , b

You need gradient descent



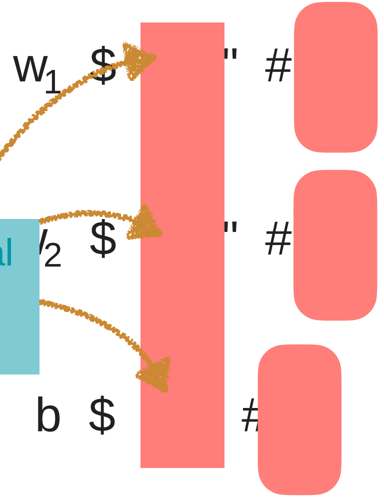
"Classification With a Perceptron



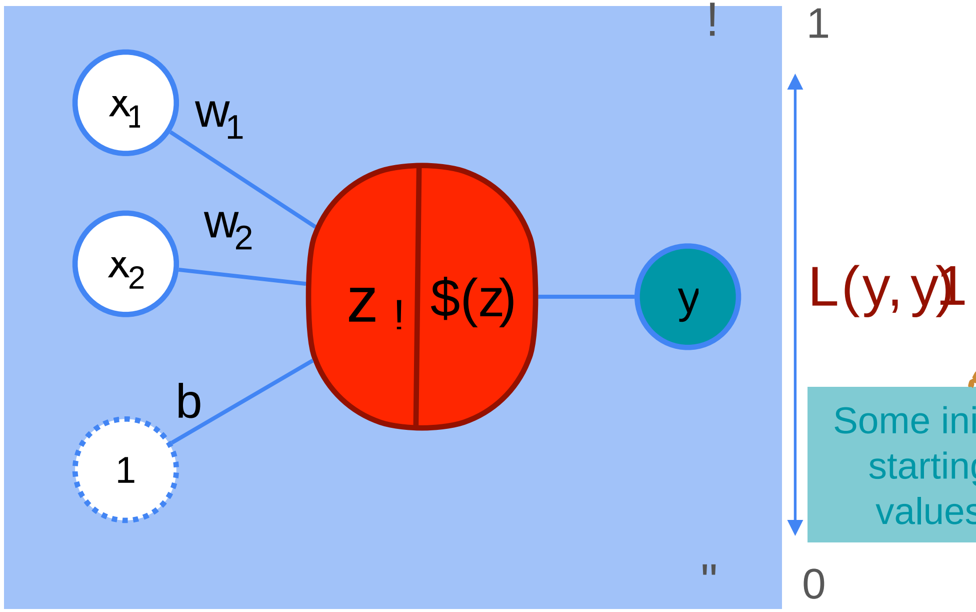
To find optimal values for:

w_1 , w_2 , b

You need gradient descent



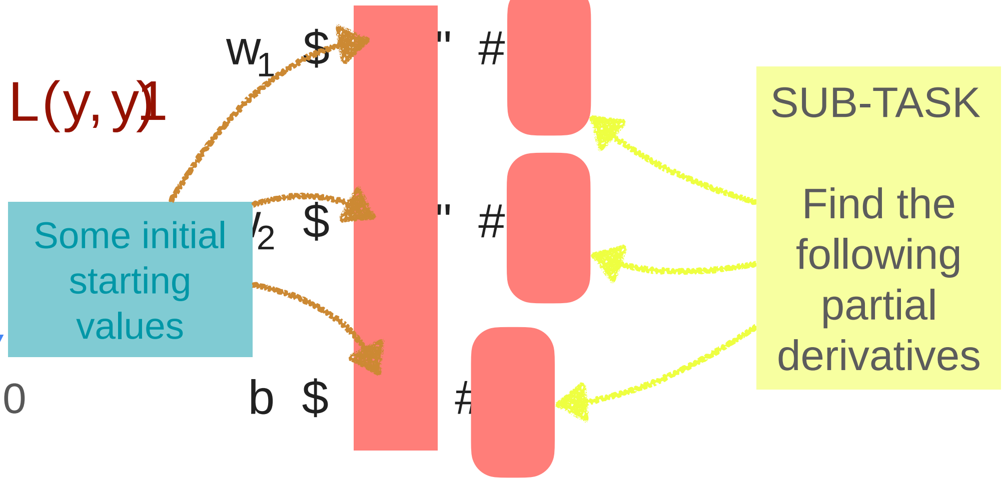
"Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

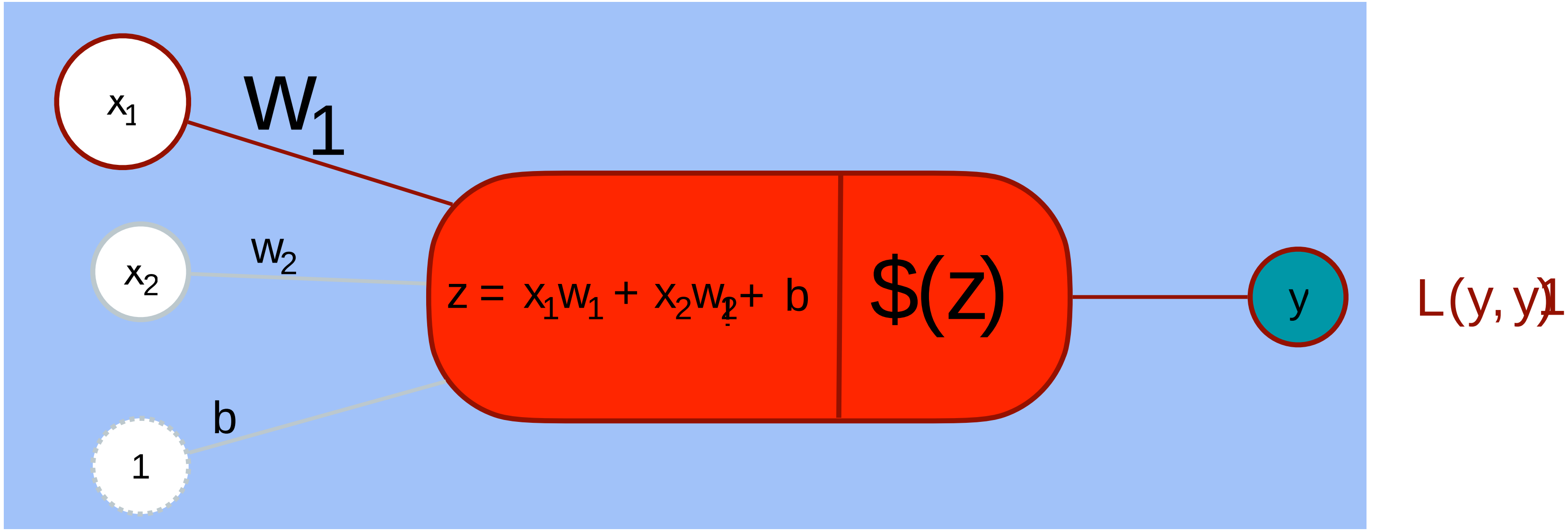
You need gradient descent



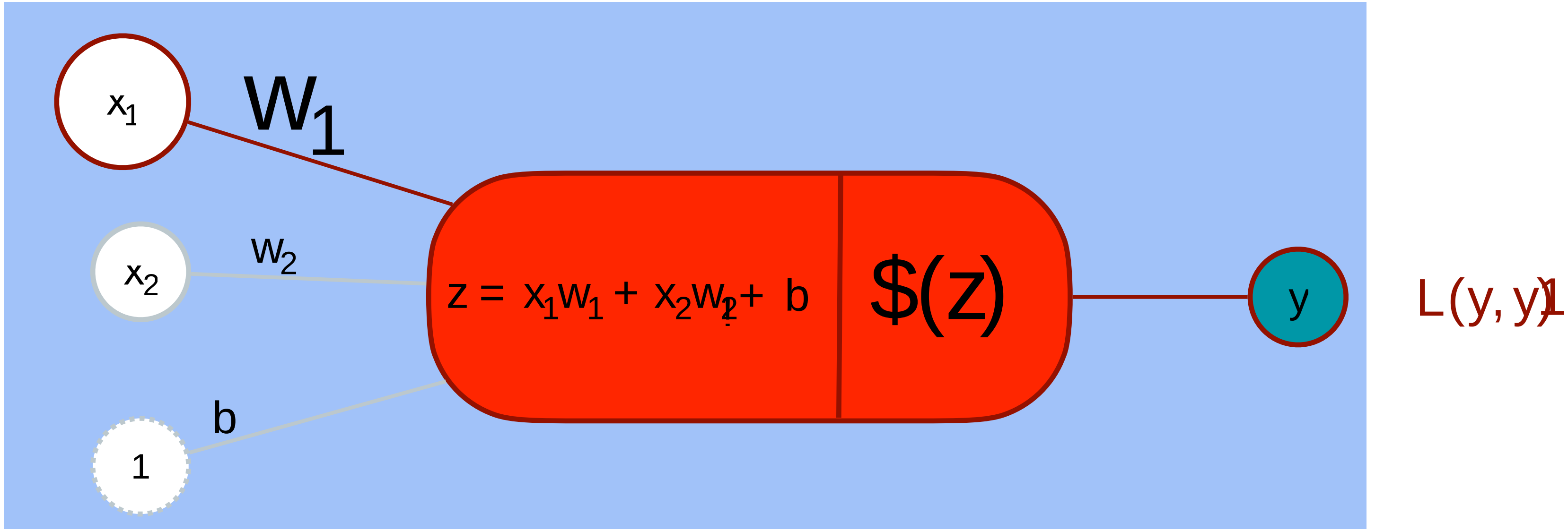
Optimization in Neural Networks and NewtonÕs Method

Classification with a perceptron: Calculating the derivatives

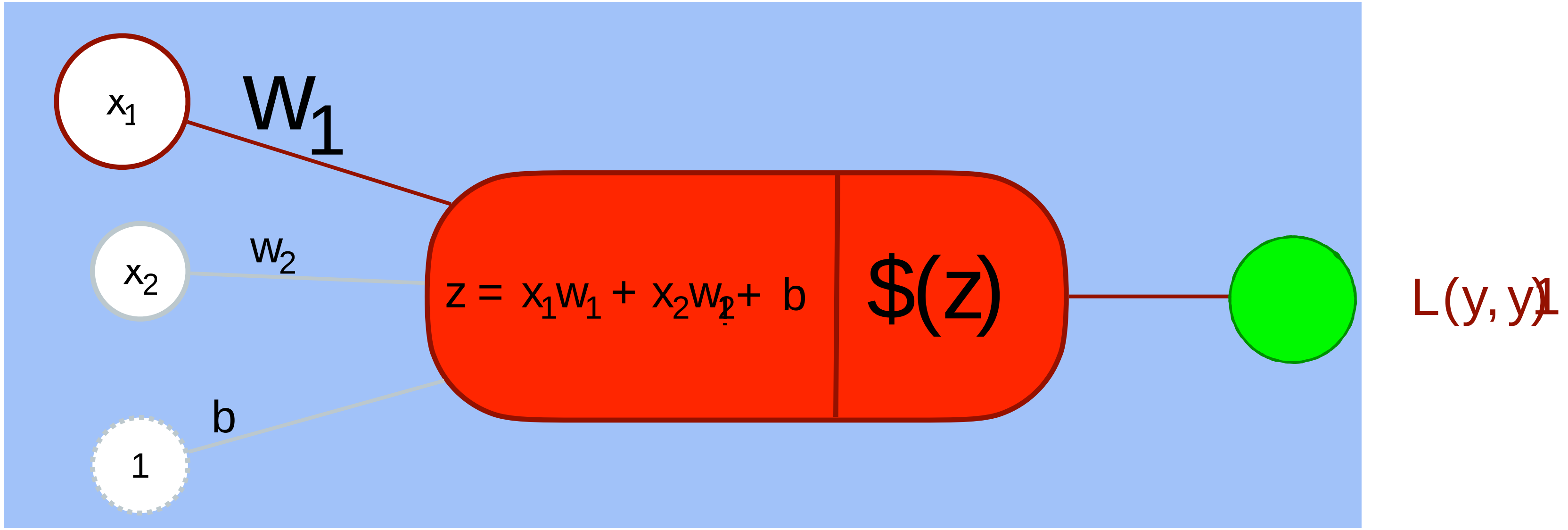
"Classification With a Perceptron



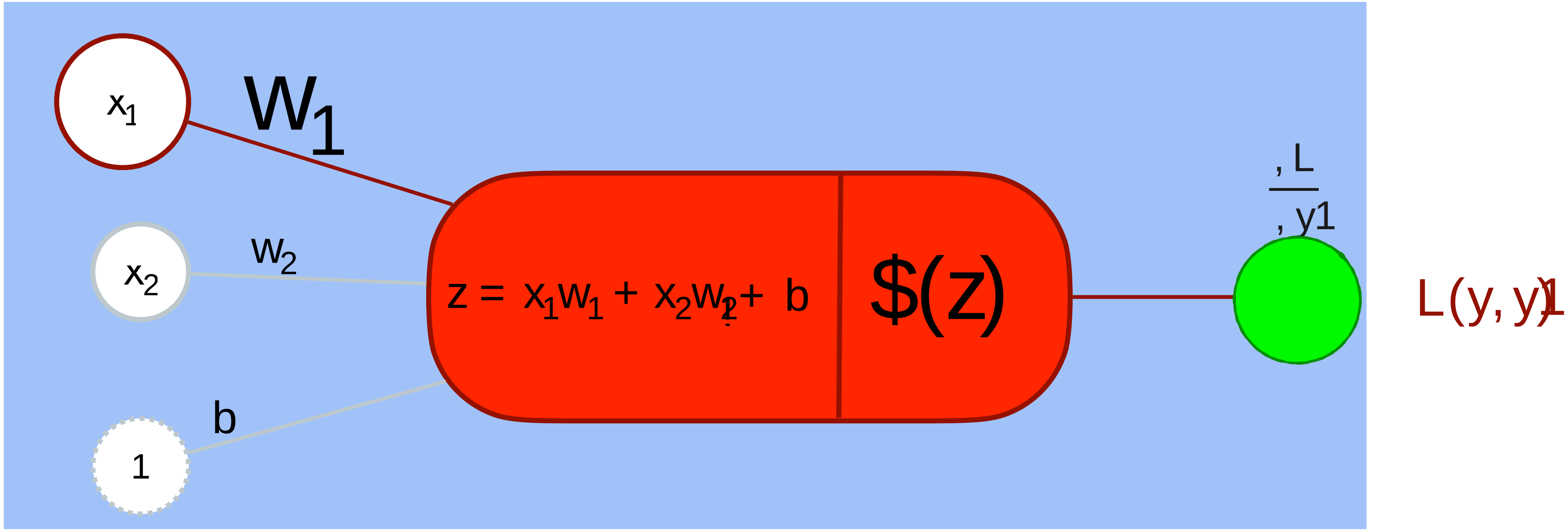
"Classification With a Perceptron



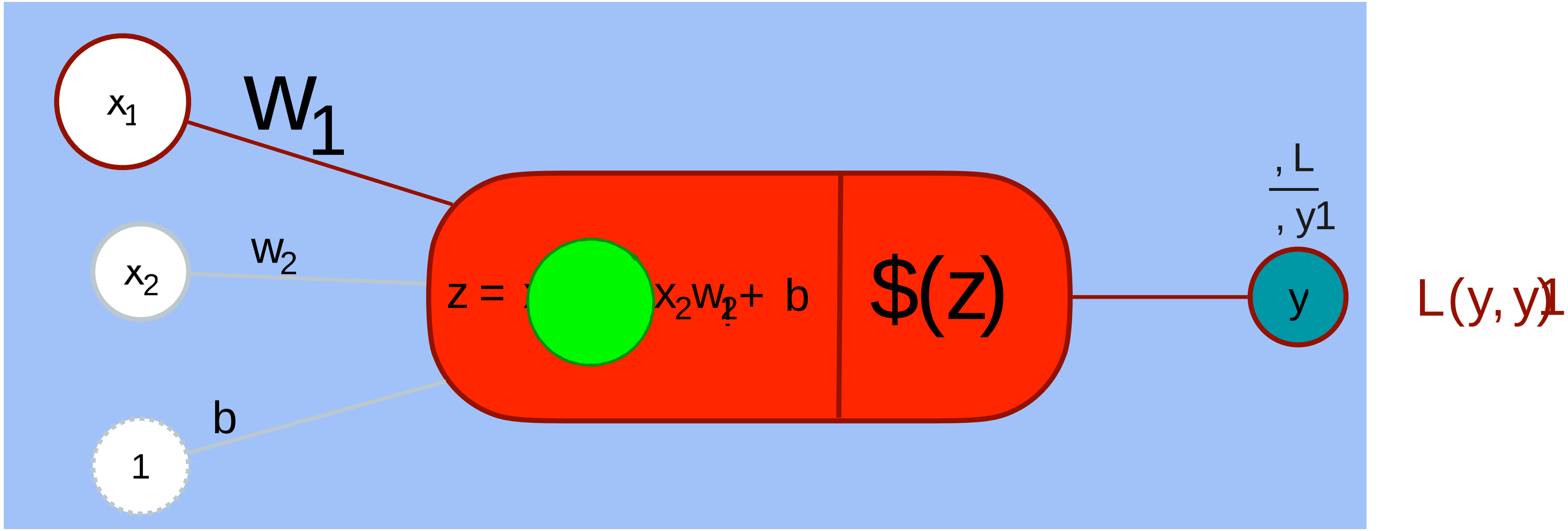
"Classification With a Perceptron



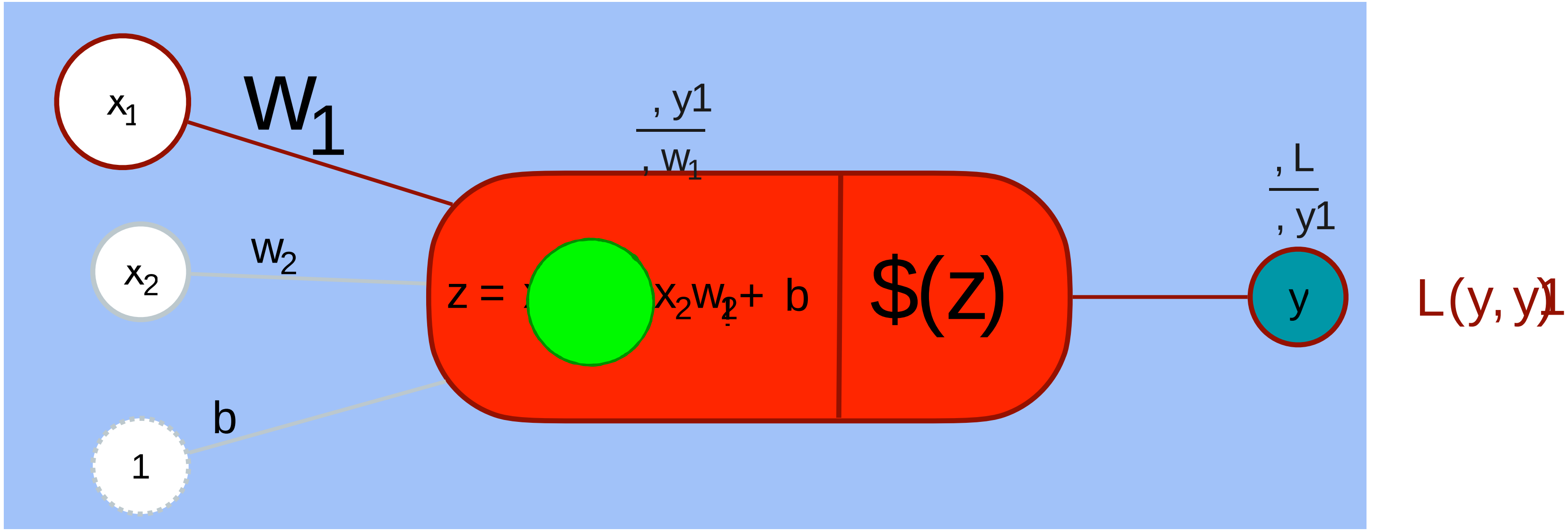
"Classification With a Perceptron



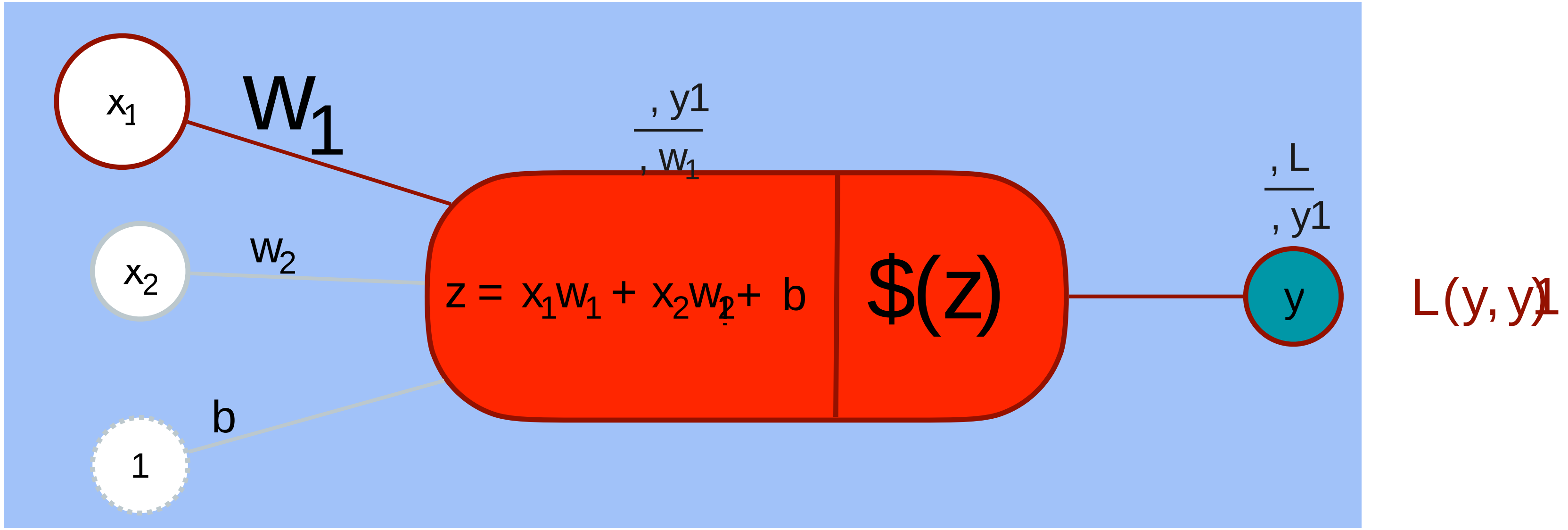
"Classification With a Perceptron



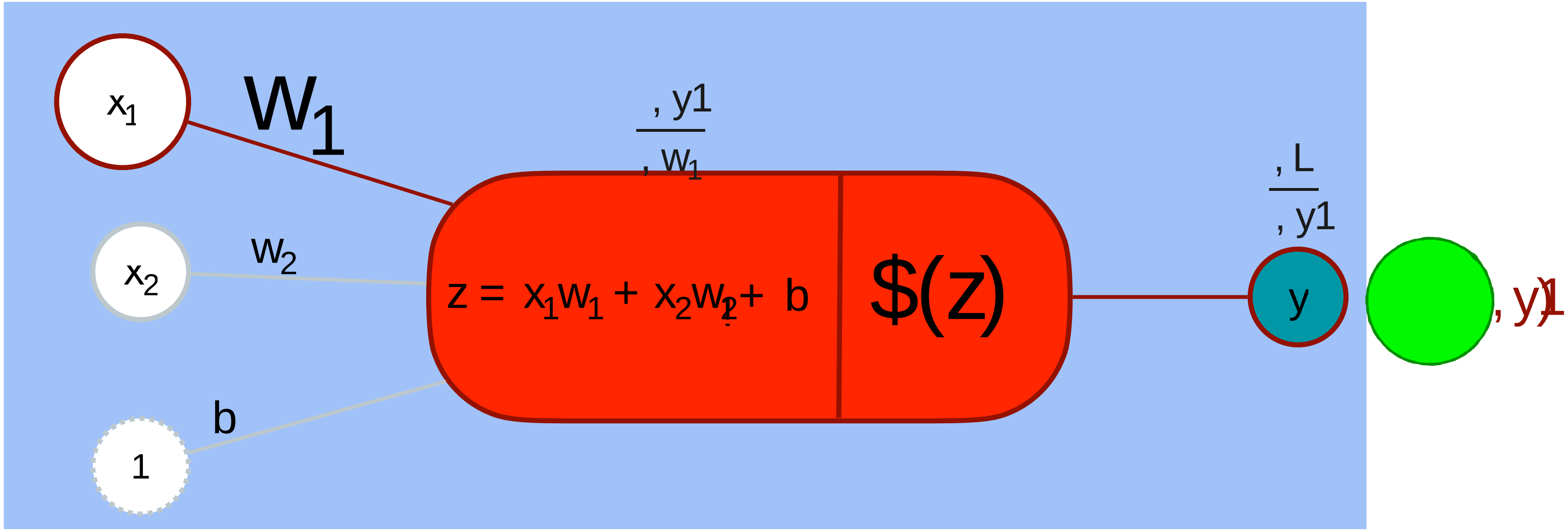
"Classification With a Perceptron



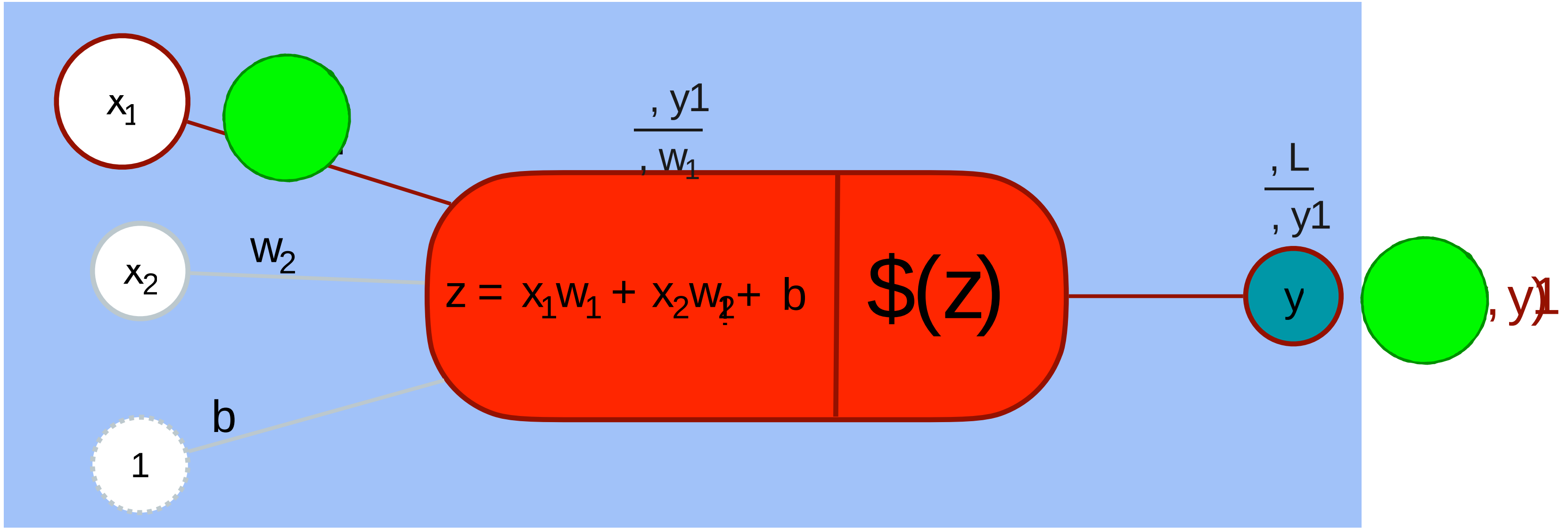
"Classification With a Perceptron



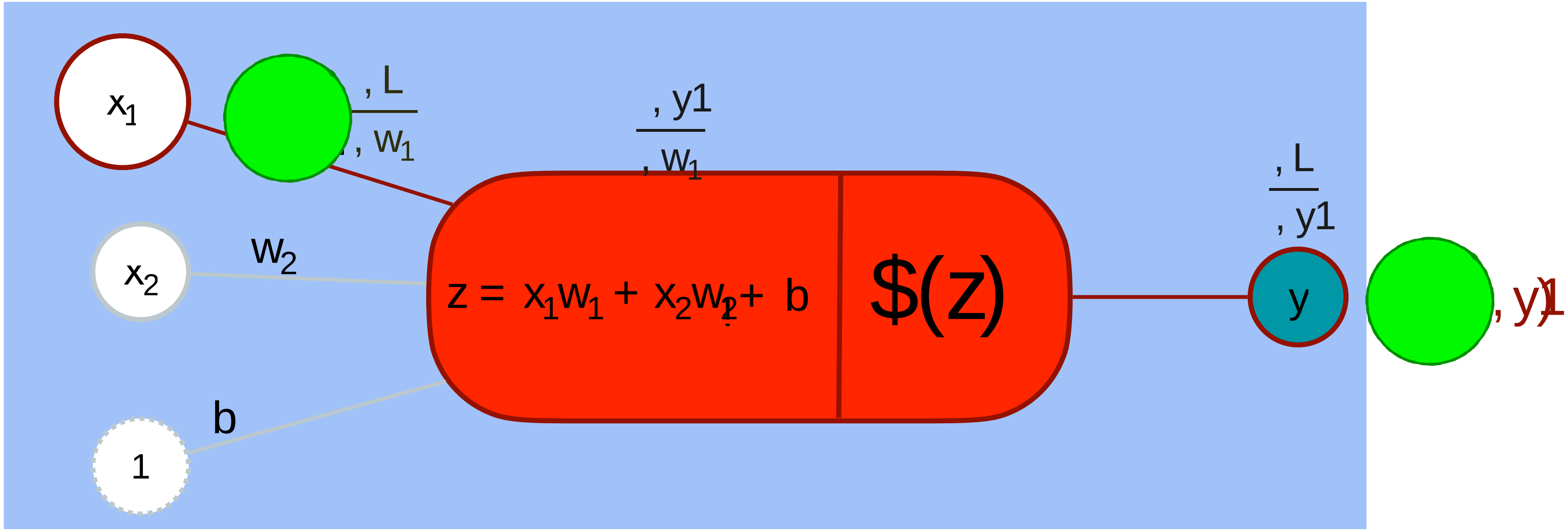
"Classification With a Perceptron



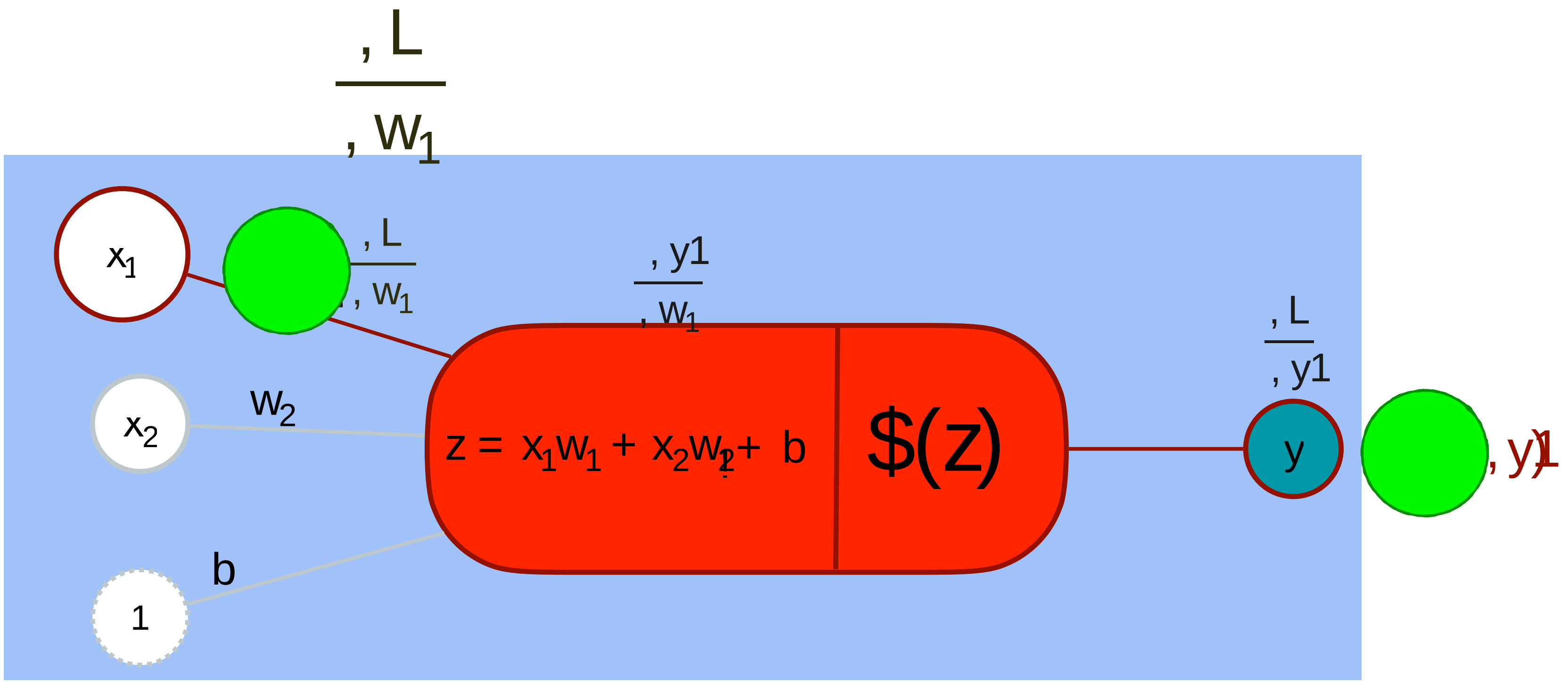
"Classification With a Perceptron



"Classification With a Perceptron

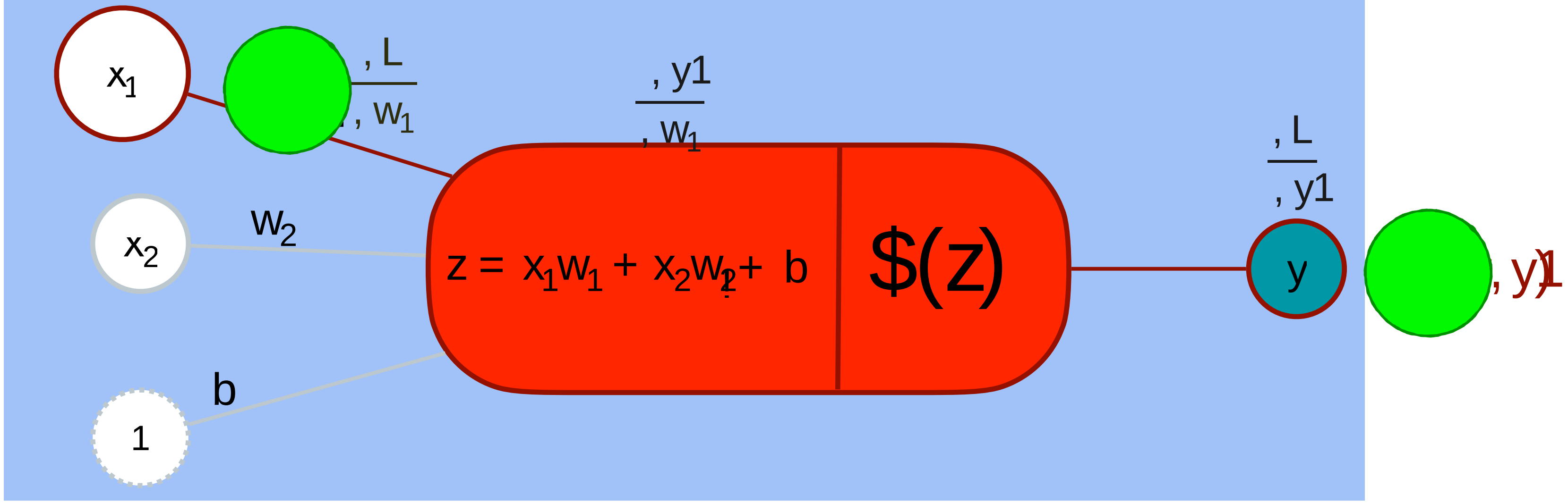


"Classification With a Perceptron



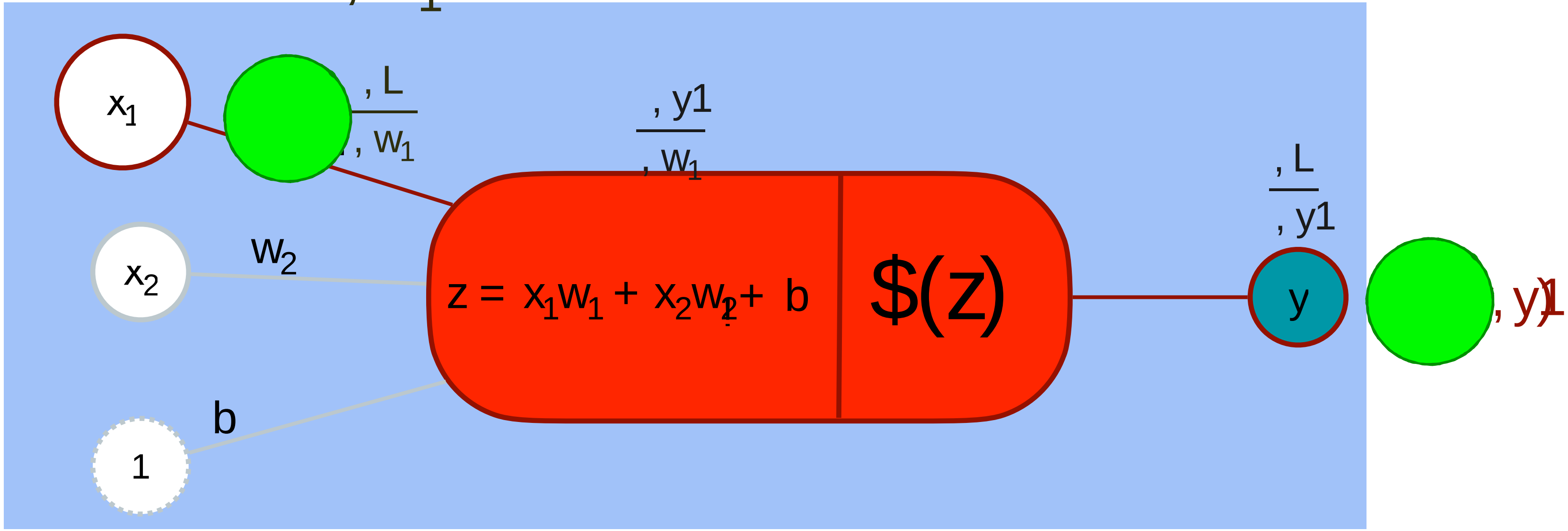
"Classification With a Perceptron

$$\frac{,\ L}{,\ w_1} =$$



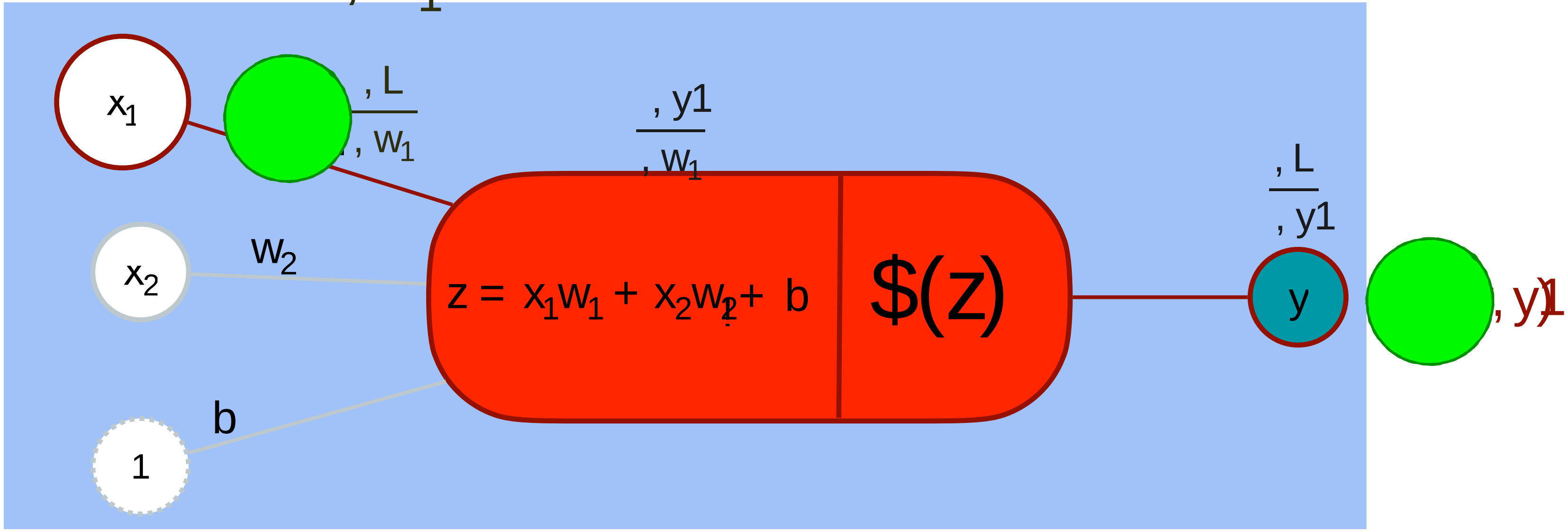
"Classification With a Perceptron

$$\frac{, L}{, w_1} = \frac{, L}{, y1}$$



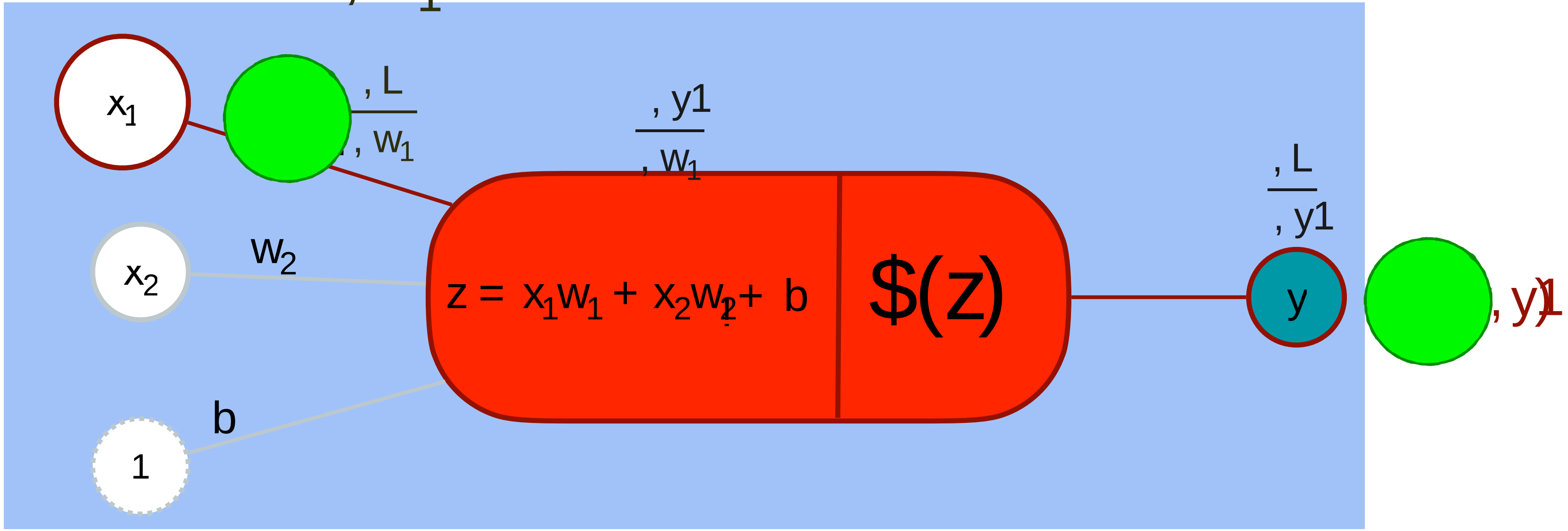
"Classification With a Perceptron

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot$$

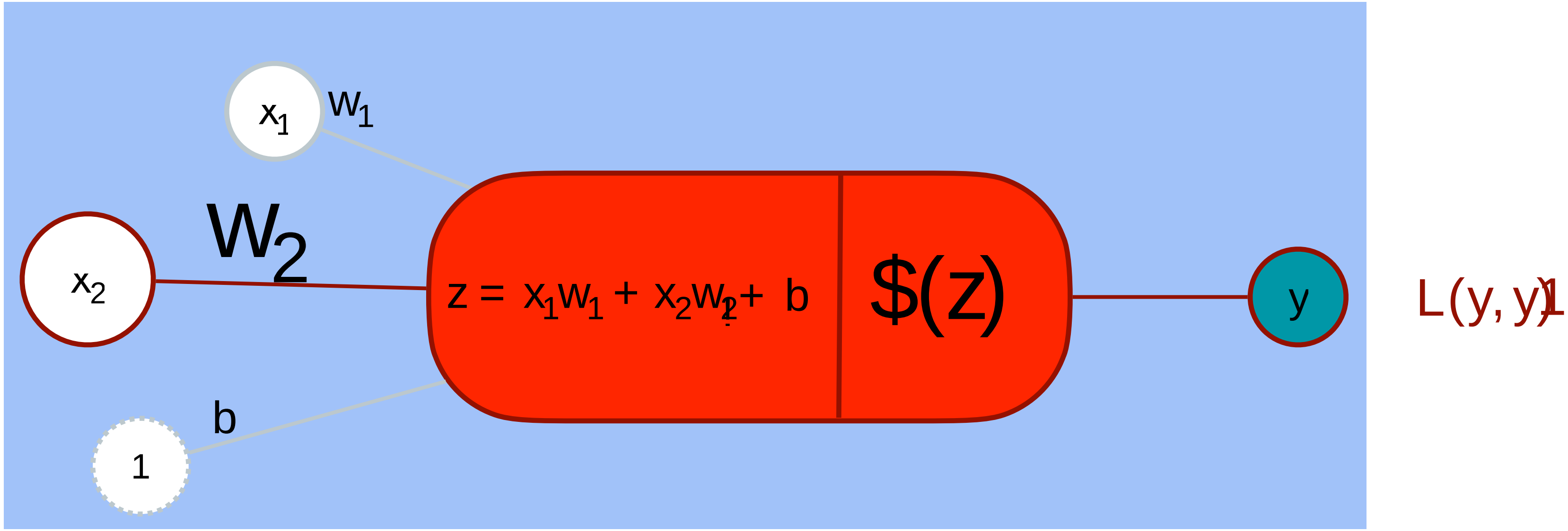


"Classification With a Perceptron

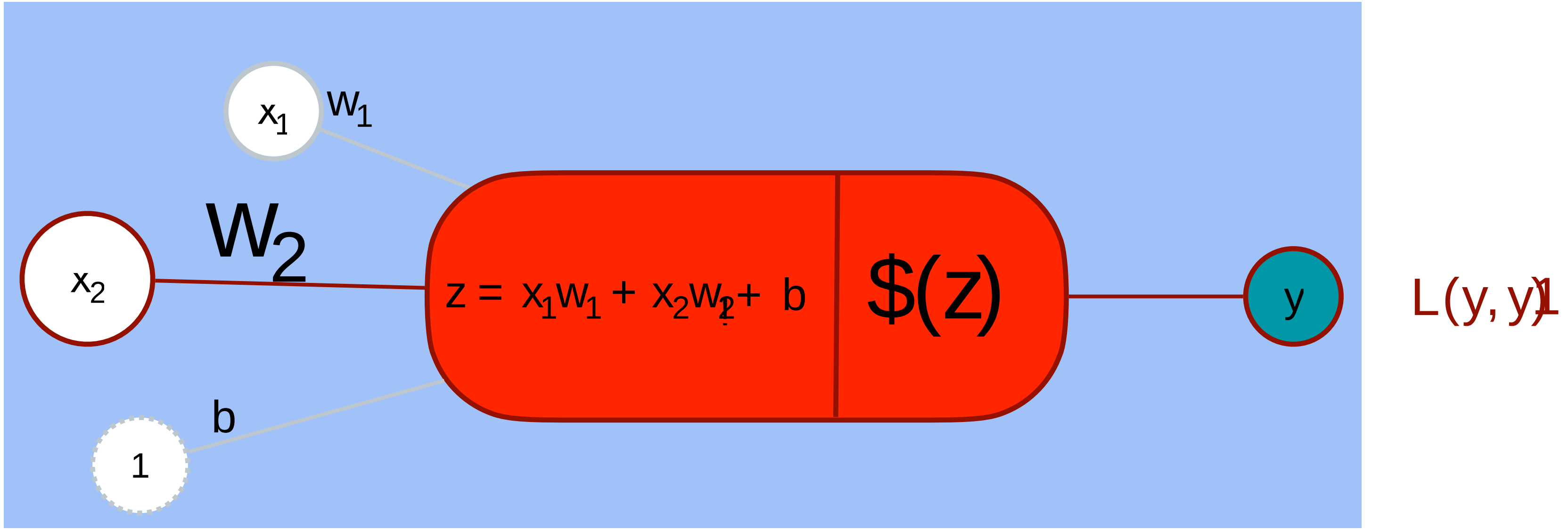
$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$



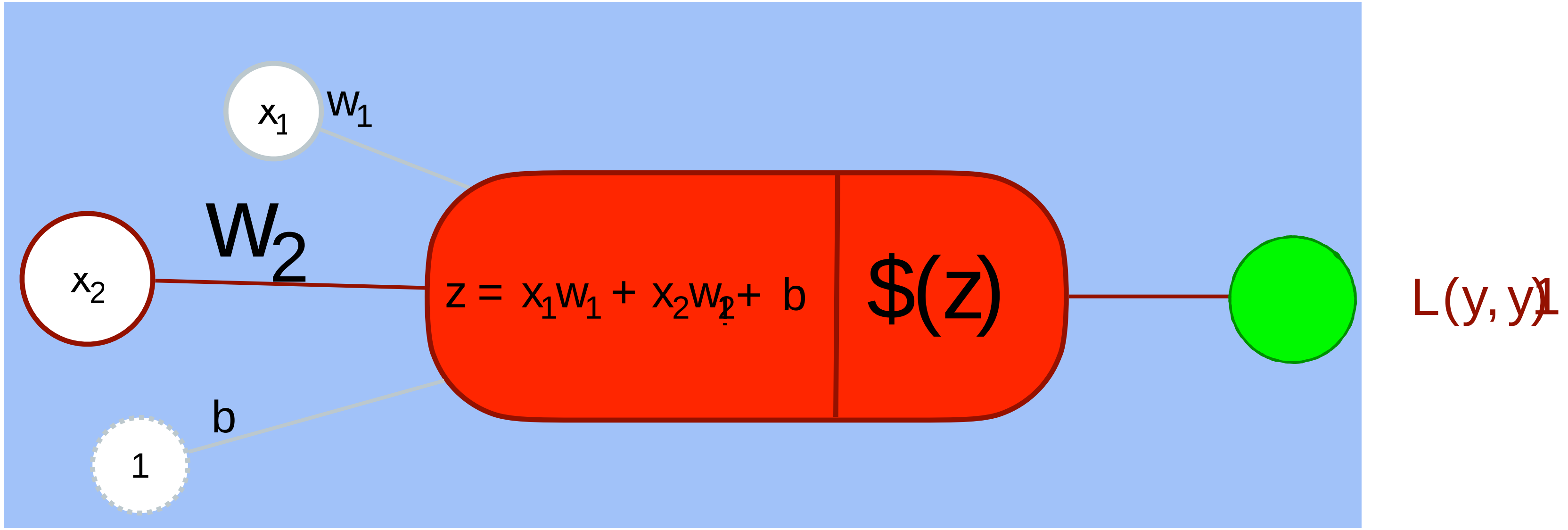
"Classification With a Perceptron



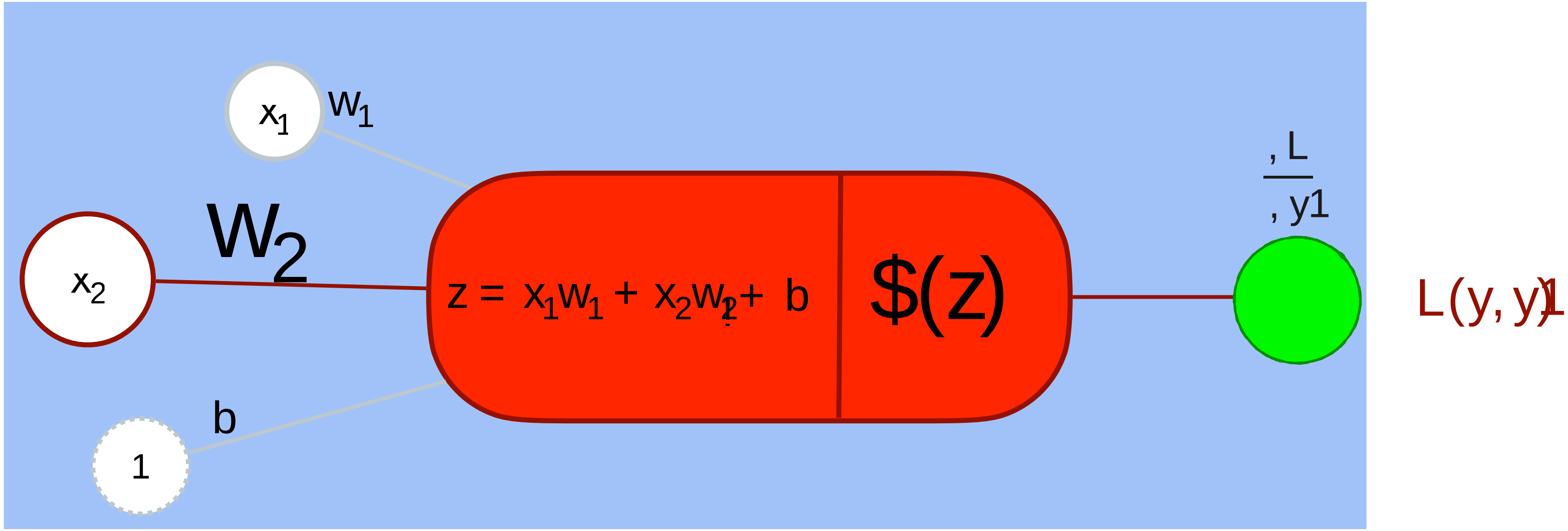
"Classification With a Perceptron



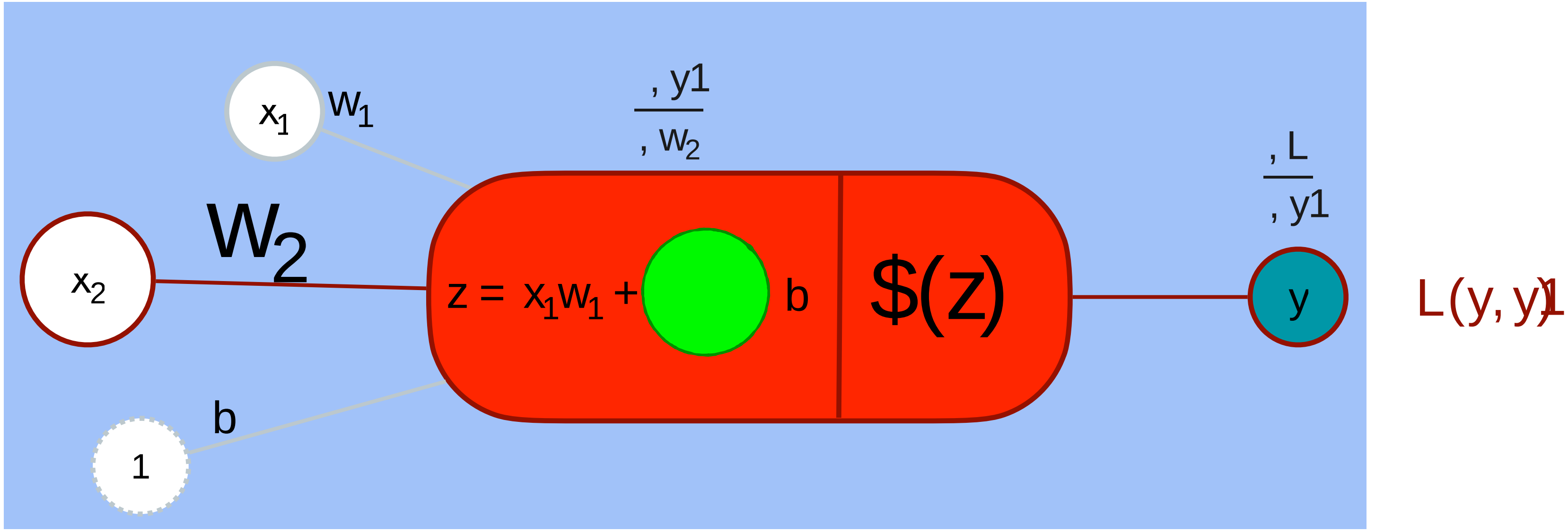
"Classification With a Perceptron



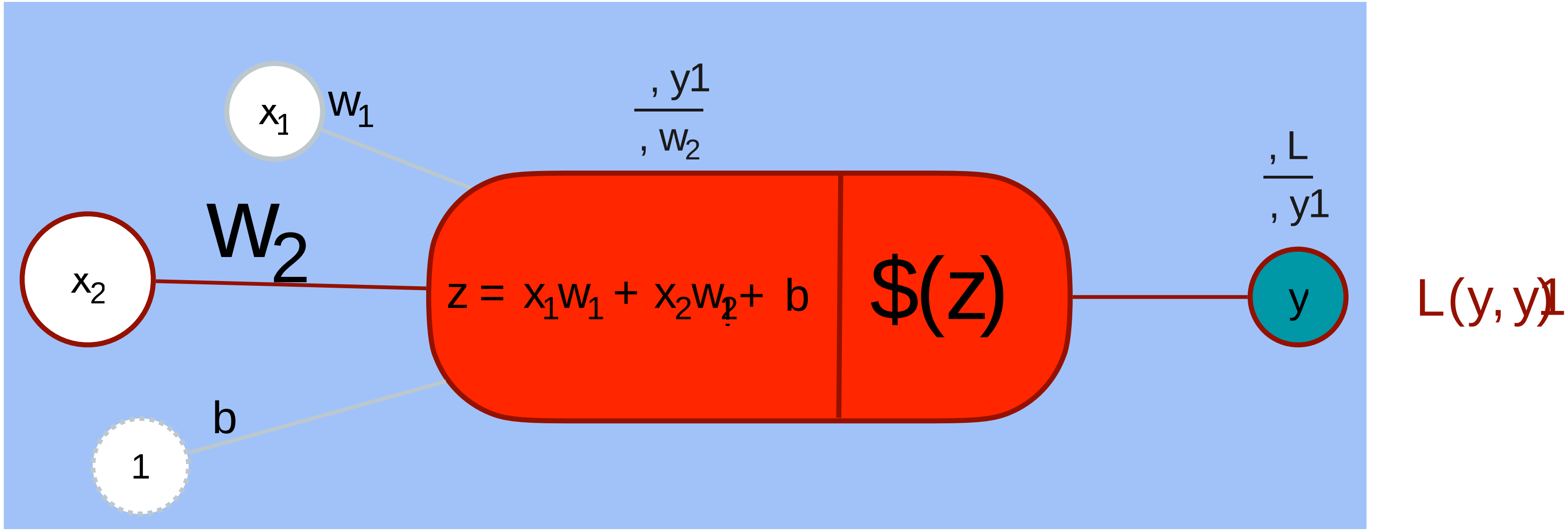
"Classification With a Perceptron



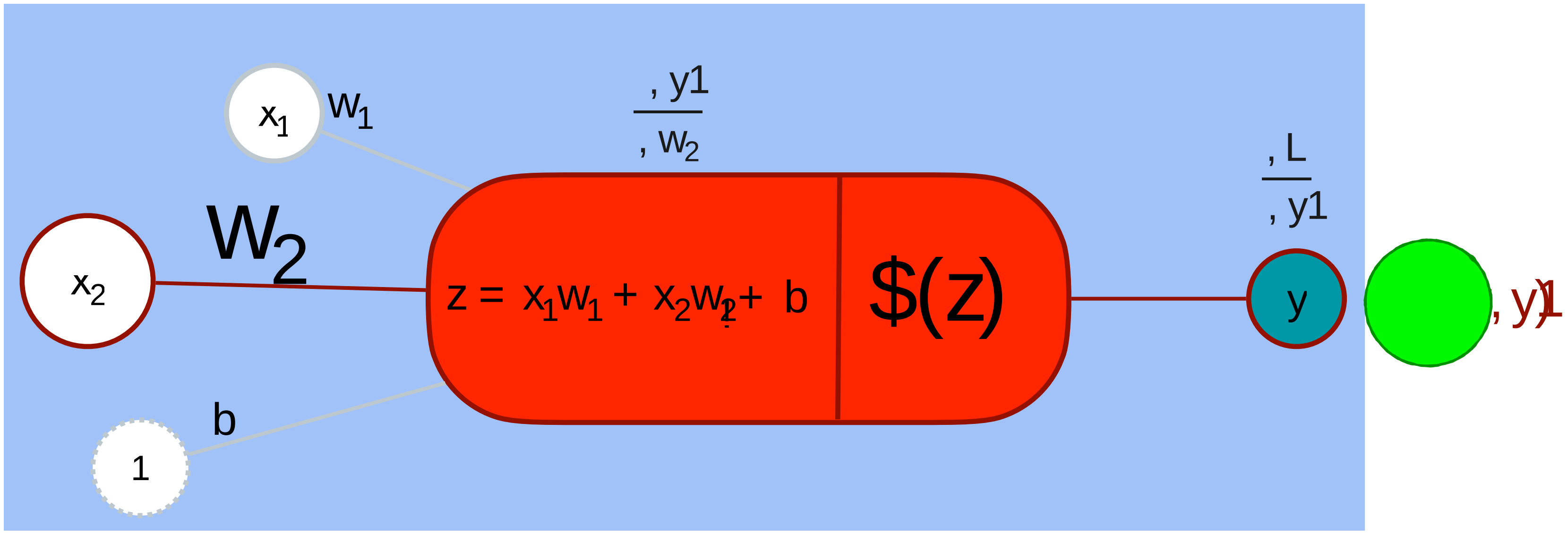
"Classification With a Perceptron



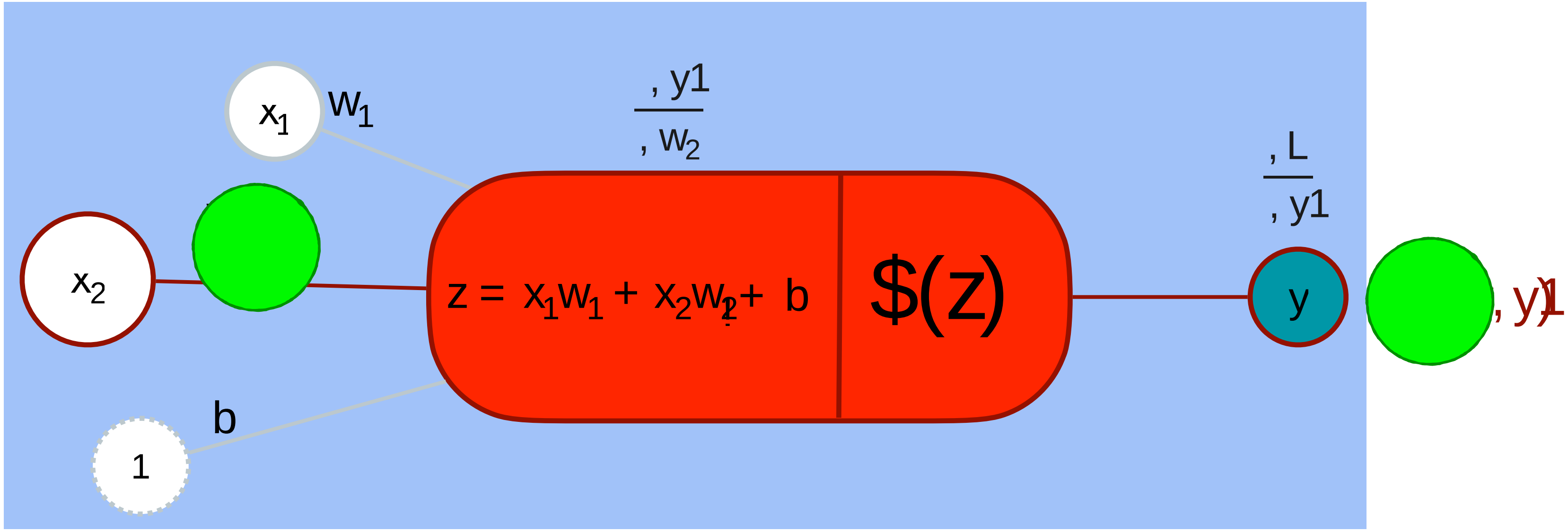
"Classification With a Perceptron



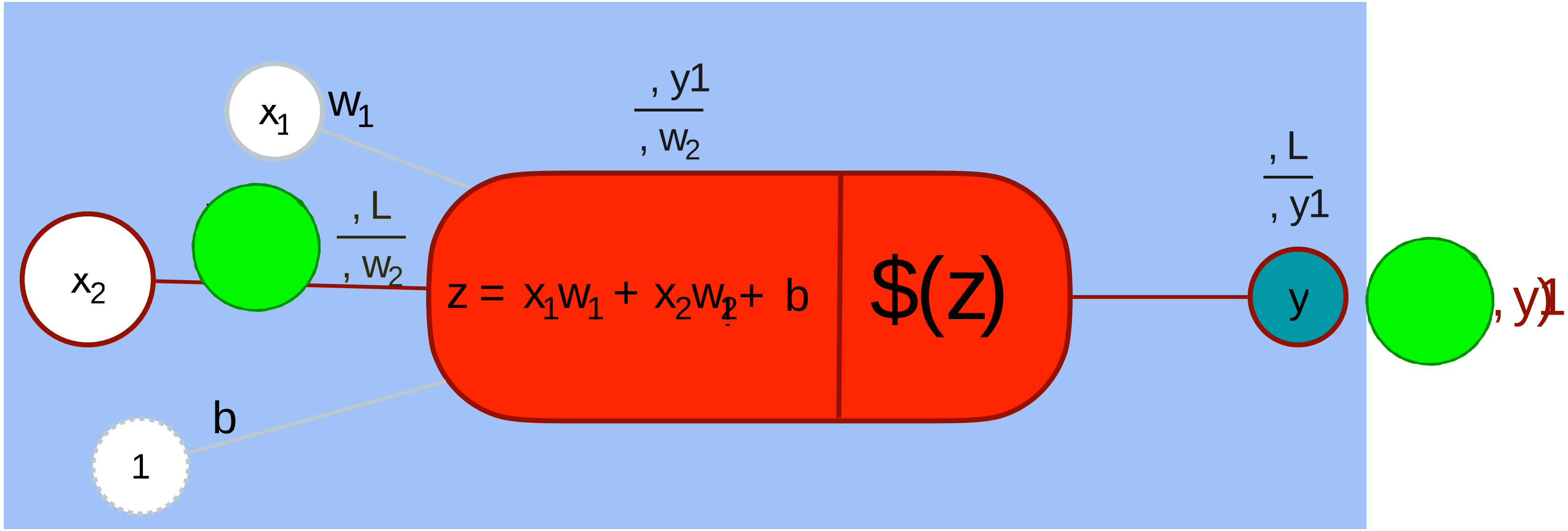
"Classification With a Perceptron



"Classification With a Perceptron

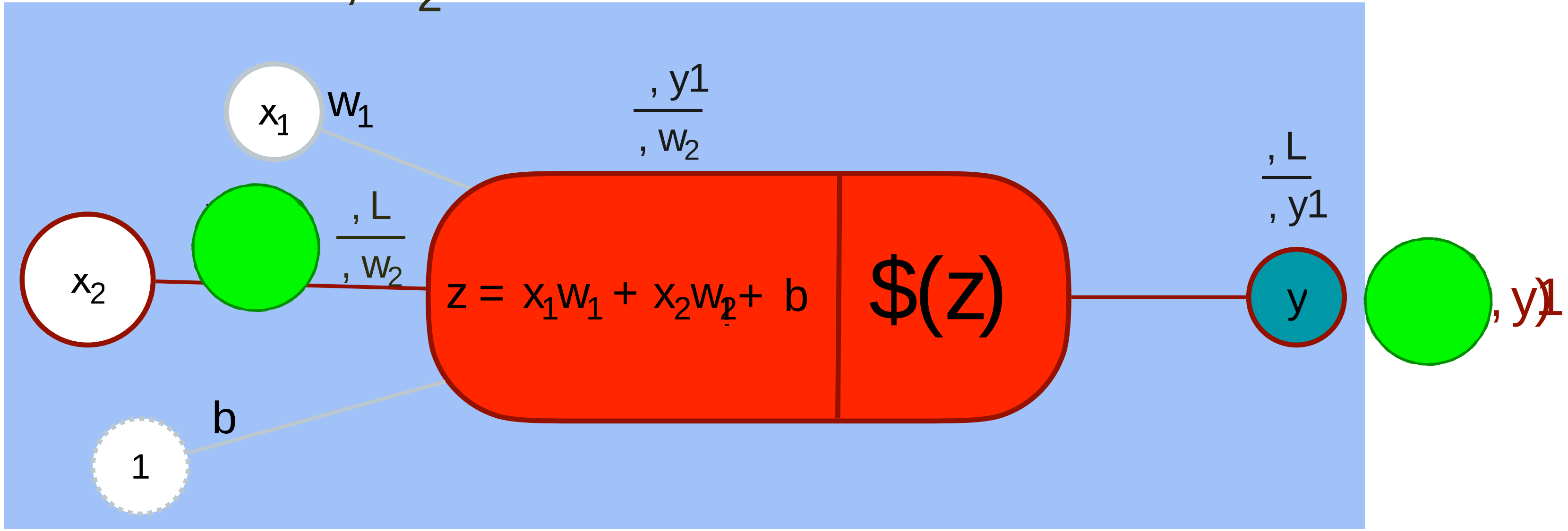


"Classification With a Perceptron

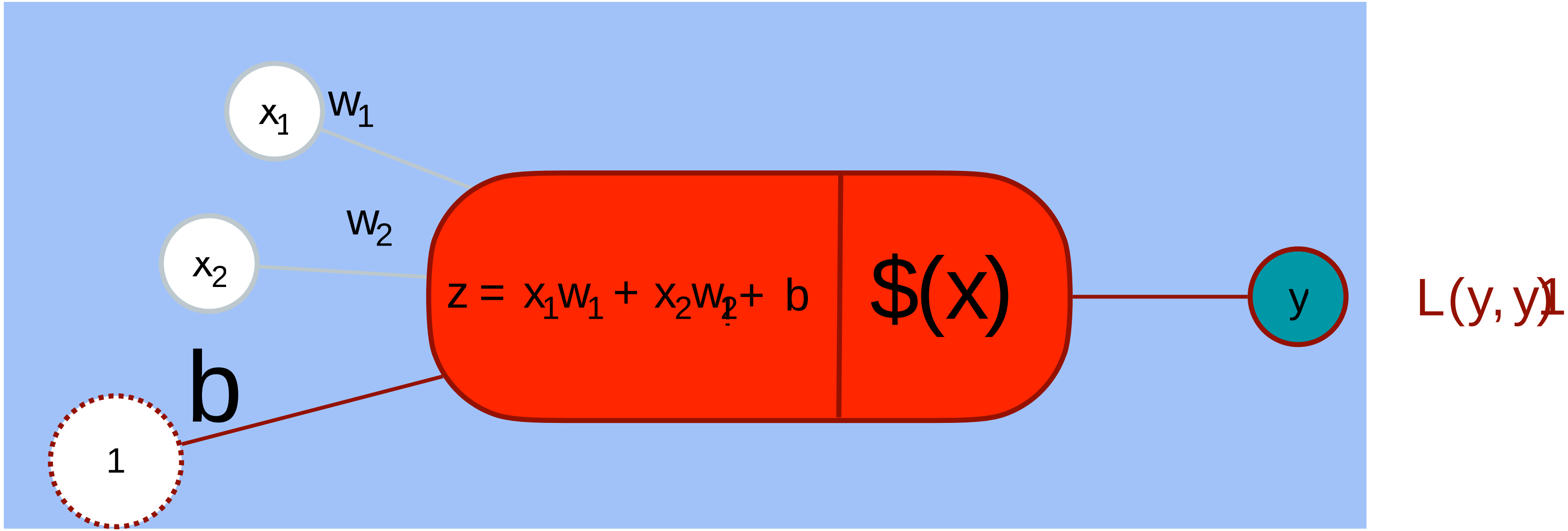


"Classification With a Perceptron

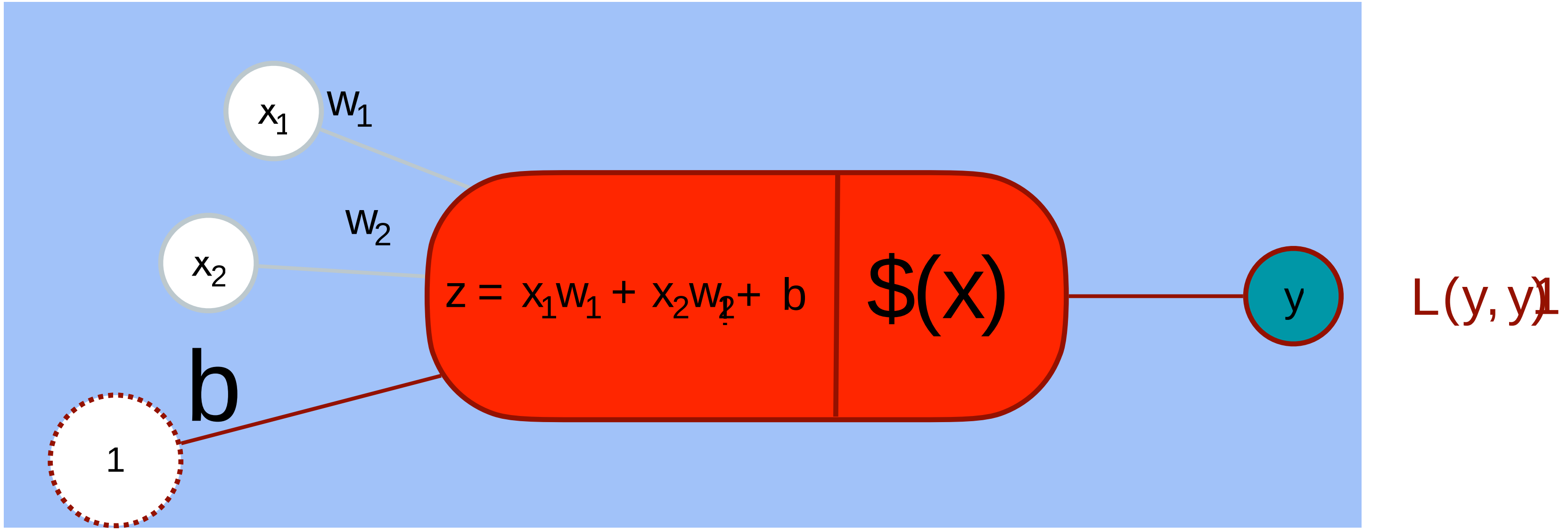
$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$



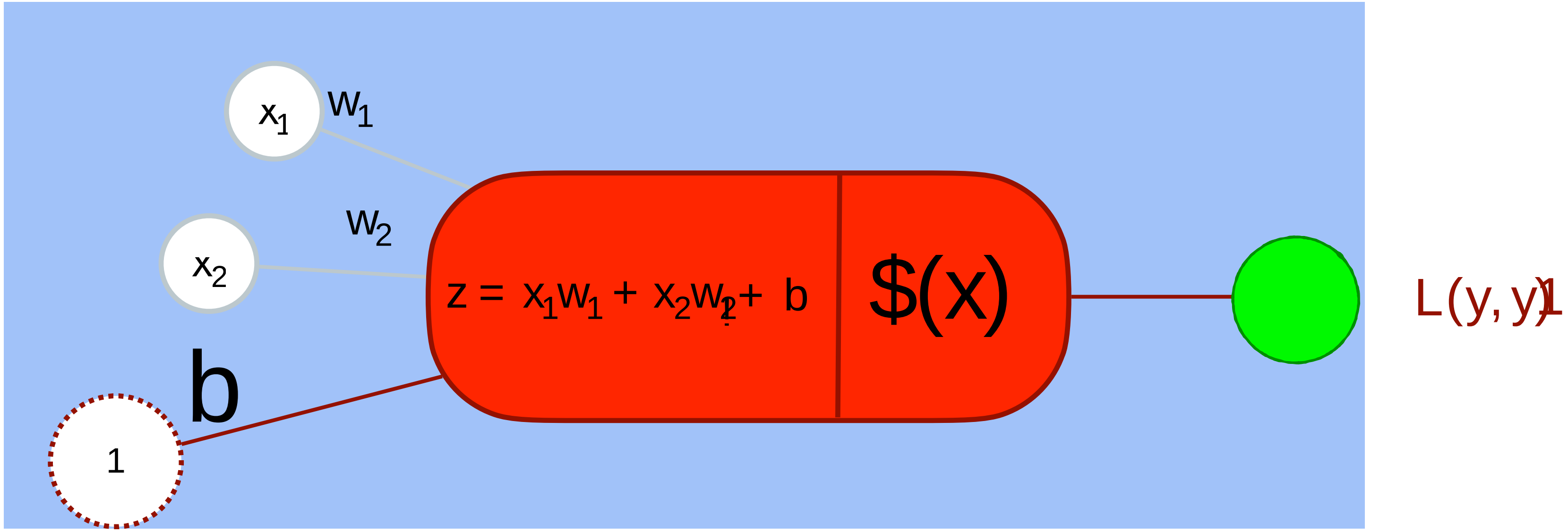
"Classification With a Perceptron



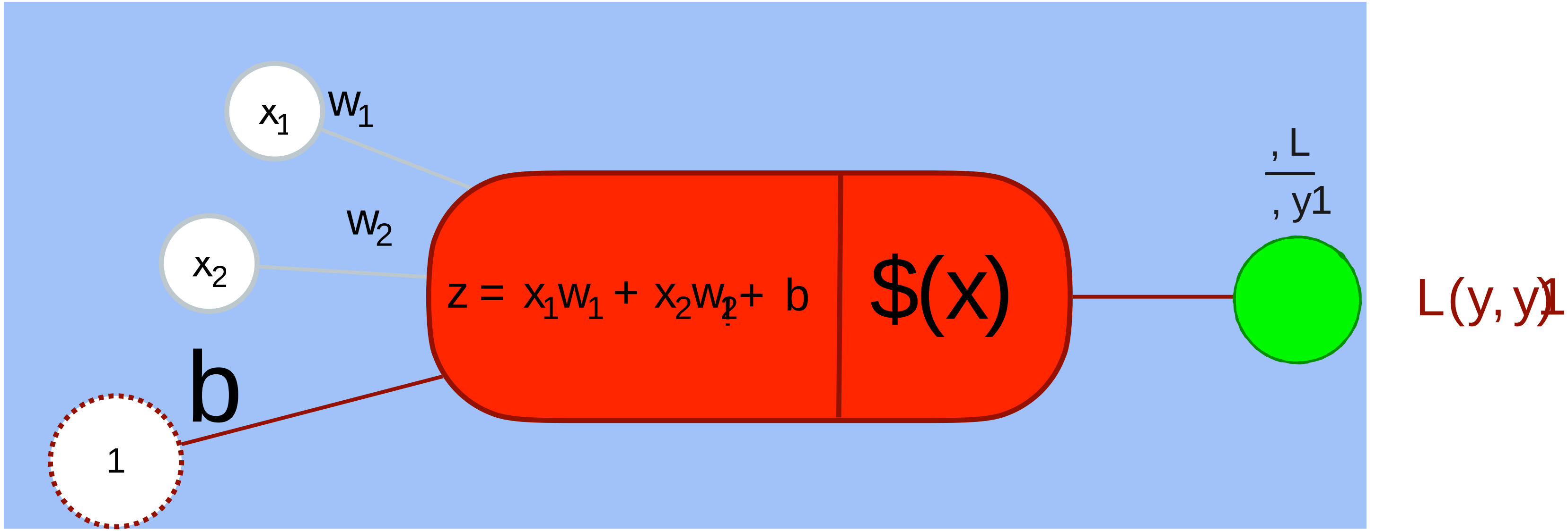
"Classification With a Perceptron



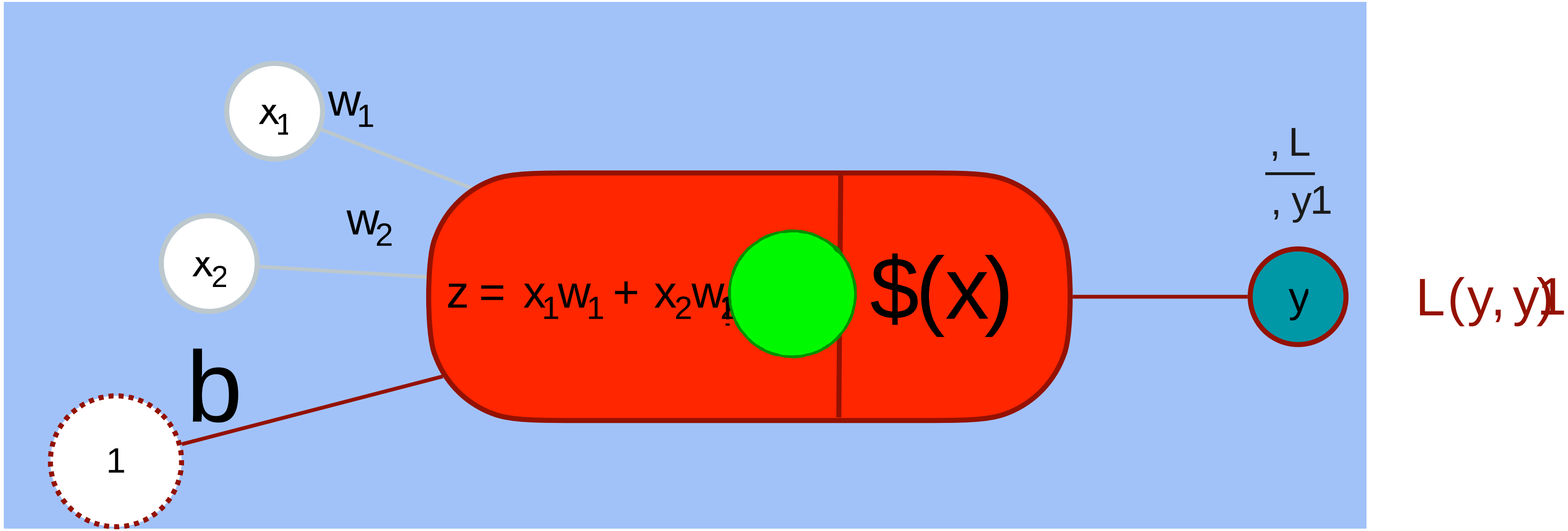
"Classification With a Perceptron



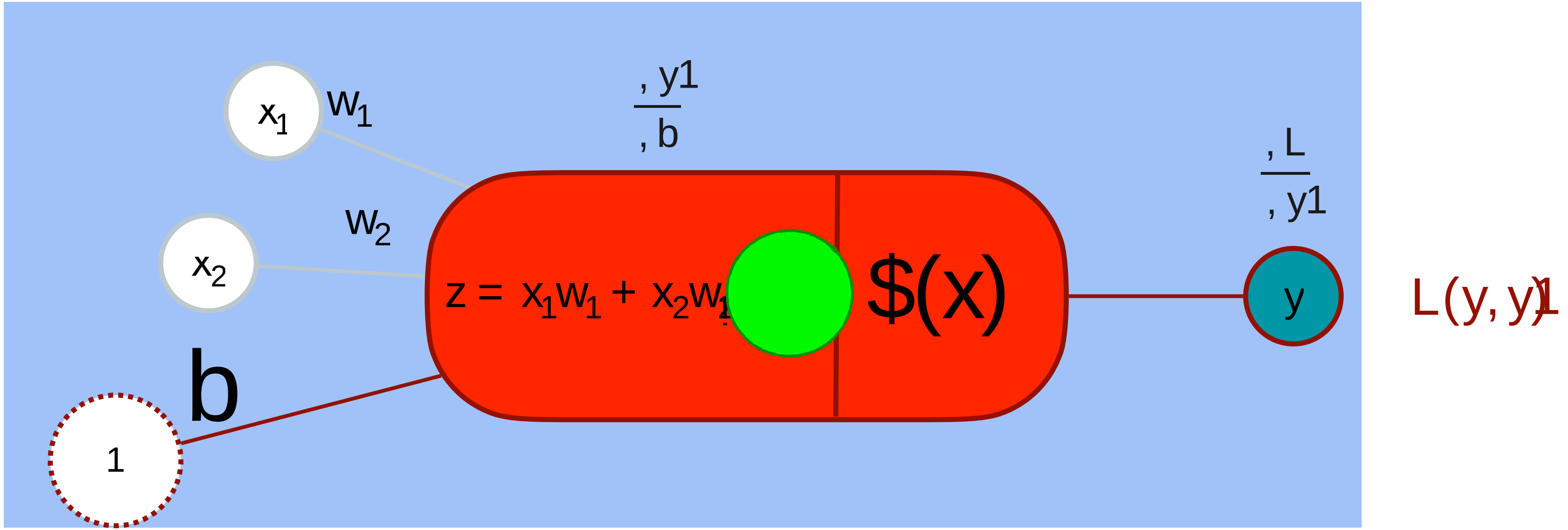
"Classification With a Perceptron



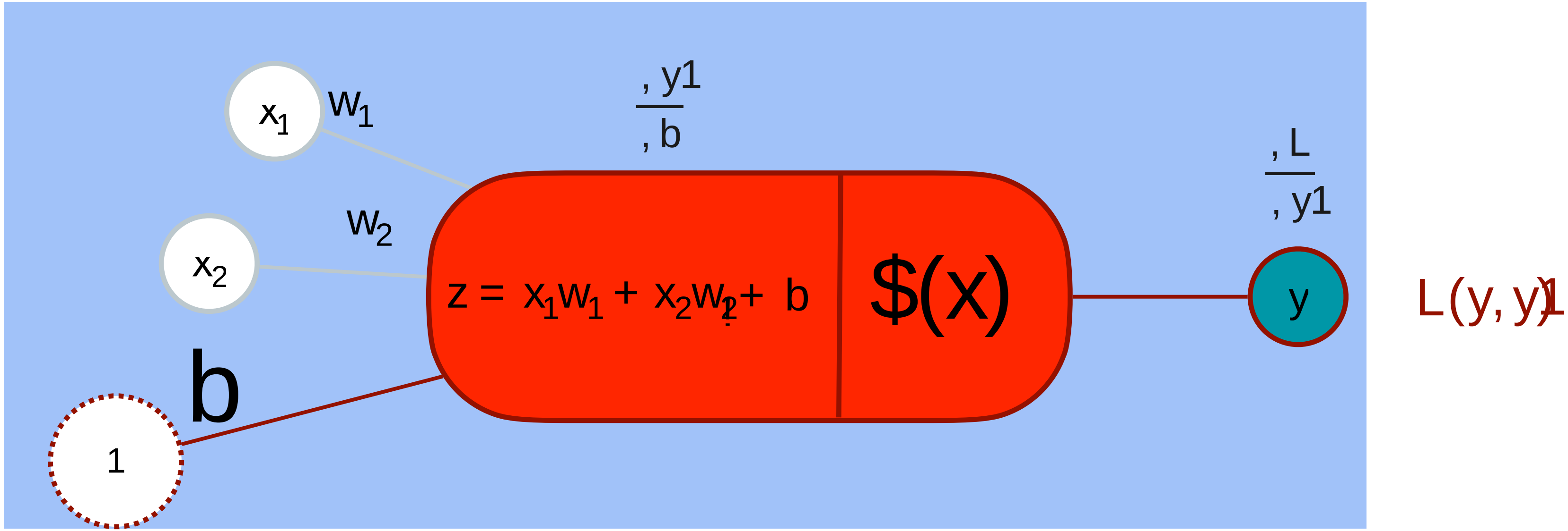
"Classification With a Perceptron



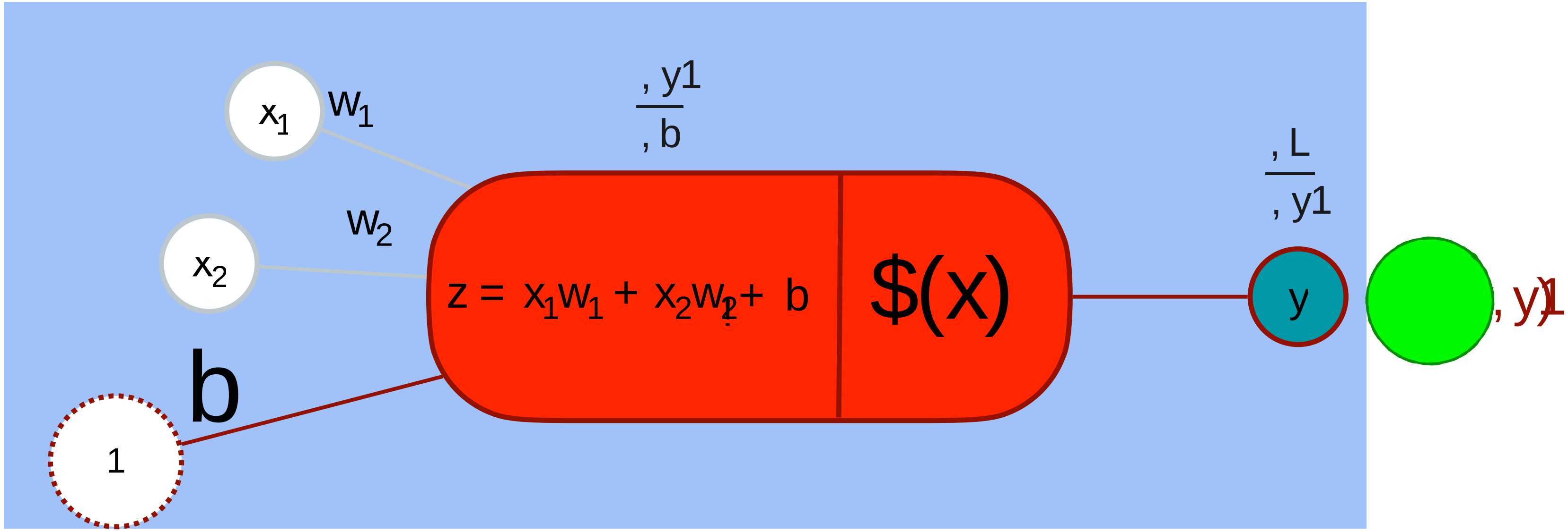
"Classification With a Perceptron



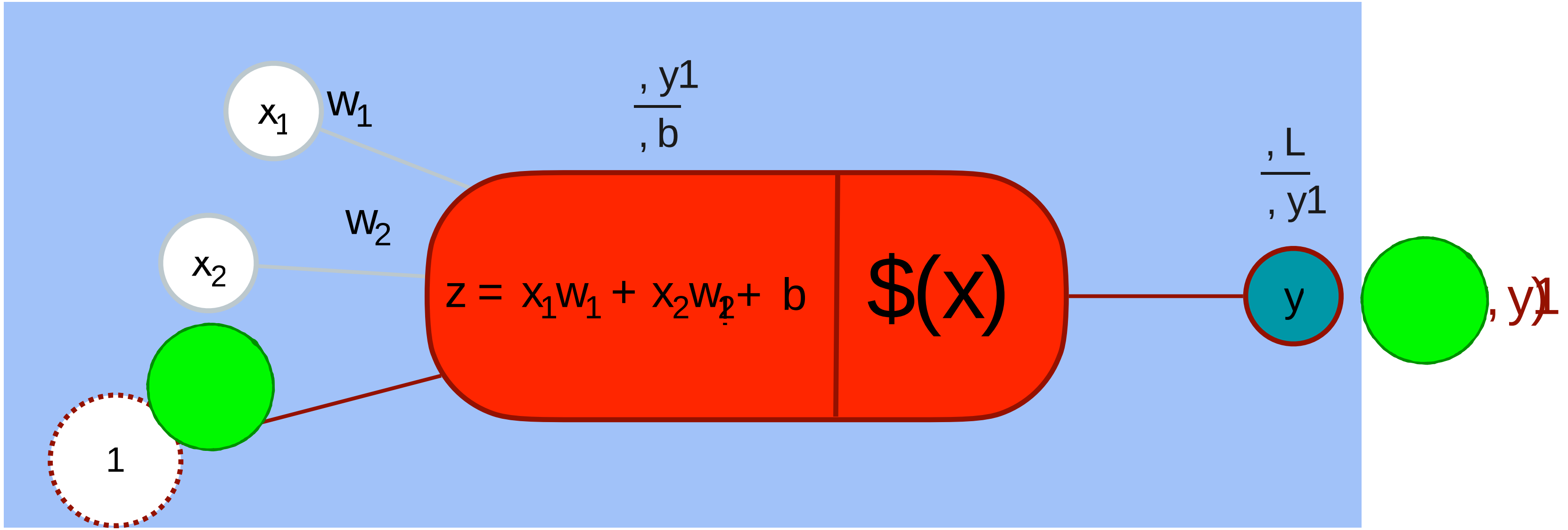
"Classification With a Perceptron



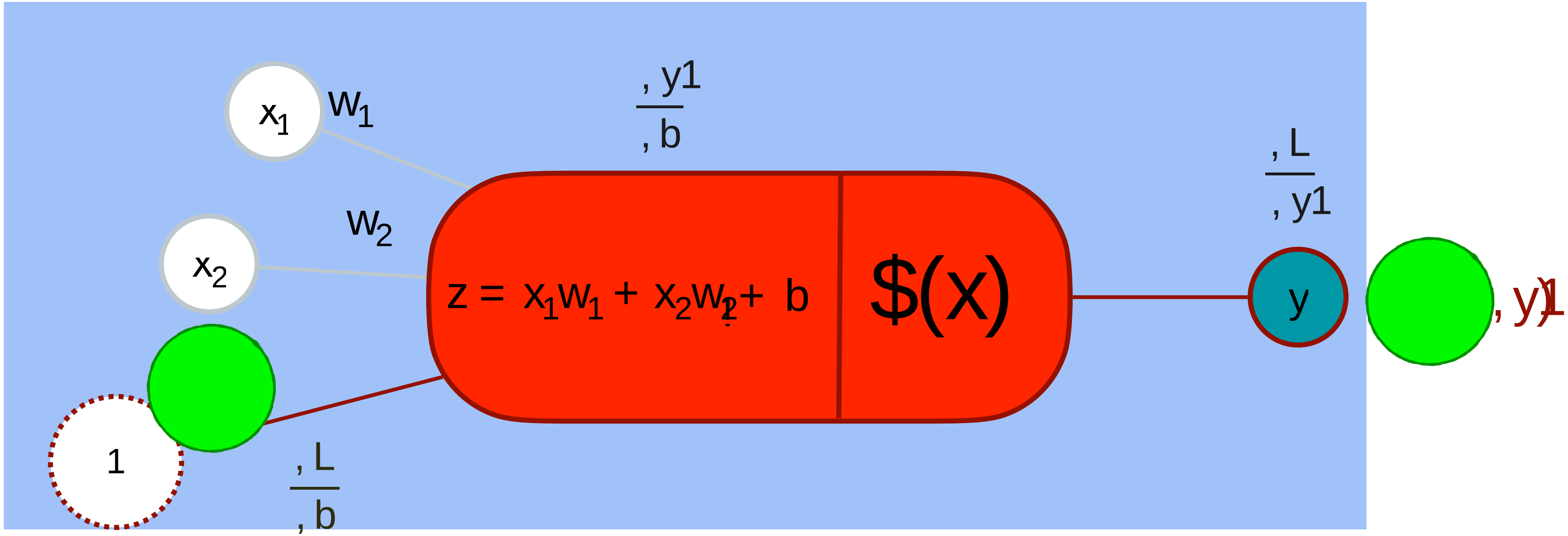
"Classification With a Perceptron



"Classification With a Perceptron

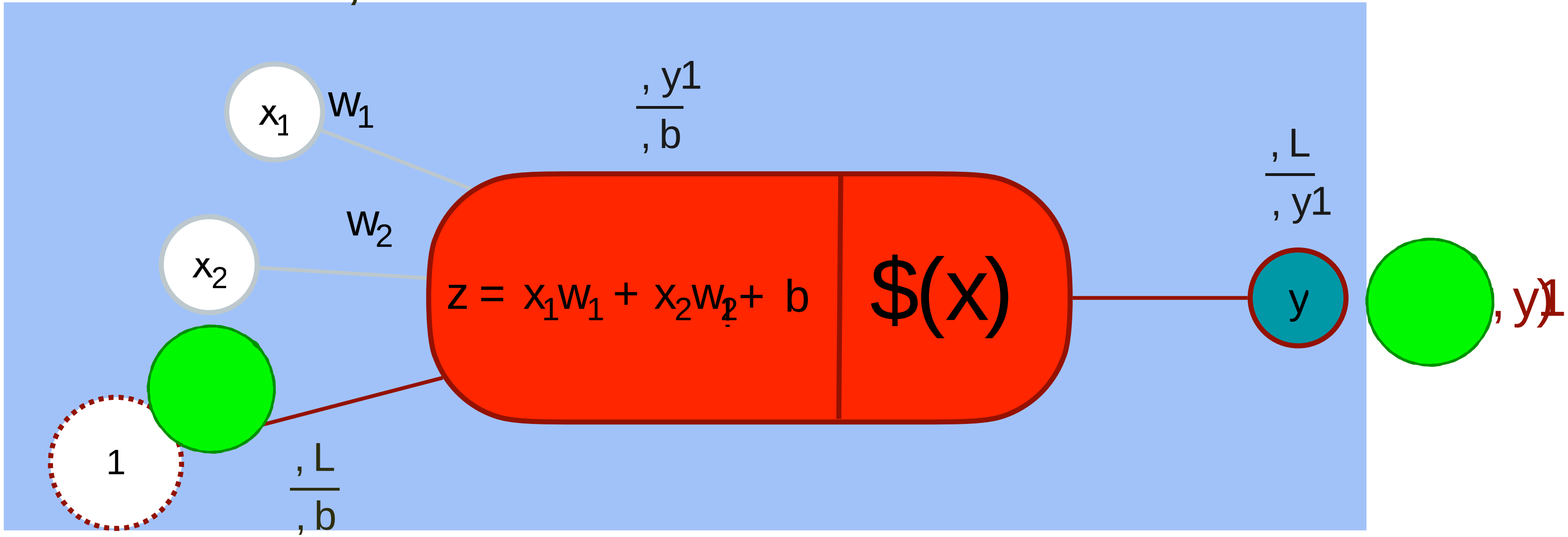


"Classification With a Perceptron



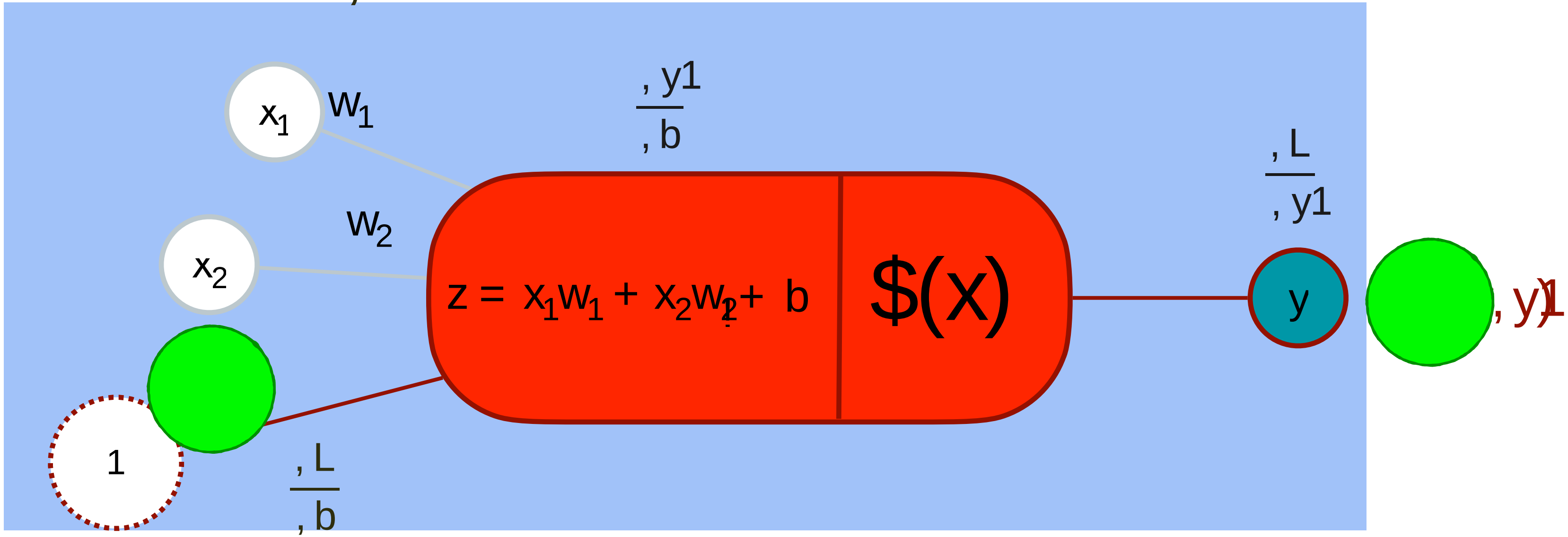
"Classification With a Perceptron

$$\frac{, L}{, b} =$$



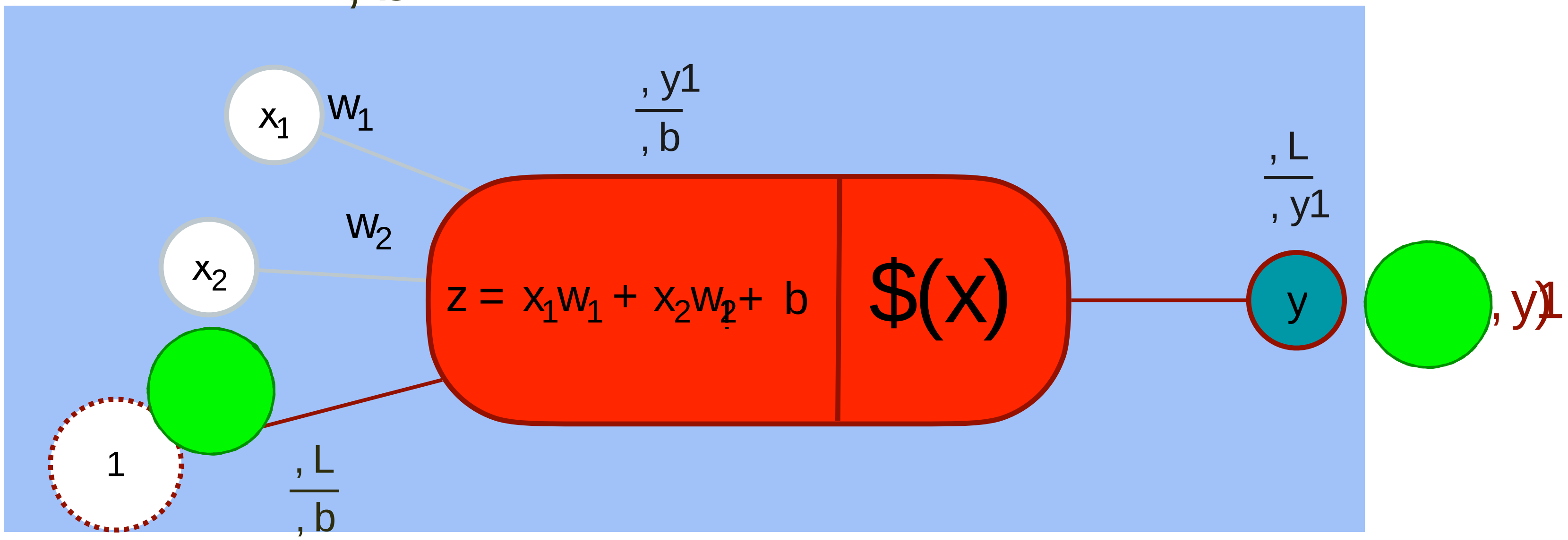
"Classification With a Perceptron

$$\frac{, L}{, b} = \frac{, L}{, y1}$$



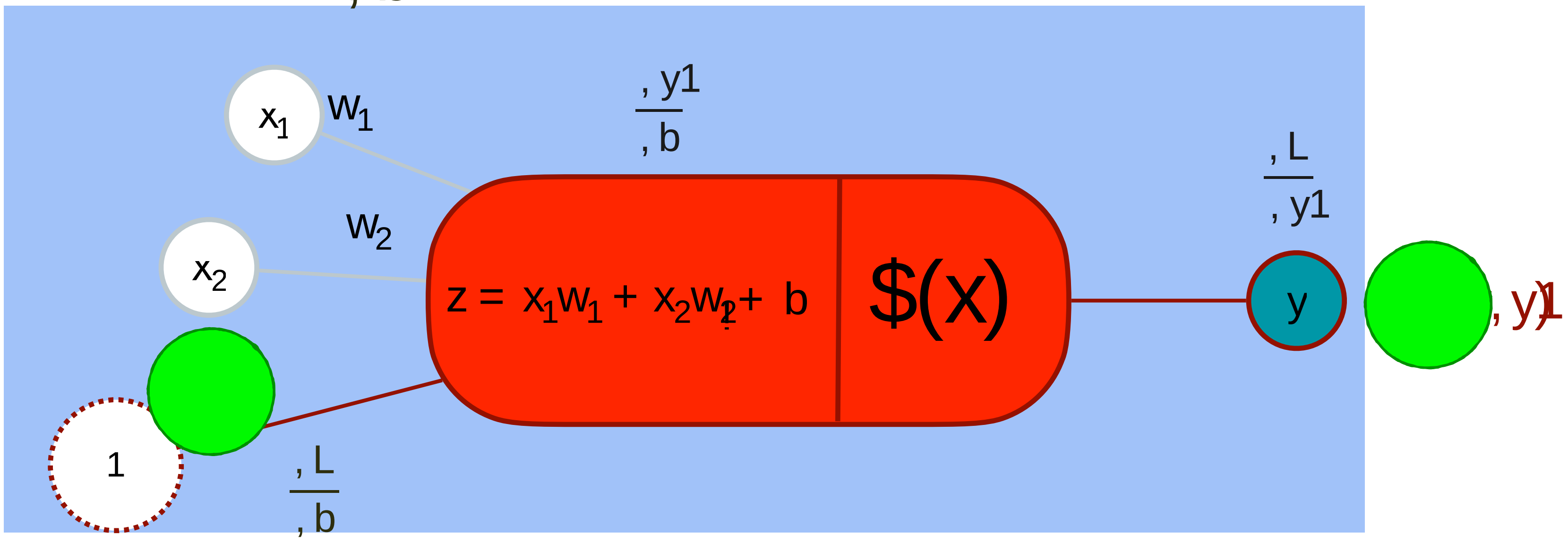
"Classification With a Perceptron

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot$$



"Classification With a Perceptron

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$



"Classification With a Perceptron

$$\frac{,\ L}{,\ b} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ b}$$

"Classification With a Perceptron

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

"Classification With a Perceptron

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

"Classification With a Perceptron

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

$$\frac{, L}{, y1} =$$

$$\frac{, y1}{, b} =$$

$$\frac{, y1}{, w_1} =$$

$$\frac{, y1}{, w_2} =$$

"Classification With a Perceptron

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

$$\frac{, L}{, y1} =$$

$$\frac{, y1}{, b} =$$

$$\frac{, y1}{, w_1} =$$

$$\frac{, y1}{, w_2} =$$

?

"Classification With a Perceptron

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_1} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

$$\frac{,L}{,w_2} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_2}$$

$$\frac{,L}{,y1} =$$

$$\frac{,y1}{,b} =$$

$$\frac{,y1}{,w_1} =$$

$$\frac{,y1}{,w_2} =$$

?

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

"Classification With a Perceptron

$$\frac{,L}{,b} = \frac{,L}{,y1} \cdot \frac{,y1}{,b}$$

$$\frac{,L}{,w_1} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_1}$$

$$\frac{,L}{,w_2} = \frac{,L}{,y1} \cdot \frac{,y1}{,w_2}$$

$$\frac{,L}{,y1} =$$

$$\frac{,y1}{,b} =$$

$$\frac{,y1}{,w_1} =$$

$$\frac{,y1}{,w_2} =$$

?

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

$$L(y, y1) = -y\ln(y1) " (1 " y)\ln(1 " y1)$$

"Classification With a Perceptron

$$\frac{, L}{, y1}$$

$$L(y, y1) = -y \ln(y1) " (1 " y) \ln 1 " y1$$

$$\frac{, L}{, y1}$$

"Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}}$$

$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1 - y) \ln(1 - \hat{y})$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{-y}{\hat{y}}$$

"Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1 - y)\ln(1 - \hat{y})$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{-y}{\hat{y}}$$

"Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1 - y) \ln(1 - \hat{y})$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{-y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}$$

"Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1 - y)\ln(1 - \hat{y})$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{-y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}$$

$$= \frac{-y + y\hat{y} + \hat{y} - y\hat{y}}{\hat{y}(1 - \hat{y})}$$

"Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1 - y)\ln(1 - \hat{y})$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{-y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}$$

$$= \frac{-y \cancel{+ y\hat{y}} + \hat{y} - y\hat{y}}{\hat{y}(1 - \hat{y})}$$

Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1 - y) \ln(1 - \hat{y})$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{-y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}$$

$$= \frac{-y + \cancel{y\hat{y}} + \hat{y} - \cancel{y\hat{y}}}{\hat{y}(1 - \hat{y})}$$

"Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1 - y)\ln(1 - \hat{y})$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{-y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}$$

$$= \frac{-y + \cancel{y\hat{y}} + \hat{y} - \cancel{y\hat{y}}}{\hat{y}(1 - \hat{y})}$$

$$= \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

"Classification With a Perceptron

"Classification With a Perceptron

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

"Classification With a Perceptron

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

"Classification With a Perceptron

$$y1 = \$(w_1 x_1 + w_2 x_2 + b)$$

$\frac{, y1}{, w_1}$	

"Classification With a Perceptron

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

$\frac{, y1}{, w_1}$	$= y1(1 " y1)x_1$

"Classification With a Perceptron

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

$\frac{, y1}{, w_1}$	$= y1(1 \text{"} y1)x_1$
$\frac{, y1}{, w_2}$	

"Classification With a Perceptron

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

$\frac{, y1}{, w_1}$	$= y(1 \text{"} y)x_1$
$\frac{, y1}{, w_2}$	$= y(1 \text{"} y)x_2$

"Classification With a Perceptron

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

$\frac{, y1}{, w_1}$	$= y(1 " y)x_1$
$\frac{, y1}{, w_2}$	$= y(1 " y)x_2$
$\frac{, y1}{, b}$	

"Classification With a Perceptron

$$y1 = \$(w_1x_1 + w_2x_2 + b)$$

$\frac{, y1}{, w_1}$	$= y(1 " y)x_1$
$\frac{, y1}{, w_2}$	$= y(1 " y)x_2$
$\frac{, y1}{, b}$	$= y(1 " y)$

Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1 - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1 - \hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1 - \hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{\partial L}{\partial \hat{y}} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1 - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1 - \hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1 - \hat{y})x_2$$

$$= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{, y1}{, b} = y(1 \text{ " } y)$$

$$\frac{, y1}{, w_1} = y(1 \text{ " } y)x_1$$

$$\frac{, y1}{, w_2} = y(1 \text{ " } y)x_2$$

$$= L \cdot \frac{, y1}{, b}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$= L \cdot \frac{\partial \hat{y}}{\partial b} \qquad = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{, y1}{, b} = y(1 \text{ " } y1)$$

$$\frac{, y1}{, w_1} = y(1 \text{ " } y1)x_1$$

$$\frac{, y1}{, w_2} = y(1 \text{ " } y1)x_2$$

$$= \frac{, L}{, y1} \cdot \frac{, y1}{, b} = \frac{-(y \text{ " } y1)}{y(1 \text{ " } y1)}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

Classification With a Perceptron

$$\frac{, y1}{, w_1} = y(1 \text{ " } y)x_1$$

$$\frac{, y1}{, w_2} = y(1 \text{ " } y)x_2$$

$$= \frac{, L}{, y1} \cdot \quad L$$

$$= \frac{-(y \text{ " } y1)}{y(1 \text{ " } y)1}$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

Classification With a Perceptron

$$\frac{, y1}{, w_1} = y1(1 \text{ " } y1)x_1$$

$$\frac{, y1}{, w_2} = y1(1 \text{ " } y1)x_2$$

$$= \frac{, L}{, y1} \cdot \quad L$$

$$= \frac{-(y \text{ " } y1)}{y1(1 \text{ " } y1)} y1(1 \text{ " } y1)$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

Classification With a Perceptron

$$\frac{\partial y_1}{\partial b} = y_1(1 - y_1)$$

$$\frac{\partial y_1}{\partial w_1} = y_1(1 - y_1)x_1$$

$$\frac{\partial y_1}{\partial w_2} = y_1(1 - y_1)x_2$$

$$= \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial b} = \frac{-(y - y_1)}{y_1(1 - y_1)} y_1(1 - y_1)$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_2}$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\,\hat{y}(1-\hat{y})$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})$$

$$= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$\frac{, y1}{, b}$ = y(1 " y)1

$\frac{, y1}{, w_1}$ = y(1 " y)1x_1

$\frac{, y1}{, w_2}$ = y(1 " y)1x_2

$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b}$ = $\frac{-(y \text{ " } y)1}{y(1 \text{ " } y)1}$ y(1 " y)1

= $\cdot \frac{, y1}{, w_1}$

$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$

Classification With a Perceptron

$$\frac{,\ y1}{,\ b} = y(1\text{ " } y)1$$

$$\frac{,\ y1}{,\ w_1} = y(1\text{ " } y)1x_1$$

$$\frac{,\ y1}{,\ w_2} = y(1\text{ " } y)1x_2$$

$$\frac{,\ L}{,\ b} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ b} = \frac{-(y\text{ " } y)1}{y(1\text{ " } y)1}\, y(1\text{ " } y)1$$

$$= \quad \cdot \frac{,\ y1}{,\ w_1} = \frac{-(y\text{ " } y)1}{y(1\text{ " } y)1}$$

$$\frac{,\ L}{,\ w_2} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ w_2}$$

Classification With a Perceptron

$$\frac{, y1}{, b} = y1(1 \text{ " } y1)$$

$$\frac{, y1}{, w_1} = y1(1 \text{ " } y1)x_1$$

$$\frac{, y1}{, w_2} = y1(1 \text{ " } y1)x_2$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b} = \frac{-(y \text{ " } y1)}{y1(1 \text{ " } y1)} \, y1(1 \text{ " } y1)$$

$$= \frac{, L}{, y1} \cdot \frac{, y1}{, w_1} = \frac{-(y \text{ " } y1)}{y1(1 \text{ " } y1)}$$

$$\frac{, L}{, w_2} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_2}$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1 - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1 - \hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})} \hat{y}(1 - \hat{y})$$

$$= \frac{\partial L}{\partial \hat{y}} \cdot \quad = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{\partial y1}{\partial b} = y1(1 - y1)$$

$$\frac{\partial y1}{\partial w_2} = y1(1 - y1)x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial y1} \cdot \frac{\partial y1}{\partial b} = \frac{-(y - y1)}{y1(1 - y1)}\, y1(1 - y1)$$

$$= \frac{\partial L}{\partial y1} \cdot \quad = \frac{-(y - y1)}{y1(1 - y1)}\, y1(1 - y1)x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial y1} \cdot \frac{\partial y1}{\partial w_2}$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})$$

$$= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})} \hat{y}(1-\hat{y})$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})} \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_1$$

$$= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{,\ y1}{,\ b} = y(1\ ''\ y)$$

$$\frac{,\ y1}{,\ w_1} = y(1\ ''\ y)x_1$$

$$\frac{,\ y1}{,\ w_2} = y(1\ ''\ y)x_2$$

$$\frac{,\ L}{,\ b} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ b} = \frac{-(y\ ''\ y1)}{y(1\ ''\ y)}\ y(1\ ''\ y)$$

$$\frac{,\ L}{,\ w_1} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ w_1} = \frac{-(y\ ''\ y1)}{y(1\ ''\ y)}\ y(1\ ''\ y)x_1$$

$$= \ L \cdot \frac{,\ y1}{,\ w_2}$$

Classification With a Perceptron

$$\frac{,\ y1}{,\ b} = y(1\ "\ \ y)1$$

$$\frac{,\ y1}{,\ w_1} = y(1\ "\ \ y)1 x_1$$

$$\frac{,\ y1}{,\ w_2} = y(1\ "\ \ y)1 x_2$$

$$\frac{,\ L}{,\ b} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ b} = \frac{-(y\ "\ \ y)1}{y(1\ "\ \ y)1}\, y(1\ "\ \ y)1$$

$$\frac{,\ L}{,\ w_1} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ w_1} = \frac{-(y\ "\ \ y)1}{y(1\ "\ \ y)1}\, y(1\ "\ \ y)1 x_1$$

$$= \ L \cdot \frac{,\ y1}{,\ w_2} = \frac{-(y\ "\ \ y)1}{y(1\ "\ \ y)1}$$

Classification With a Perceptron

$$\frac{,\ y1}{,\ b} = y(1\ ''\ y1)$$

$$\frac{,\ y1}{,\ w_1} = y(1\ ''\ y1)x_1$$

$$\frac{,\ y1}{,\ w_2} = y(1\ ''\ y1)x_2$$

$$\frac{,\ L}{,\ b} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ b} = \frac{-(y\ ''\ y1)}{y(1\ ''\ y1)}\, y(1\ ''\ y1)$$

$$\frac{,\ L}{,\ w_1} = \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ w_1} = \frac{-(y\ ''\ y1)}{y(1\ ''\ y1)}\, y(1\ ''\ y1)x_1$$

$$= \frac{,\ L}{,\ y1} \cdot \frac{,\ y1}{,\ w_2} = \frac{-(y\ ''\ y1)}{y(1\ ''\ y1)}$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1 - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1 - \hat{y})x_1$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})} \hat{y}(1 - \hat{y})$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})} \hat{y}(1 - \hat{y})x_1$$

$$= \frac{\partial L}{\partial \hat{y}} \cdot \qquad = \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

Classification With a Perceptron

$$\frac{, y1}{, b} = y(1 \text{ " } y)$$

$$\frac{, y1}{, w_1} = y(1 \text{ " } y)x_1$$

$$\frac{, L}{, b} = \frac{, L}{, y1} \cdot \frac{, y1}{, b} = \frac{-(y \text{ " } y)}{y(1 \text{ " } y)} y(1 \text{ " } y)$$

$$\frac{, L}{, w_1} = \frac{, L}{, y1} \cdot \frac{, y1}{, w_1} = \frac{-(y \text{ " } y)}{y(1 \text{ " } y)} y(1 \text{ " } y)x_1$$

$$= \frac{, L}{, y1} \cdot \quad L \quad = \frac{-(y \text{ " } y)}{y(1 \text{ " } y)} y(1 \text{ " } y)x_2$$

Classification With a Perceptron

$$\frac{\partial y_1}{\partial b} = y_1(1 - y_1)$$

$$\frac{\partial y_1}{\partial w_1} = y_1(1 - y_1)x_1$$

$$\frac{\partial y_1}{\partial w_2} = y_1(1 - y_1)x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial b} = \frac{-(y - y_1)}{y_1(1 - y_1)} y_1(1 - y_1)$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_1} = \frac{-(y - y_1)}{y_1(1 - y_1)} y_1(1 - y_1)x_1$$

$$= \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_2} = \frac{-(y - y_1)}{y_1(1 - y_1)} y_1(1 - y_1)x_2$$

Classification With a Perceptron

$$\frac{\partial \hat{y}}{\partial b} = \hat{y}(1-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_2$$

Classification With a Perceptron

$$\frac{,\, L}{,\, b} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, b} = \frac{-(y \;\text{"}\; y)1}{y(1 \;\text{"}\; y)1}\, y(1 \;\text{"}\; y)1$$

$$\frac{,\, L}{,\, w_1} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_1} = \frac{-(y \;\text{"}\; y)1}{y(1 \;\text{"}\; y)1}\, y(1 \;\text{"}\; y)1 x_1$$

$$\frac{,\, L}{,\, w_2} = \frac{,\, L}{,\, y1} \cdot \frac{,\, y1}{,\, w_2} = \frac{-(y \;\text{"}\; y)1}{y(1 \;\text{"}\; y)1}\, y(1 \;\text{"}\; y)1 x_2$$

Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{-(y-\hat{y})}{\cancel{\hat{y}(1-\hat{y})}} \cancel{\hat{y}(1-\hat{y})}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y-\hat{y})}{\cancel{\hat{y}(1-\hat{y})}} \cancel{\hat{y}(1-\hat{y})}x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = \frac{-(y-\hat{y})}{\cancel{\hat{y}(1-\hat{y})}} \cancel{\hat{y}(1-\hat{y})}x_2$$

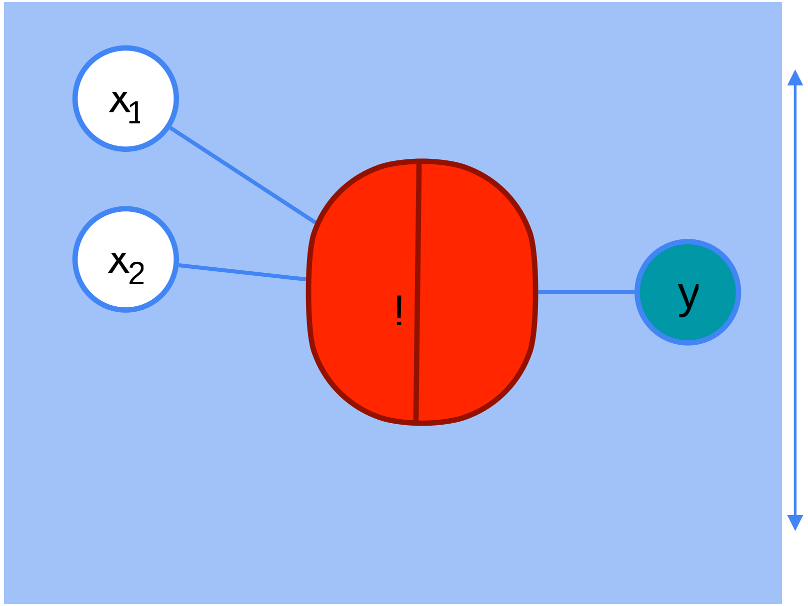
Classification With a Perceptron

$$\frac{\partial L}{\partial b} = -(y - \hat{y})$$

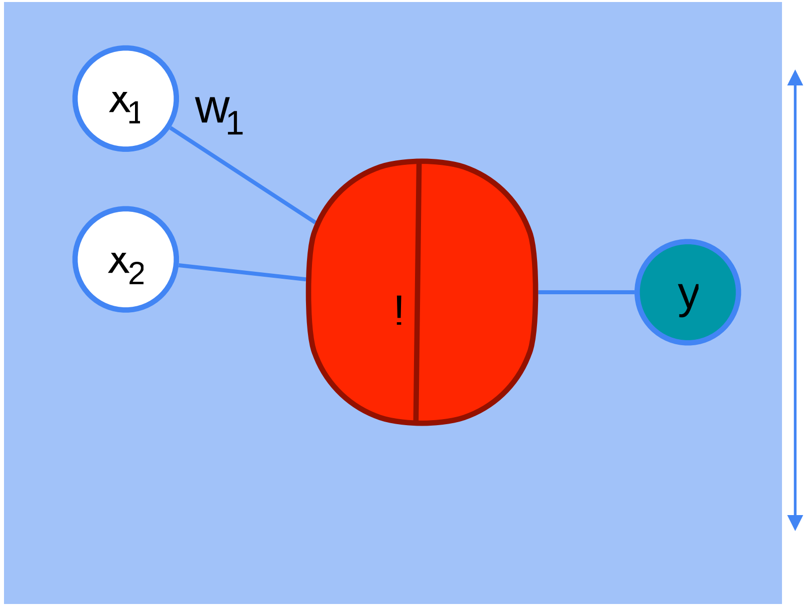
$$\frac{\partial L}{\partial w_1} = -(y - \hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = -(y - \hat{y})x_2$$

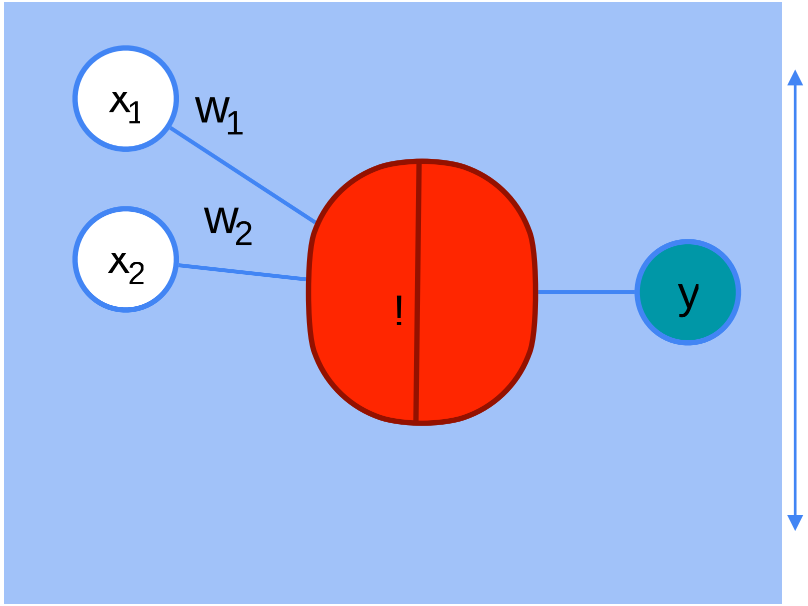
"Classification With a Perceptron



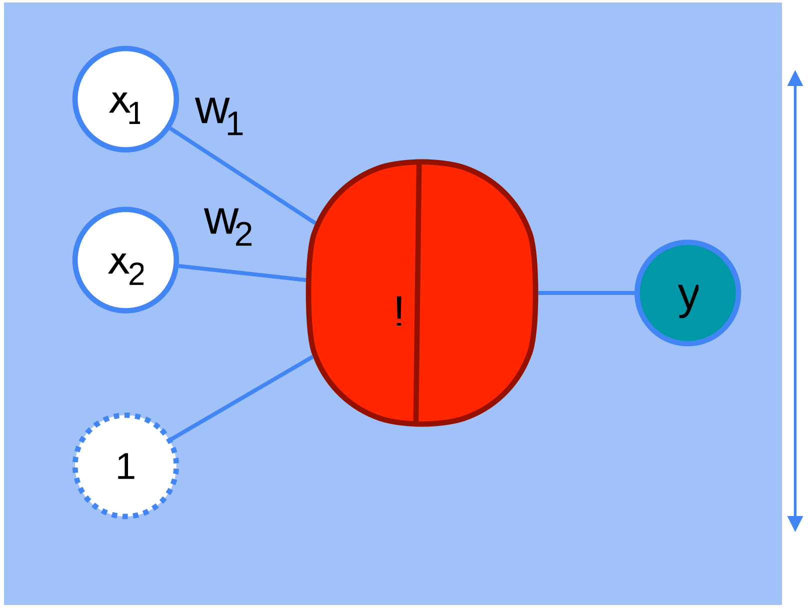
"Classification With a Perceptron



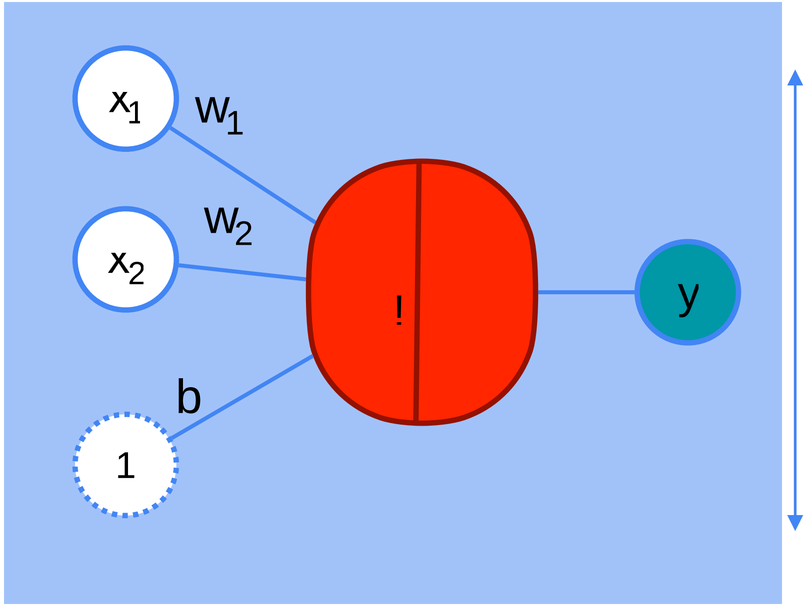
"Classification With a Perceptron



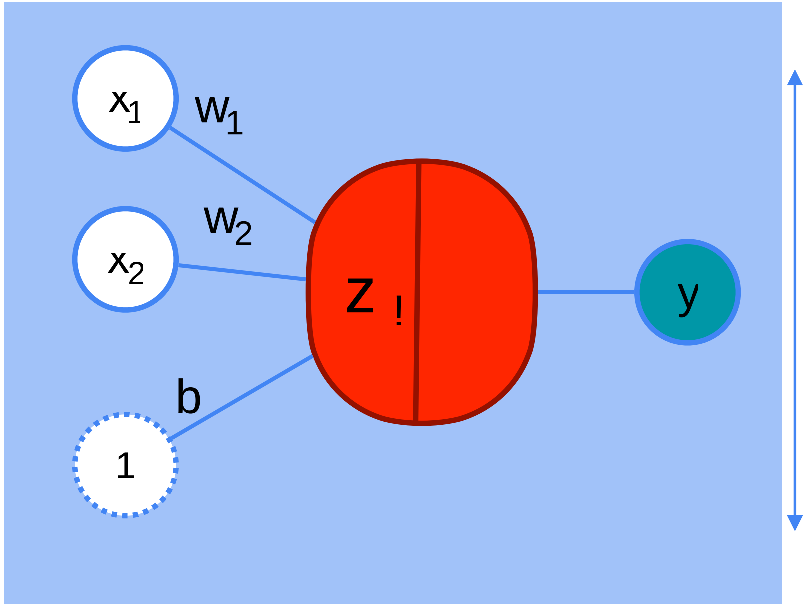
"Classification With a Perceptron



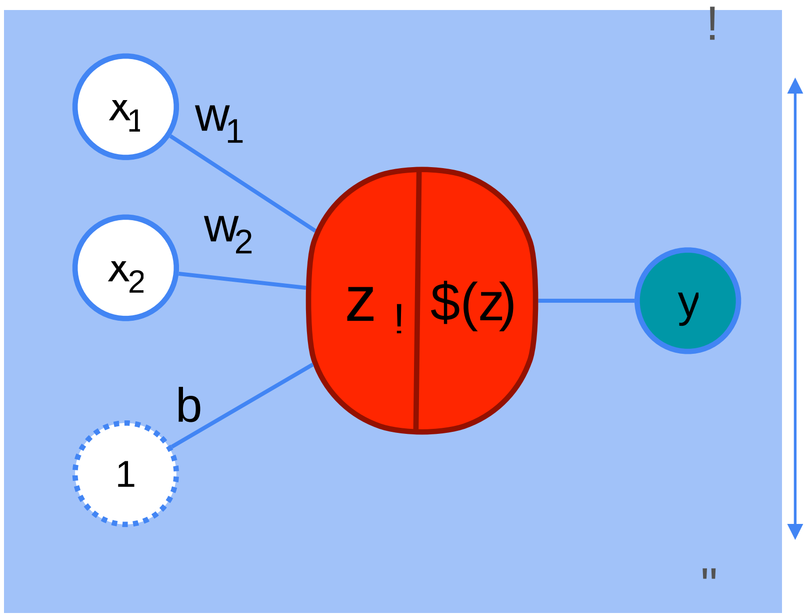
"Classification With a Perceptron



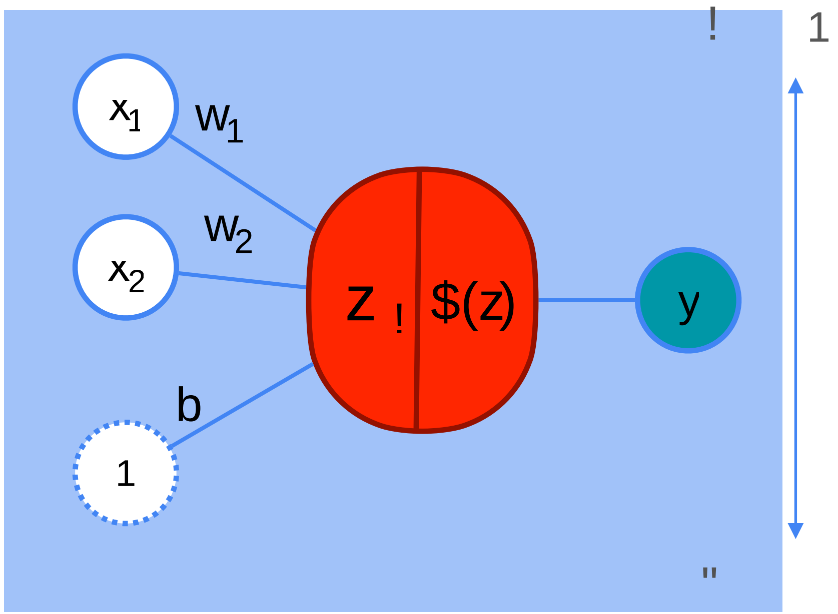
"Classification With a Perceptron



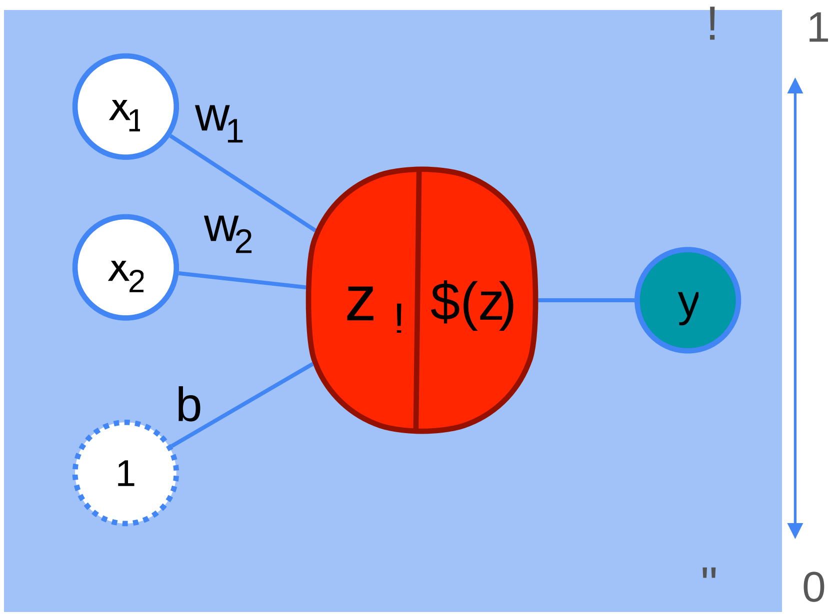
"Classification With a Perceptron



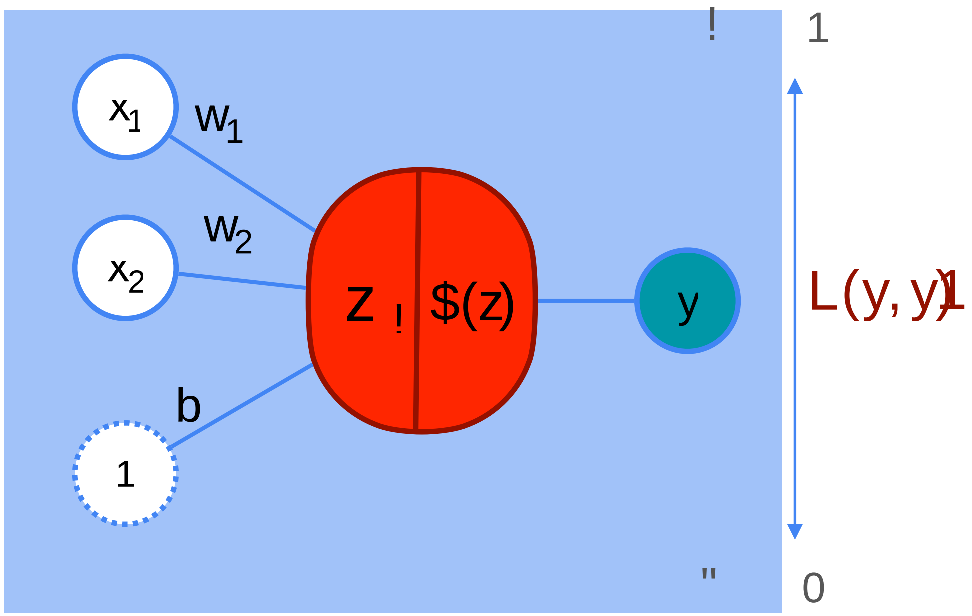
"Classification With a Perceptron



"Classification With a Perceptron

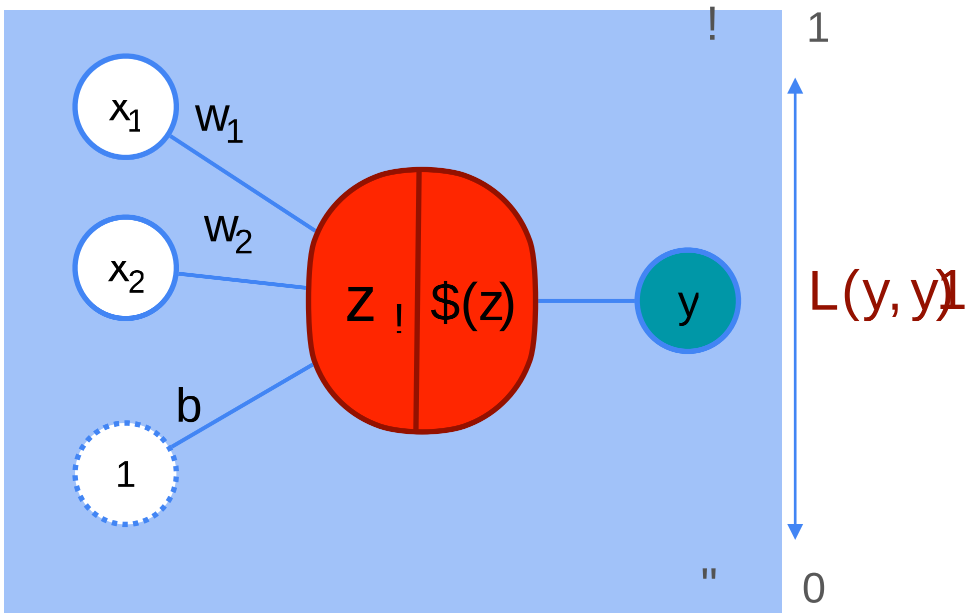


"Classification With a Perceptron

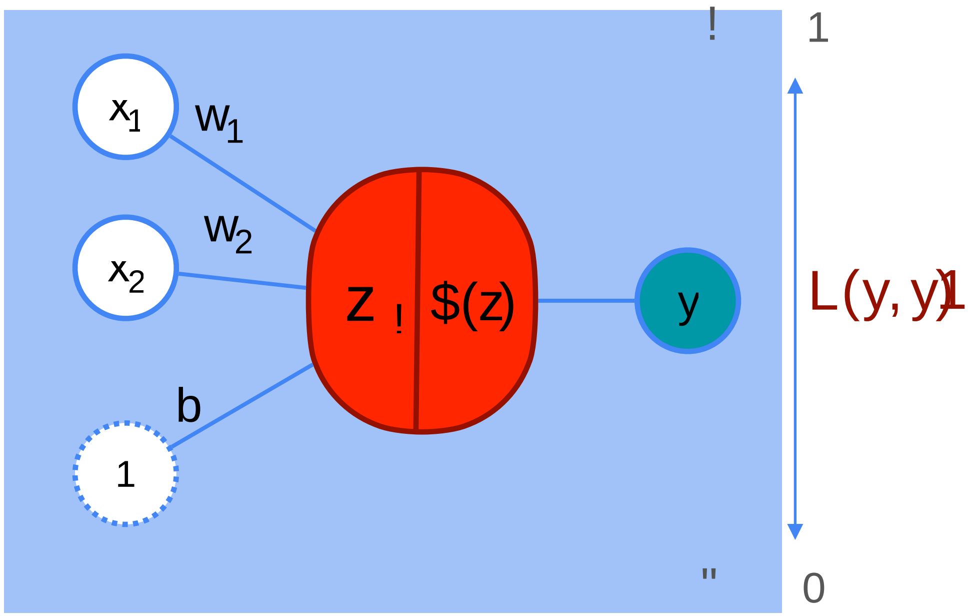


"Classification With a Perceptron

To find optimal values for:



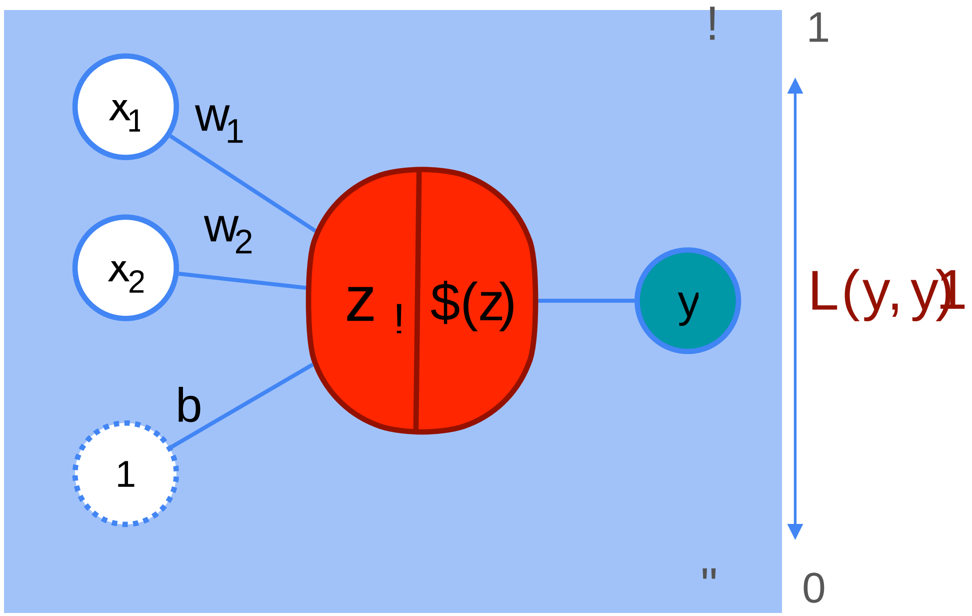
"Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

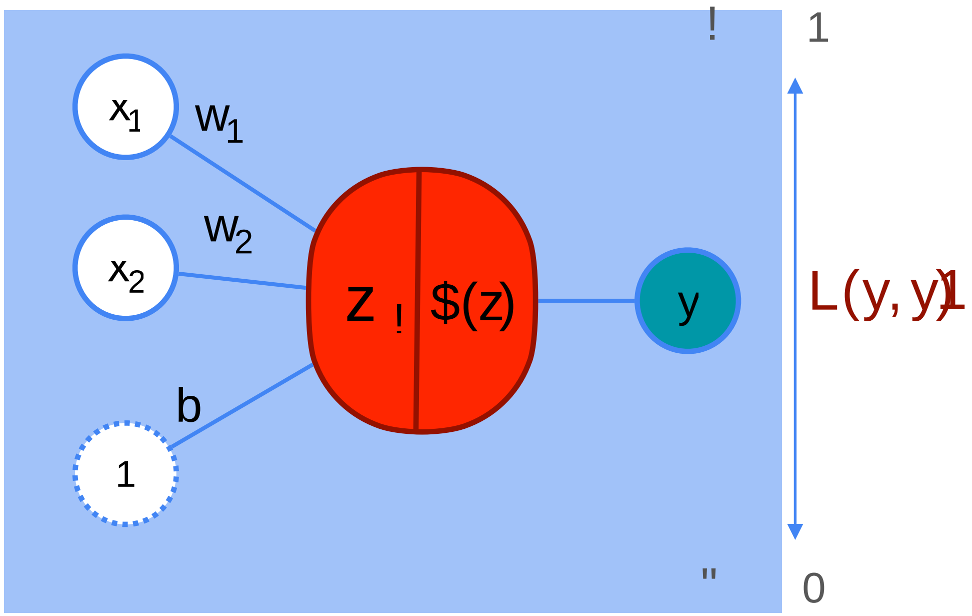
"Classification With a Perceptron



To find optimal values for:
w_1 , w_2 , b

You need gradient descent

"Classification With a Perceptron



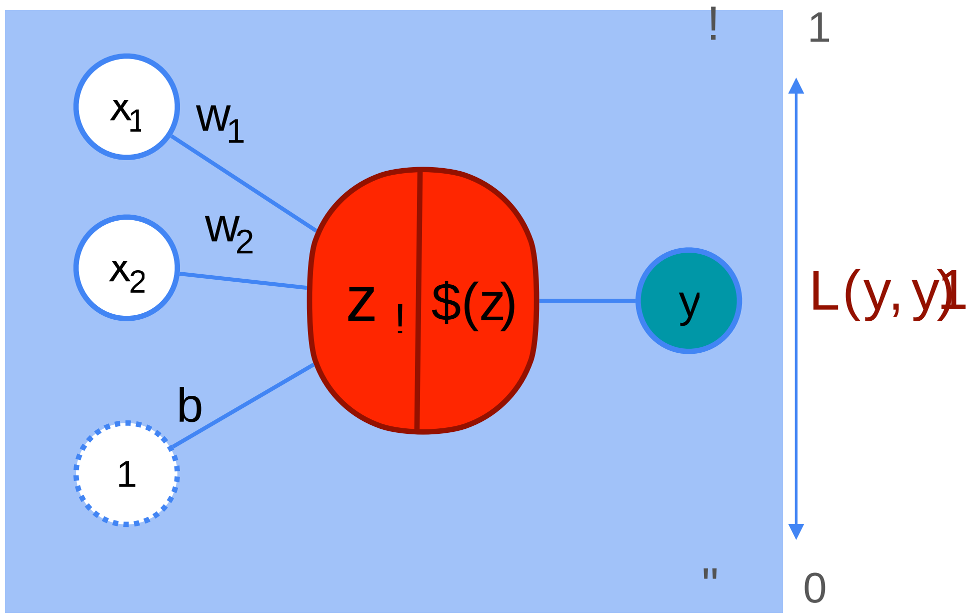
To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$$w_1 \; \$ \; w_1 \, " \, \# \frac{, L}{, w_1}$$

"Classification With a Perceptron



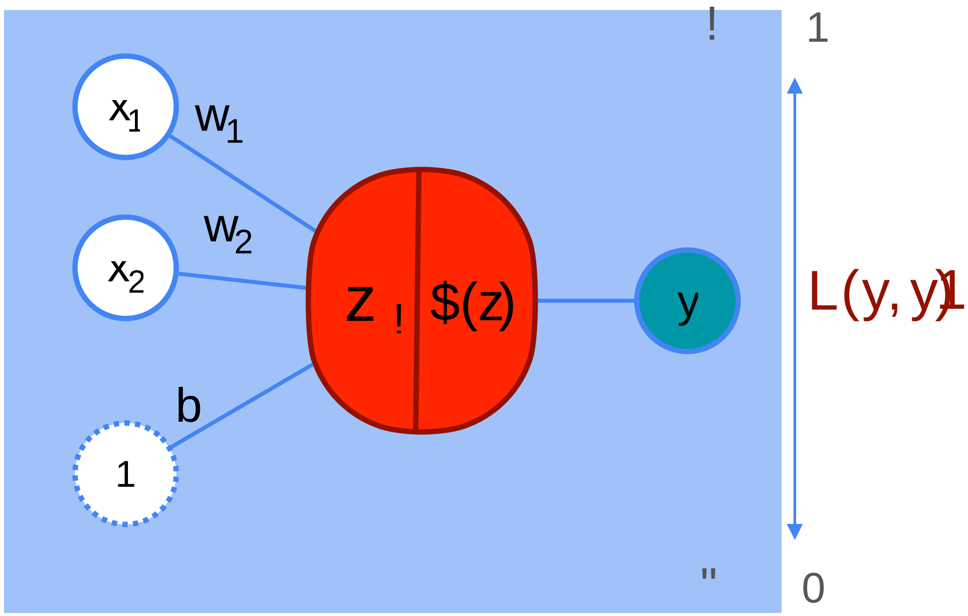
To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$w_1 \ \$ \ w_1 '' \ \#$

"Classification With a Perceptron

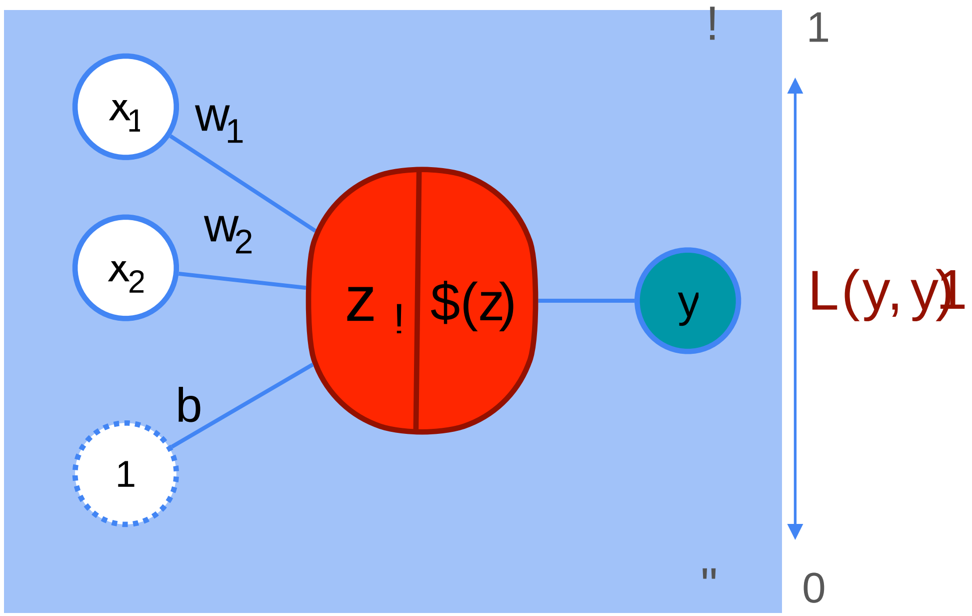


To find optimal values for:
w_1 , w_2 , b

You need gradient descent

w_1 $ w_1 " #(-x_1(y " y))

"Classification With a Perceptron



To find optimal values for:

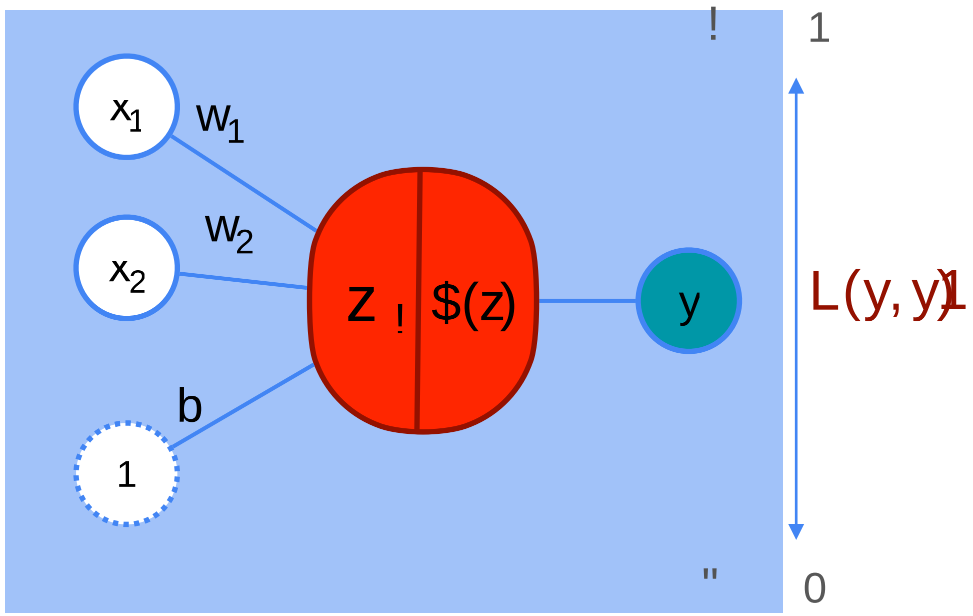
w_1 , w_2 , b

You need gradient descent

$$w_1 \ \$ \ w_1 \ " \ \#(-x_1(y \ " \ y))$$

$$w_2 \ \$ \ w_2 \ " \ \# \frac{,L}{,w_2}$$

"Classification With a Perceptron



To find optimal values for:

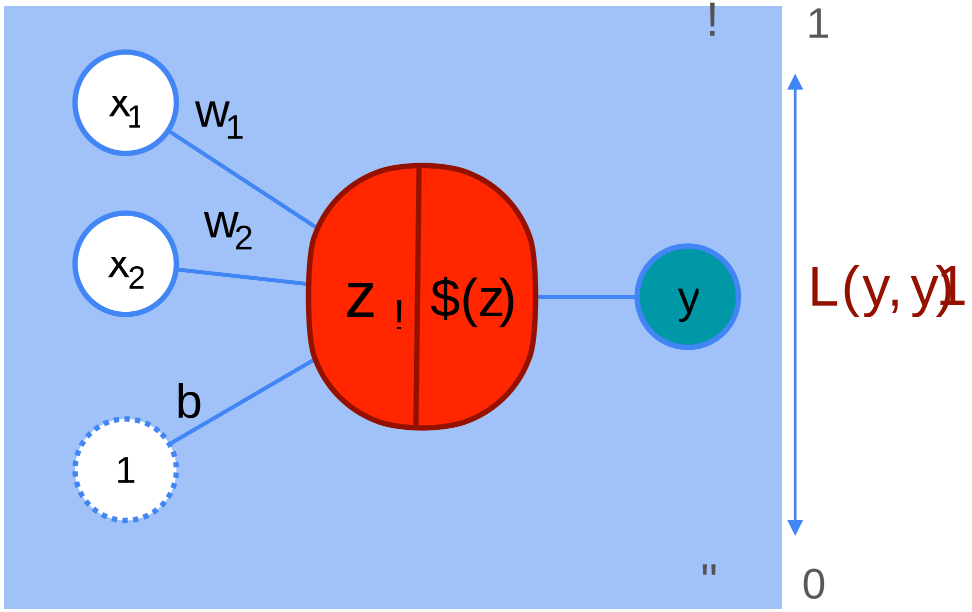
w_1 , w_2 , b

You need gradient descent

$$w_1 \; \$ \; w_1 \, " \; \#(-x_1(y \, " \; y))$$

$$w_2 \; \$ \; w_2 \, " \; \#$$

Classification With a Perceptron



To find optimal values for:

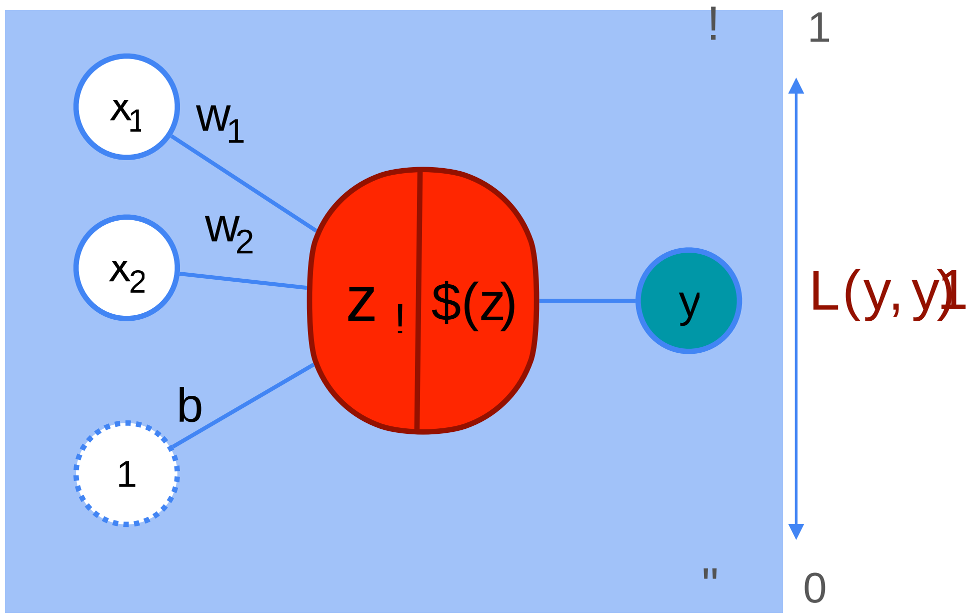
w_1 , w_2 , b

You need gradient descent

$$w_1 \leftarrow w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 \leftarrow w_2 - \alpha(-x_2(y - \hat{y}))$$

"Classification With a Perceptron



To find optimal values for:
w_1 , w_2 , b

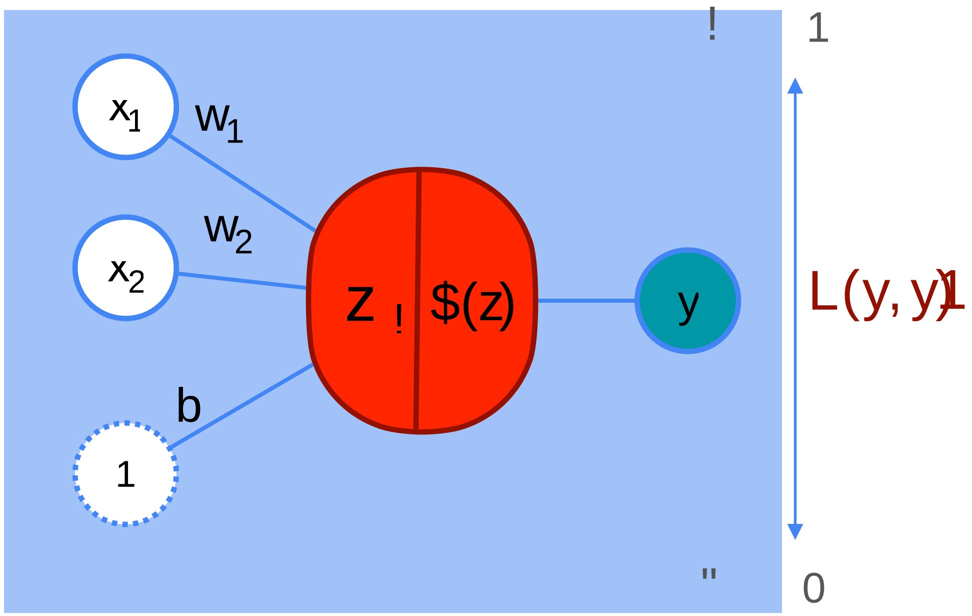
You need gradient descent

w_1 \$ w_1 " #(-x_1(y " y))

w_2 \$ w_2 " #(-x_2(y " y))

b \$ b " # ,L / ,b

"Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

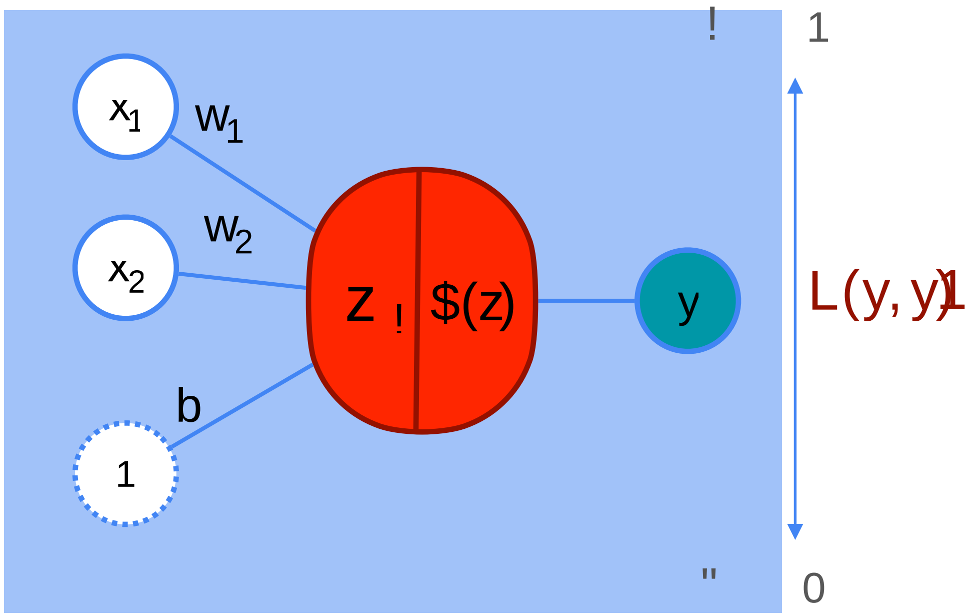
You need gradient descent

$$w_1 \; \$ \; w_1 \, " \, \#(-x_1(y \, " \, y))$$

$$w_2 \; \$ \; w_2 \, " \, \#(-x_2(y \, " \, y))$$

$$b \; \$ \; b \, " \, \#$$

"Classification With a Perceptron



To find optimal values for:
w_1 , w_2 , b

You need gradient descent

w_1 \$ w_1 " #(-x_1(y " y1))

w_2 \$ w_2 " #(-x_2(y " y1))

b \$ b " #(-(y " y1))

Optimization in Neural Networks and NewtonÕs Method

Classification with a Neural Network

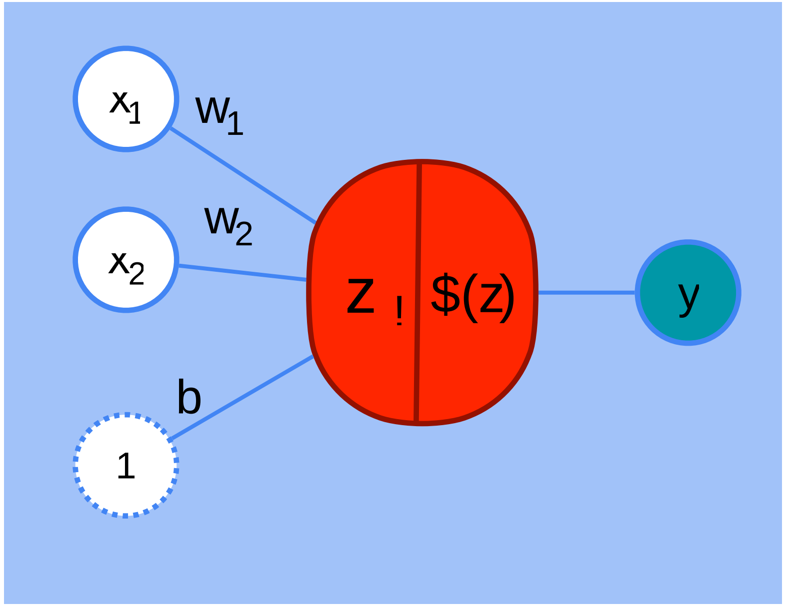
Classification Problem Motivation

Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0	2	Sad "
Aack beep beep beep!	1	3	Sad "
Aack beep aack!	2	1	Happy !

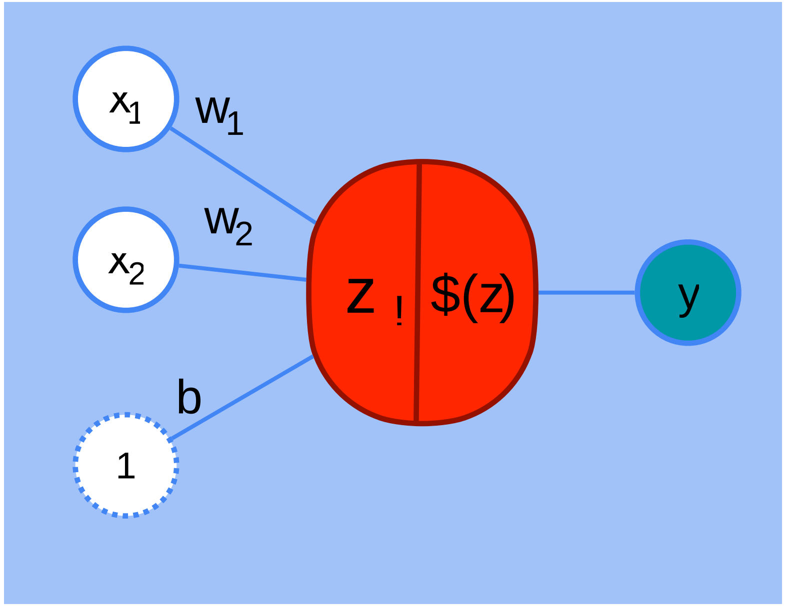
Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0	2	Sad "
Aack beep beep beep!	1	3	Sad "
Aack beep aack!	2	1	Happy !



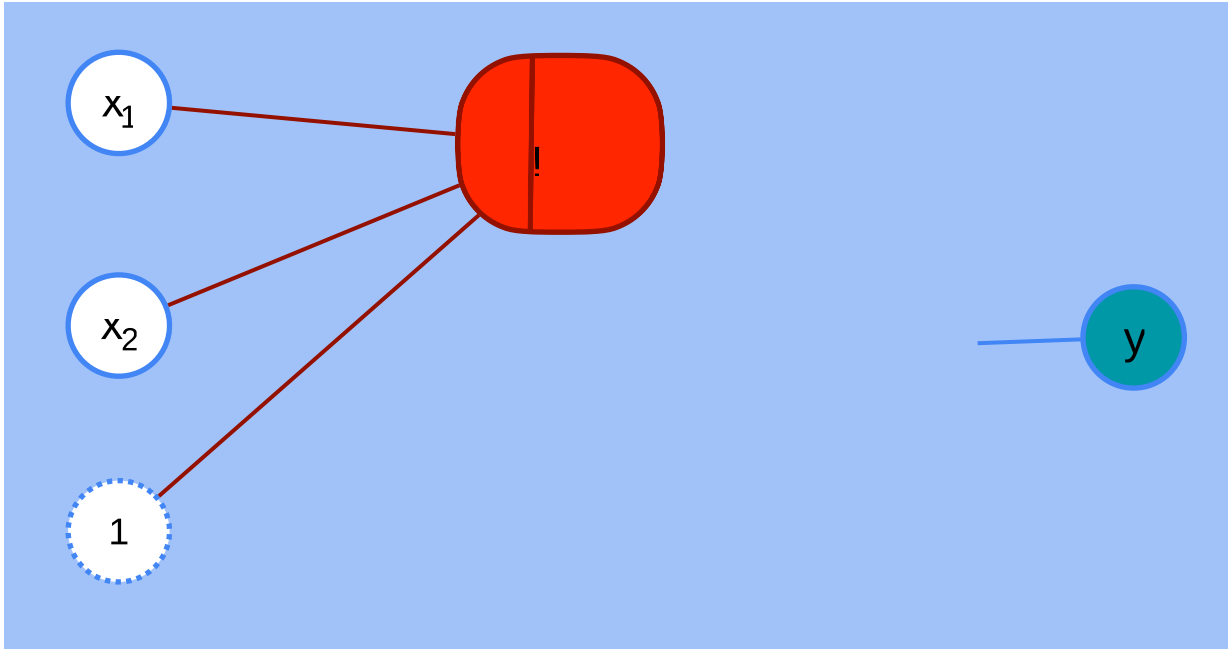
Classification Problem Motivation

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy !
Beep beep!	0	2	Sad "
Aack beep beep beep!	1	3	Sad "
Aack beep aack!	2	1	Happy !

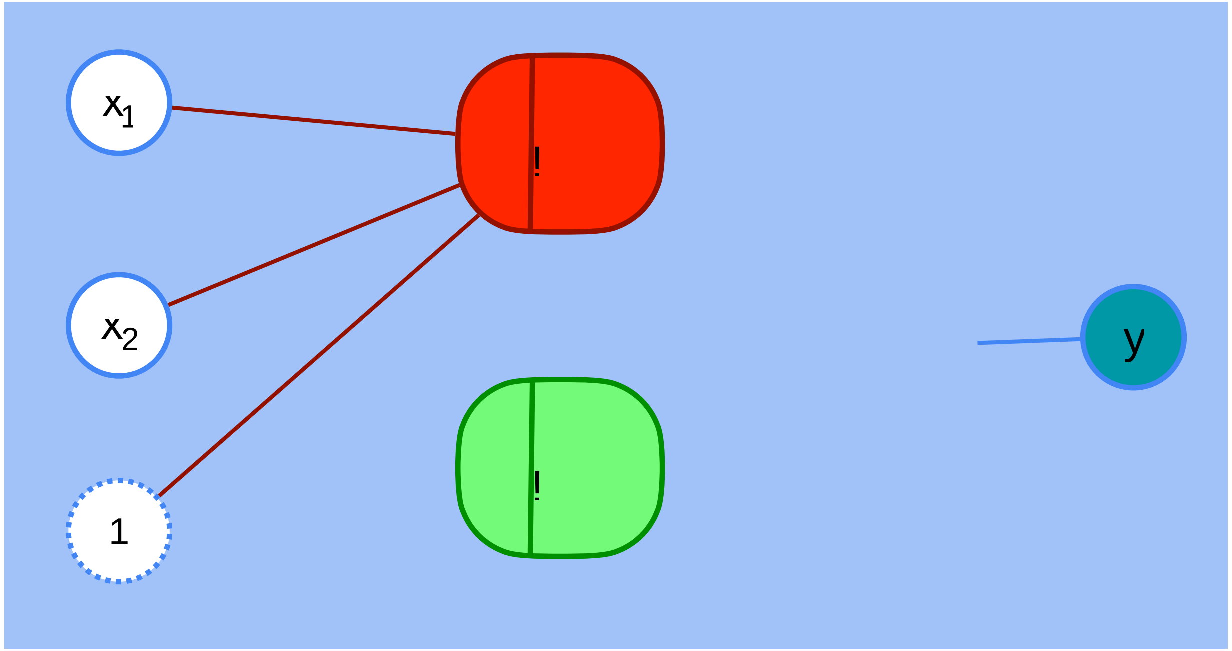


$$z = x_1 w_1 + x_2 w_2 + b$$

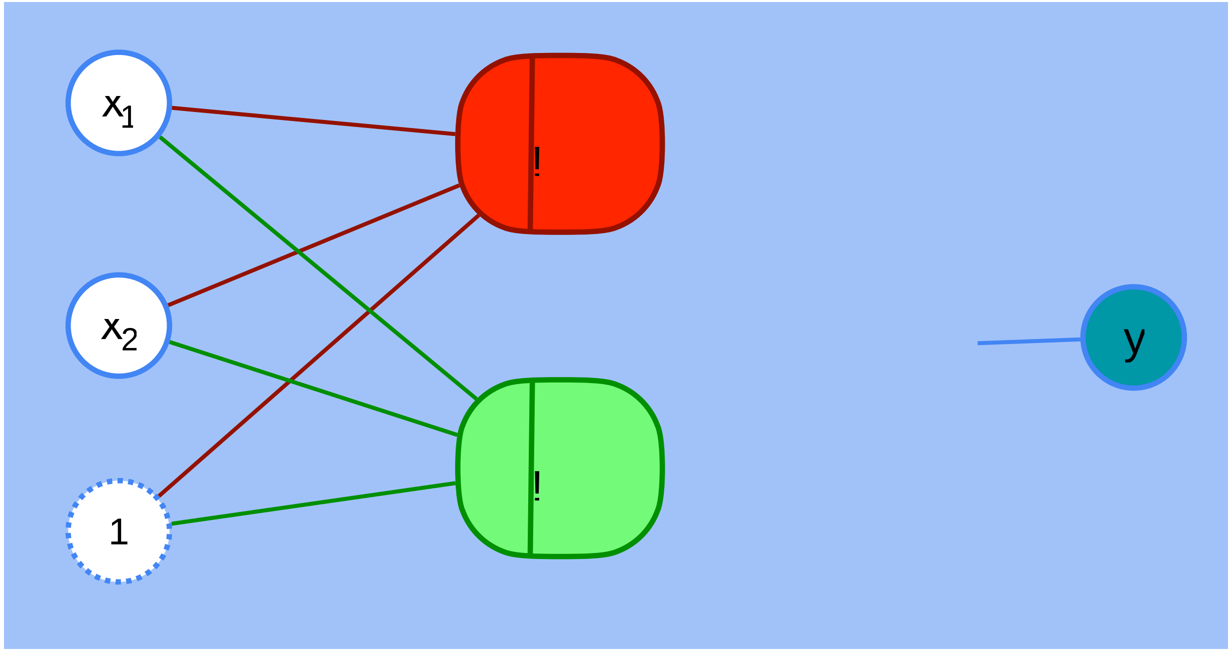
2,2,1 Neural Network



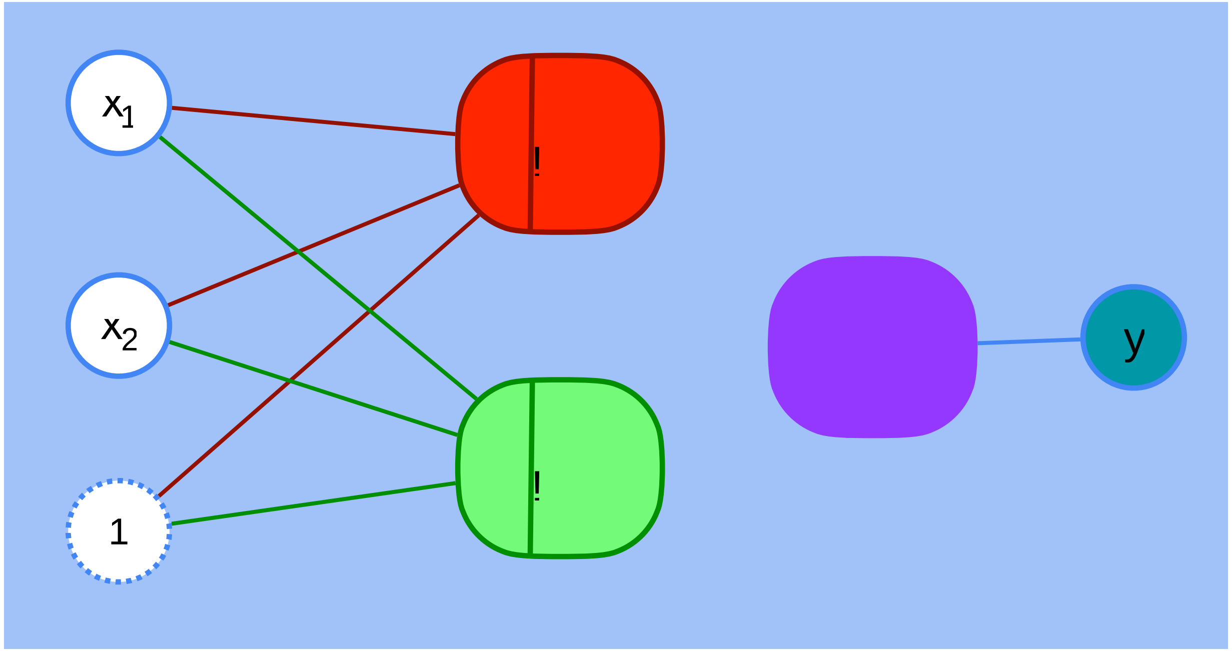
2,2,1 Neural Network



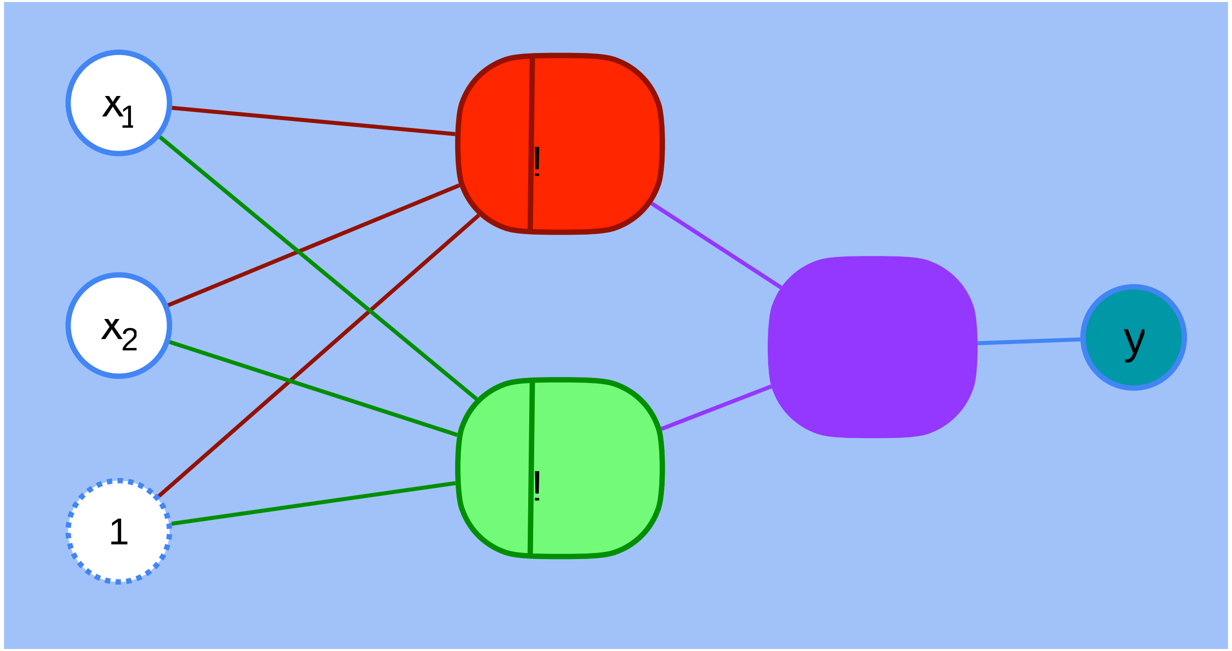
2,2,1 Neural Network



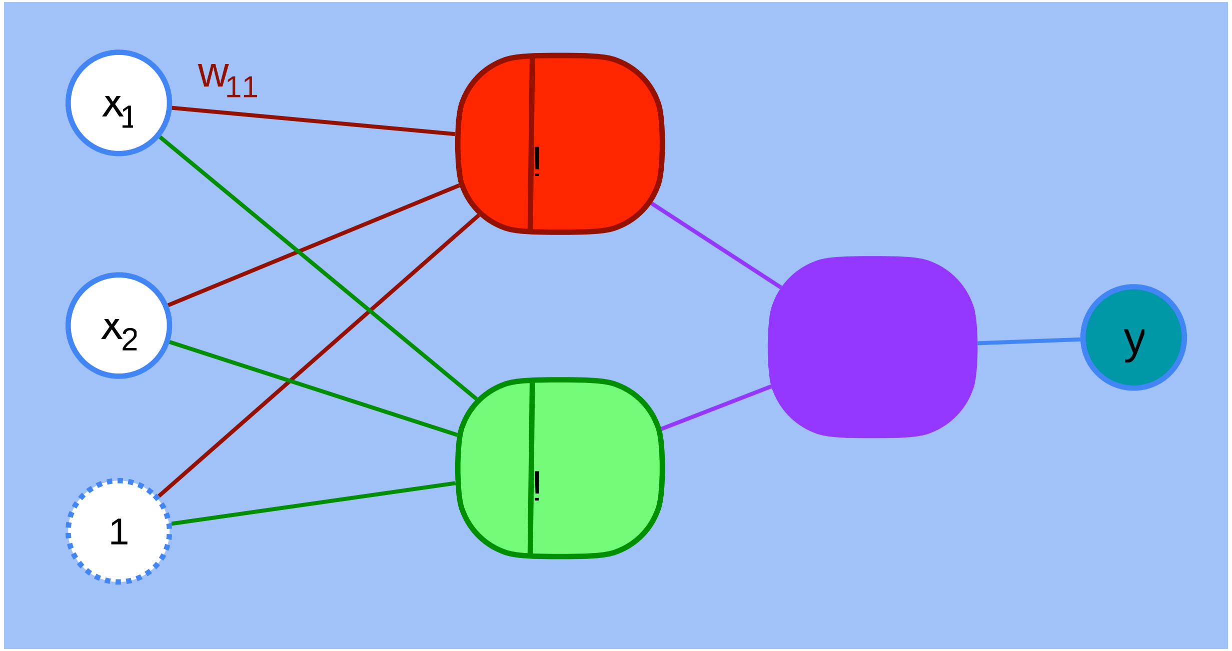
2,2,1 Neural Network



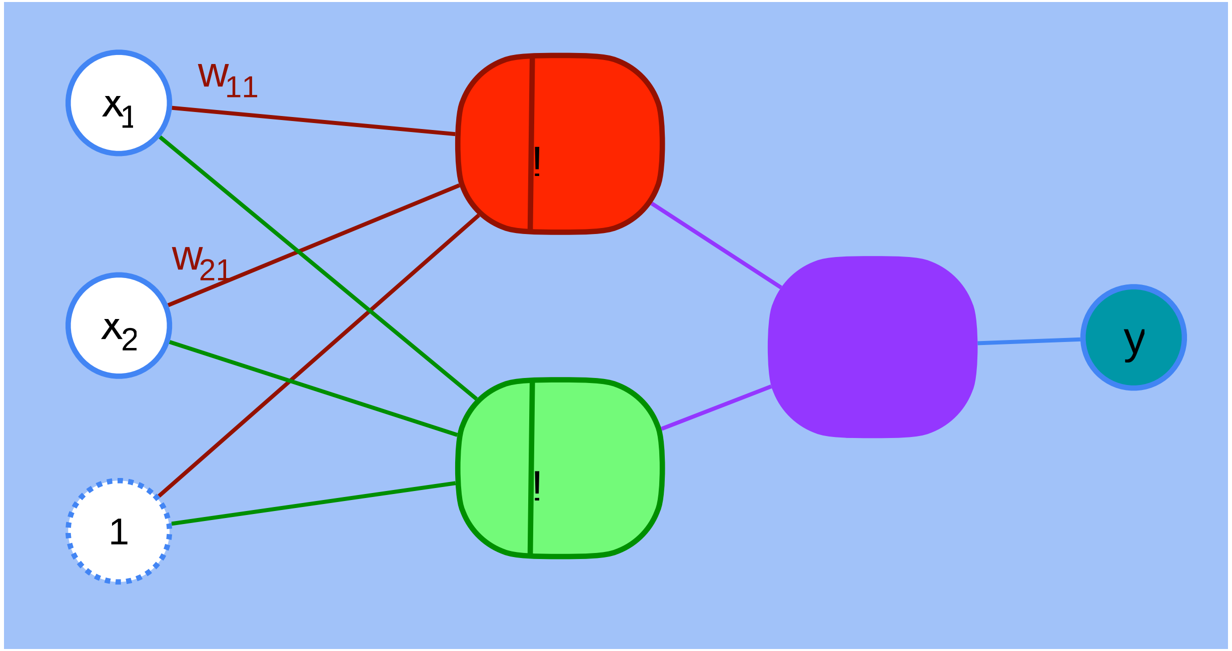
2,2,1 Neural Network



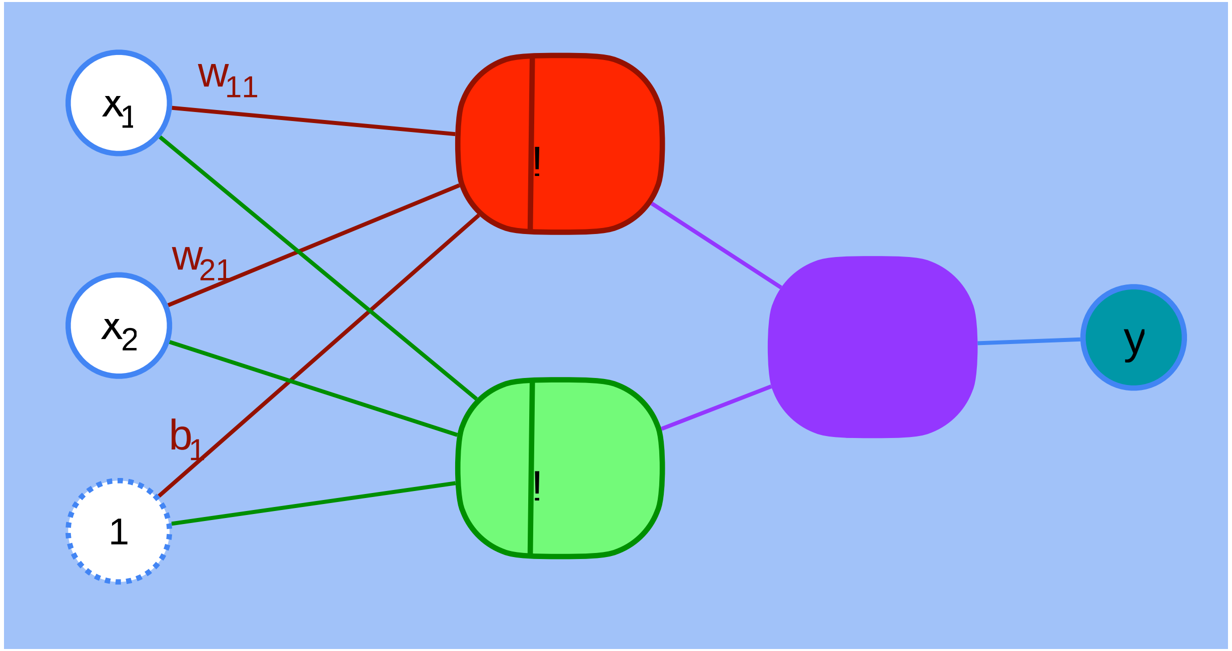
2,2,1 Neural Network



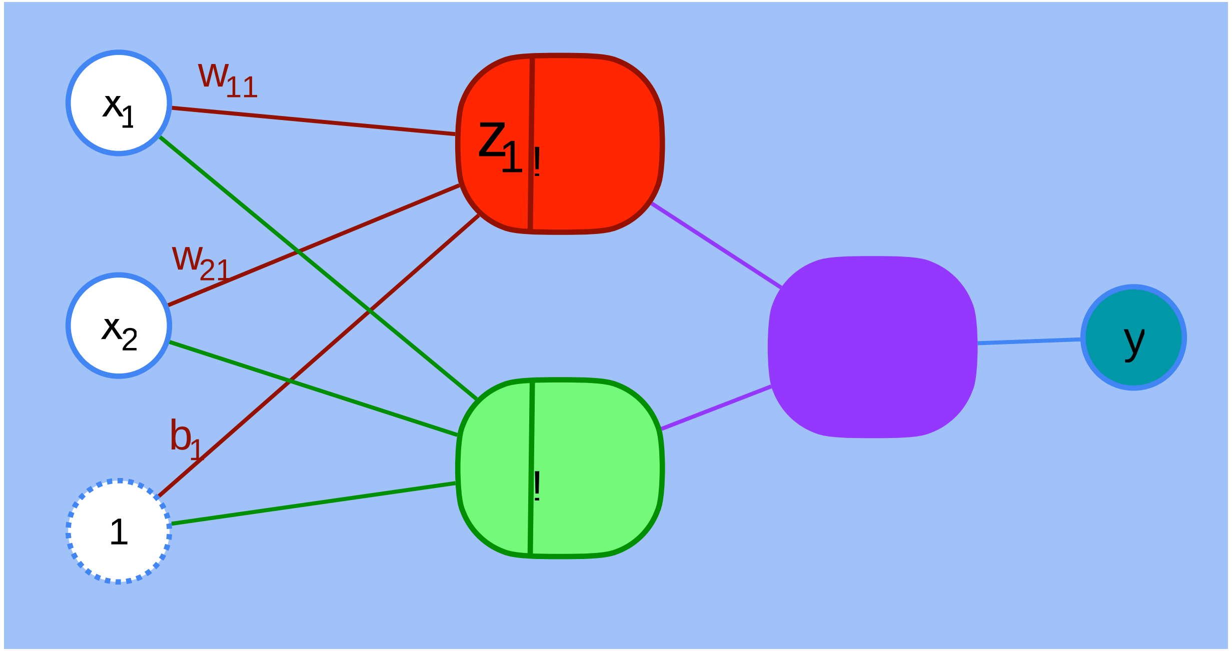
2,2,1 Neural Network



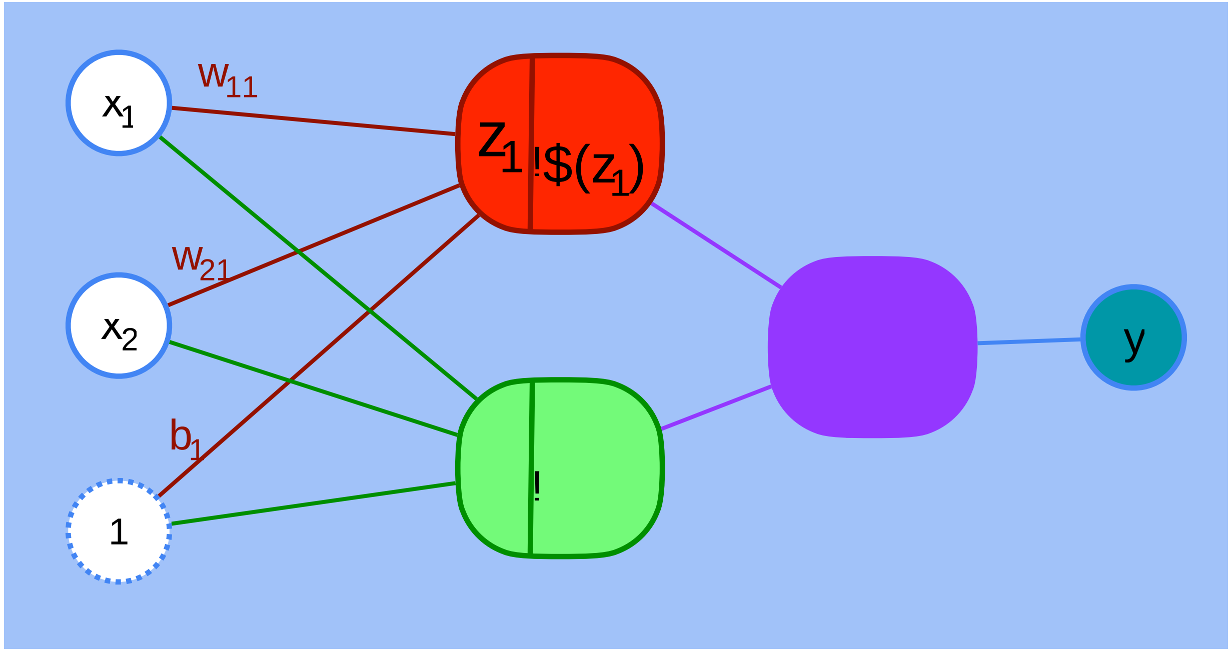
2,2,1 Neural Network



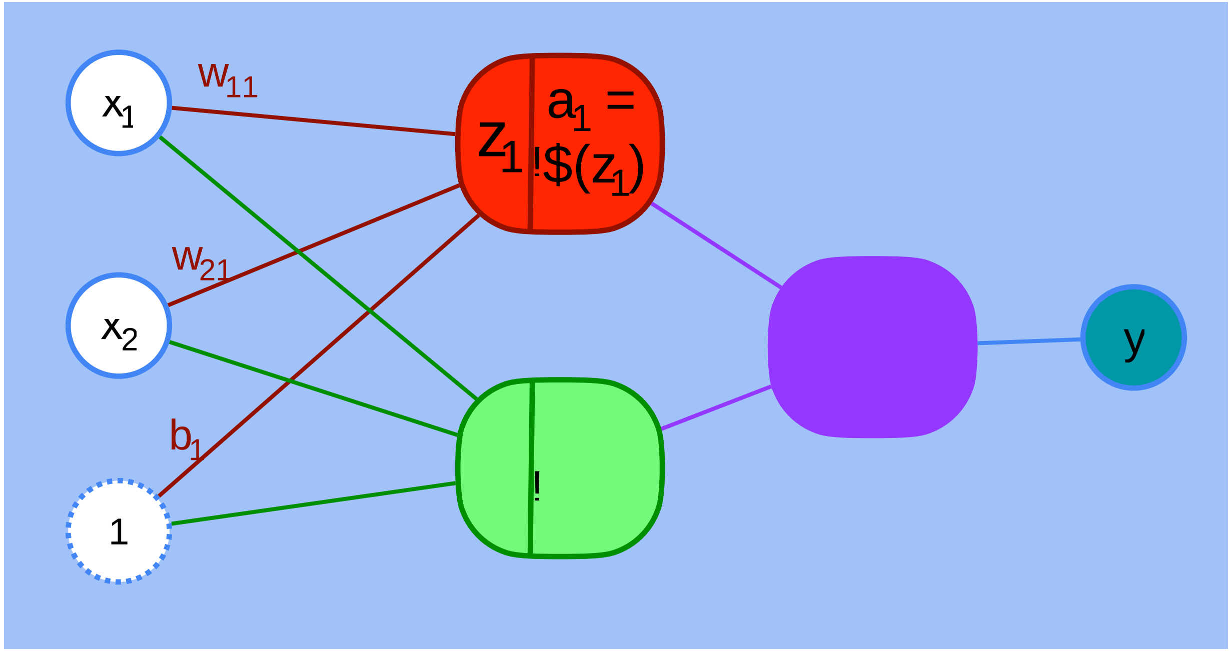
2,2,1 Neural Network



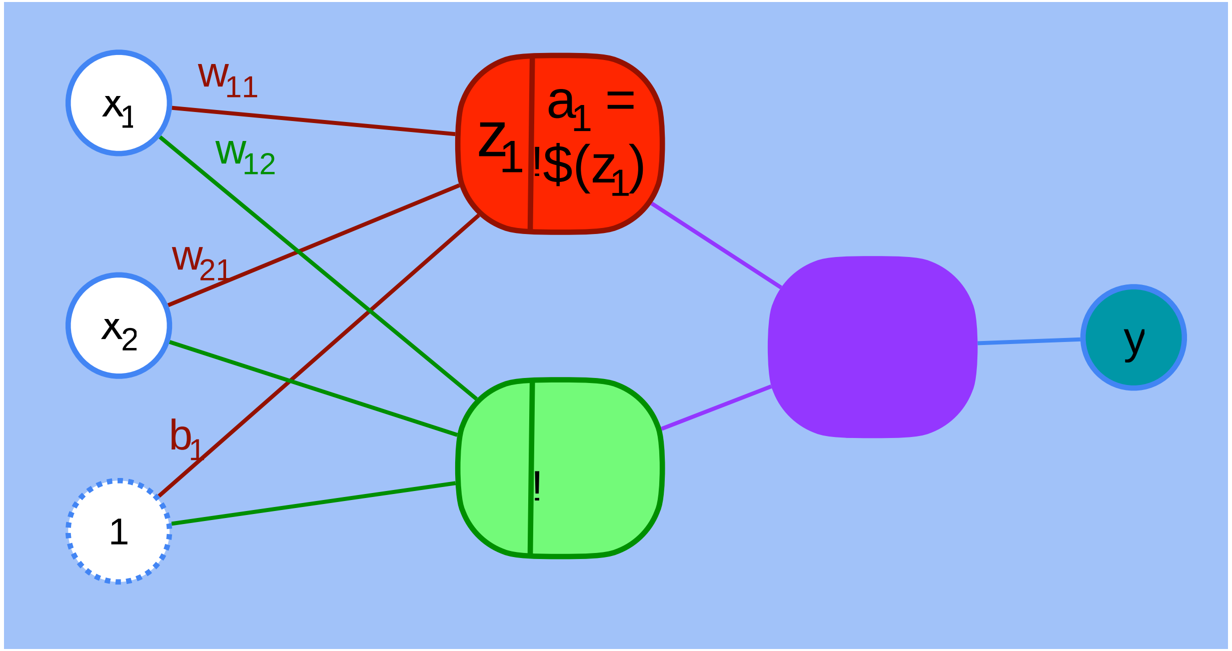
2,2,1 Neural Network



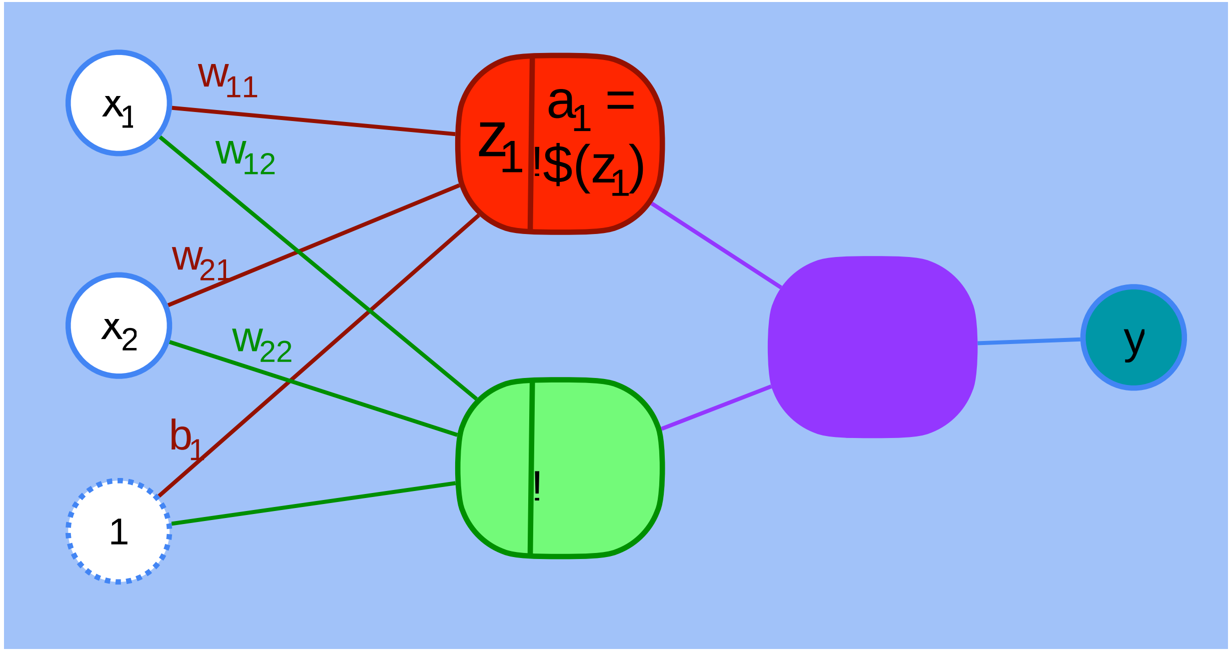
2,2,1 Neural Network



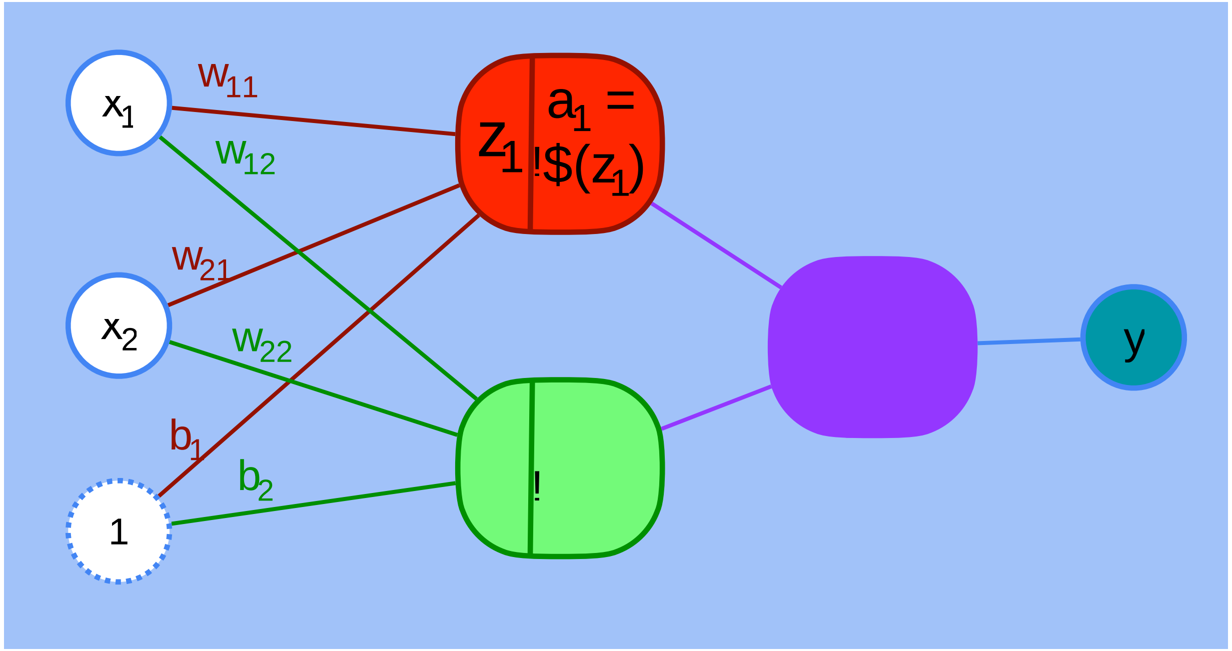
2,2,1 Neural Network



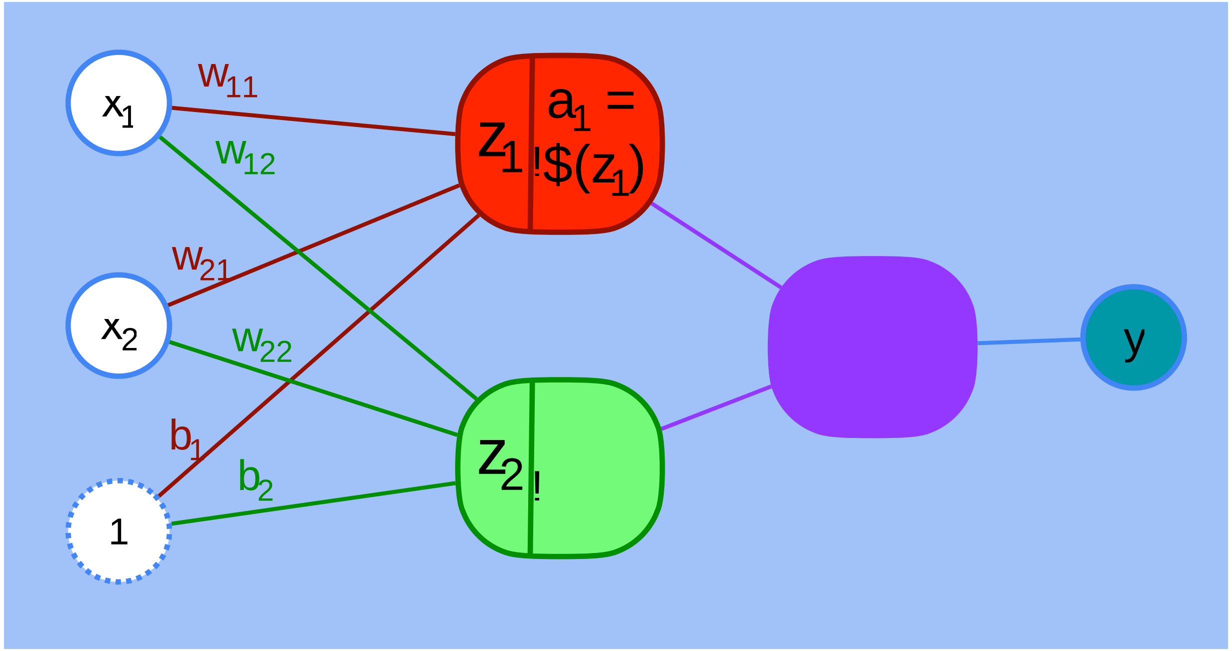
2,2,1 Neural Network



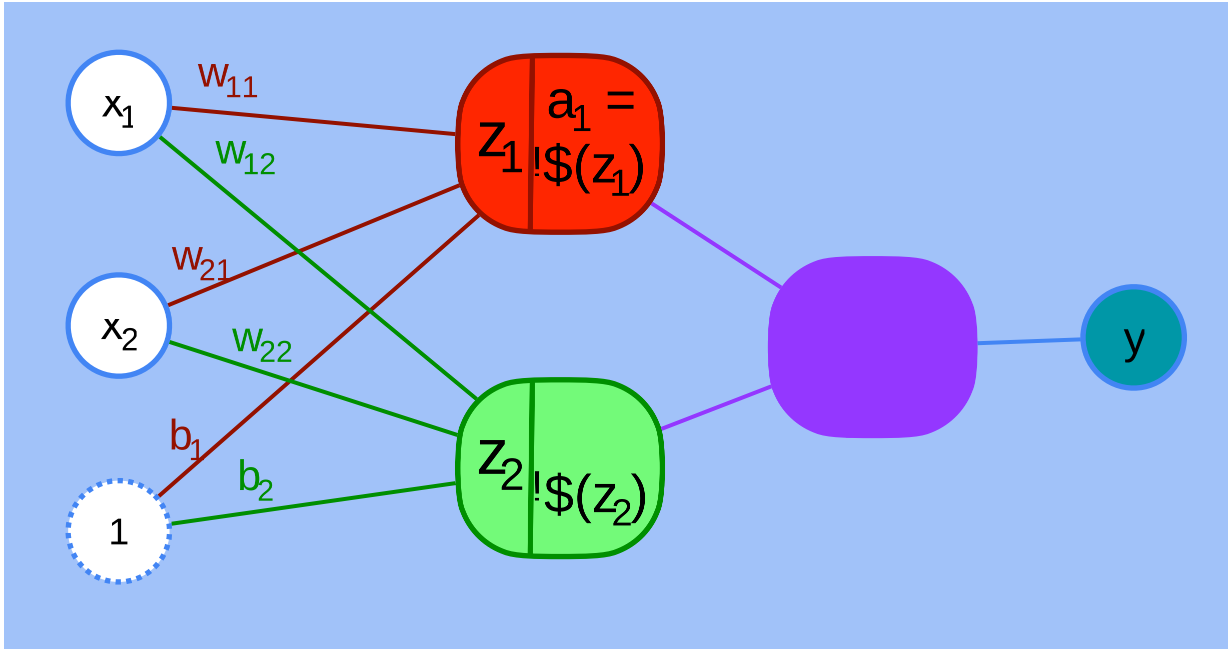
2,2,1 Neural Network



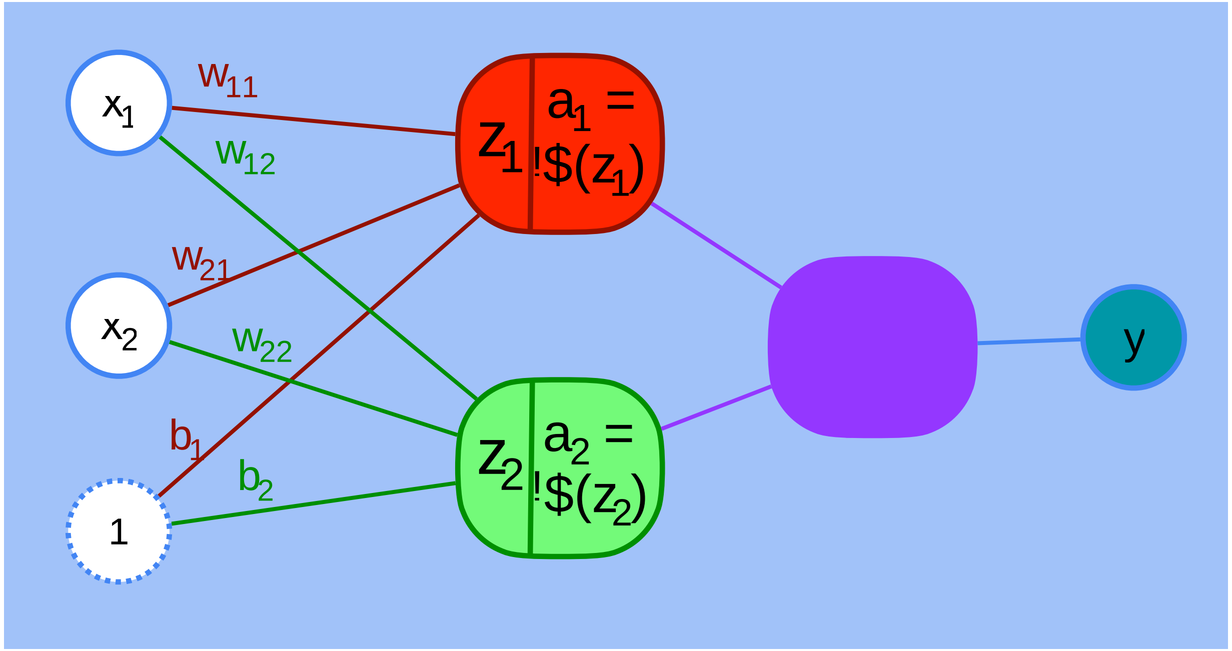
2,2,1 Neural Network



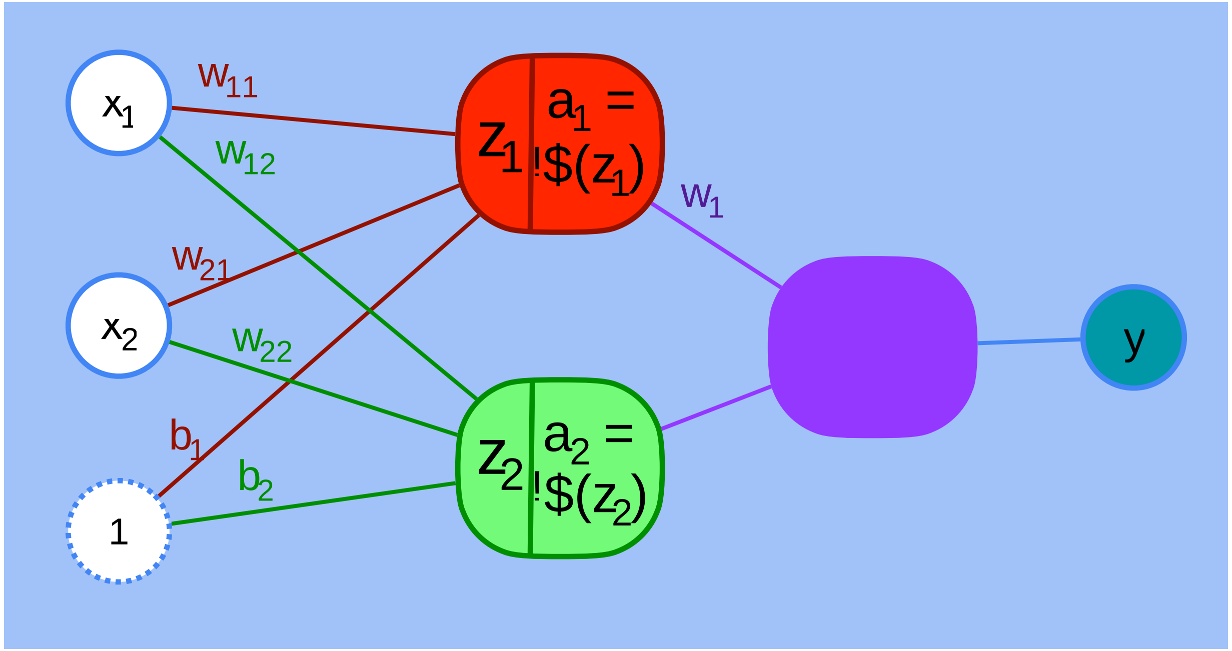
2,2,1 Neural Network



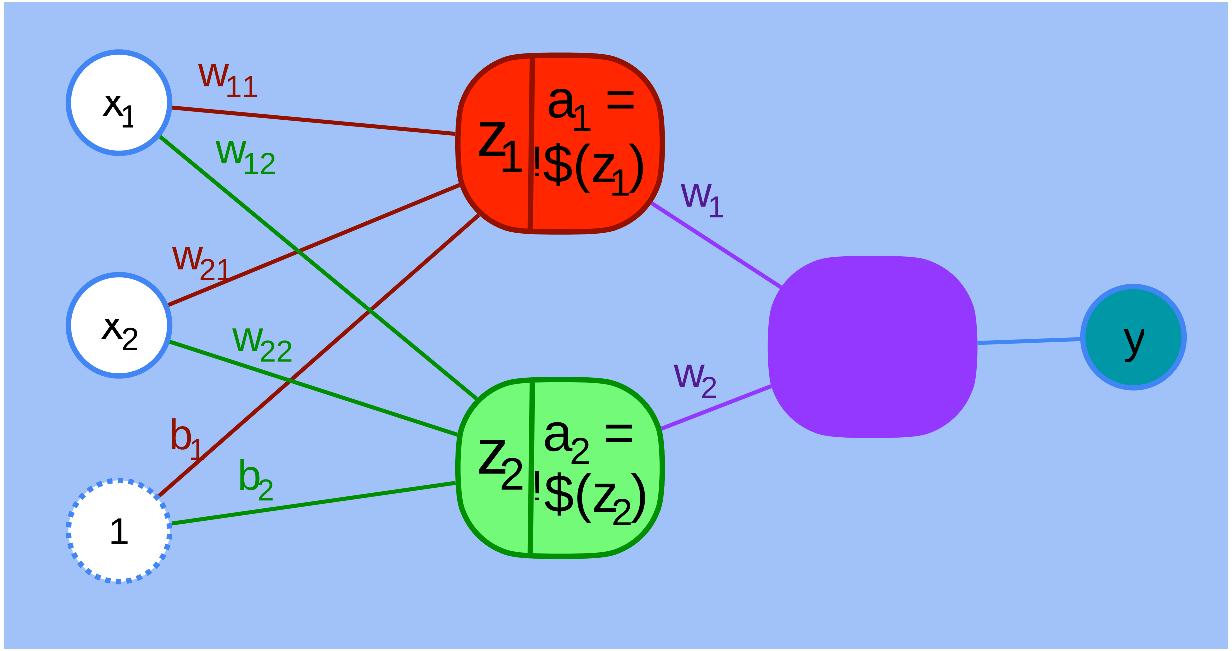
2,2,1 Neural Network



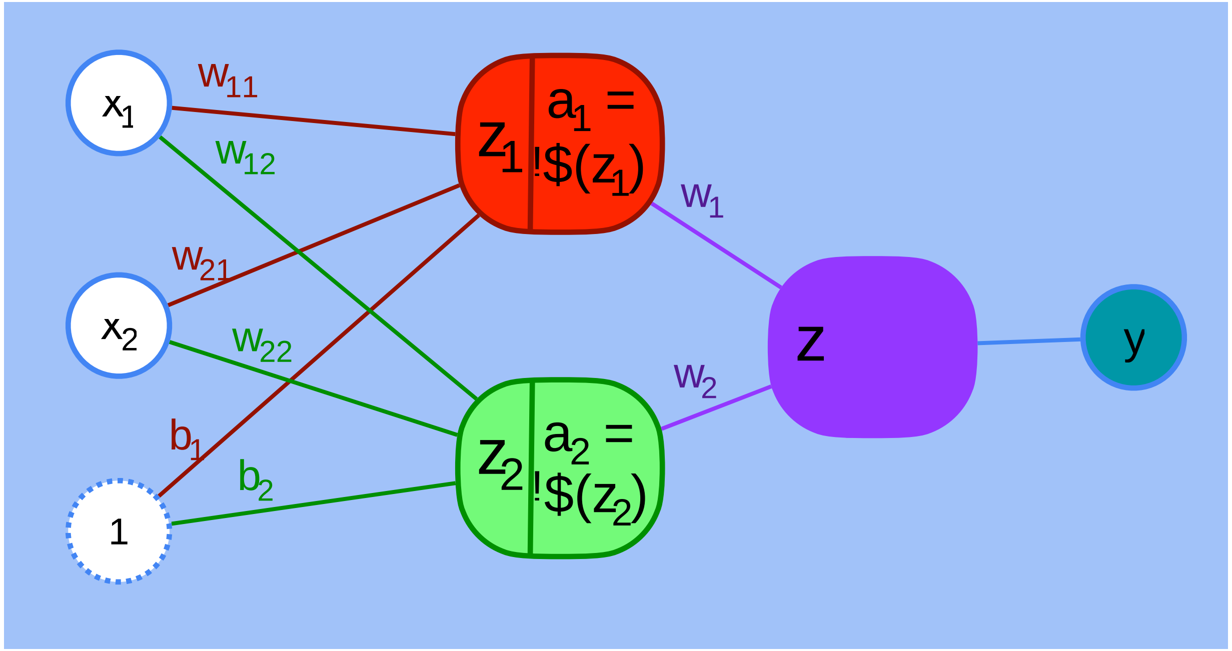
2,2,1 Neural Network



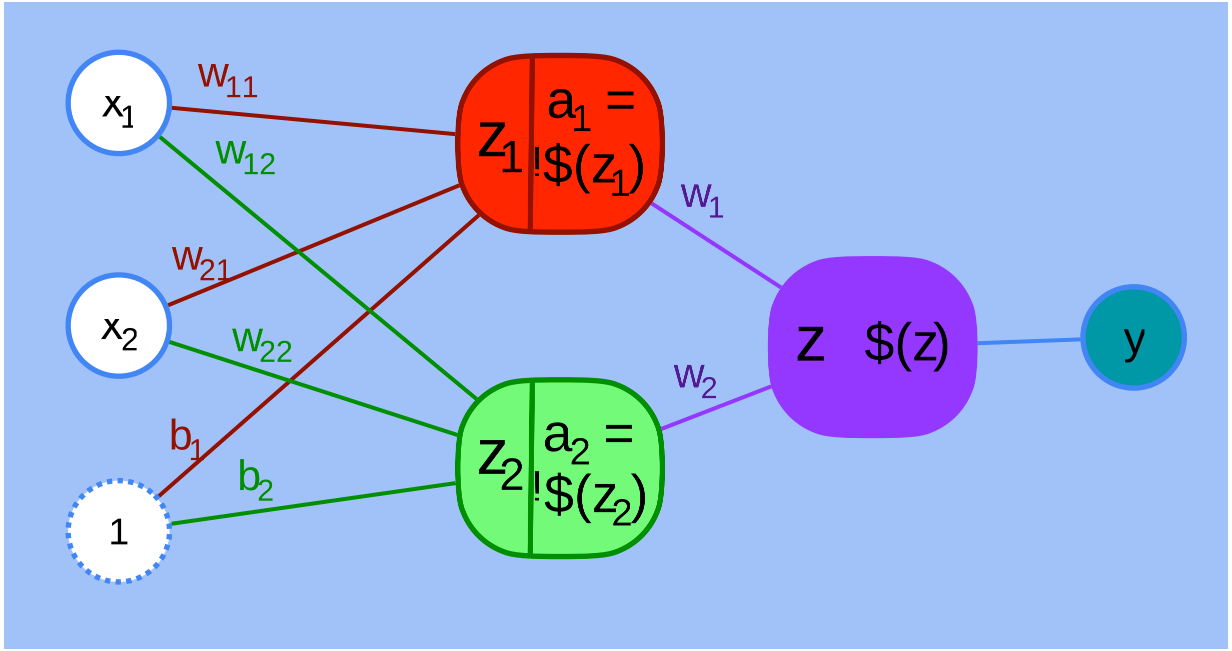
2,2,1 Neural Network



2,2,1 Neural Network



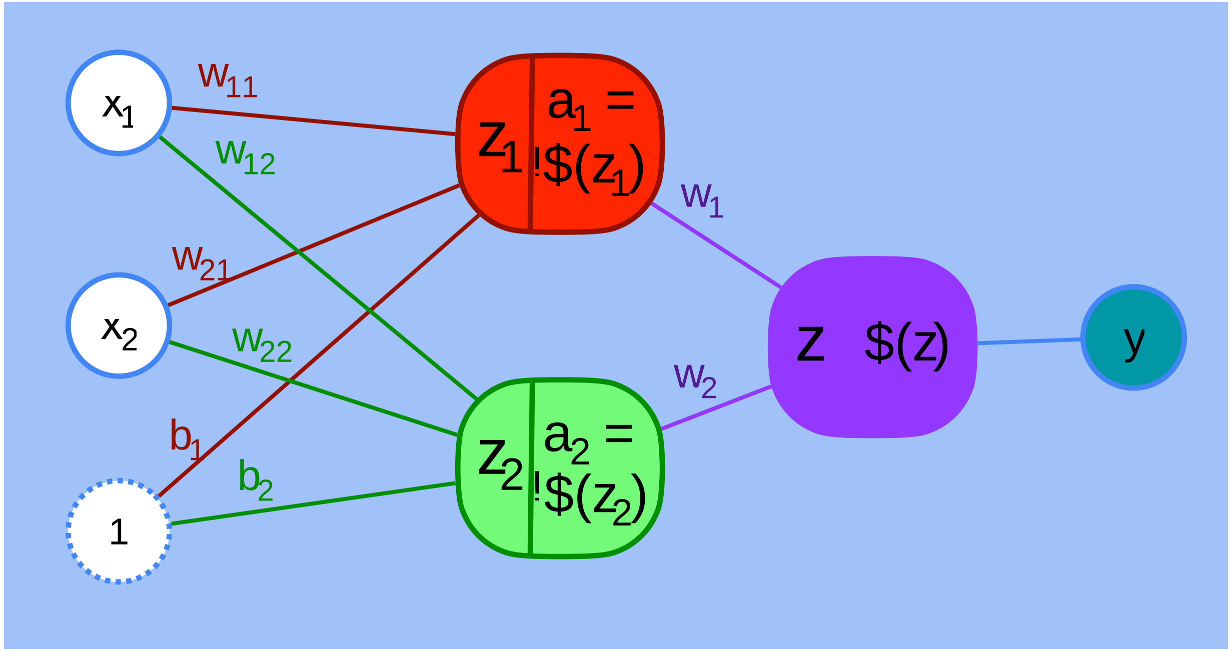
2,2,1 Neural Network



2,2,1 Neural Network

Neural network of depth 2

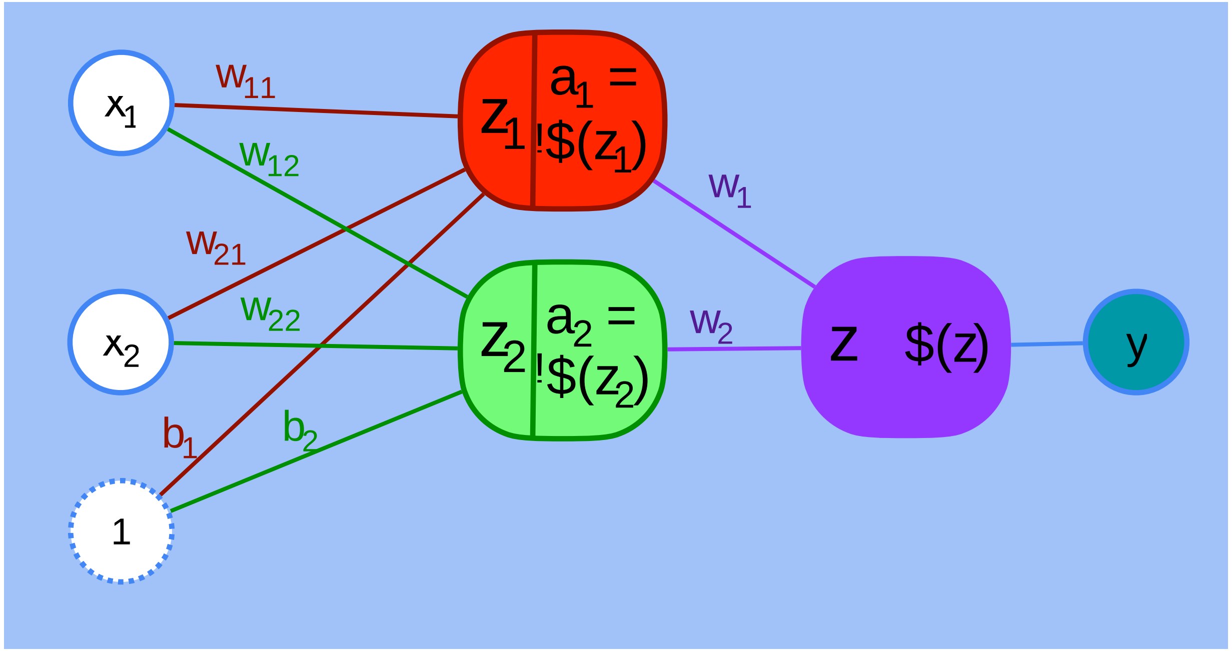
- ¥ one input layer
- ¥ one hidden layer
- ¥ one output layer



2,2,1 Neural Network

Neural network of depth 2

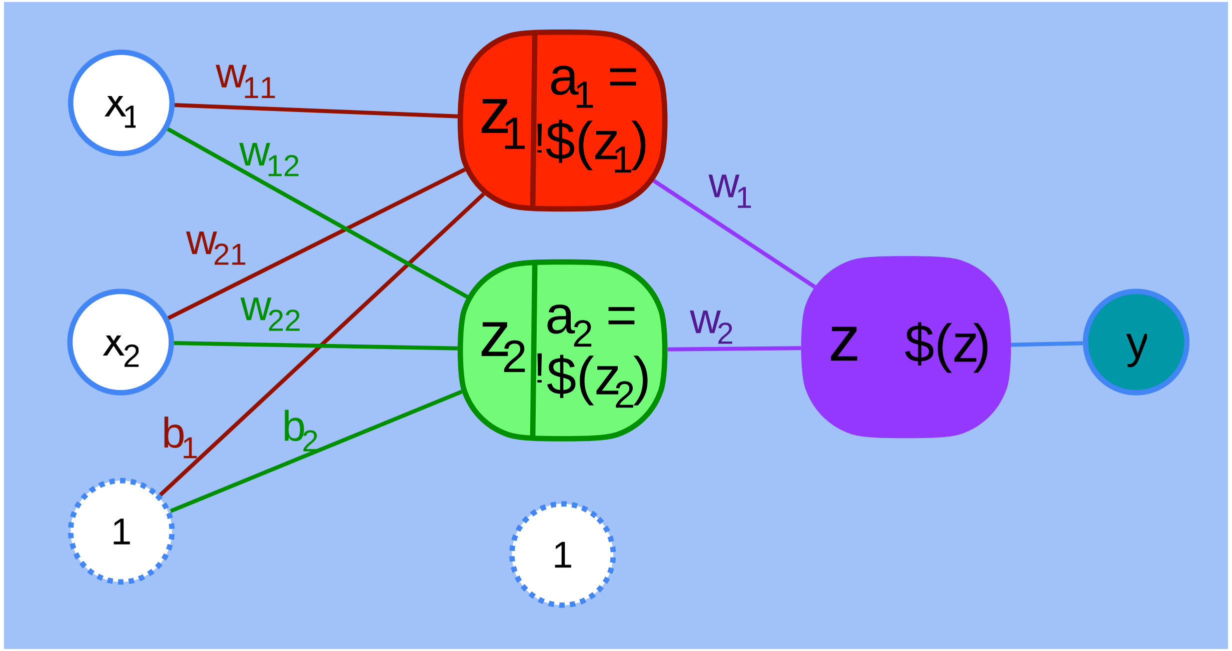
- ¥ one input layer
- ¥ one hidden layer
- ¥ one output layer



2,2,1 Neural Network

Neural network of depth 2

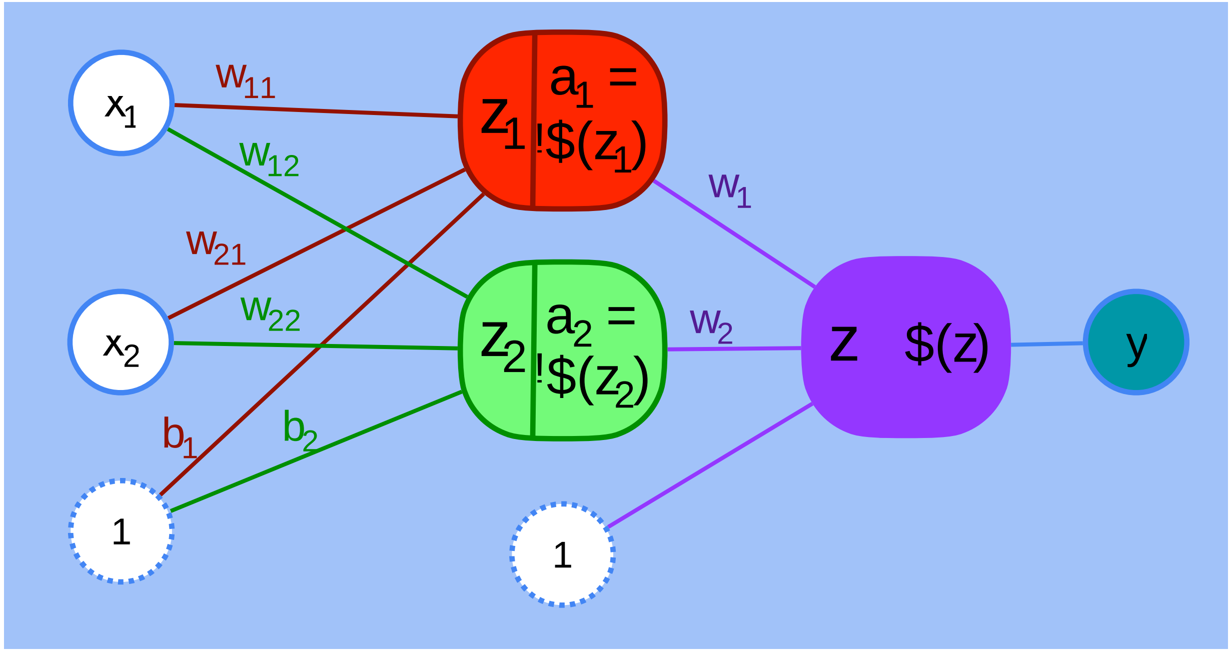
- ¥ one input layer
- ¥ one hidden layer
- ¥ one output layer



2,2,1 Neural Network

Neural network of depth 2

- ¥ one input layer
- ¥ one hidden layer
- ¥ one output layer



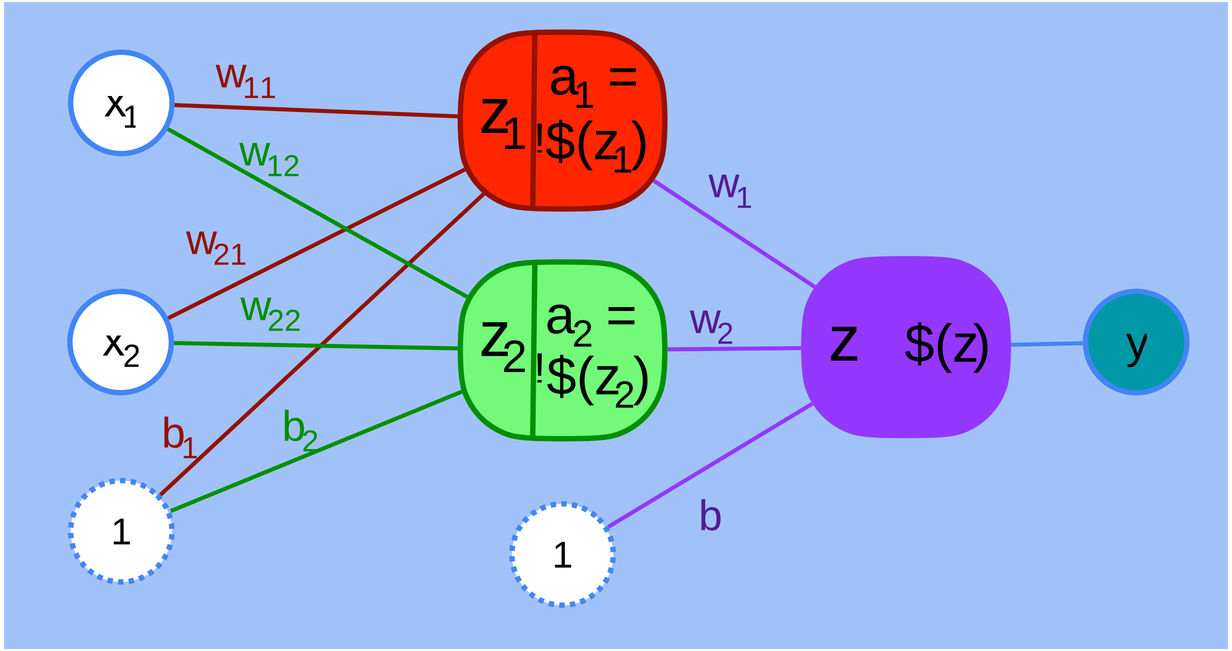
2,2,1 Neural Network

Neural network of depth 2

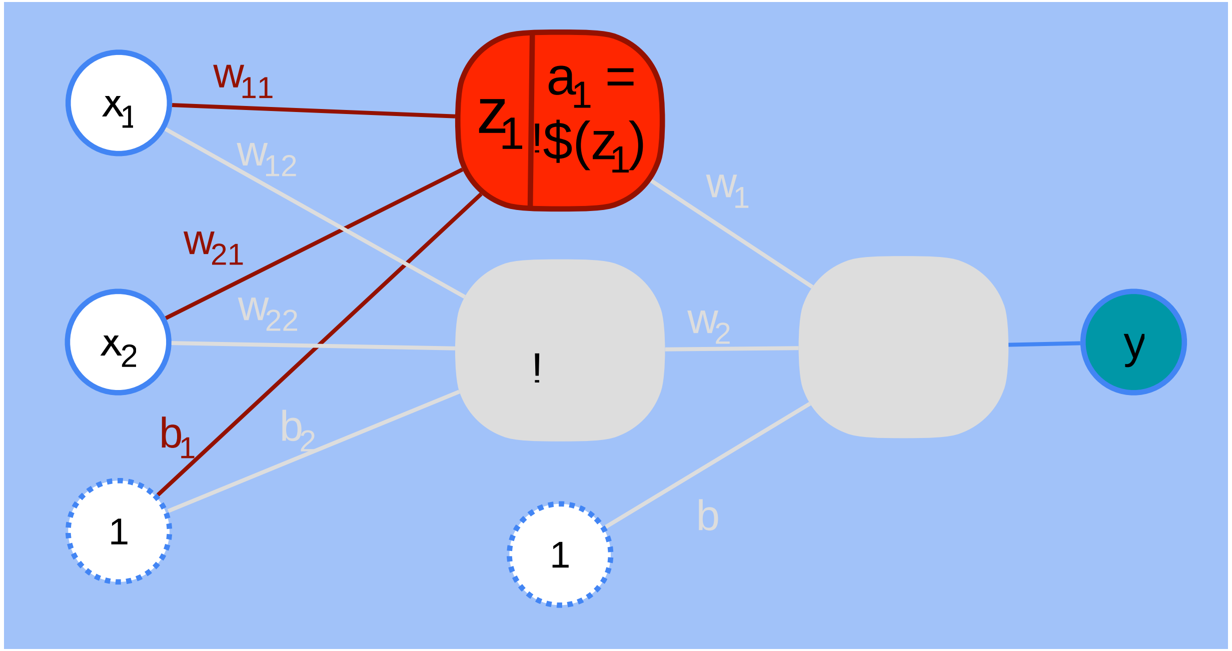
¥ one input layer

¥ one hidden layer

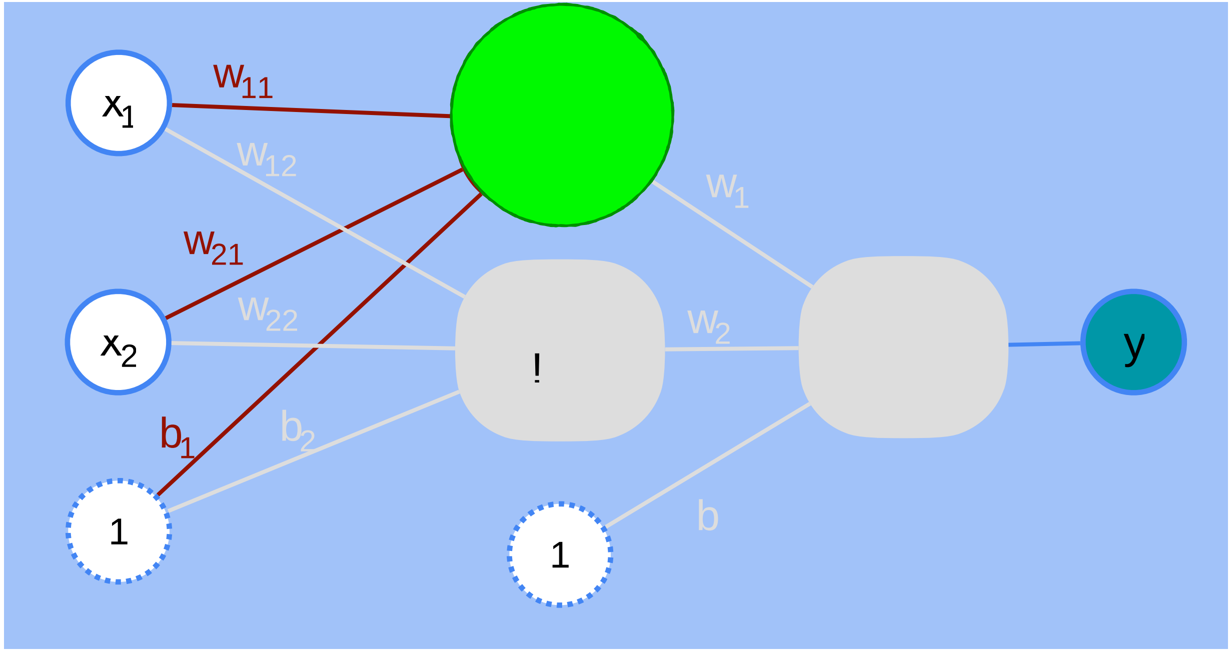
¥ one output layer



2,2,1 Neural Network

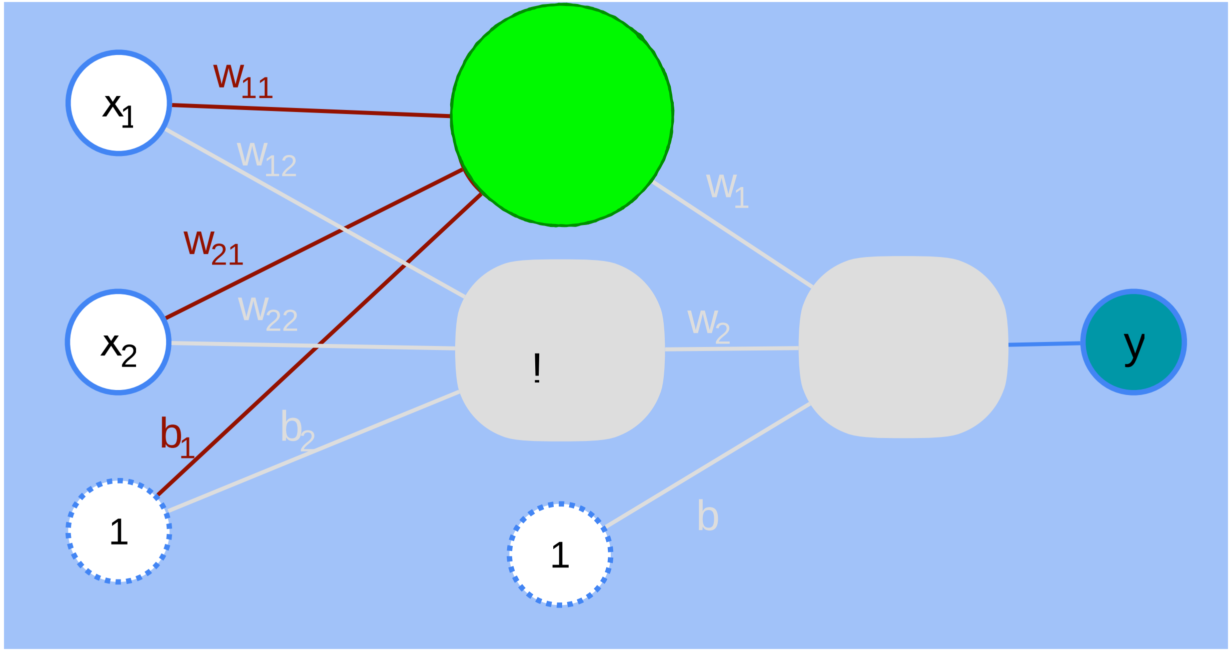


2,2,1 Neural Network



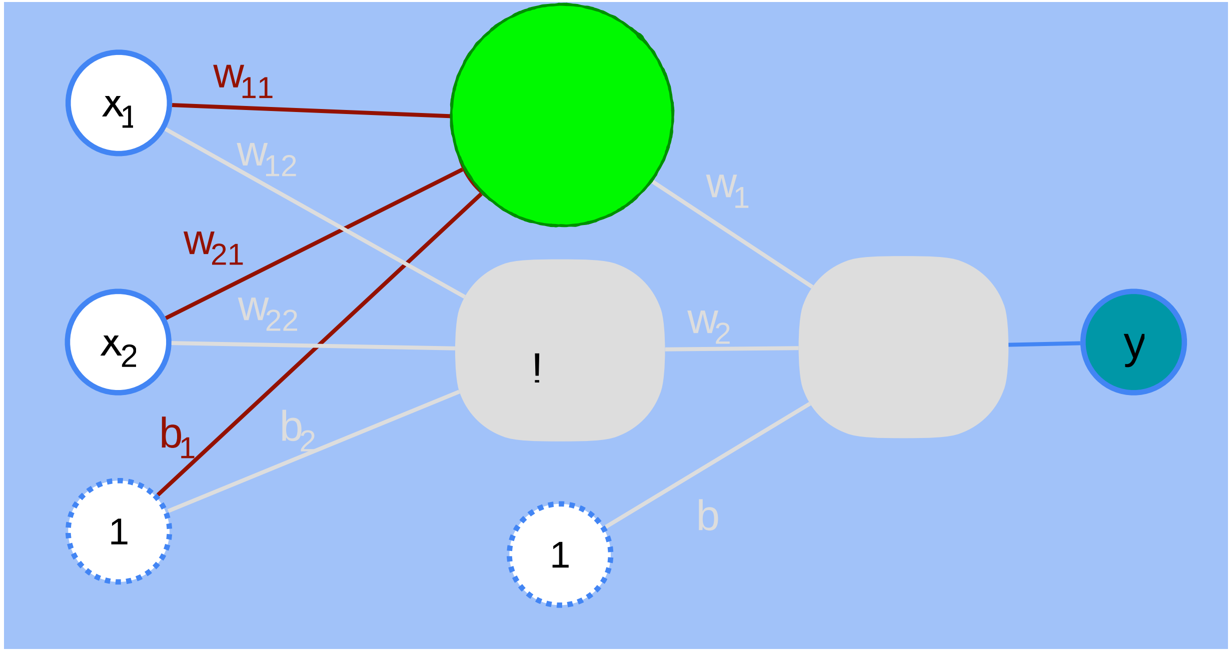
2,2,1 Neural Network

a_1



2,2,1 Neural Network

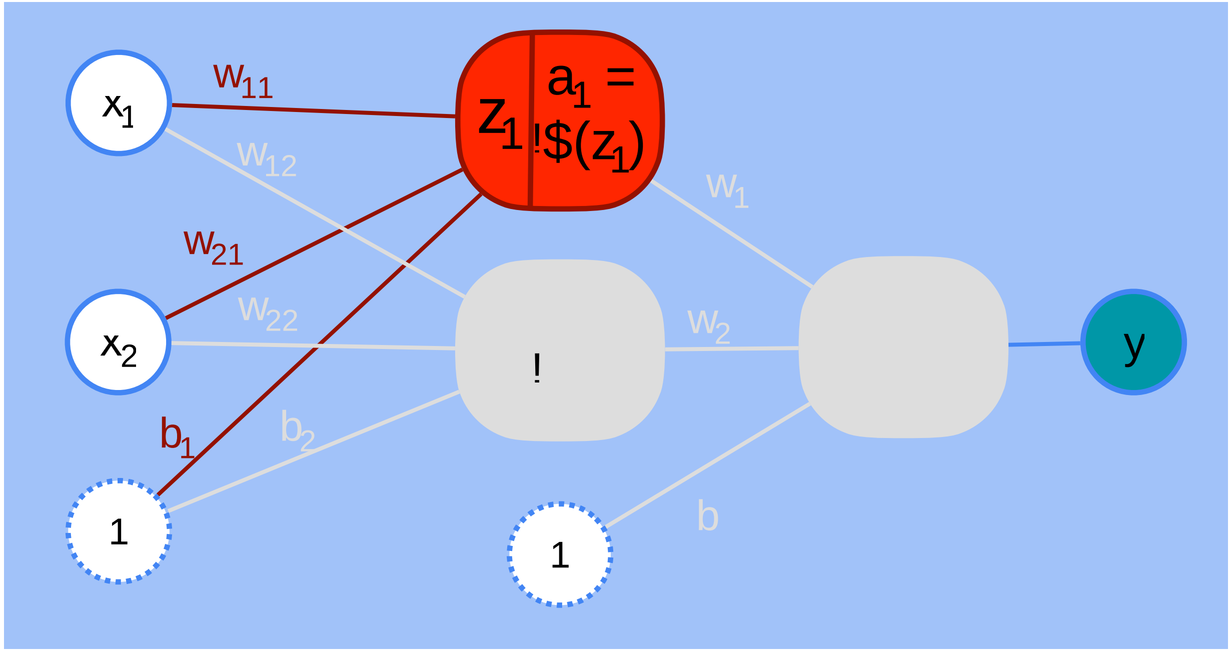
$a_1 = \$(z_1)$



2,2,1 Neural Network

$a_1 = \$(z_1)$

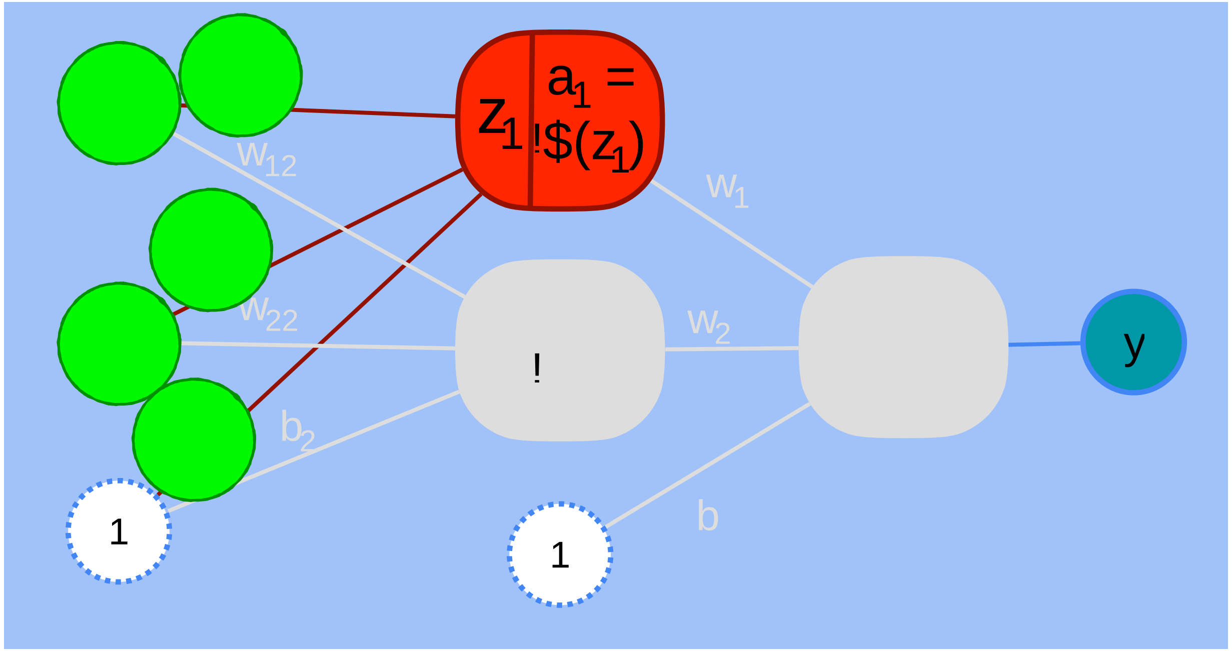
z_1



2,2,1 Neural Network

$a_1 = \$(z_1)$

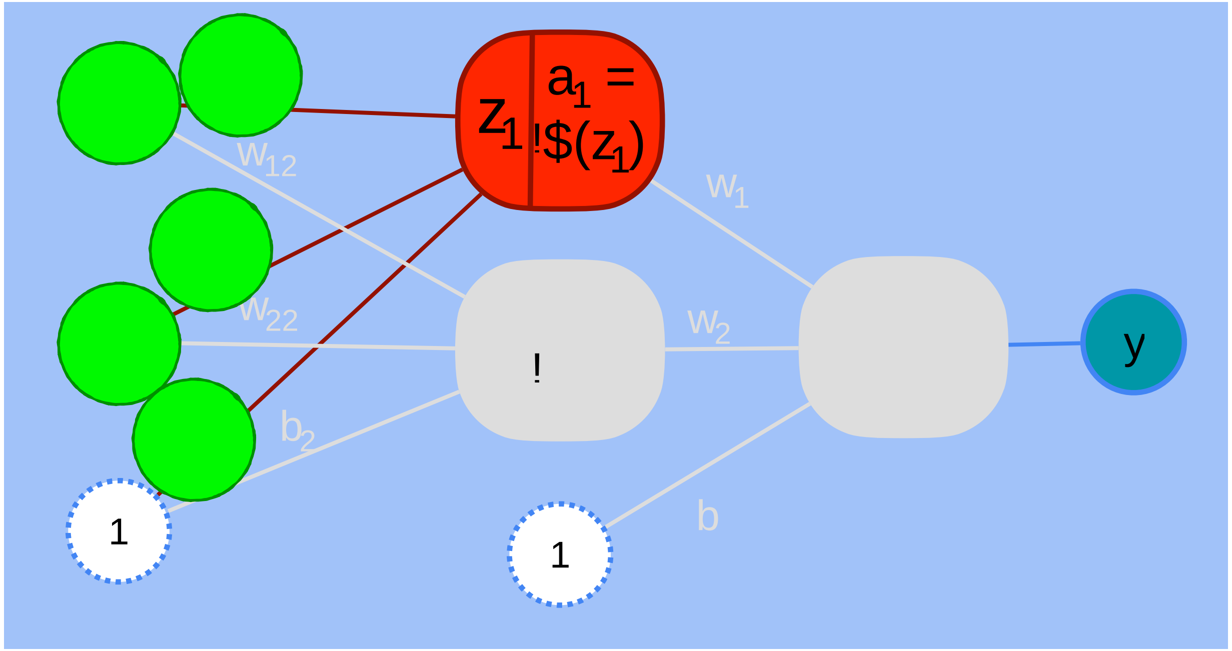
z_1



2,2,1 Neural Network

$a_1 = \$(z_1)$

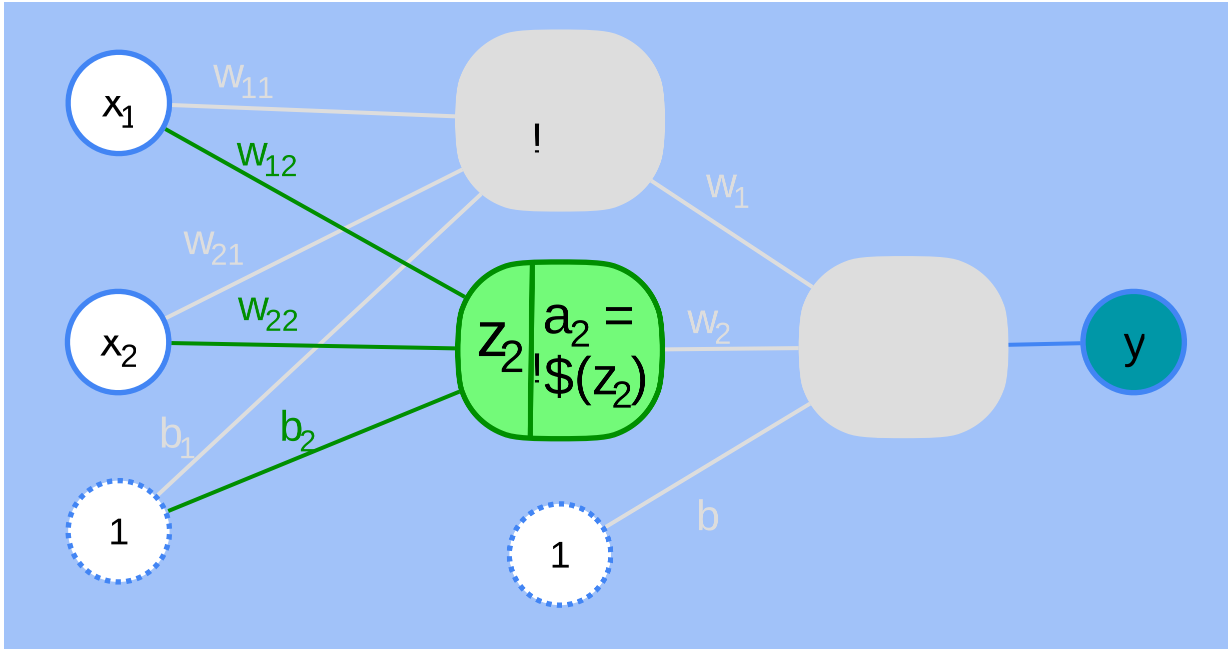
$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$



2,2,1 Neural Network

$a_1 = \$(z_1)$

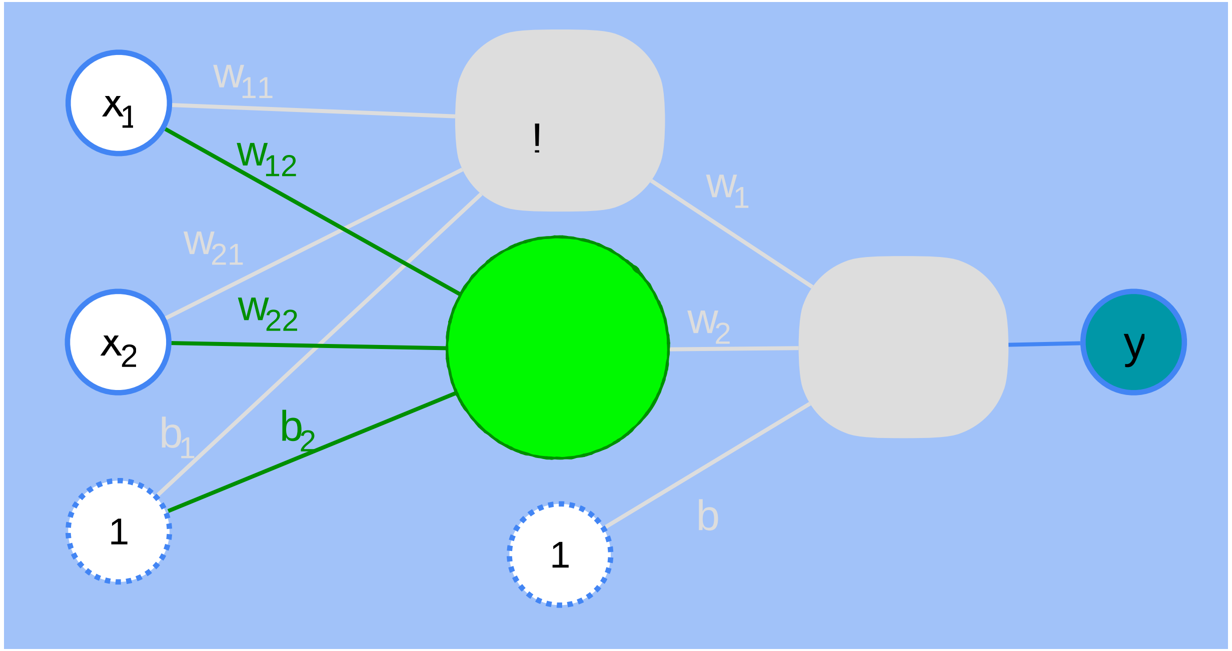
$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$



2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

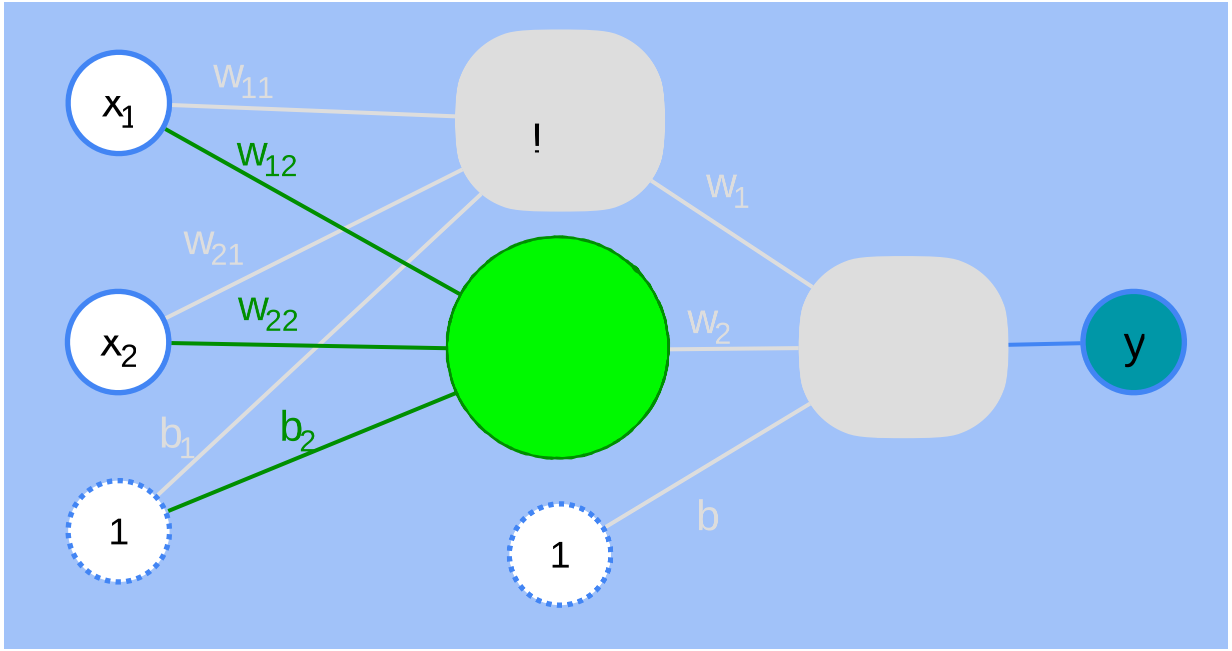


2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

a_2

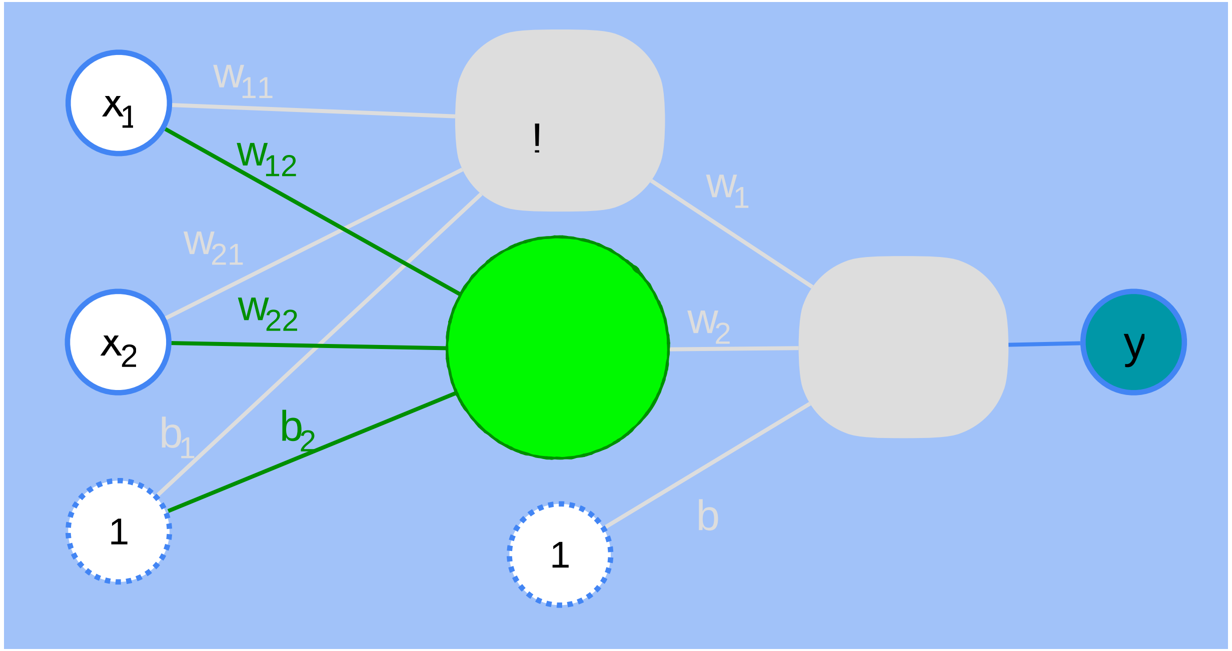


2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} \ + \ x_2 w_{21} \ + \ b_1$

$a_2 = \$(z_2)$



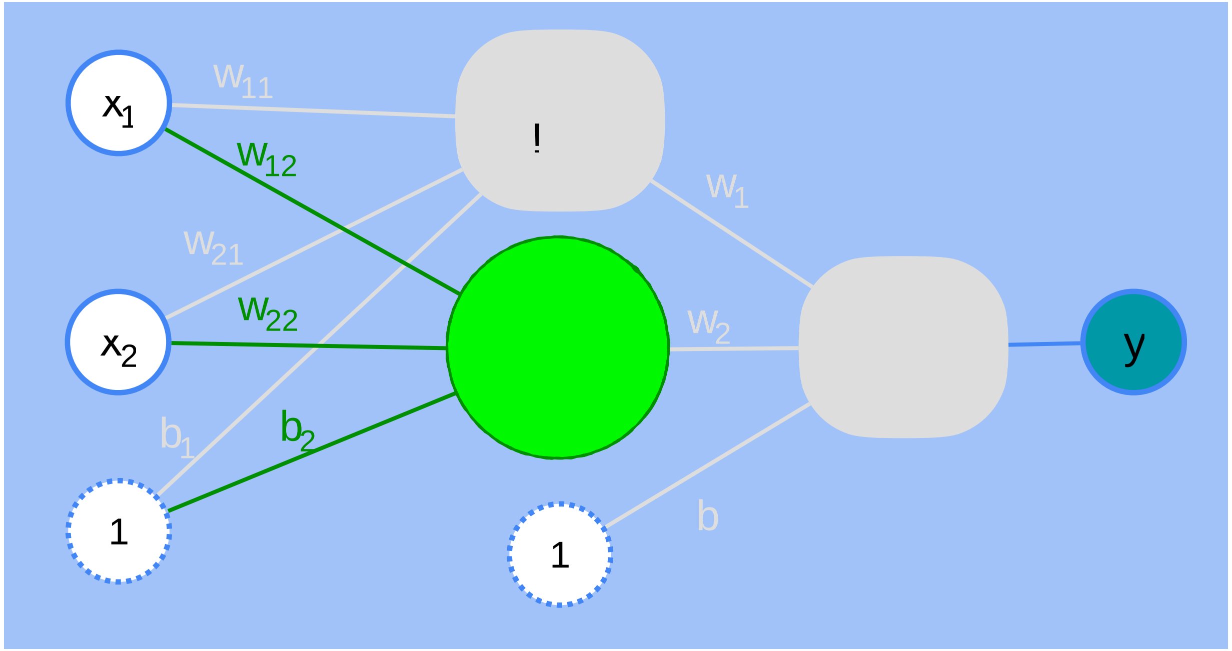
2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} \ + \ x_2 w_{21} \ + \ b_1$

$a_2 = \$(z_2)$

z_2



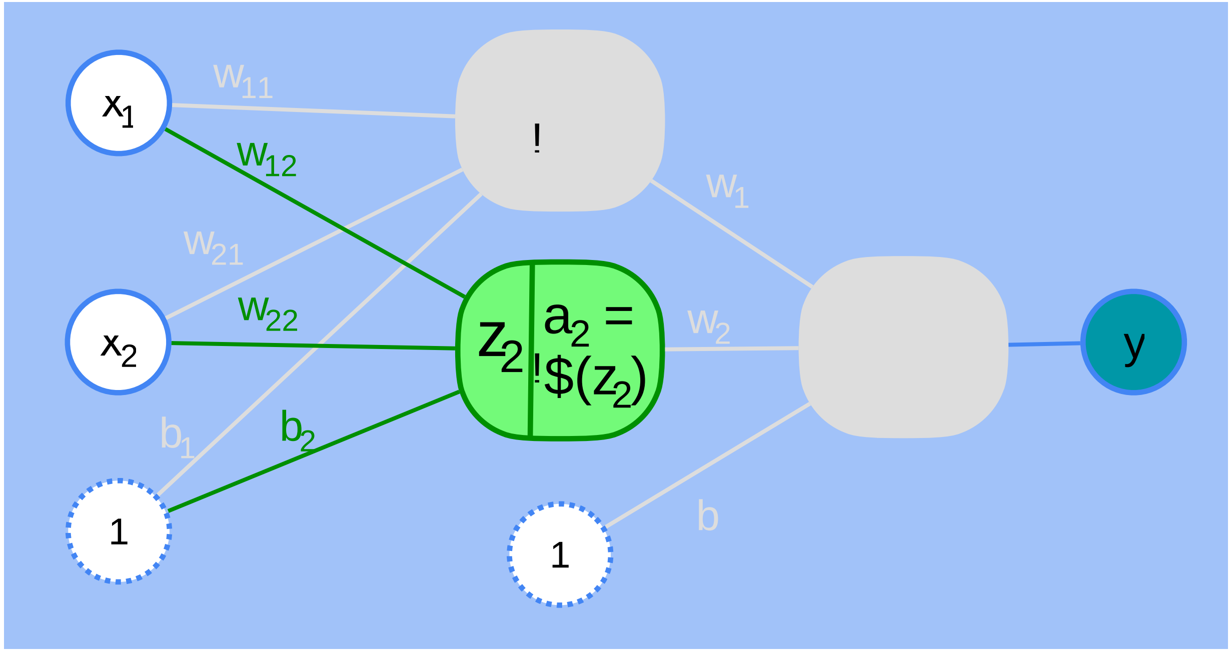
2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$a_2 = \$(z_2)$

z_2



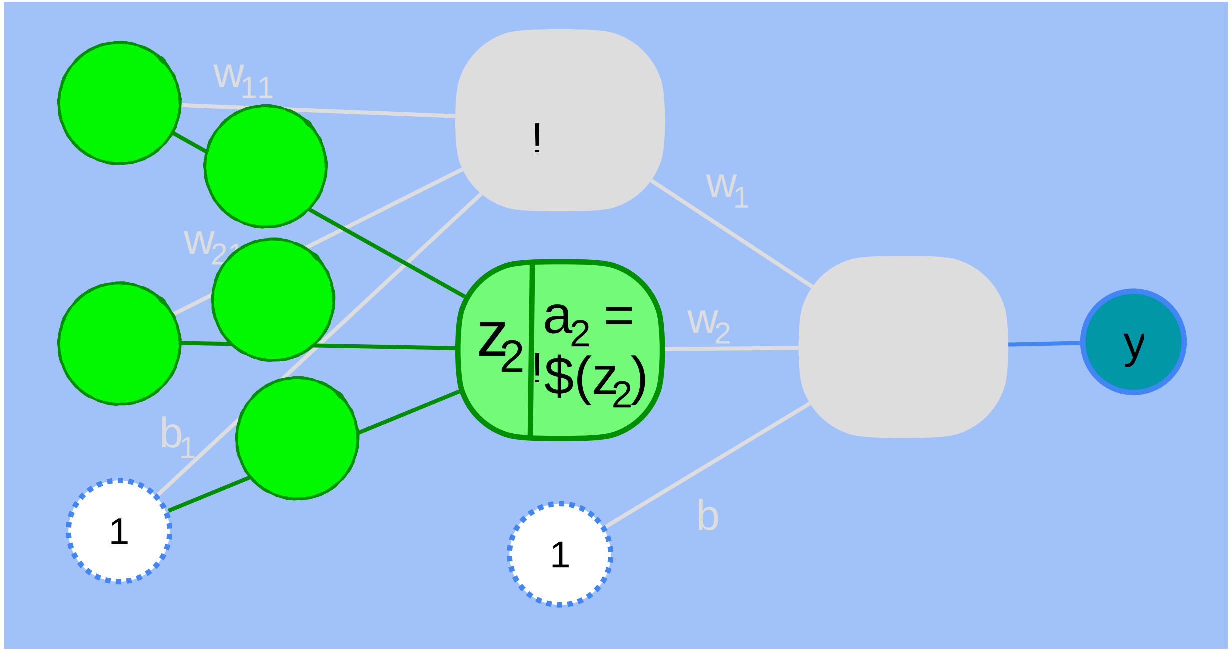
2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$a_2 = \$(z_2)$

z_2



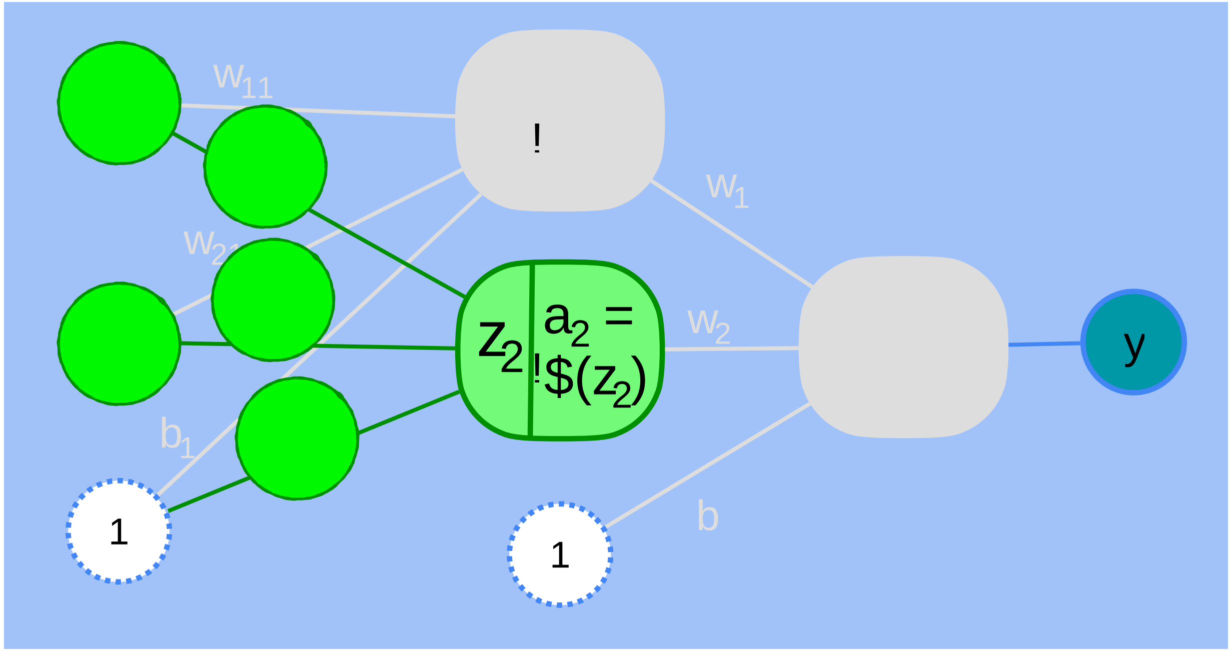
2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$a_2 = \$(z_2)$

$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$



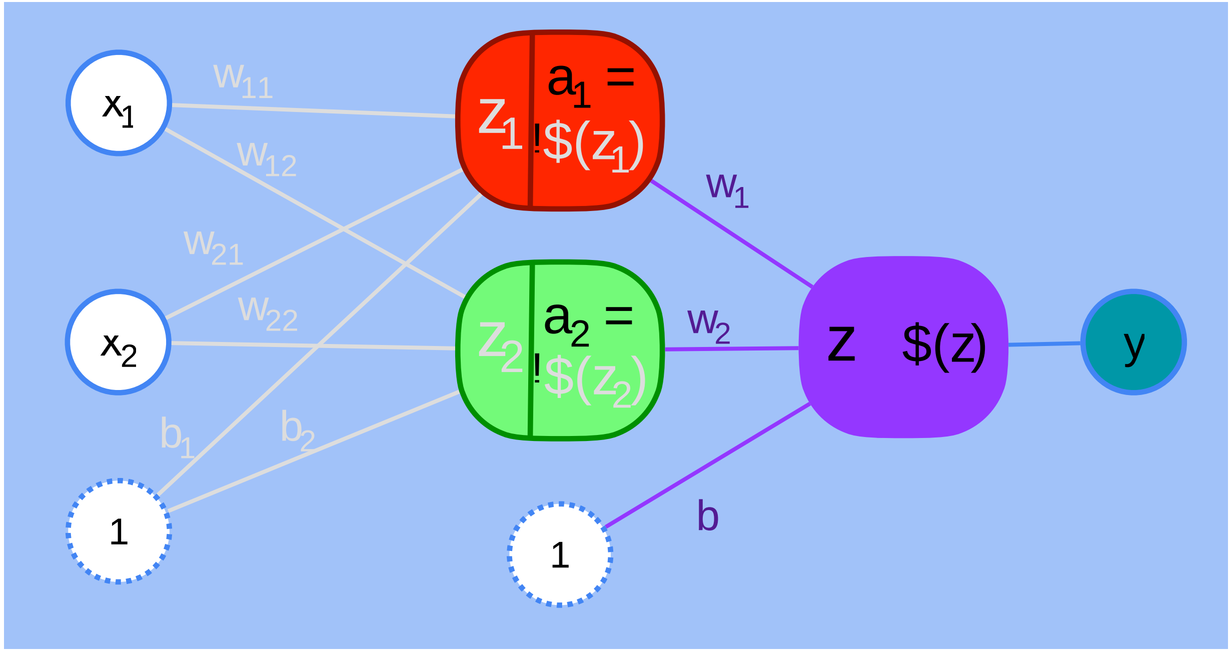
2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$a_2 = \$(z_2)$

$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$



2,2,1 Neural Network

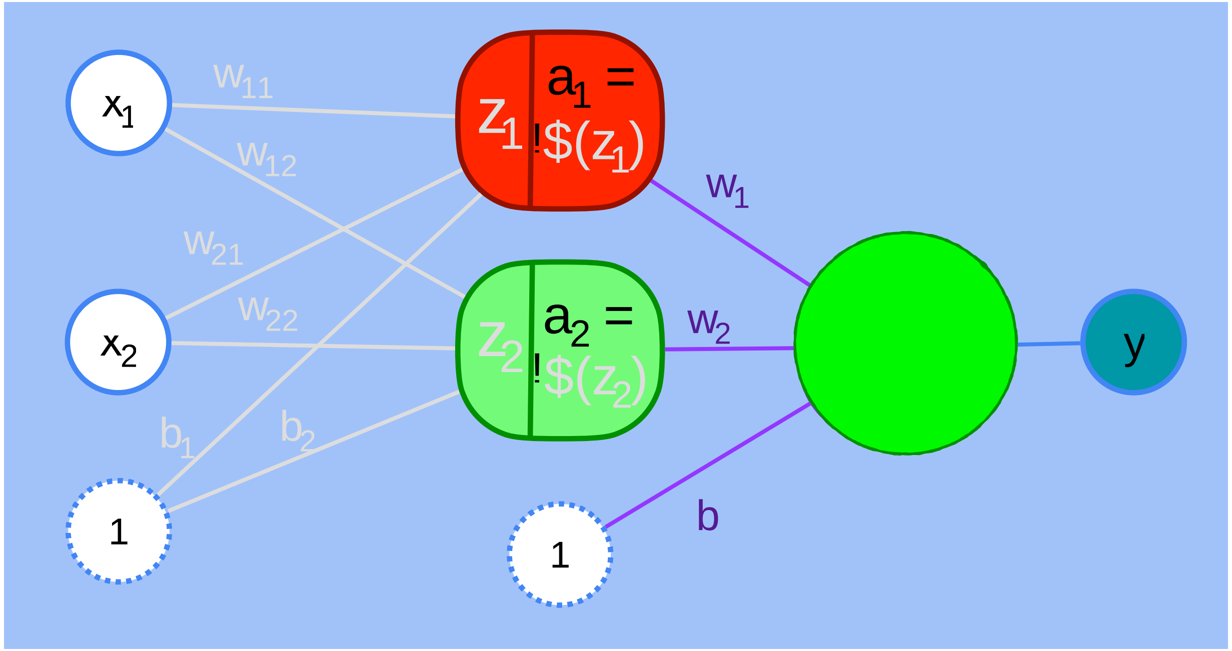
$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$a_2 = \$(z_2)$

$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$

$y1= \$(z)$



2,2,1 Neural Network

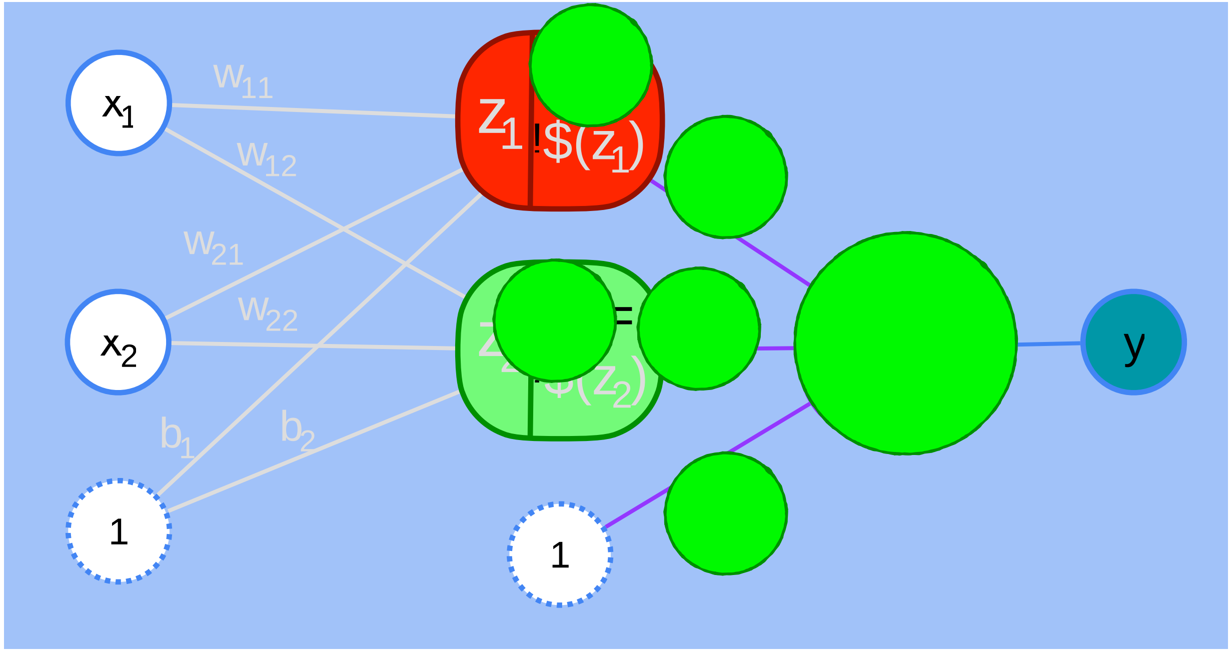
$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$a_2 = \$(z_2)$

$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$

$y1 = \$(z)$



2,2,1 Neural Network

$a_1 = \$(z_1)$

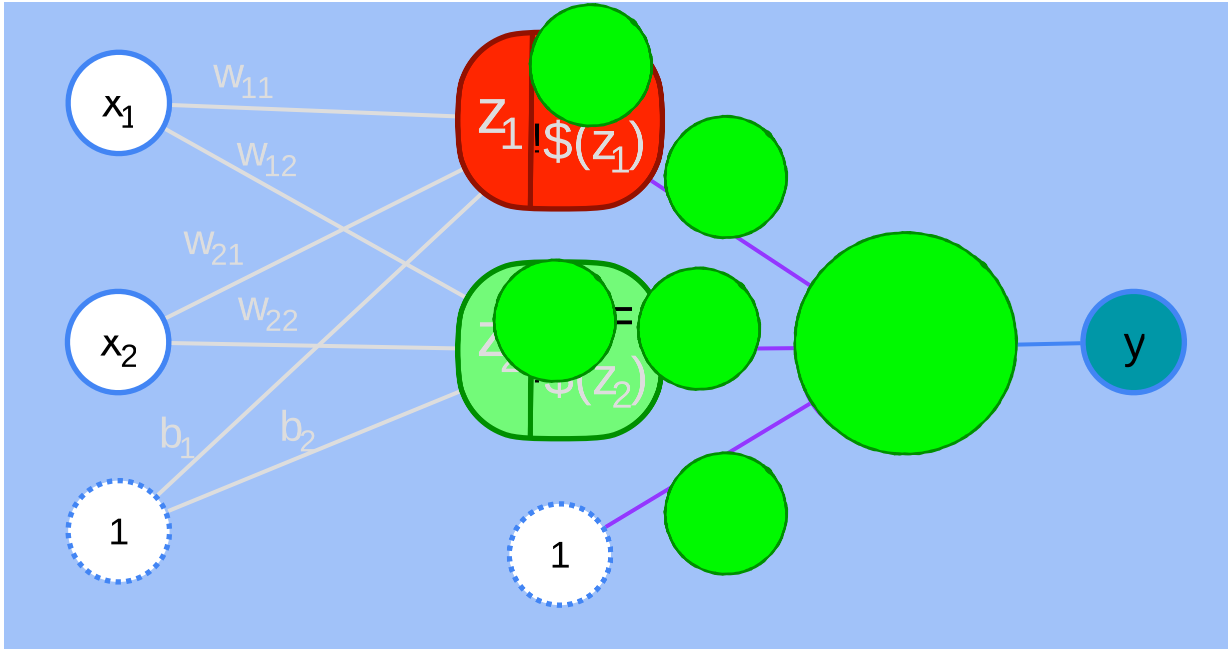
$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$a_2 = \$(z_2)$

$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$

y1= $\$(z)$

$z = a_1 w_1 + a_2 w_2 + b$



2,2,1 Neural Network

$a_1 = \$(z_1)$

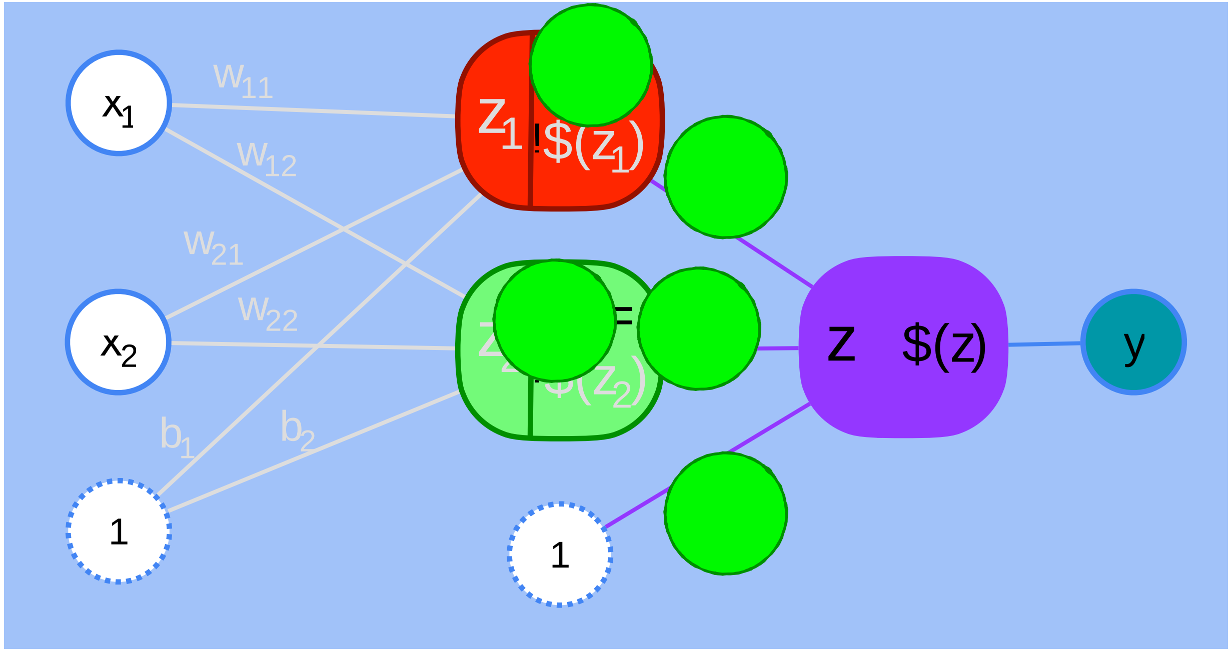
$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$a_2 = \$(z_2)$

$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$



2,2,1 Neural Network

$a_1 = \$(z_1)$

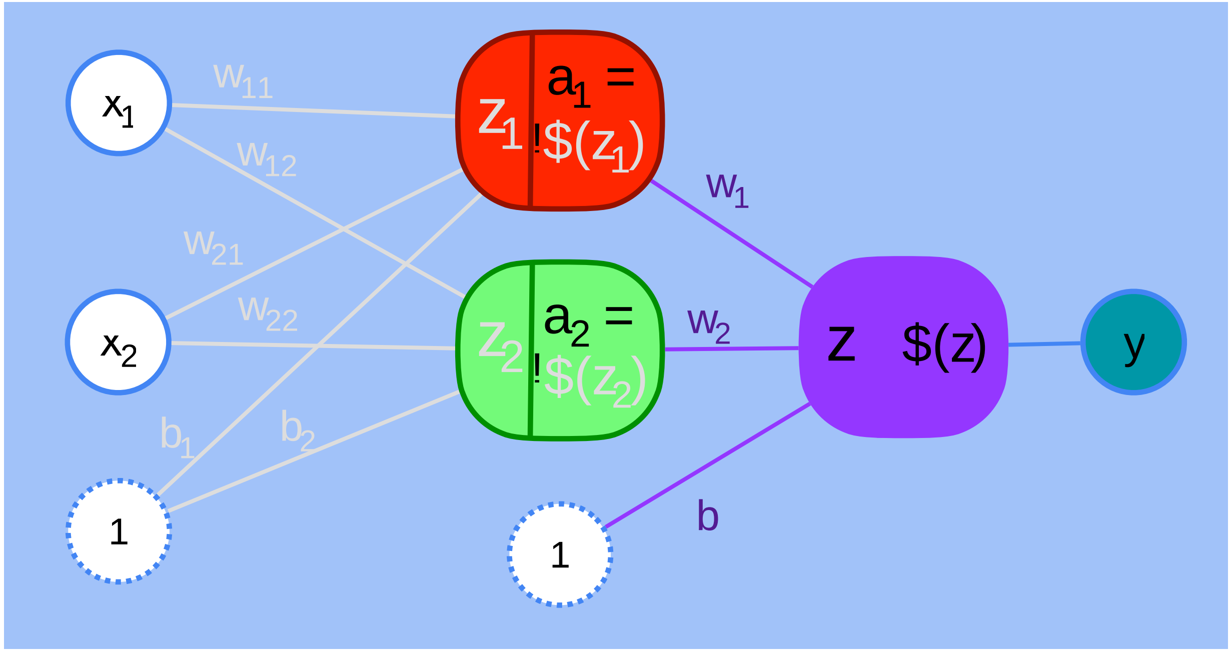
$z_1 = x_1 w_{11} \; + \; x_2 w_{21} \; + \; b_1$

$a_2 = \$(z_2)$

$z_2 = x_1 w_{12} \; + \; x_2 w_{22} \; + \; b_2$

$y1 = \$(z)$

$z = a_1 w_1 \; + \; a_2 w_2 \; + \; b$



2,2,1 Neural Network

$a_1 = \$(z_1)$

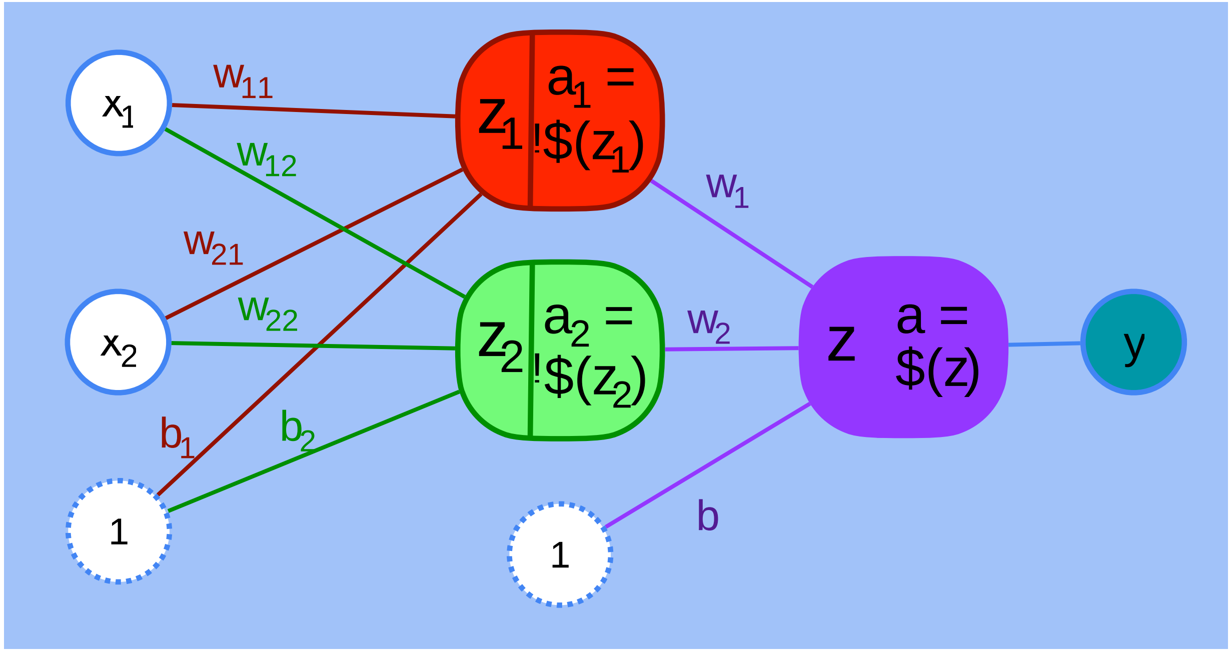
$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$a_2 = \$(z_2)$

$z_2 = x_1w_{12} + x_2w_{22} + b_2$

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$



2,2,1 Neural Network

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

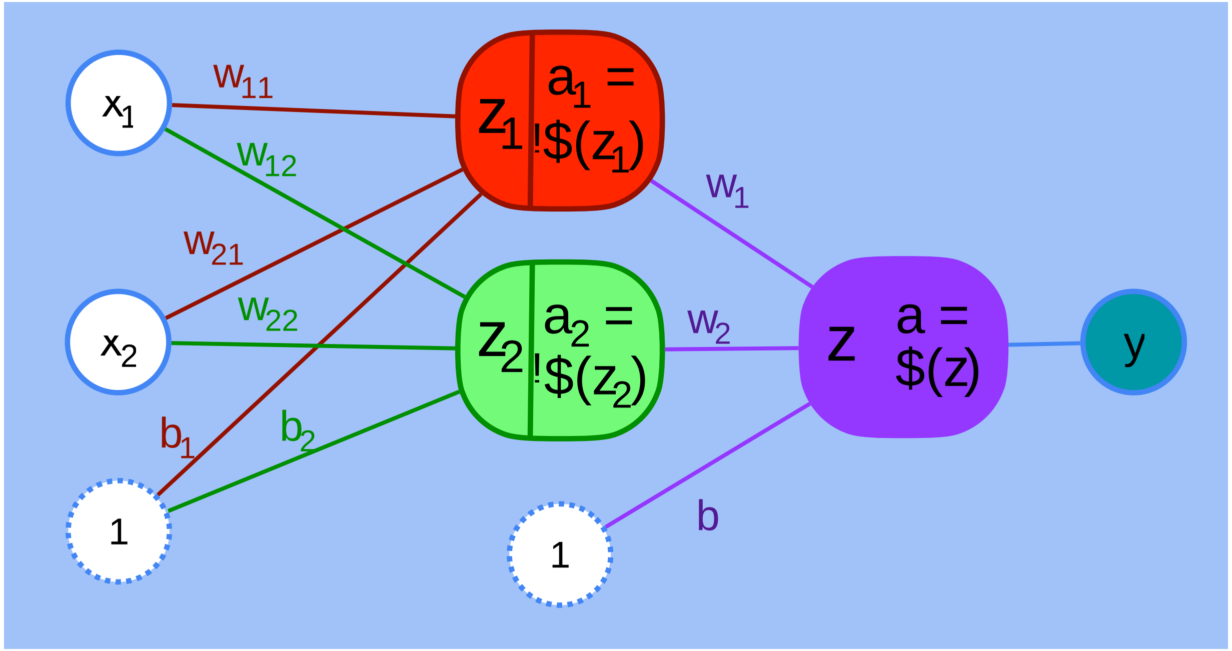
$a_2 = \$(z_2)$

$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

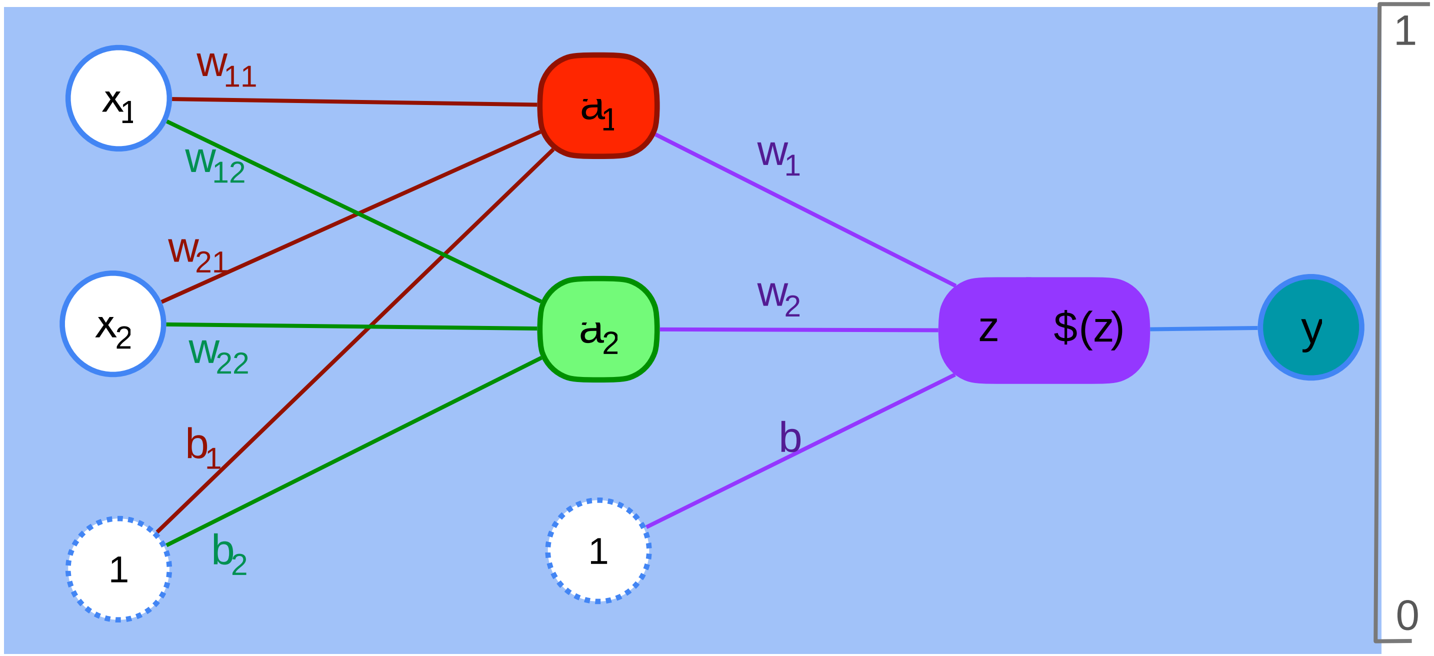
$L(y, y1) = -y\log(y1) \text{ " } (1 \text{ " } y)\log(1 \text{ " } y1)$



Optimization in Neural Networks and NewtonÕs Method

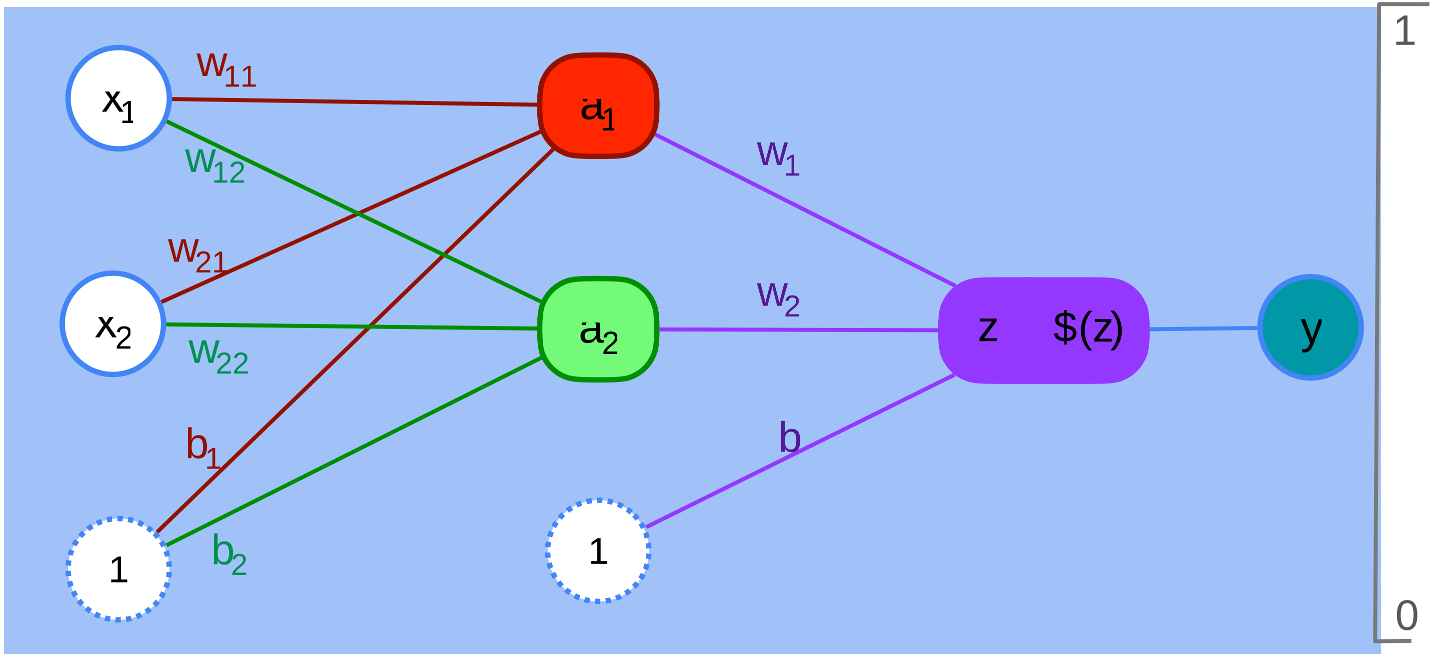
Classification with a Neural Network: Minimizing log-loss

2,2,1 Neural Network



2,2,1 Neural Network

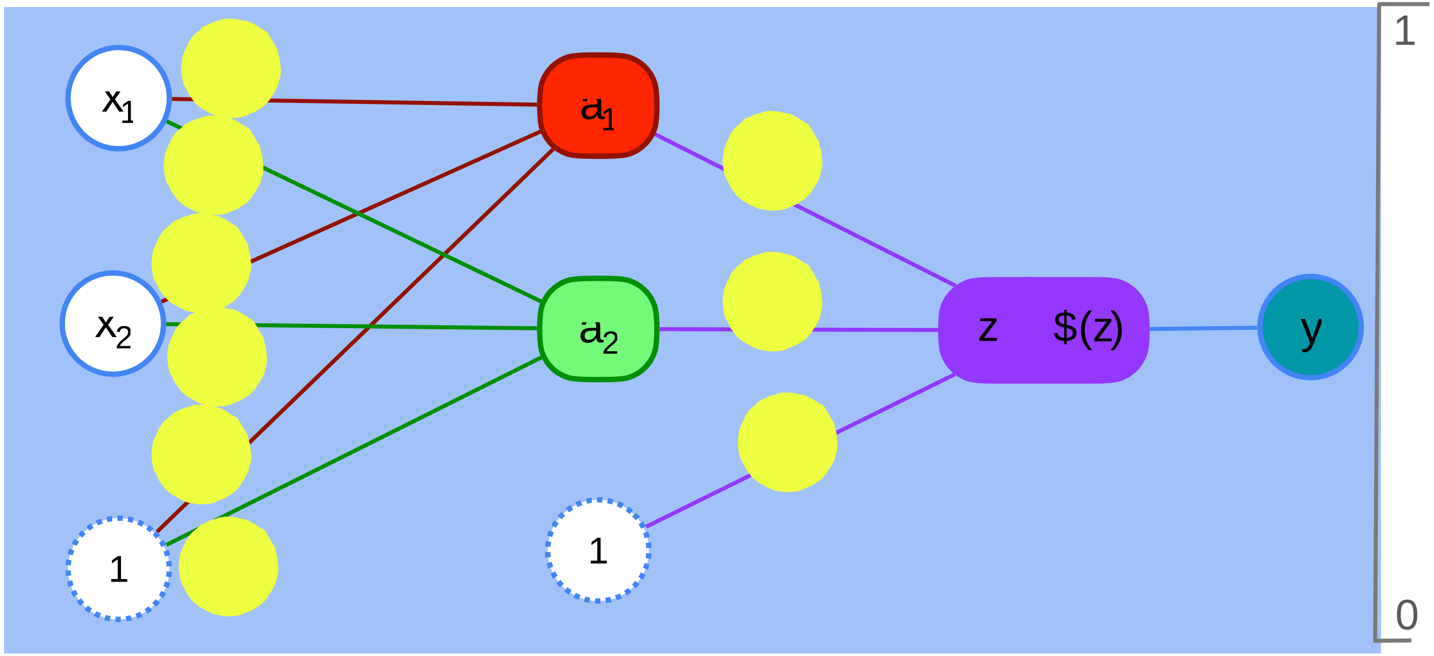
Goal



2,2,1 Neural Network

Goal

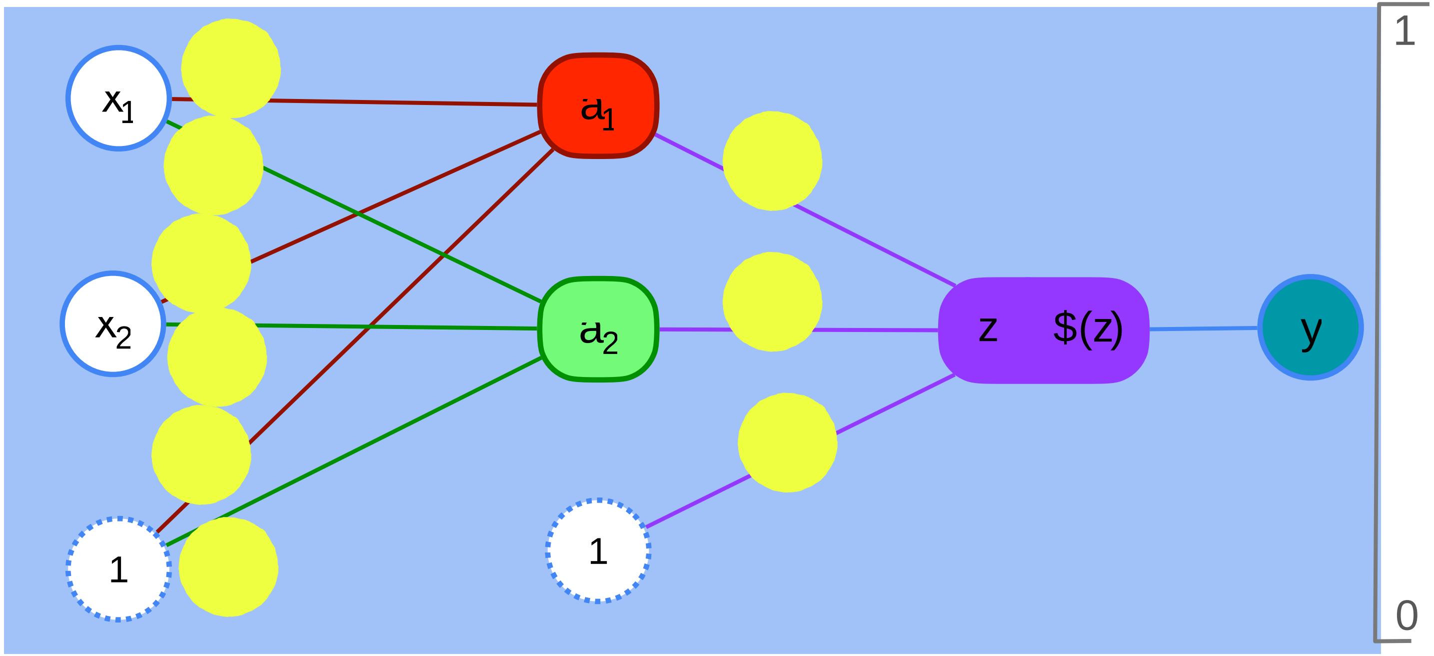
Adjust each of the highlighted weights and biases



2,2,1 Neural Network

Goal

Adjust each of the highlighted weights and biases

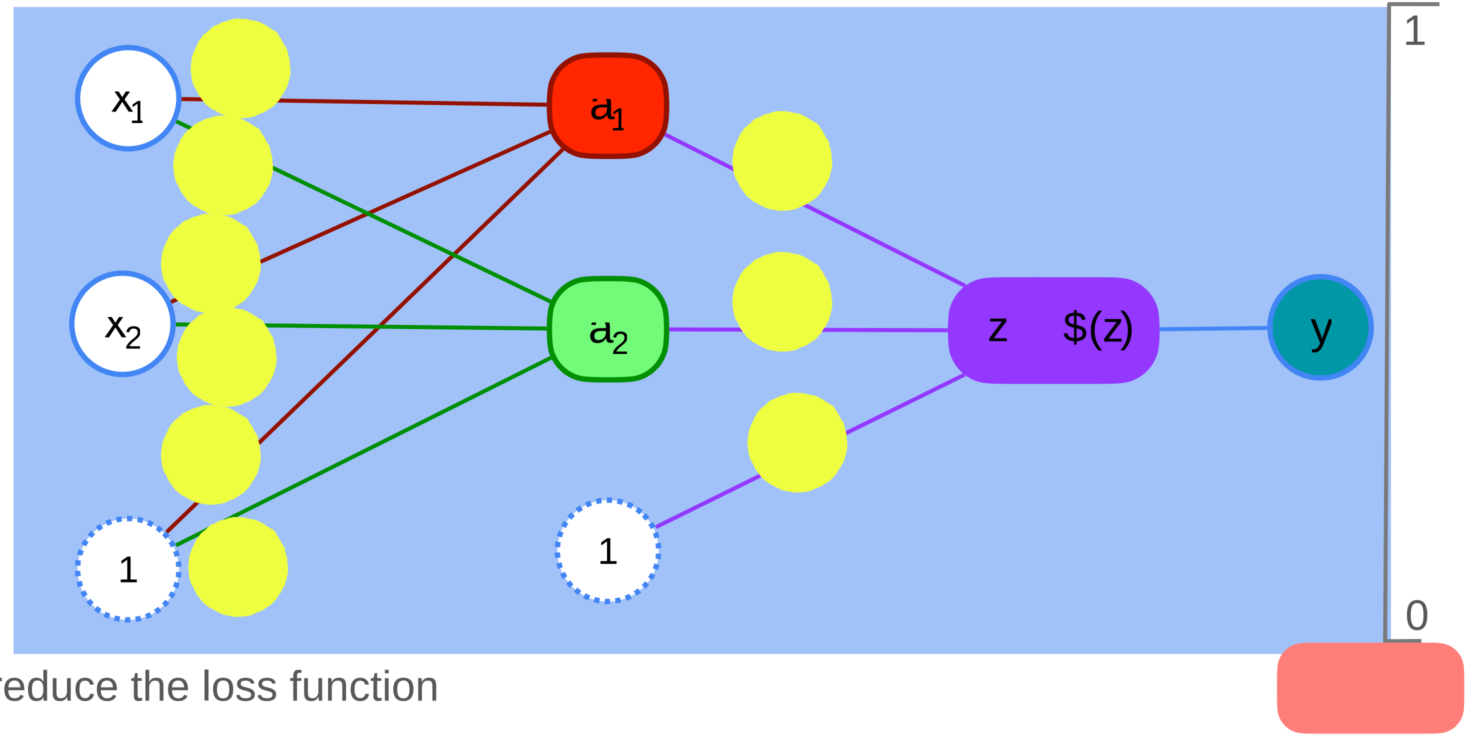


To reduce the loss function

2,2,1 Neural Network

Goal

Adjust each of the highlighted weights and biases

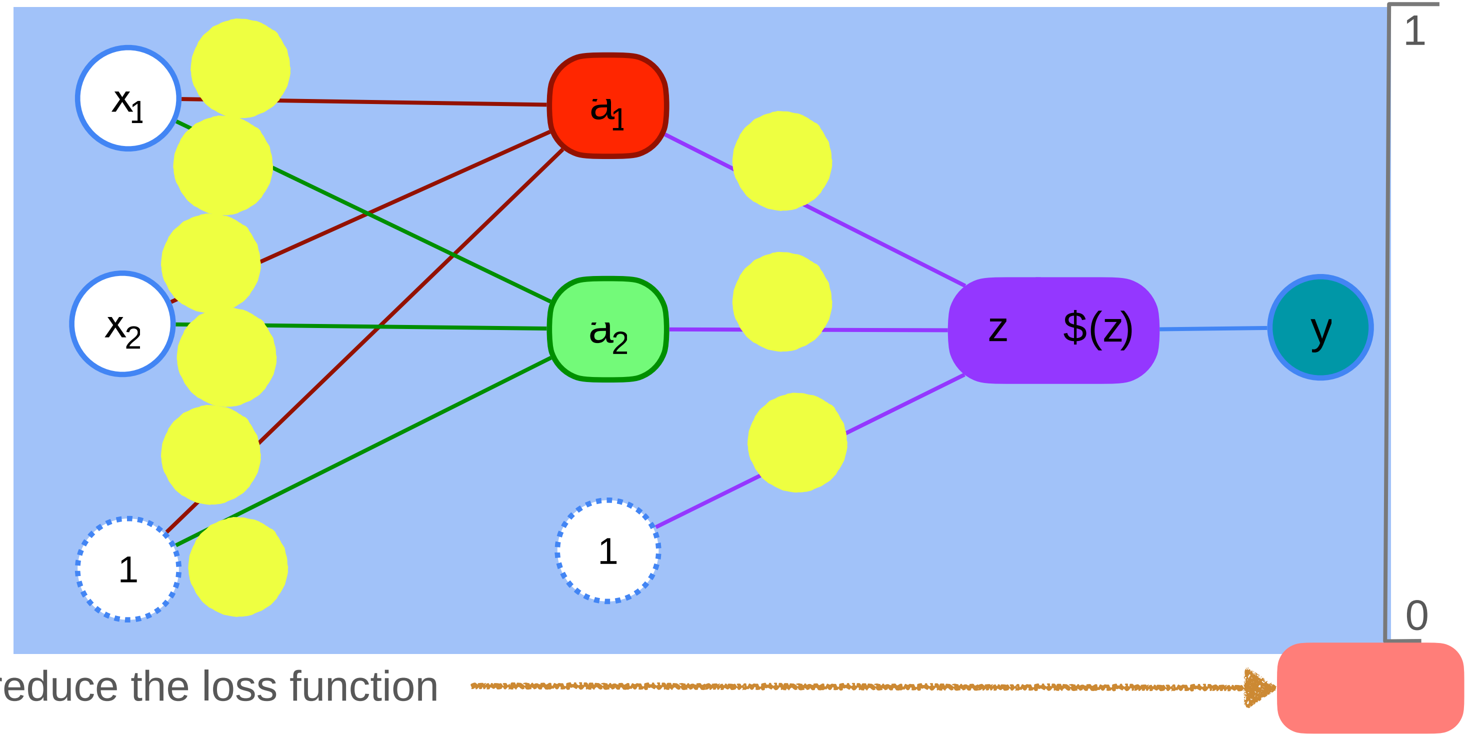


To reduce the loss function

2,2,1 Neural Network

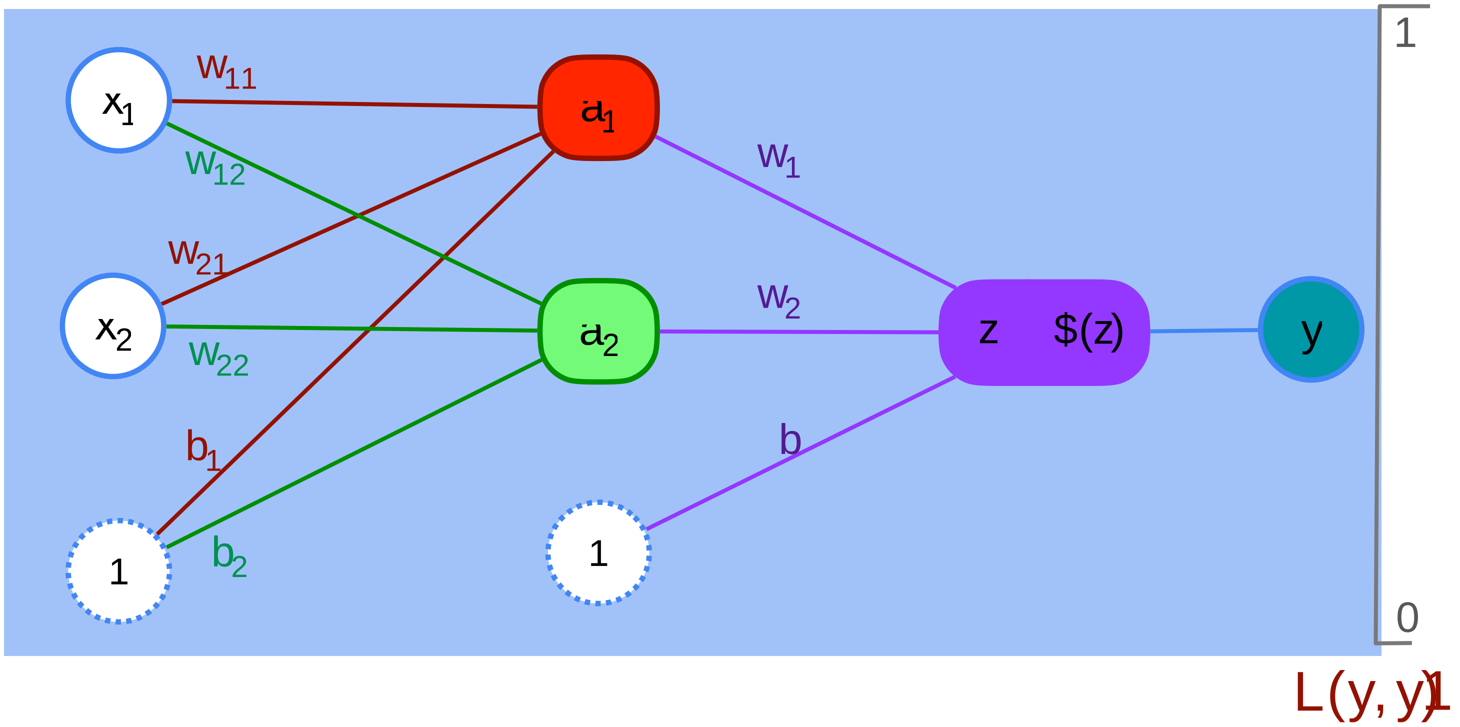
Goal

Adjust each of the highlighted weights and biases

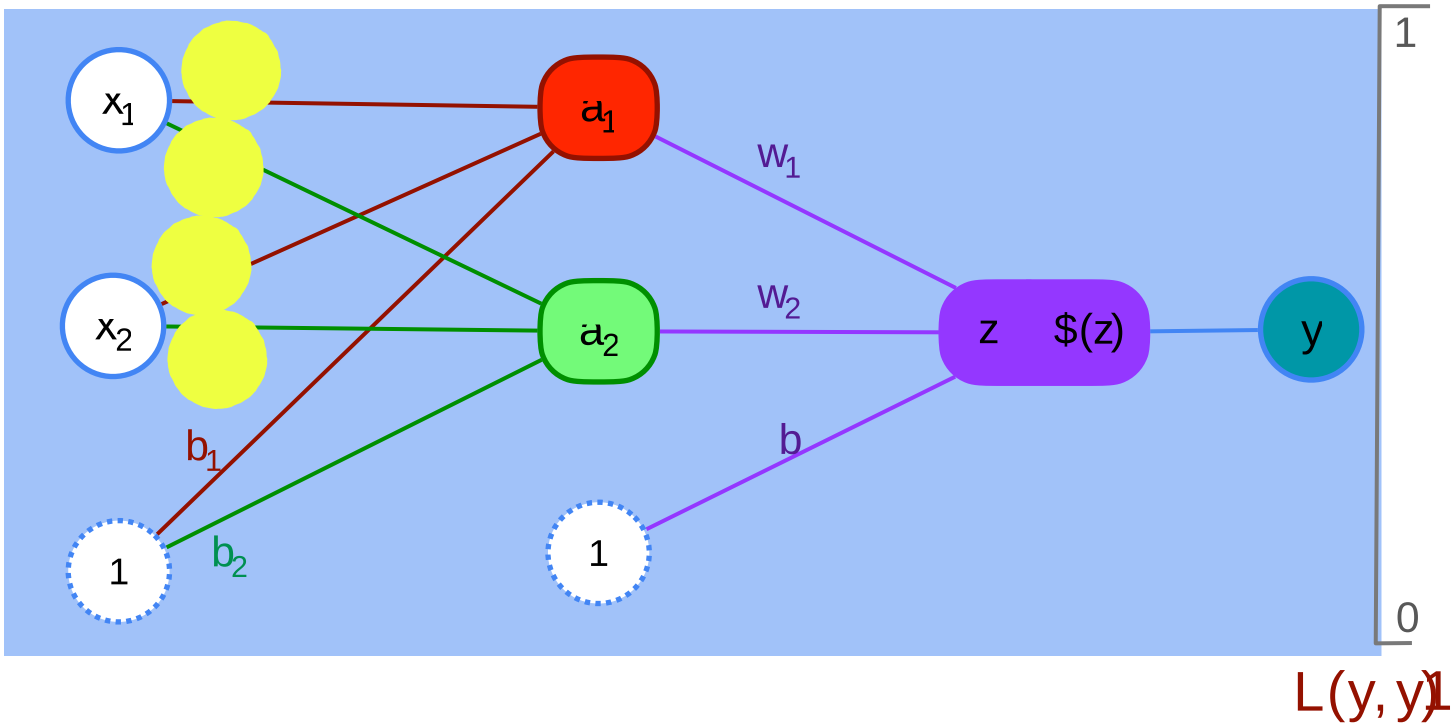


To reduce the loss function

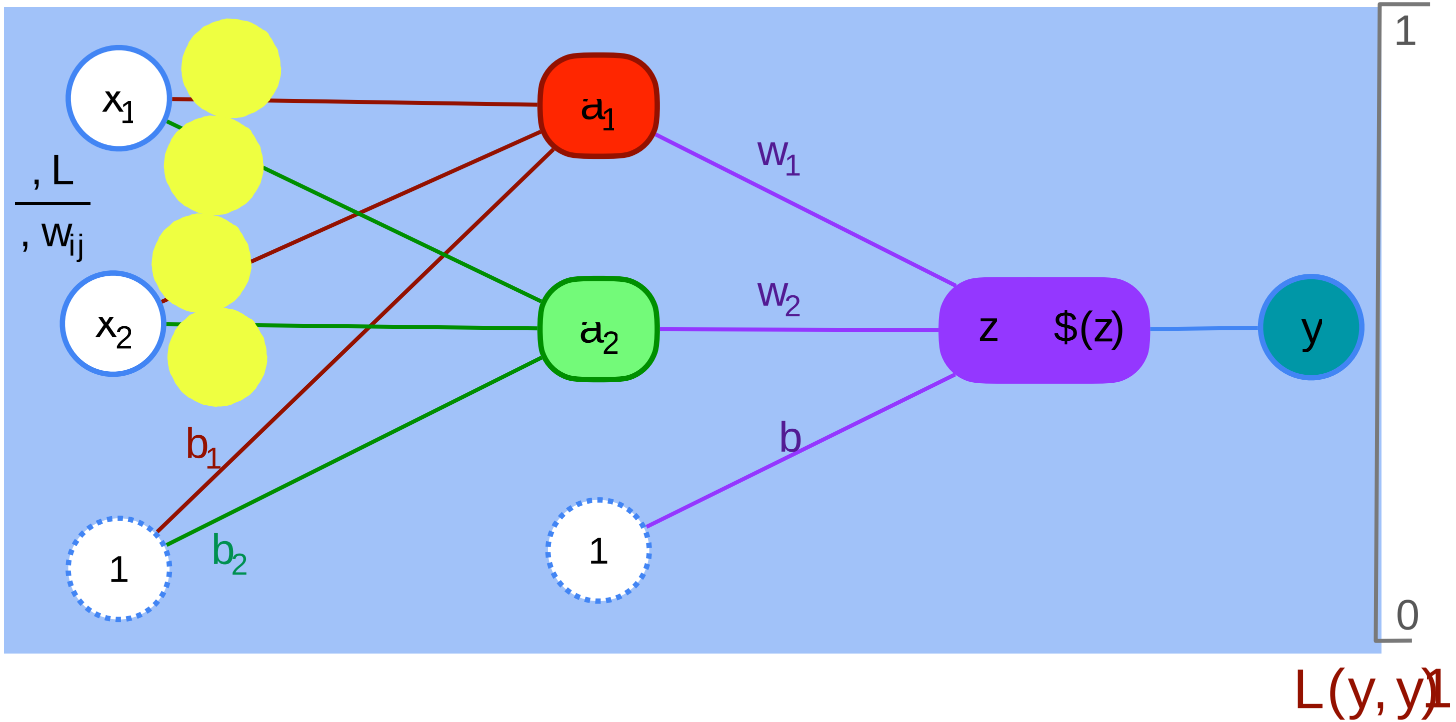
2,2,1 Neural Network



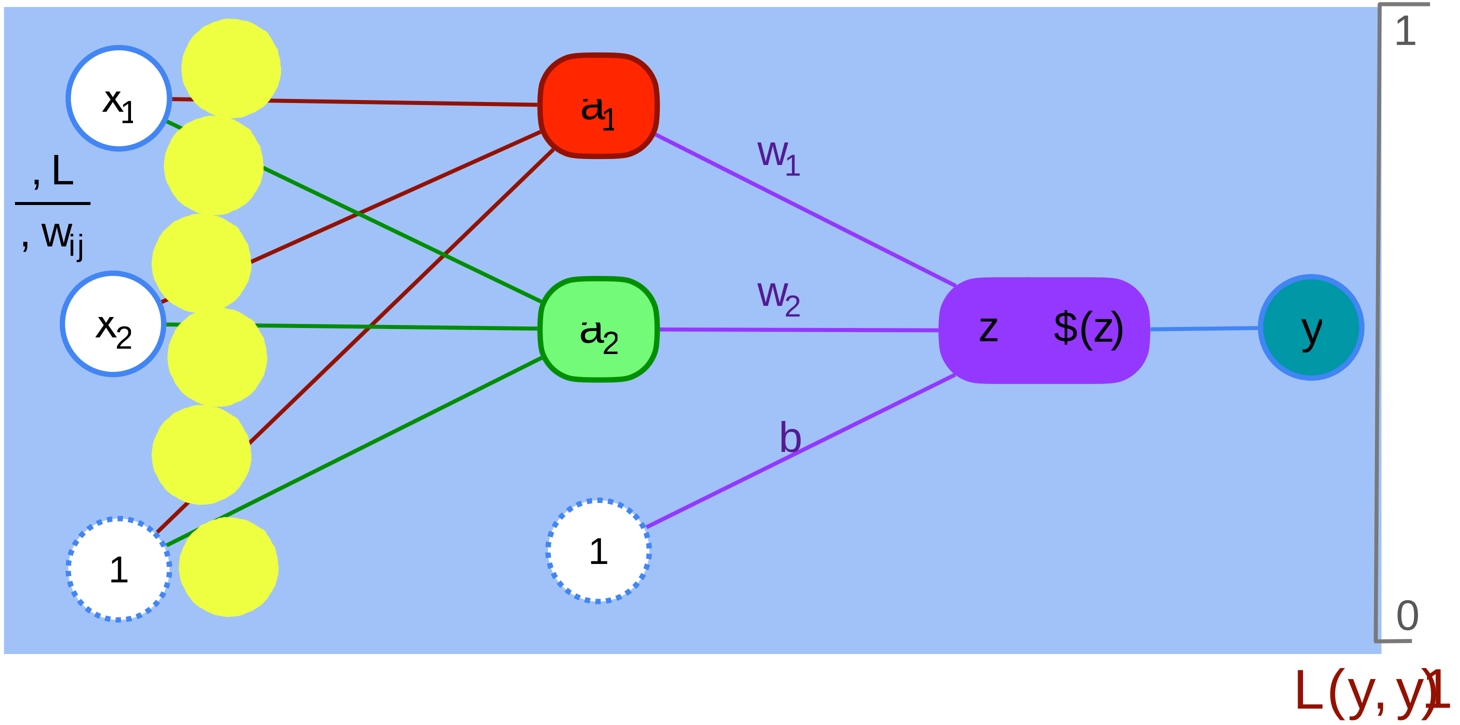
2,2,1 Neural Network



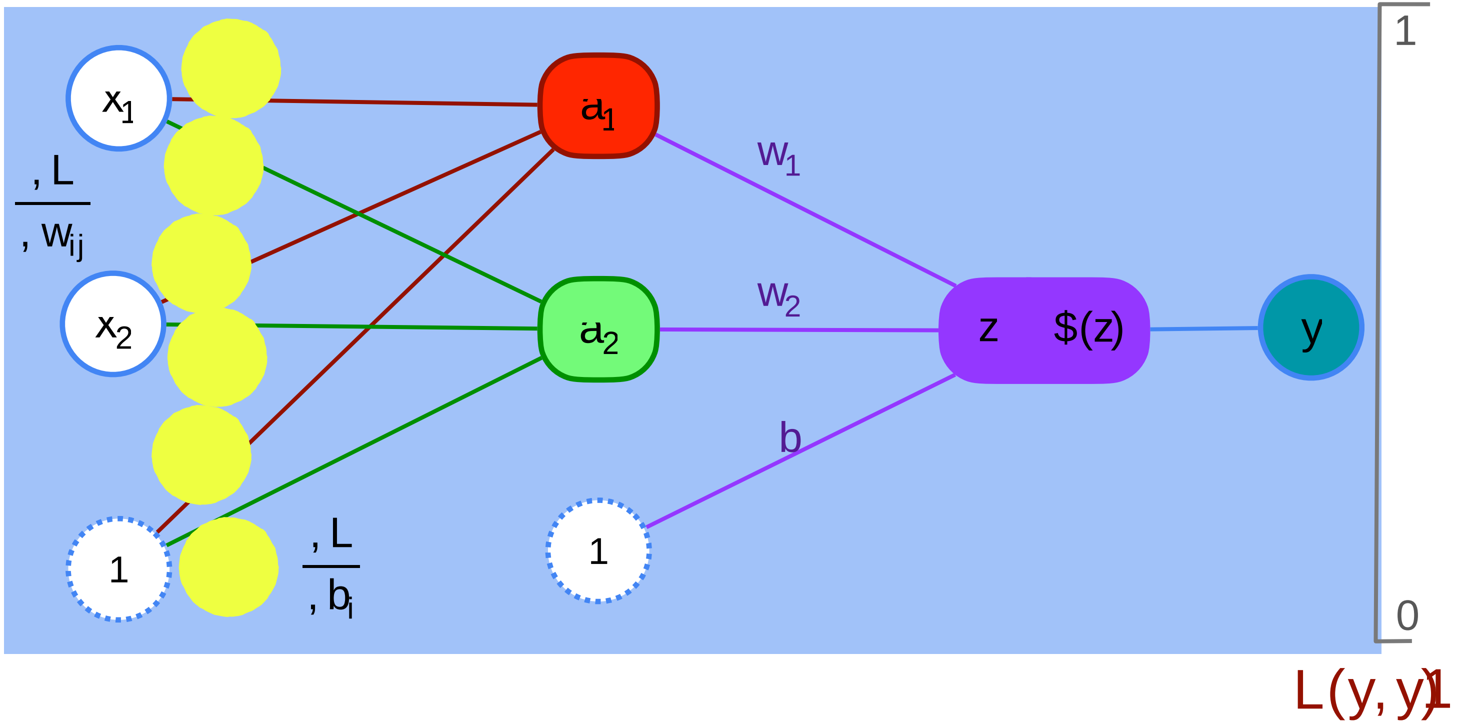
2,2,1 Neural Network



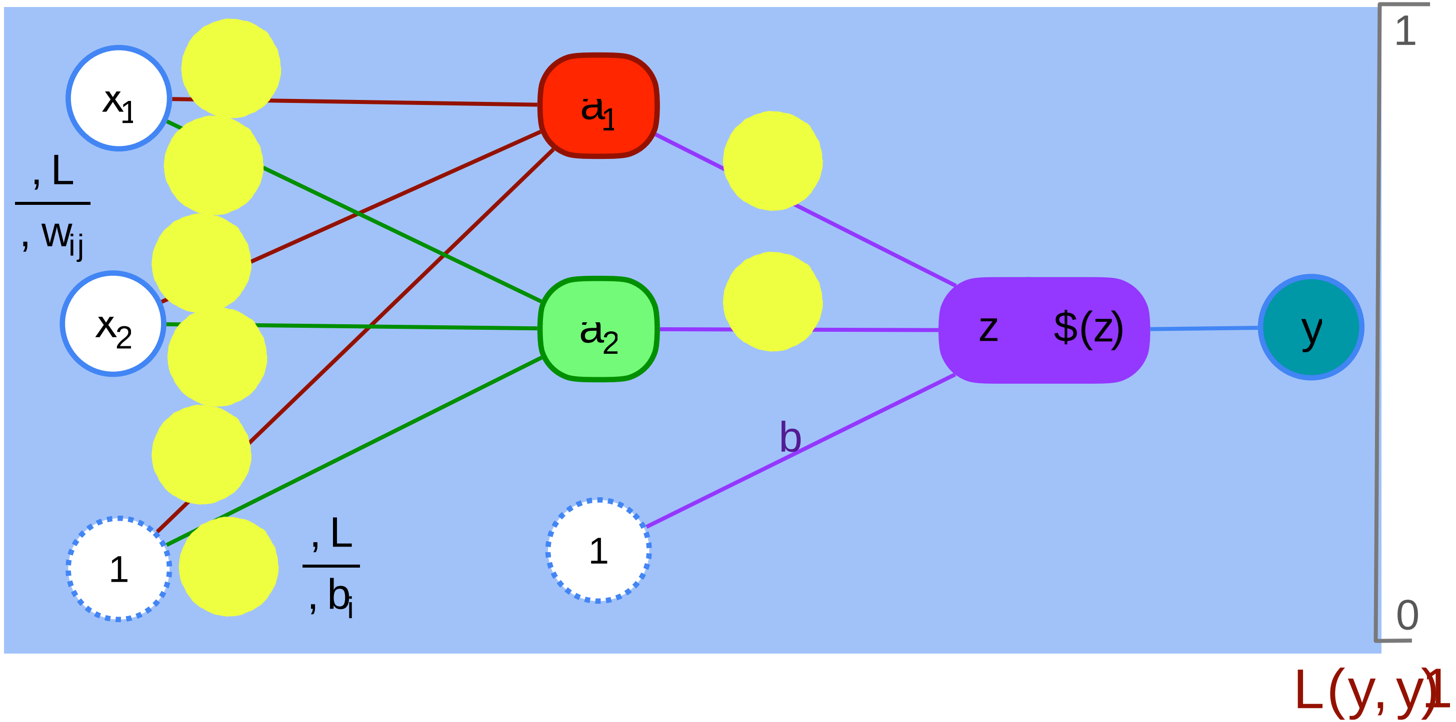
2,2,1 Neural Network



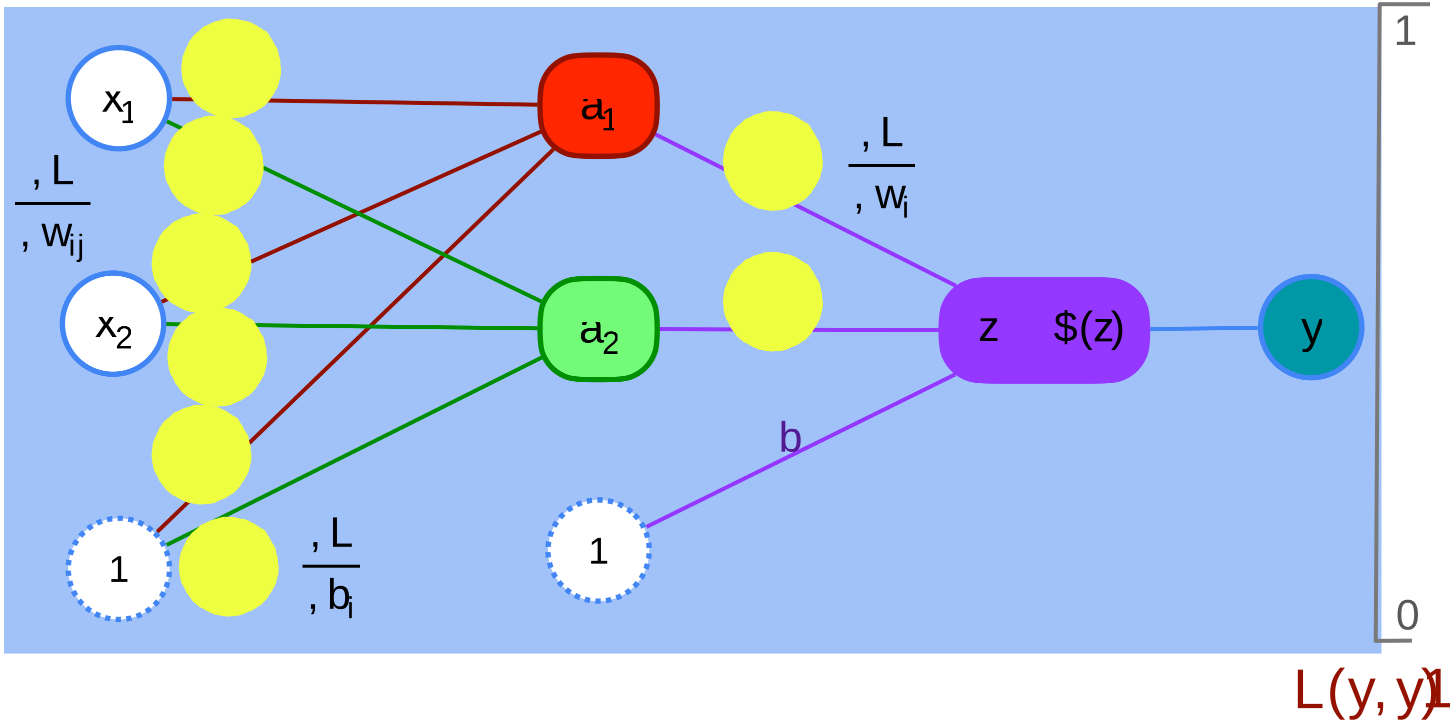
2,2,1 Neural Network



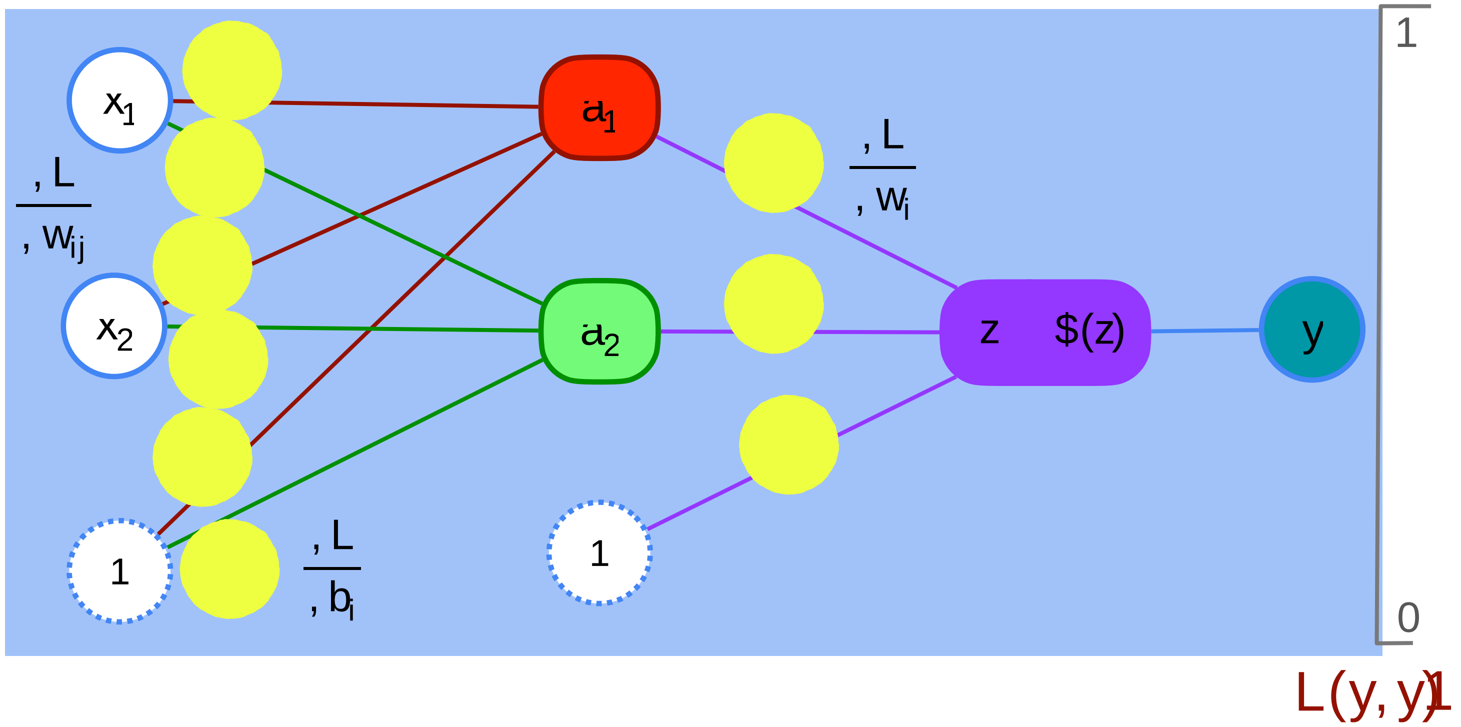
2,2,1 Neural Network



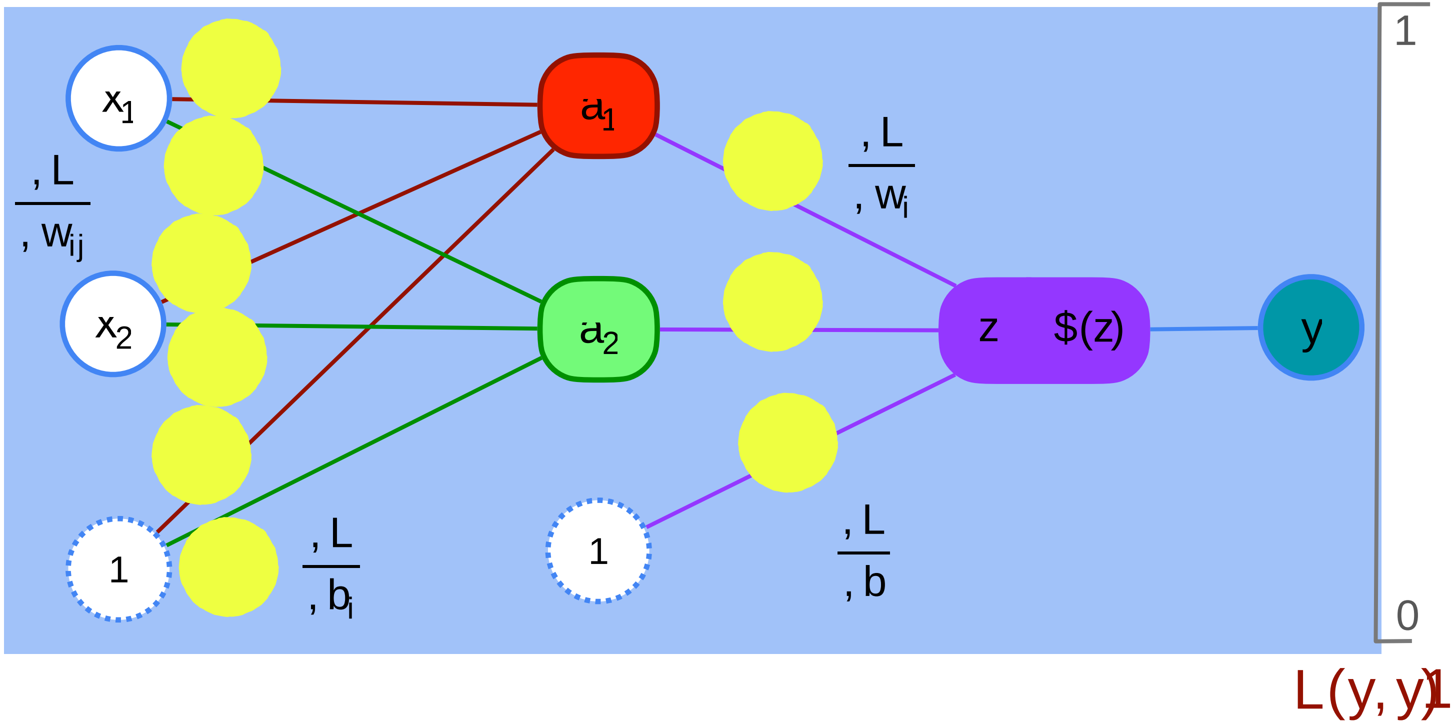
2,2,1 Neural Network



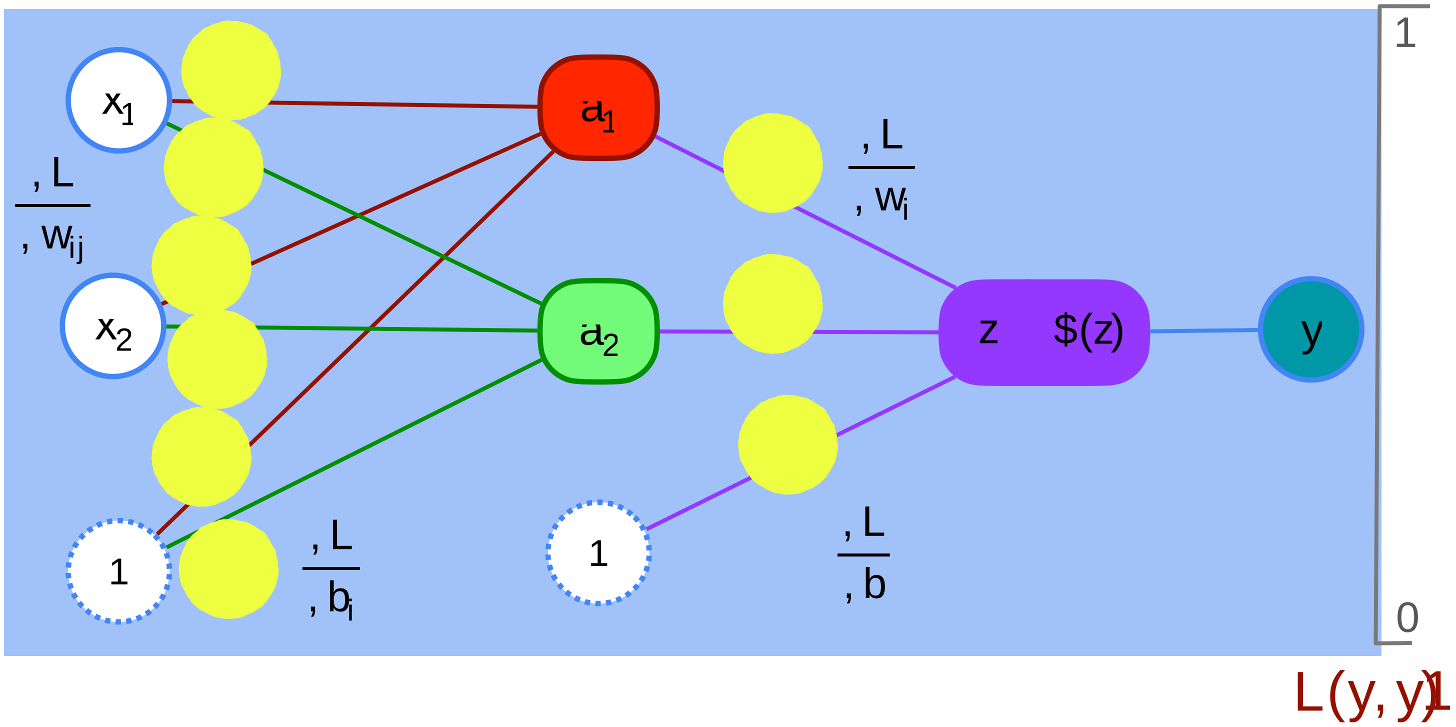
2,2,1 Neural Network



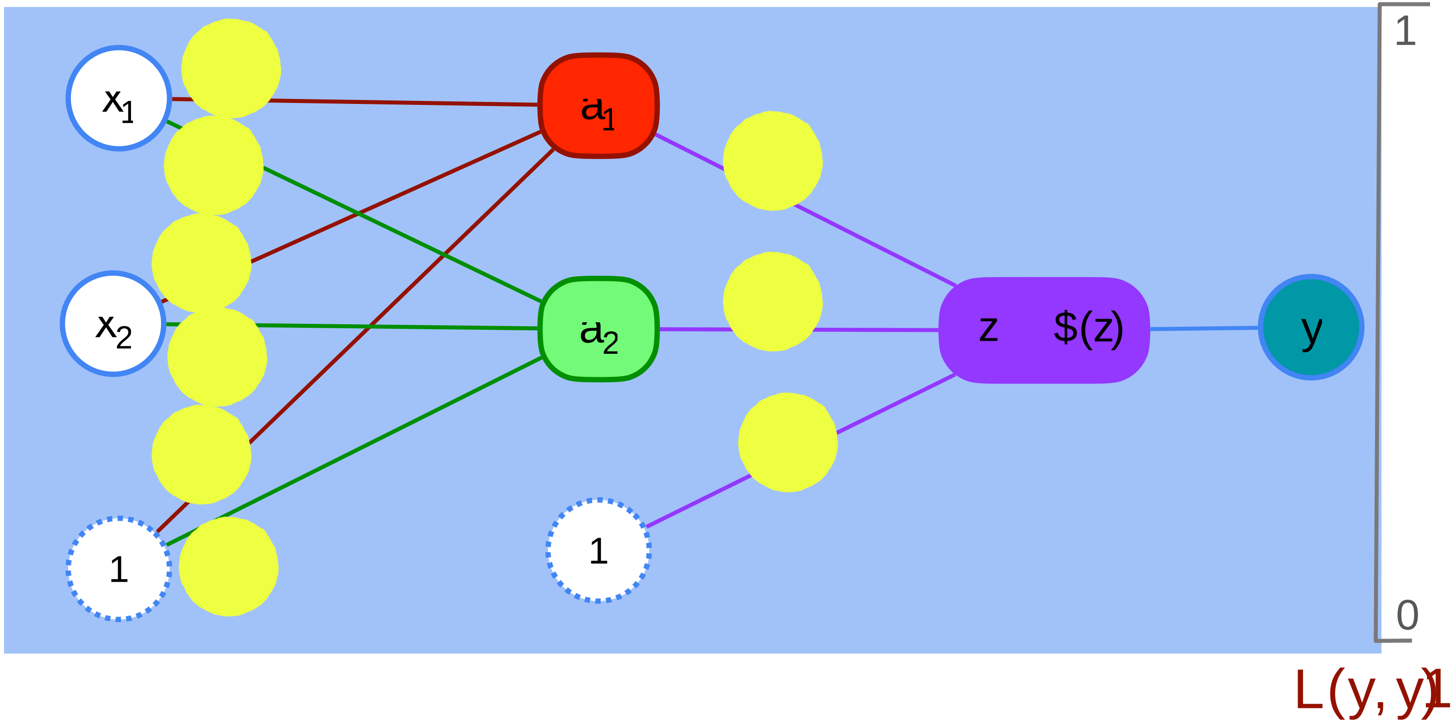
2,2,1 Neural Network



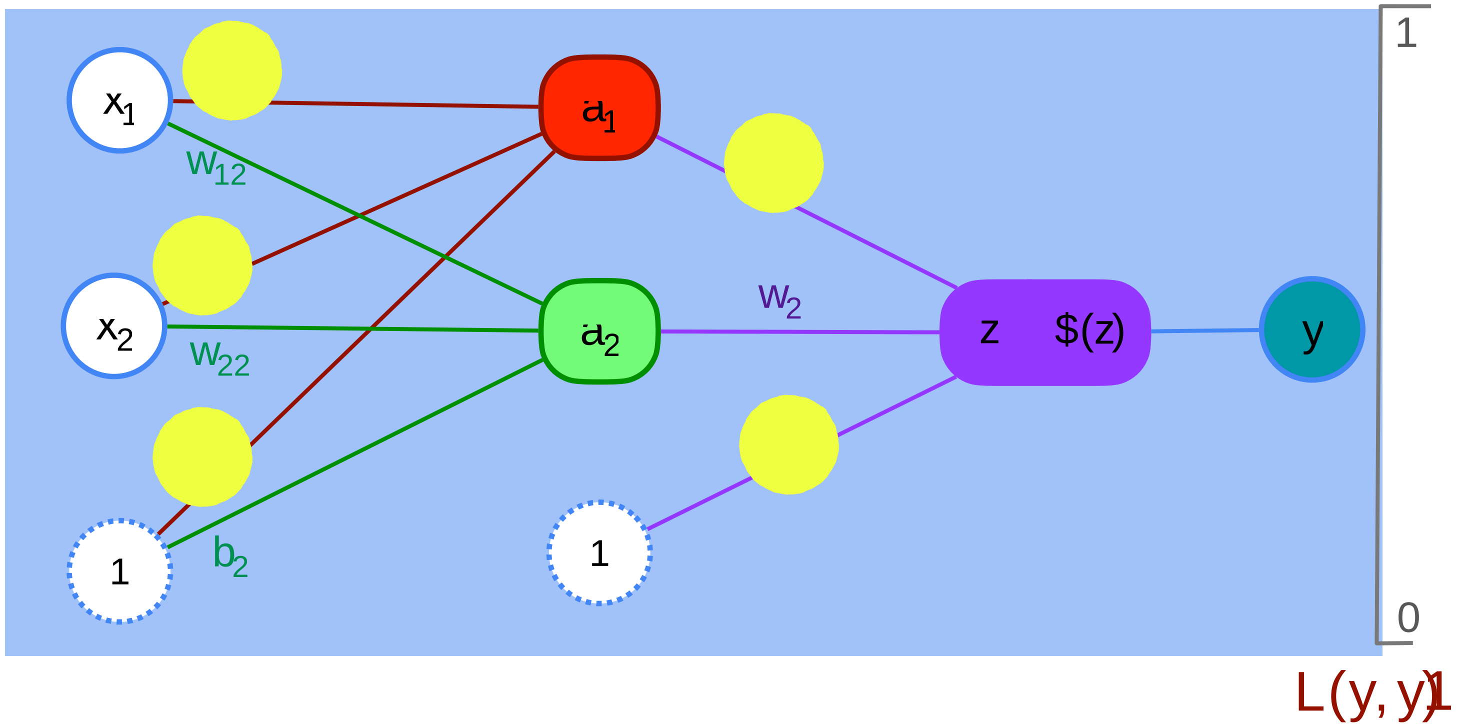
2,2,1 Neural Network



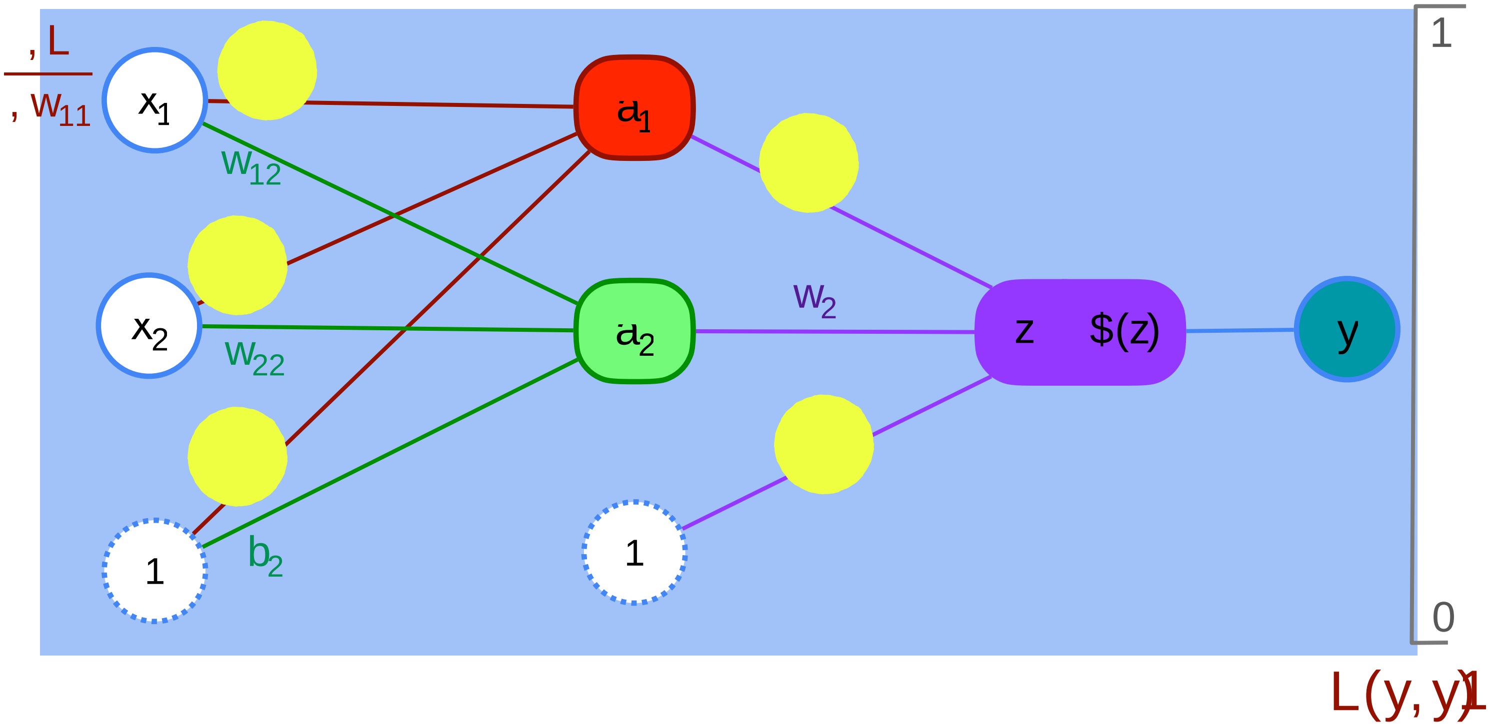
2,2,1 Neural Network



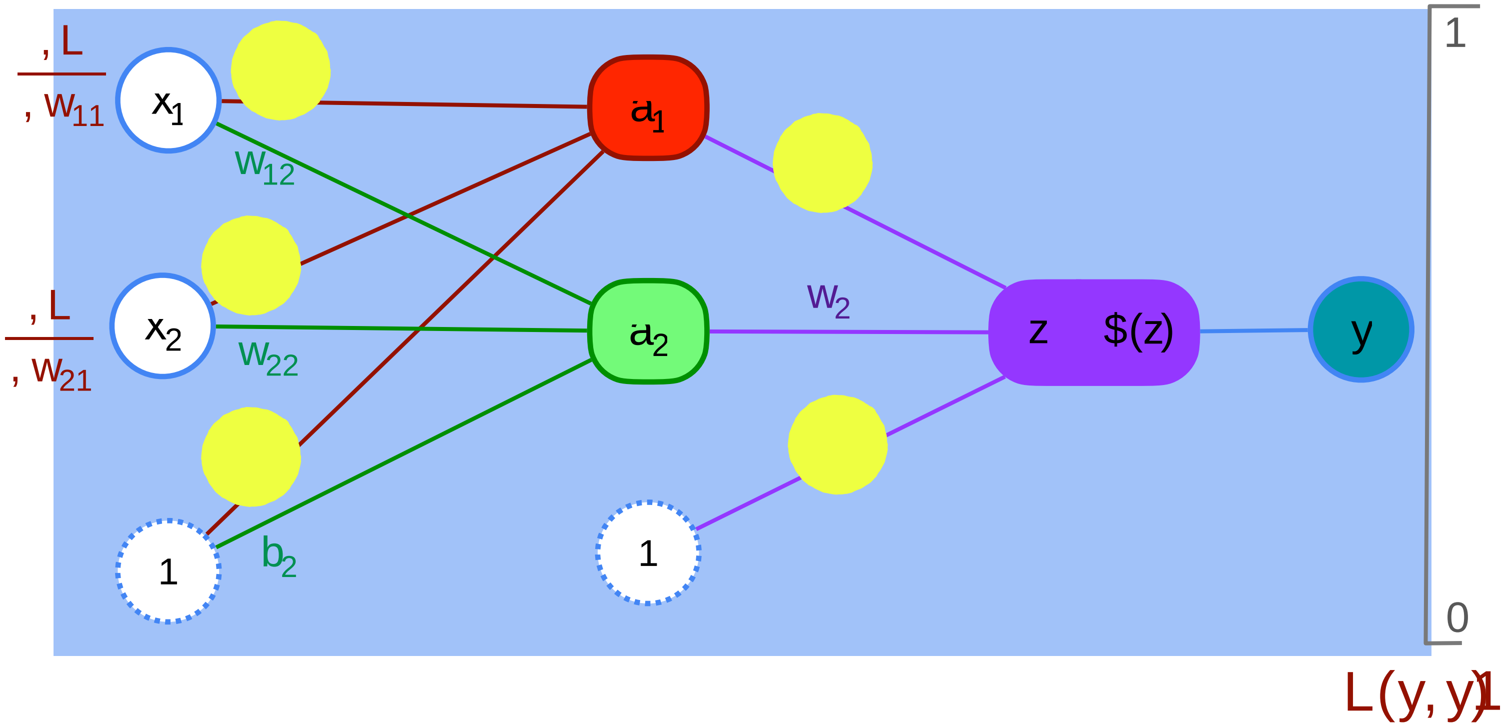
2,2,1 Neural Network



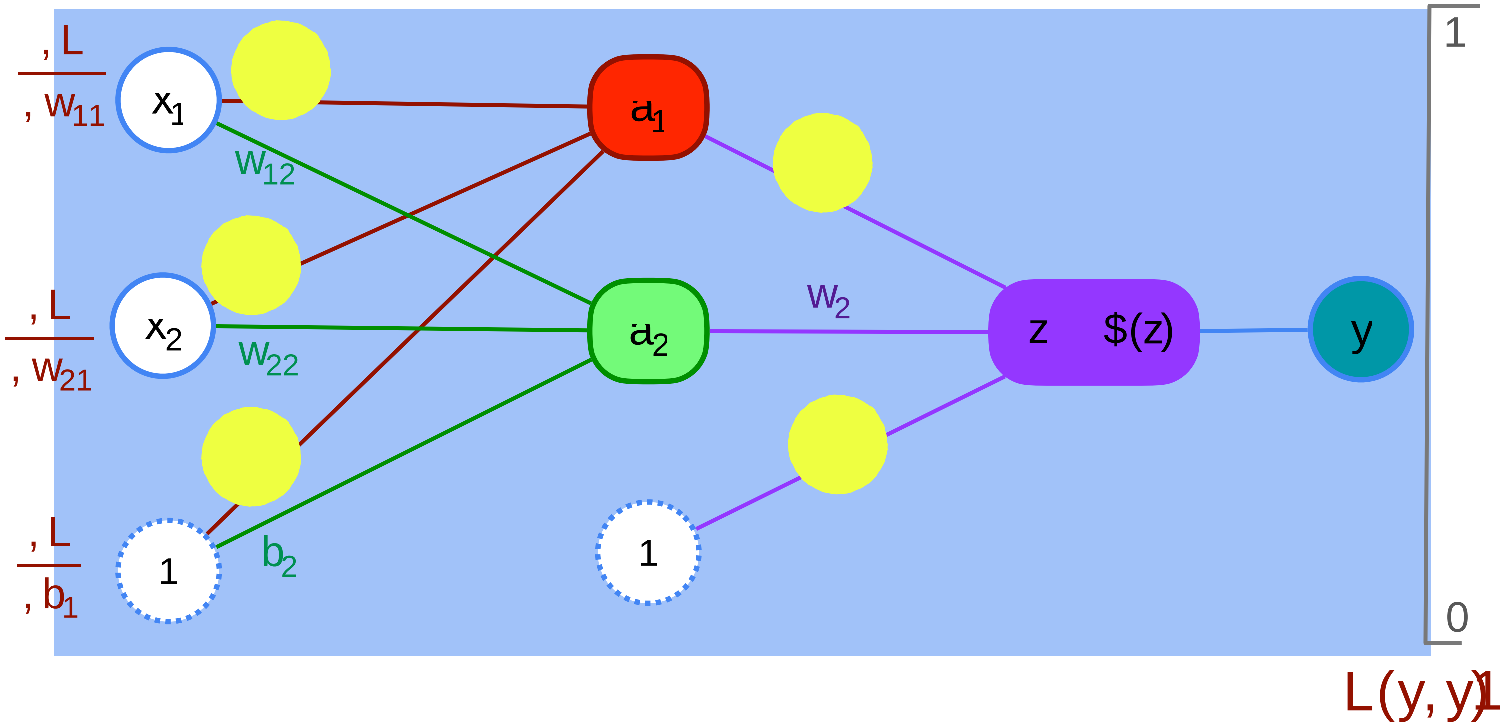
2,2,1 Neural Network



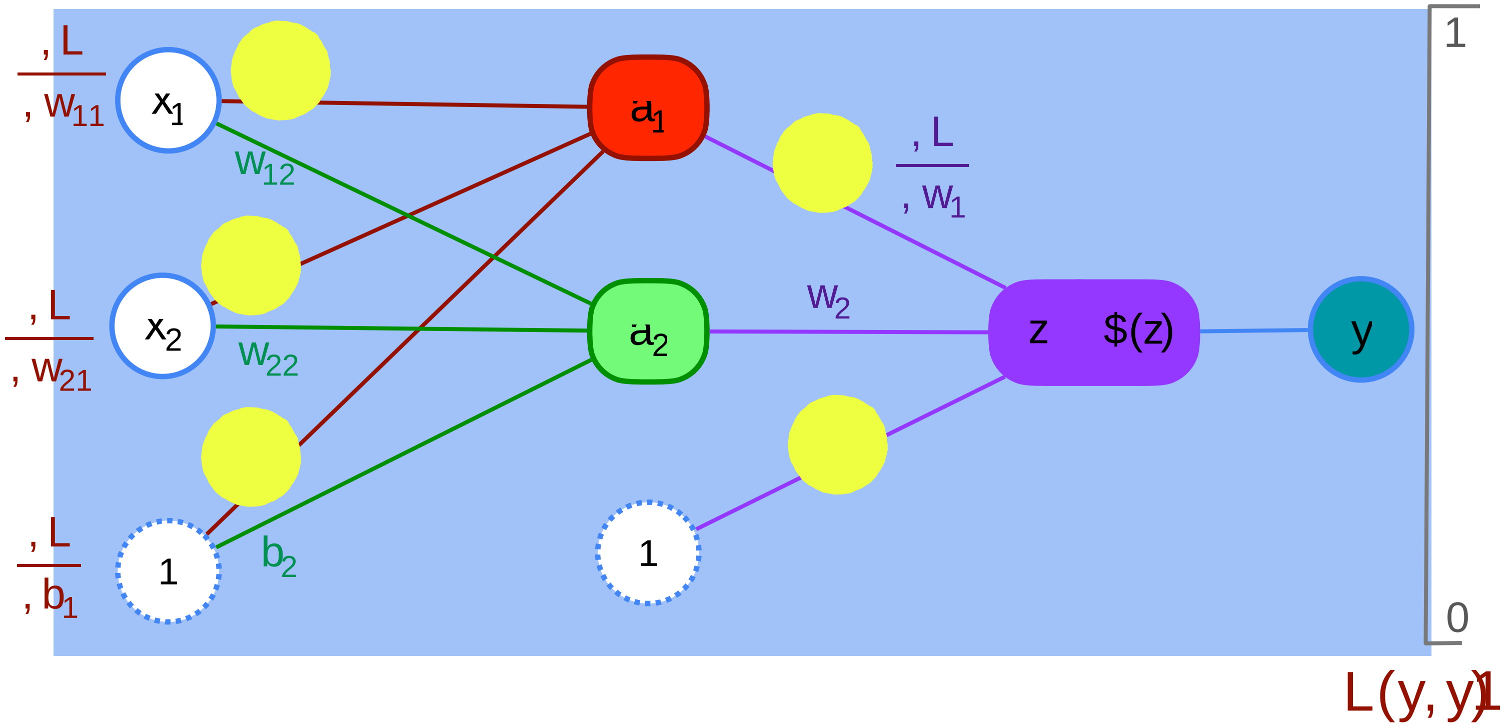
2,2,1 Neural Network



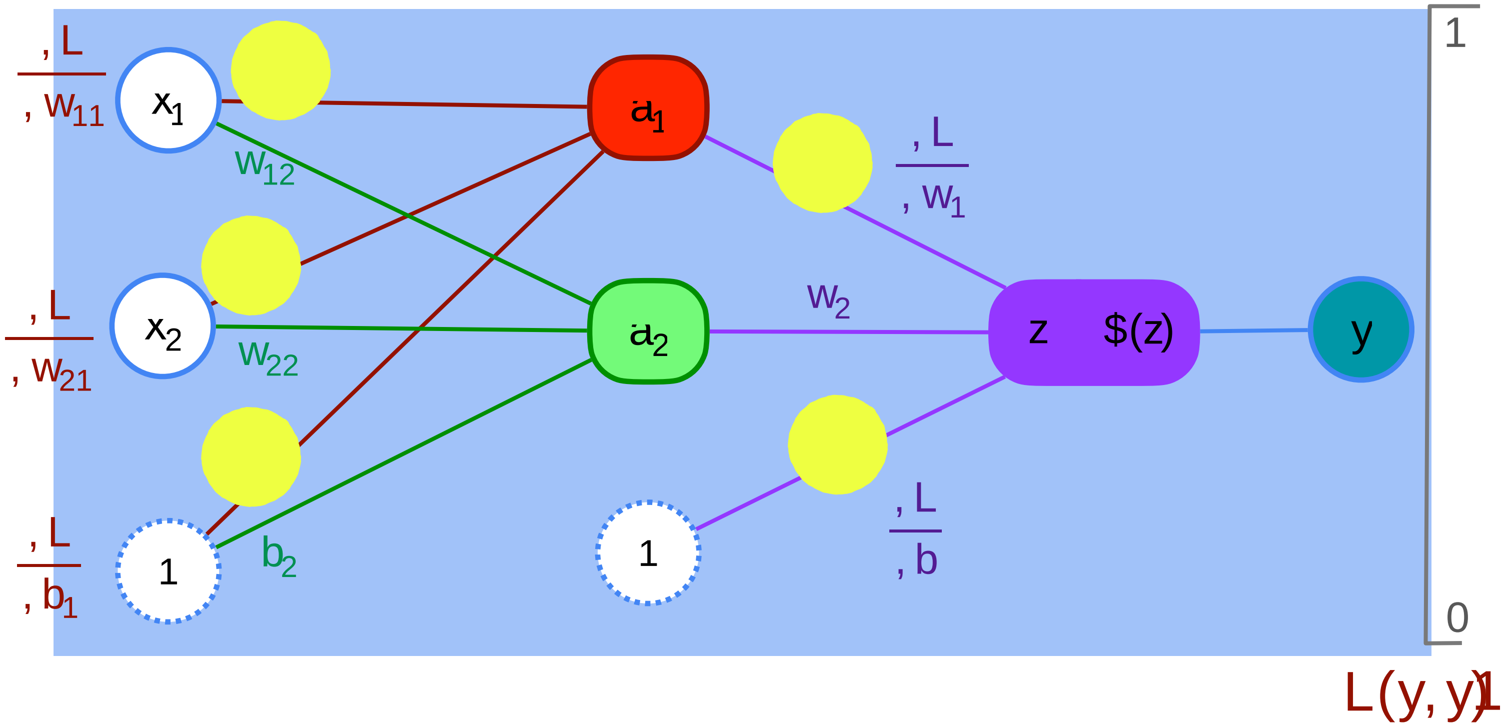
2,2,1 Neural Network



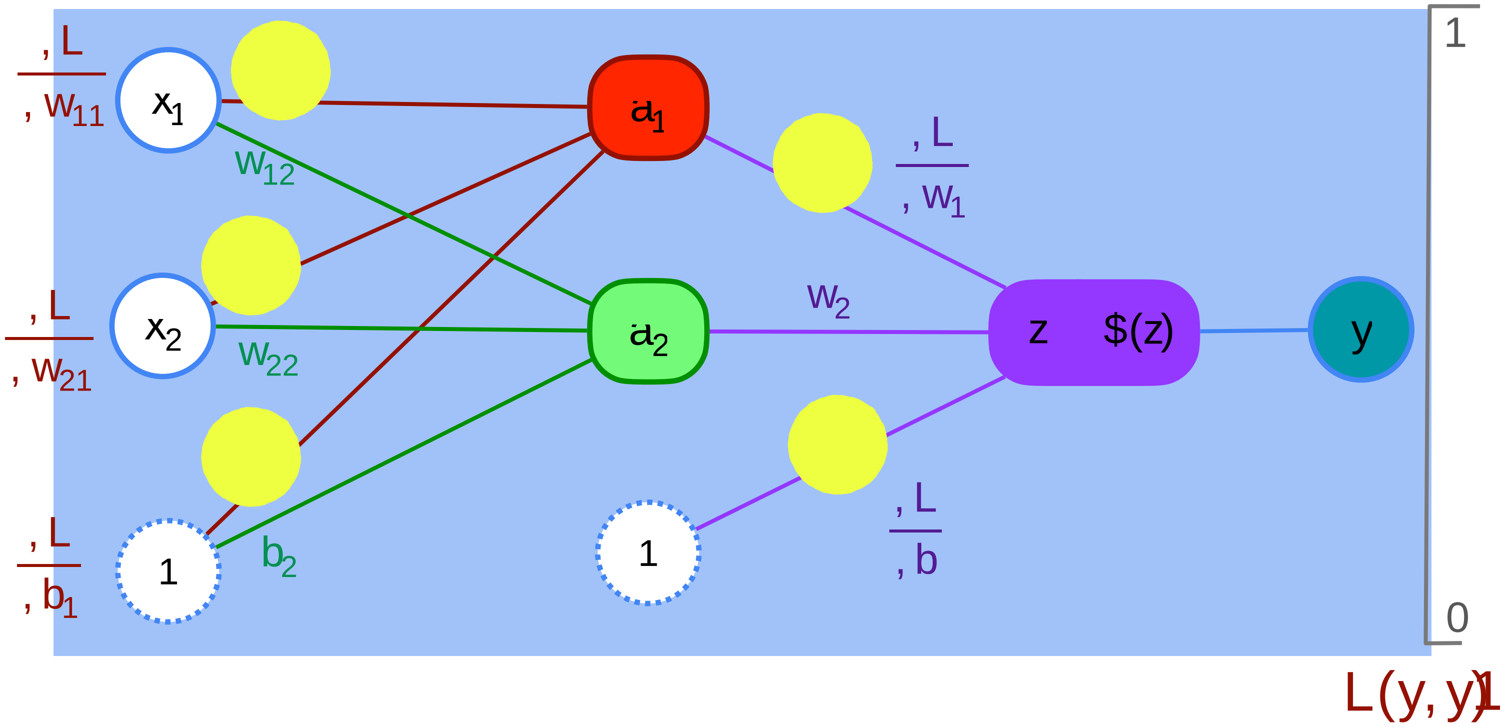
2,2,1 Neural Network



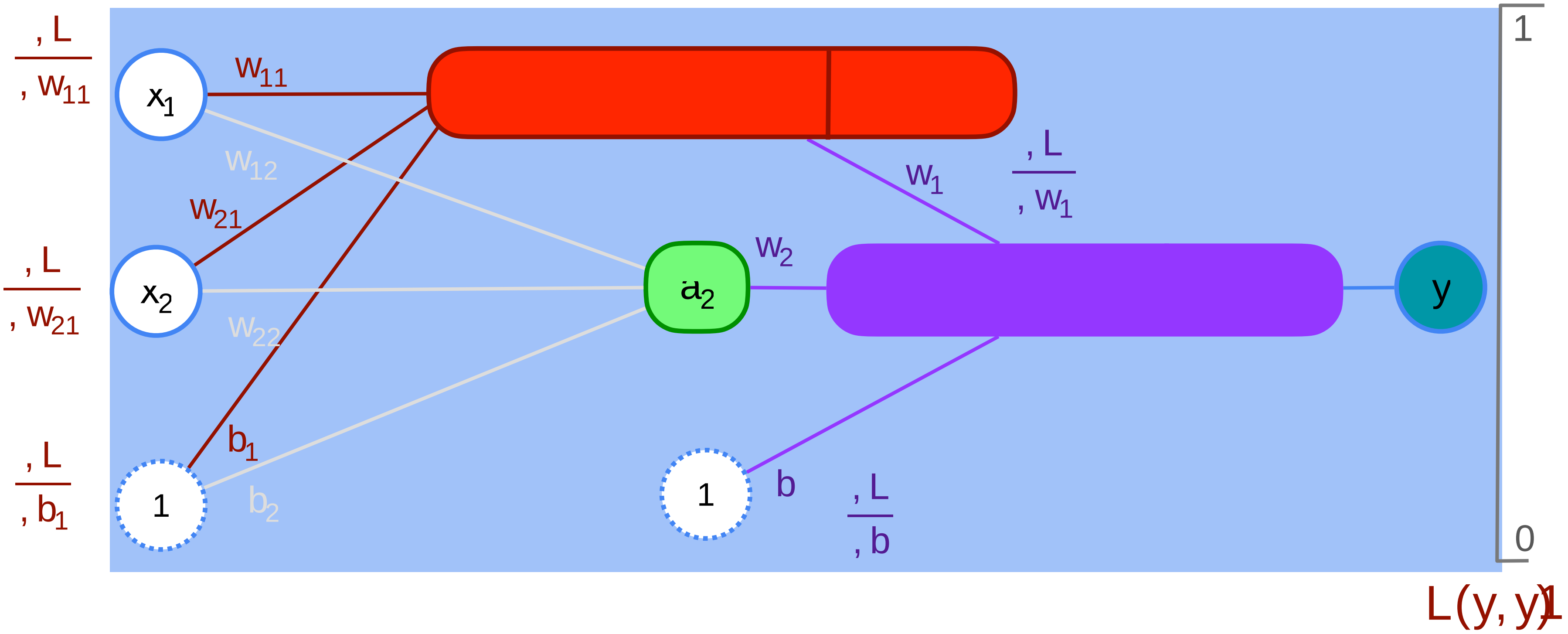
2,2,1 Neural Network



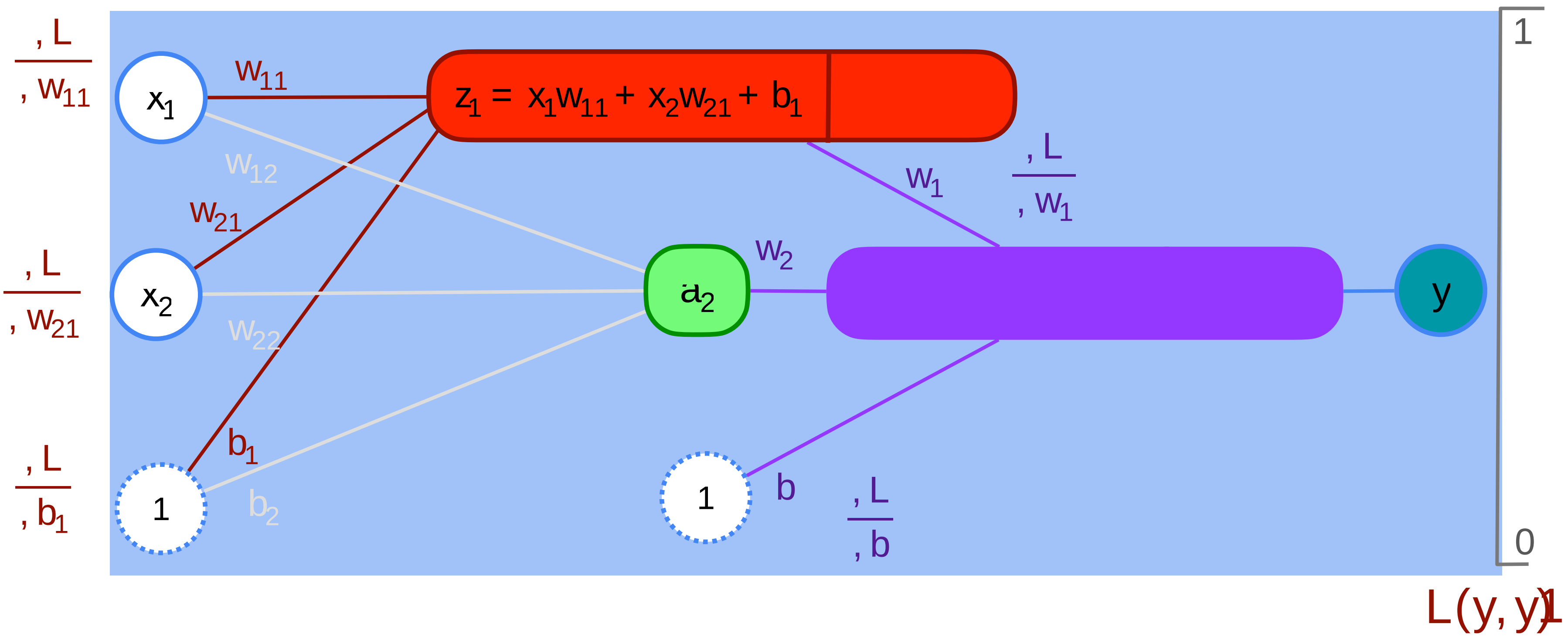
2,2,1 Neural Network



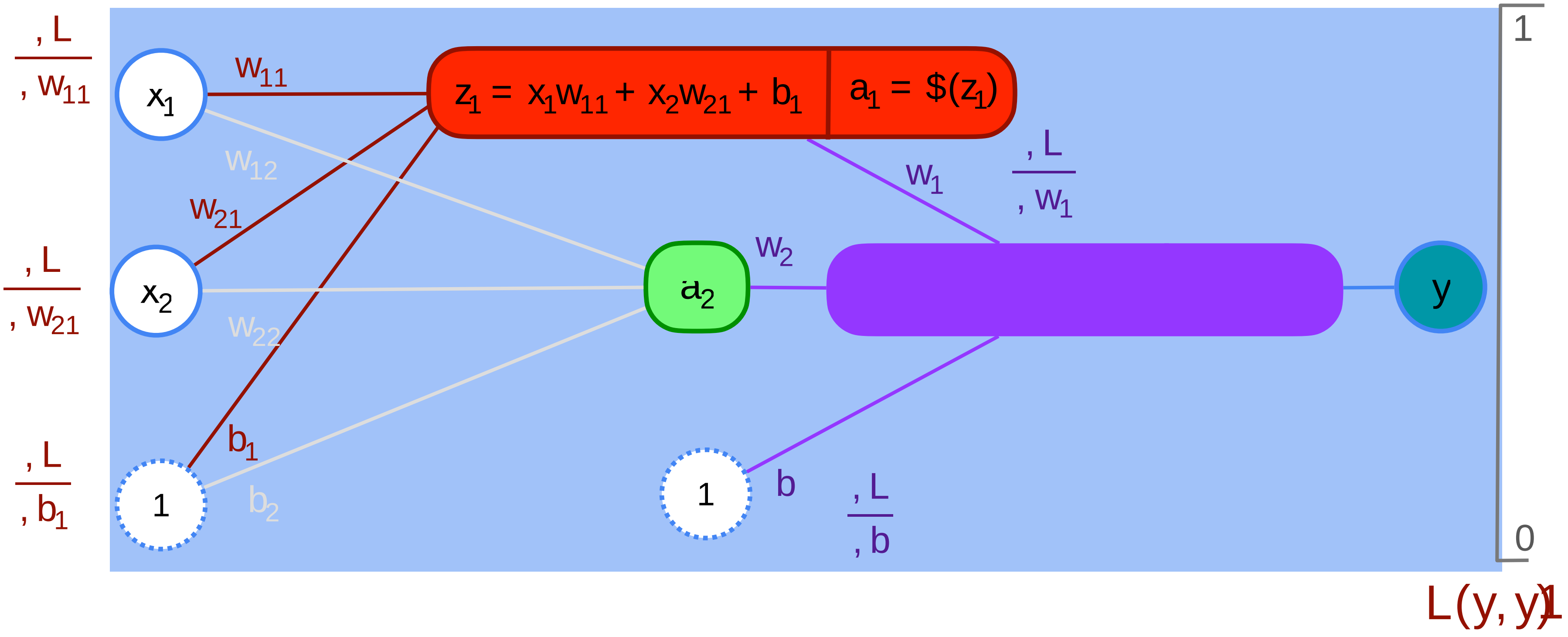
2,2,1 Neural Network



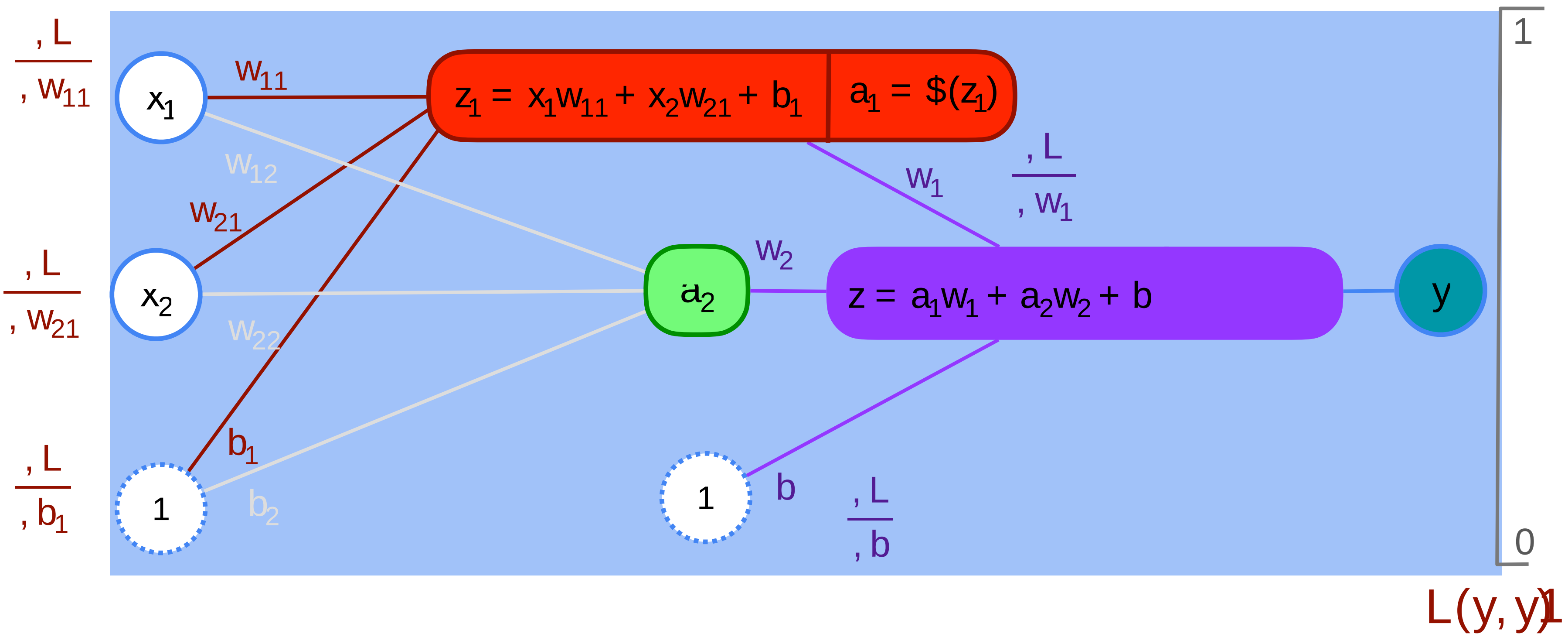
2,2,1 Neural Network



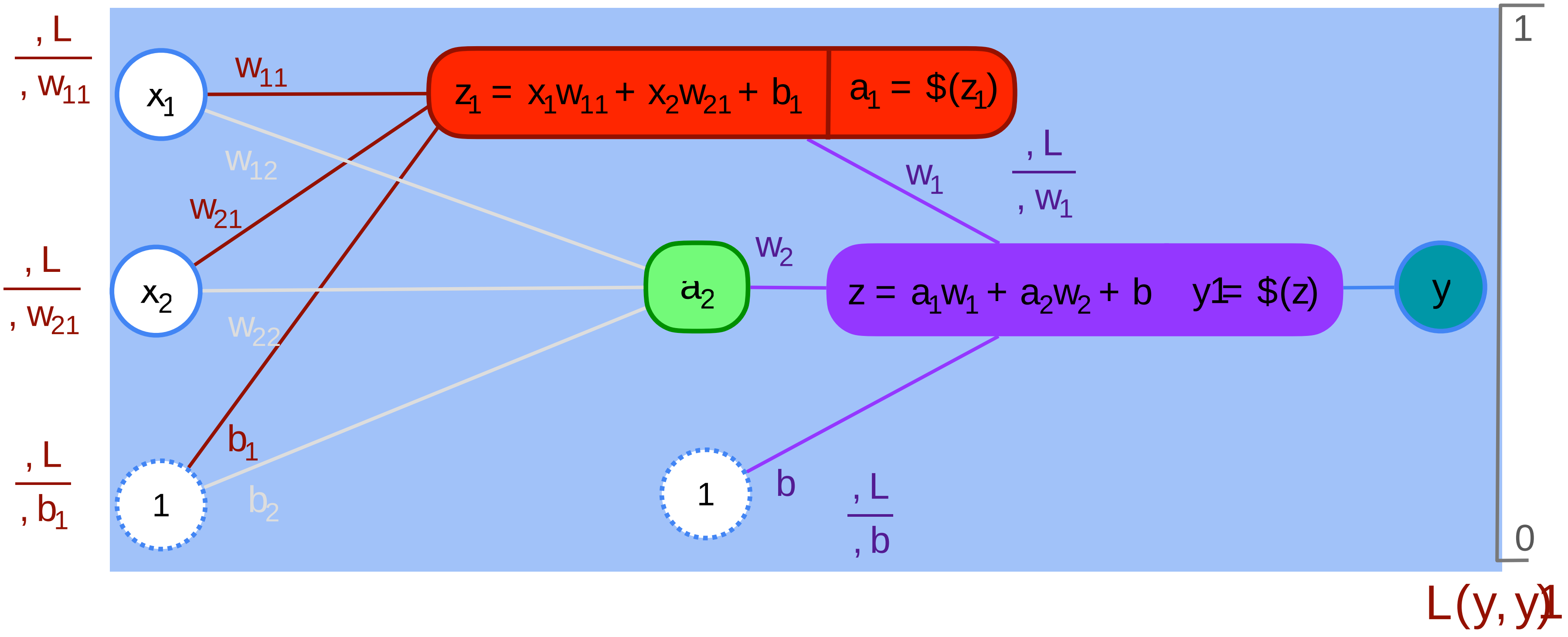
2,2,1 Neural Network



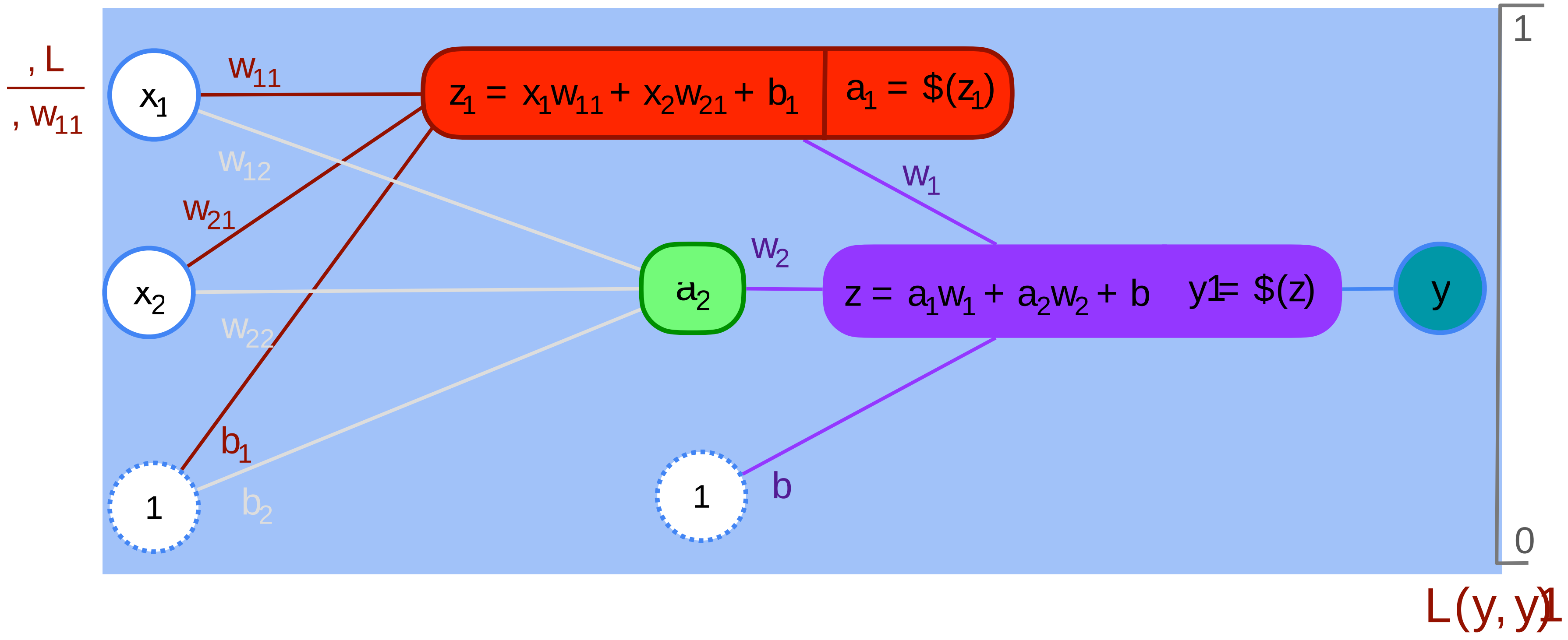
2,2,1 Neural Network



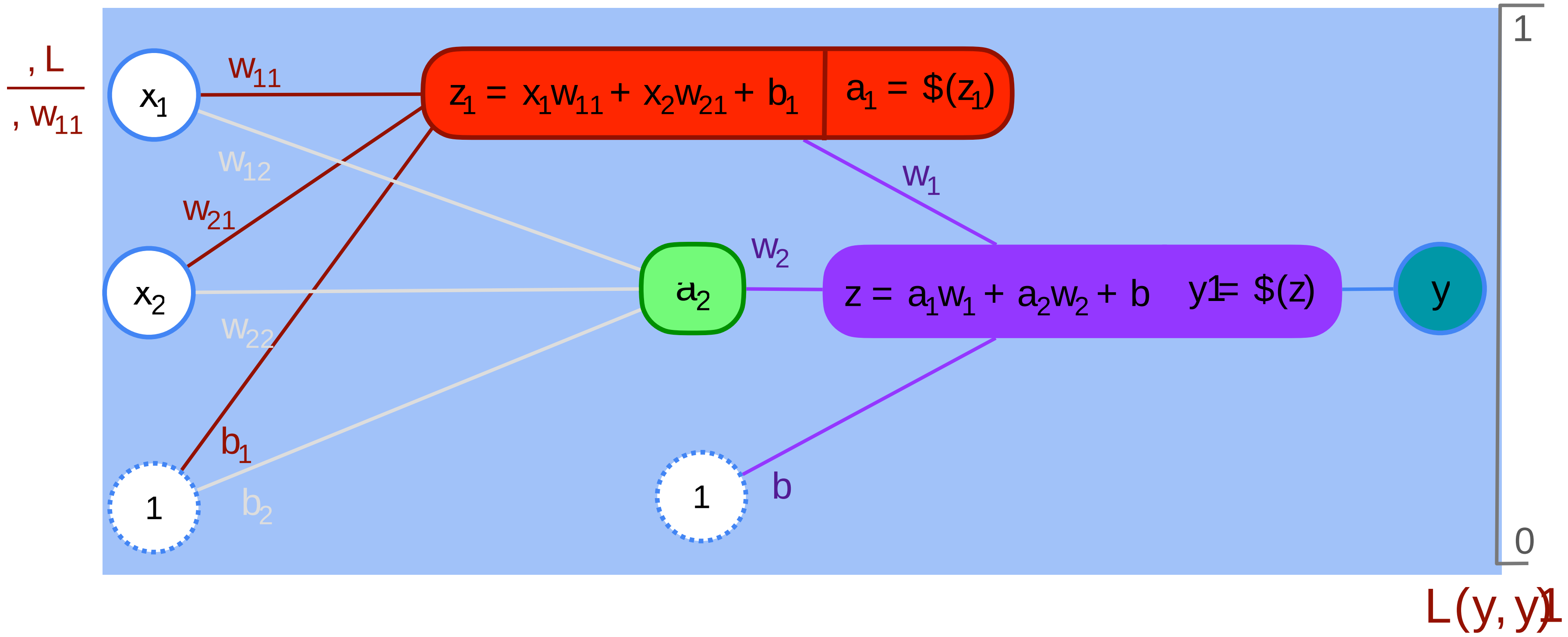
2,2,1 Neural Network



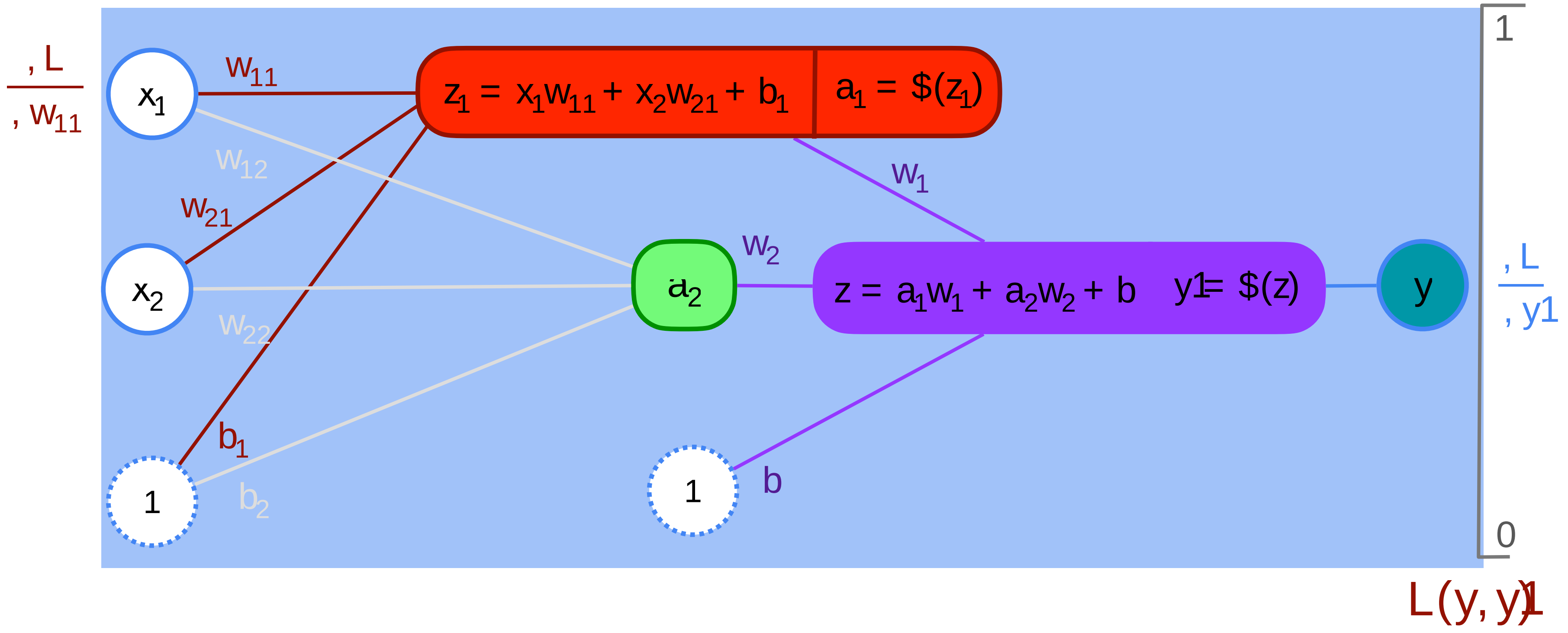
2,2,1 Neural Network



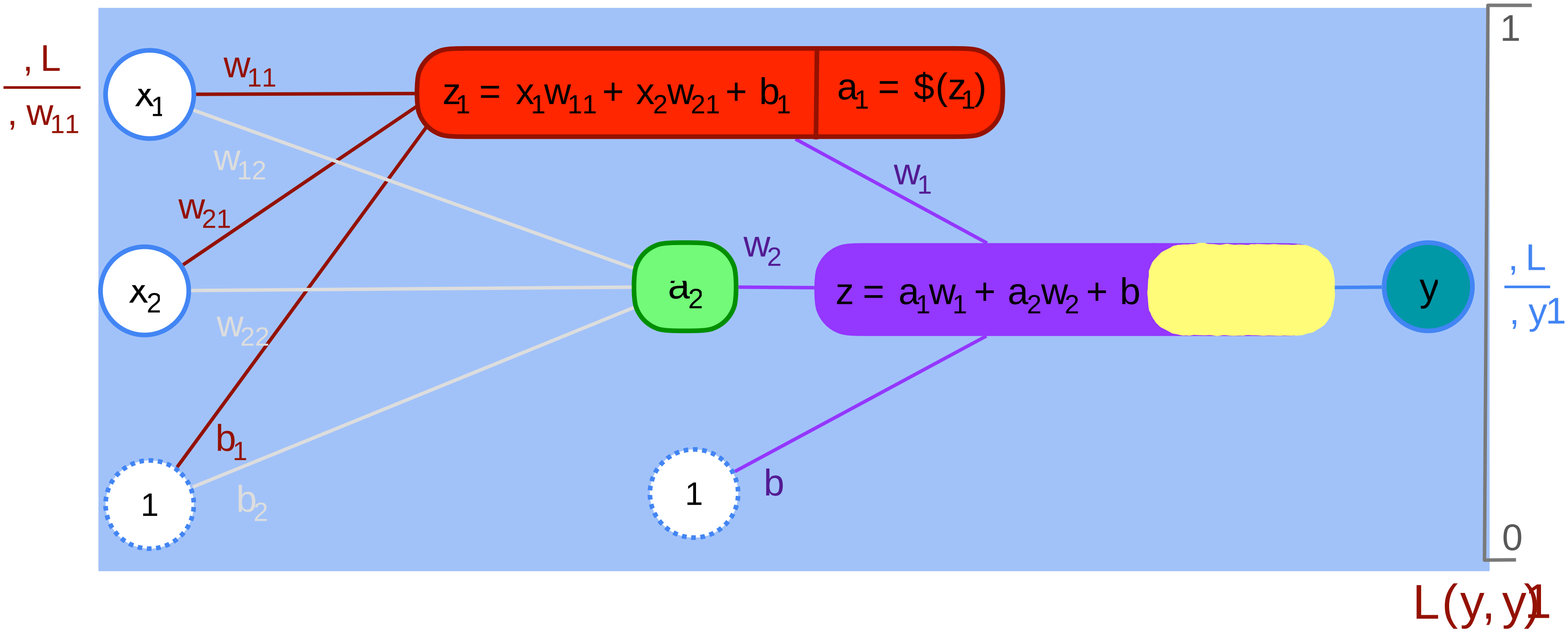
2,2,1 Neural Network



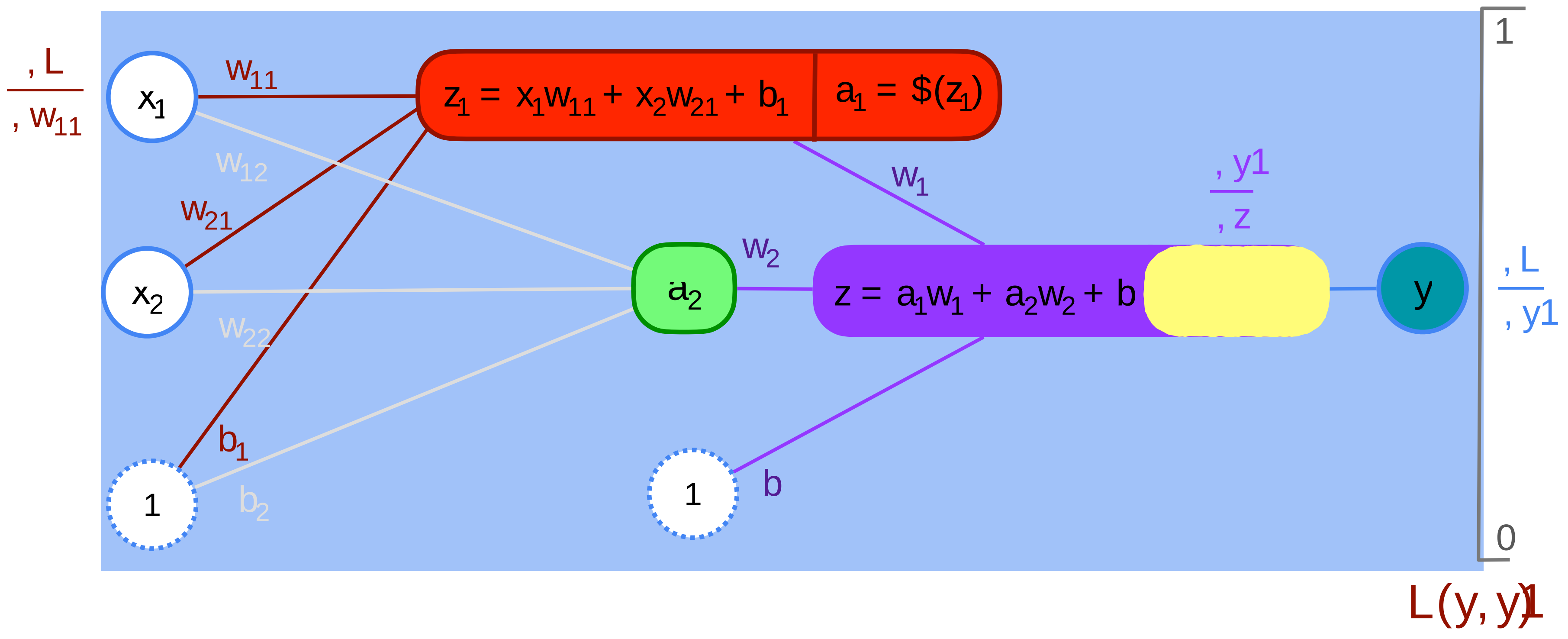
2,2,1 Neural Network



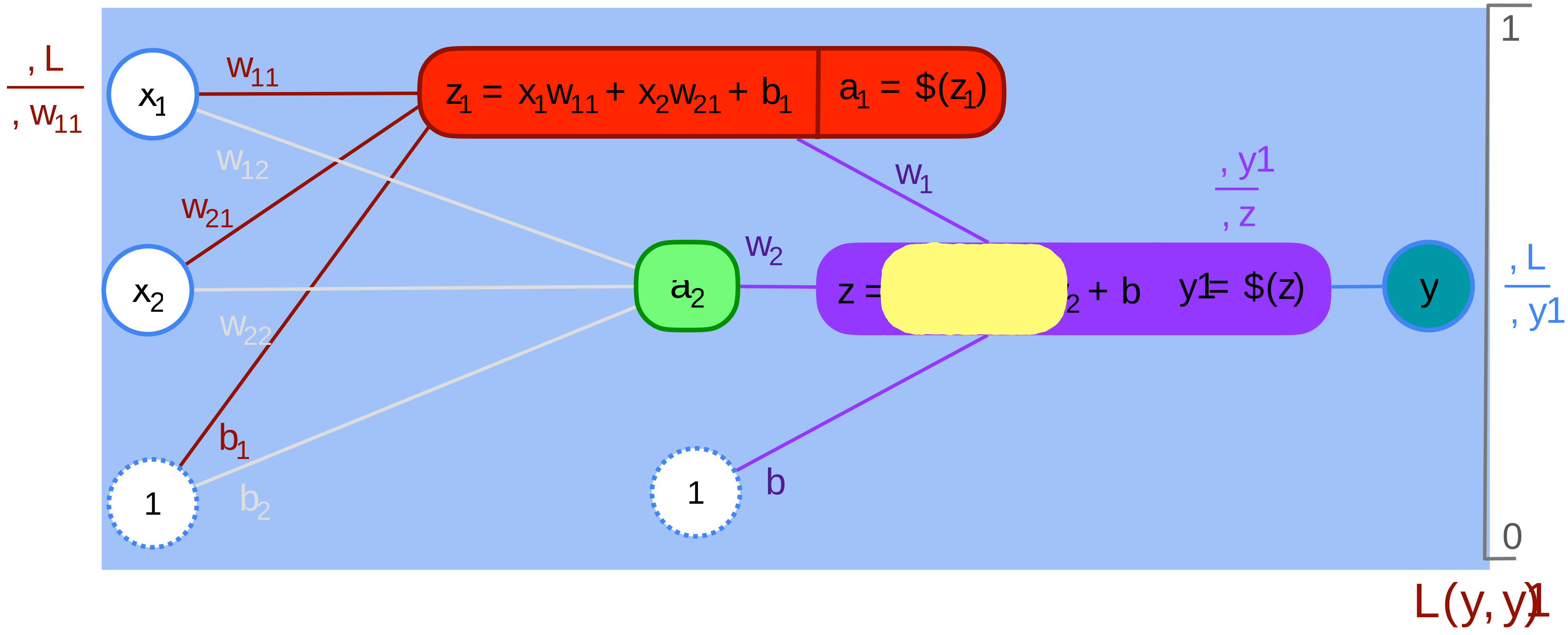
2,2,1 Neural Network



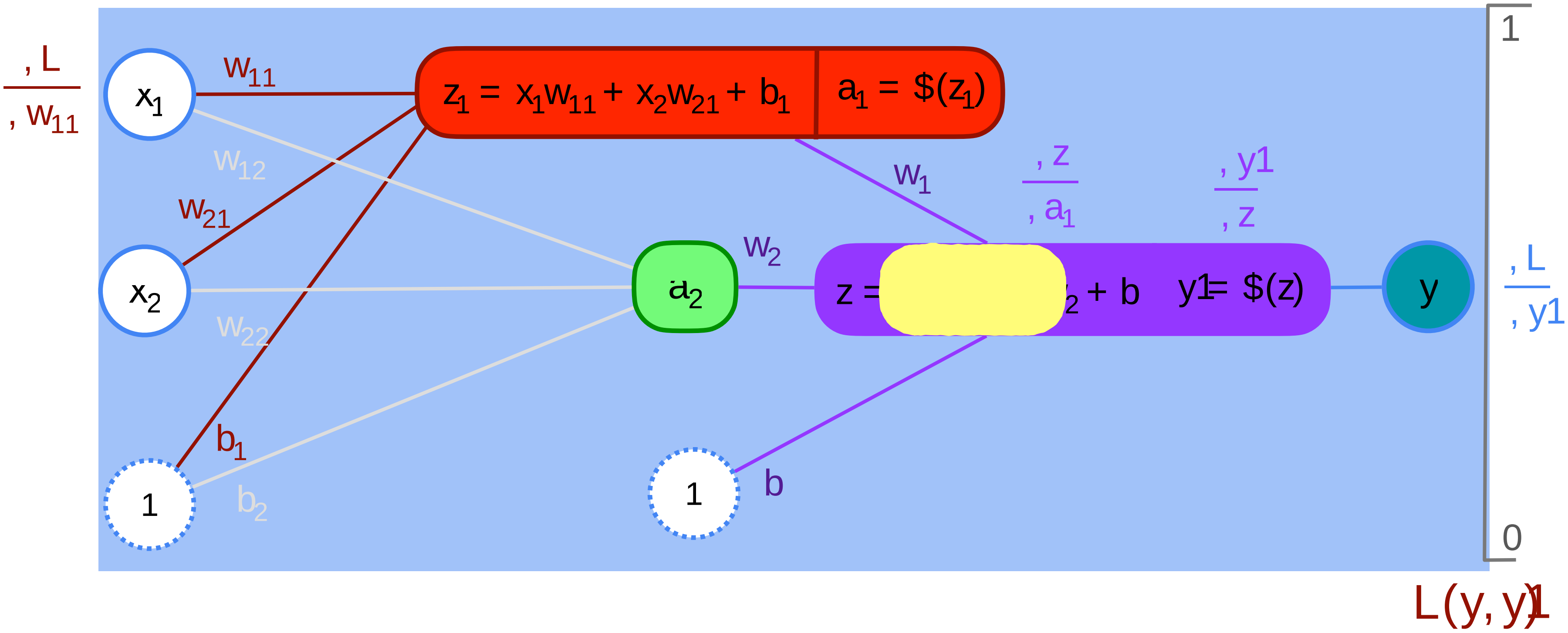
2,2,1 Neural Network



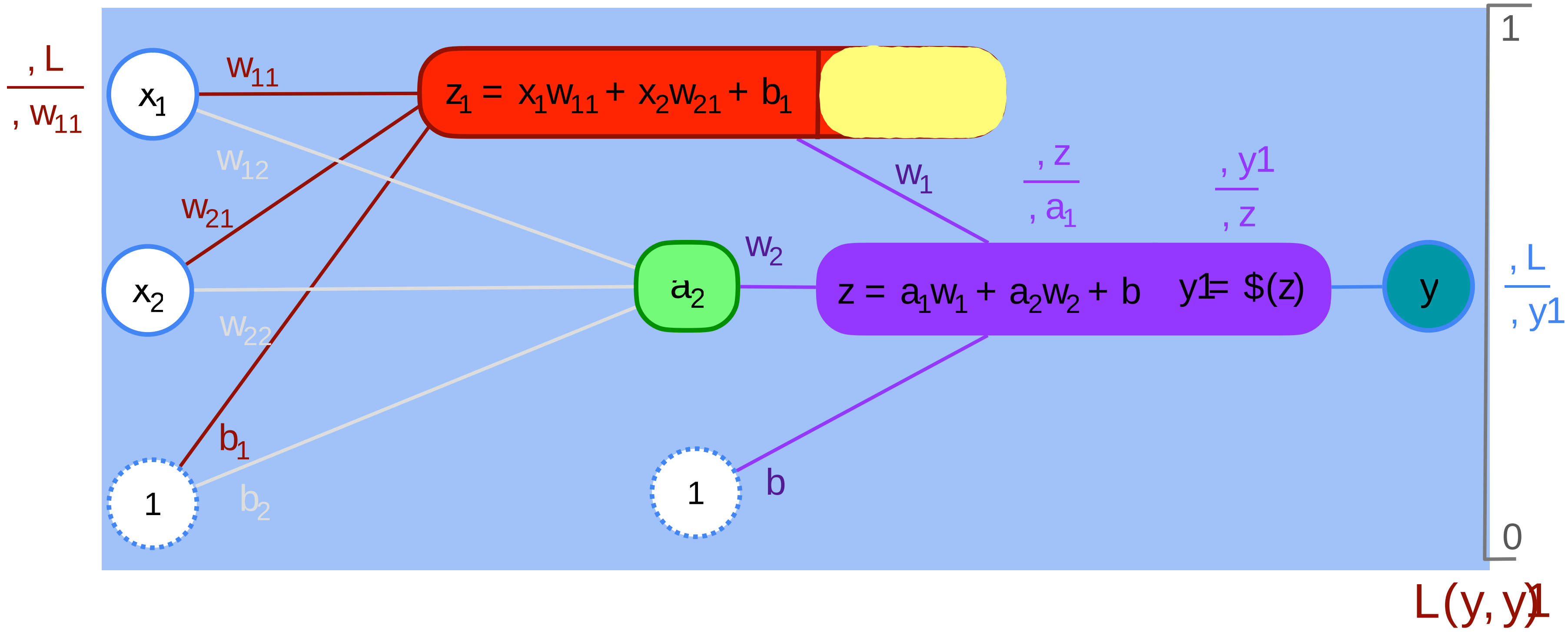
2,2,1 Neural Network



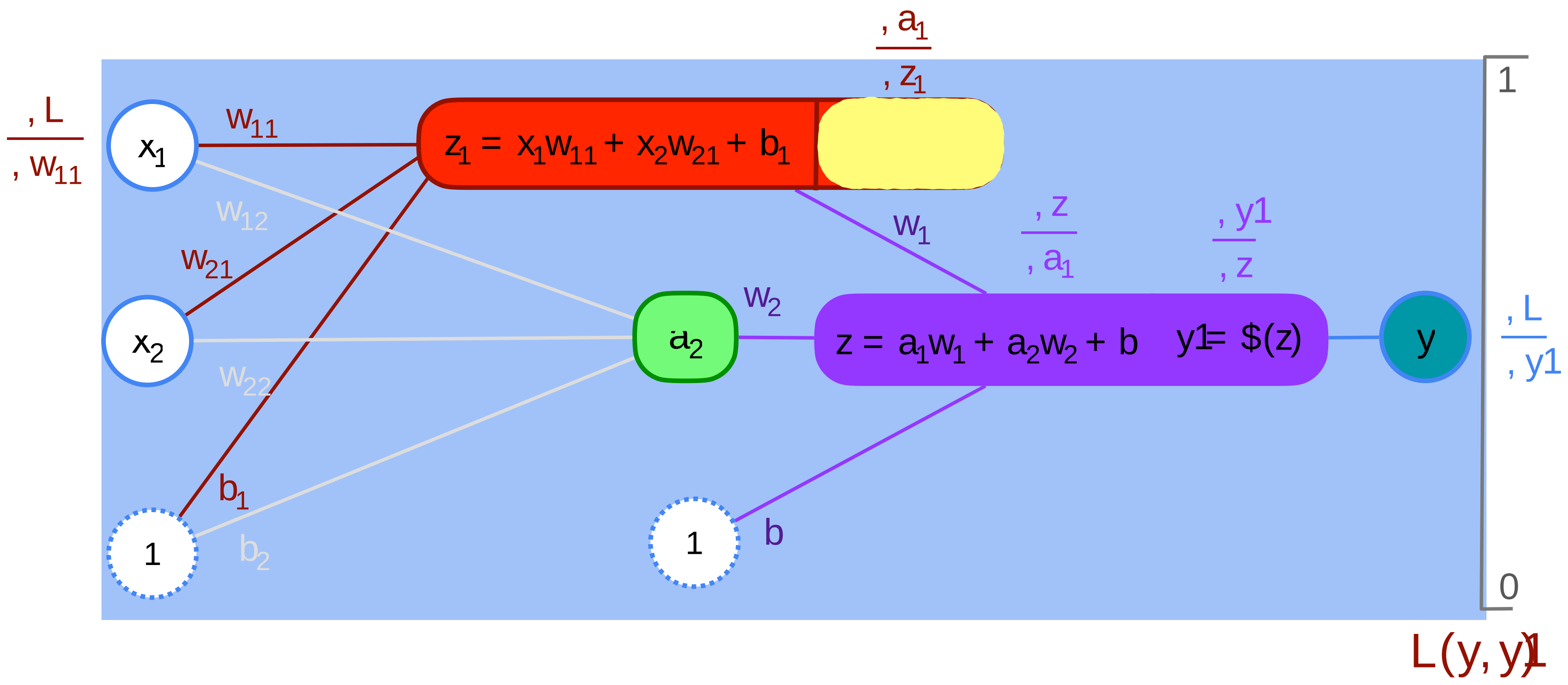
2,2,1 Neural Network



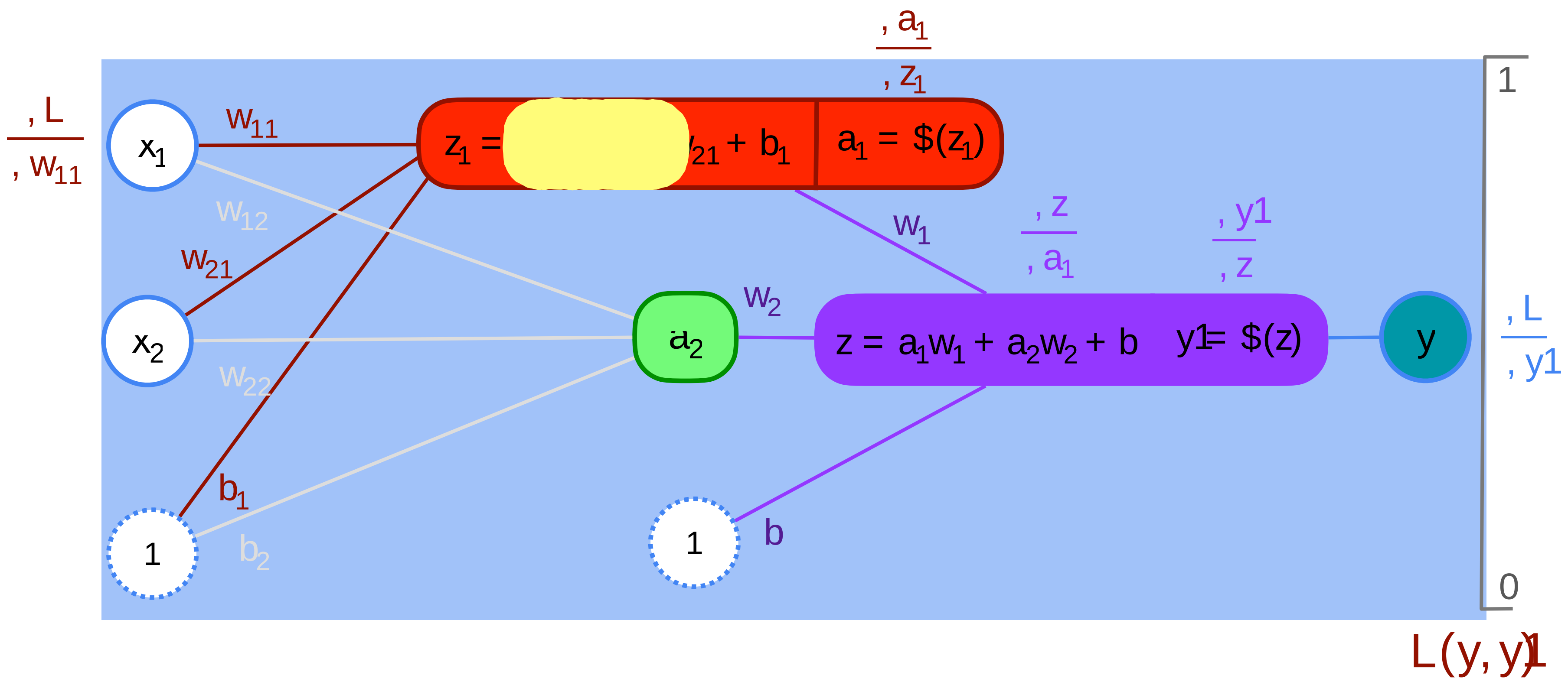
2,2,1 Neural Network



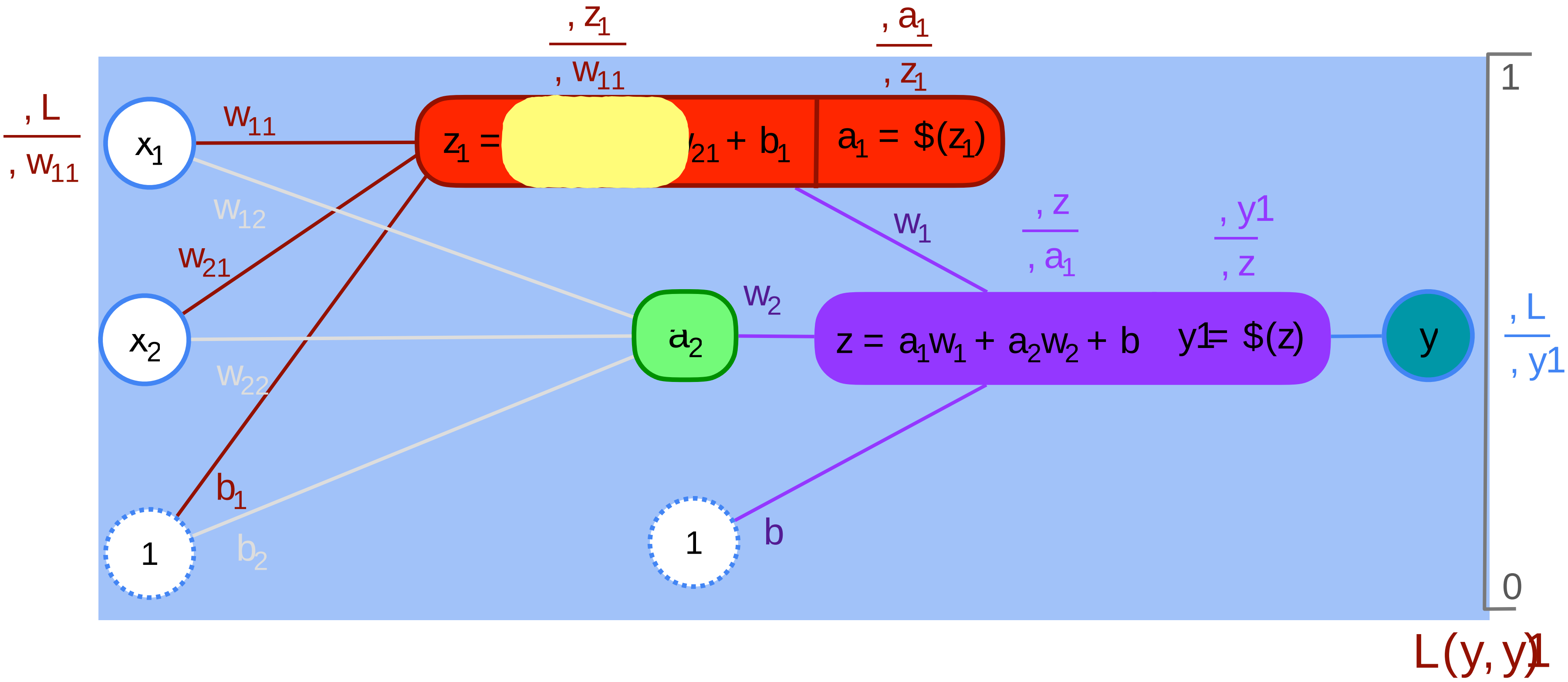
2,2,1 Neural Network



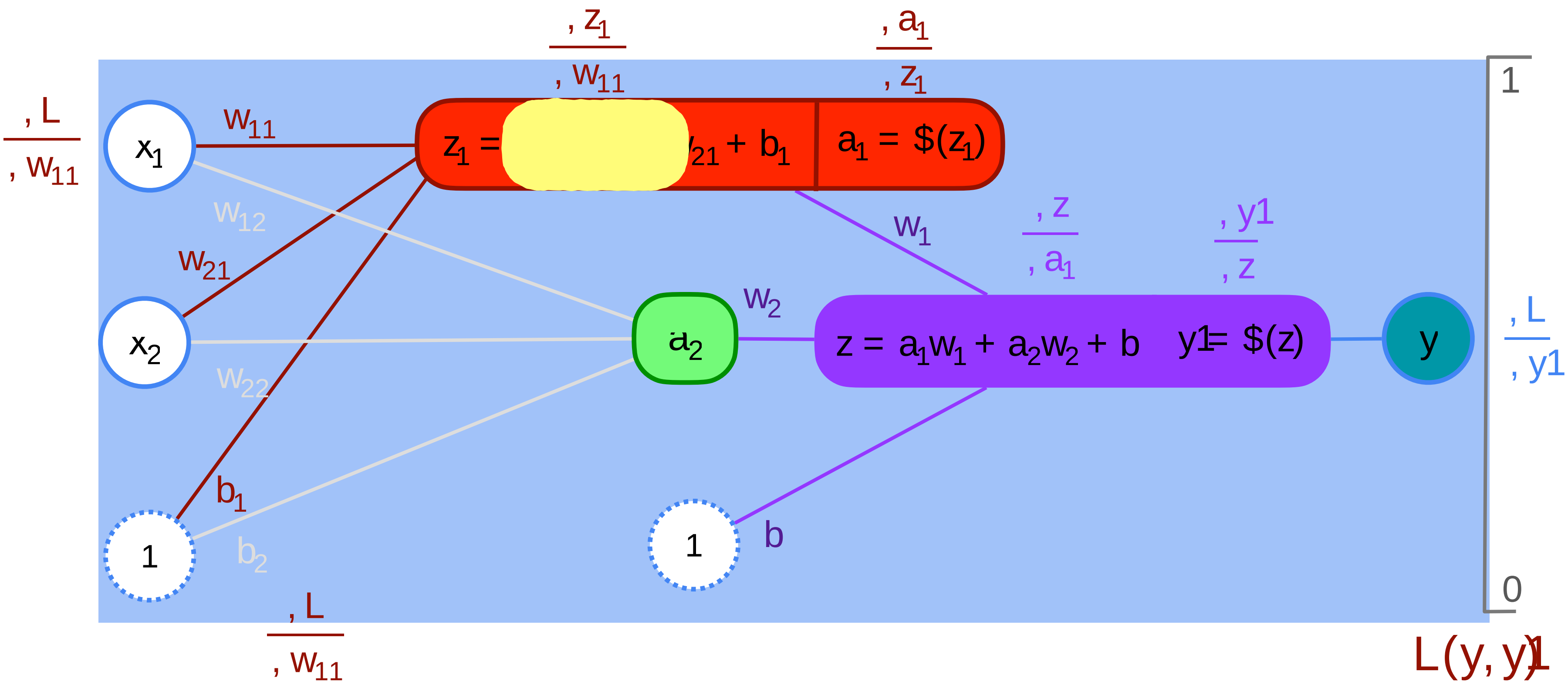
2,2,1 Neural Network



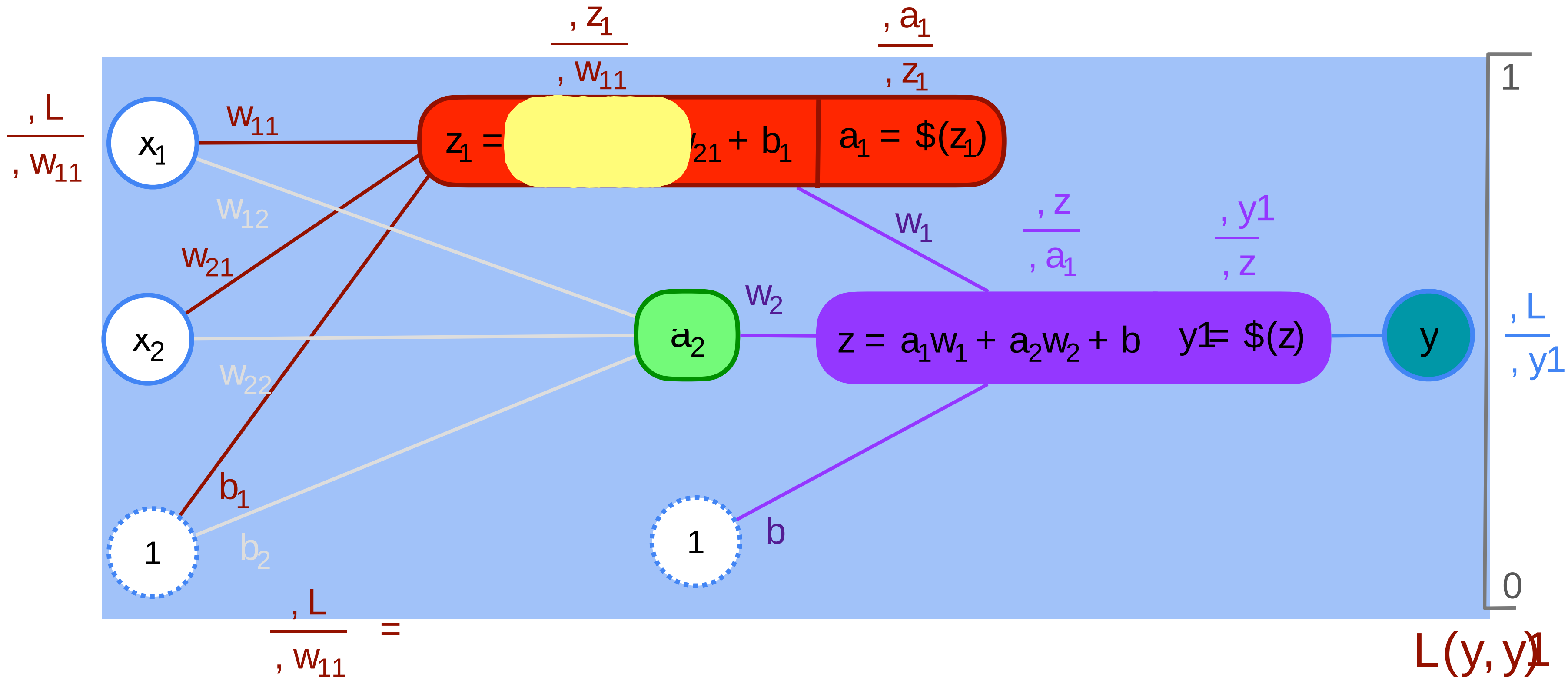
2,2,1 Neural Network



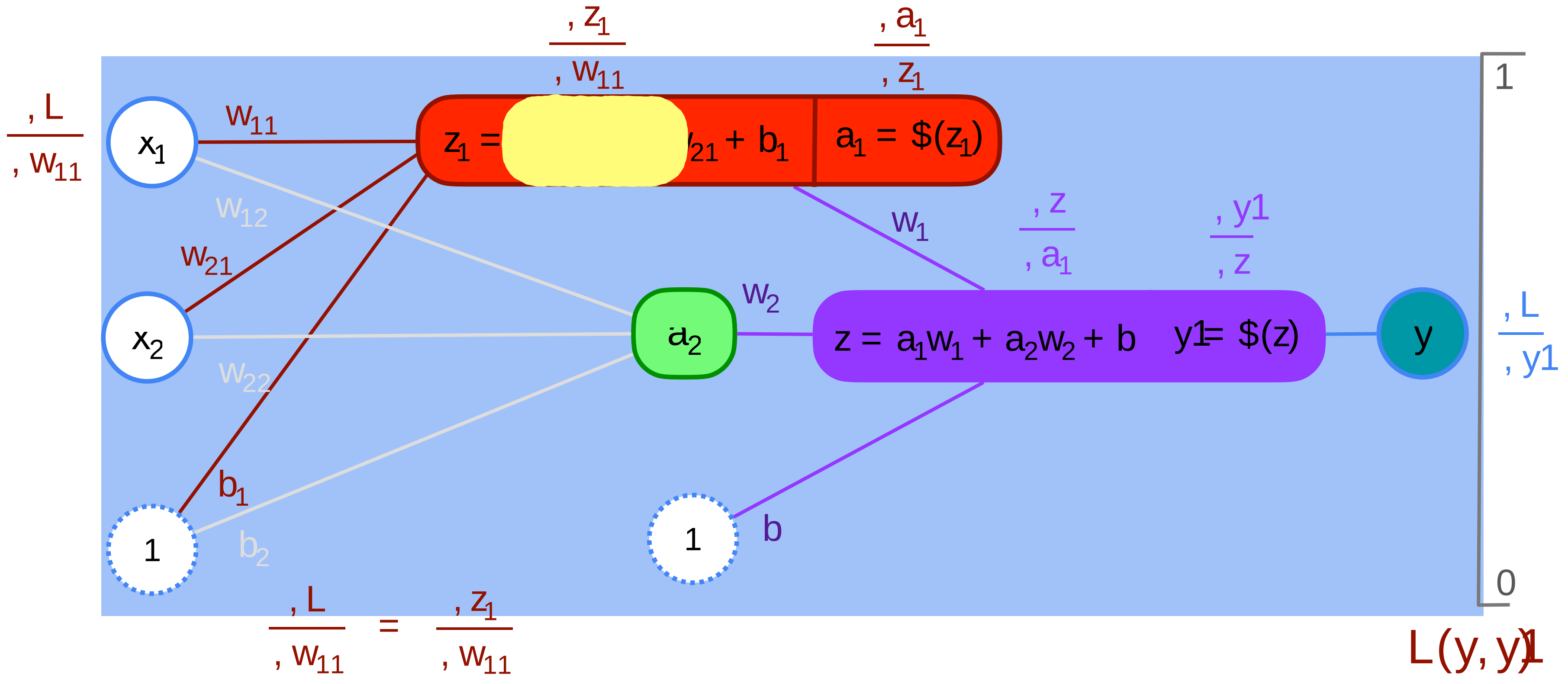
2,2,1 Neural Network



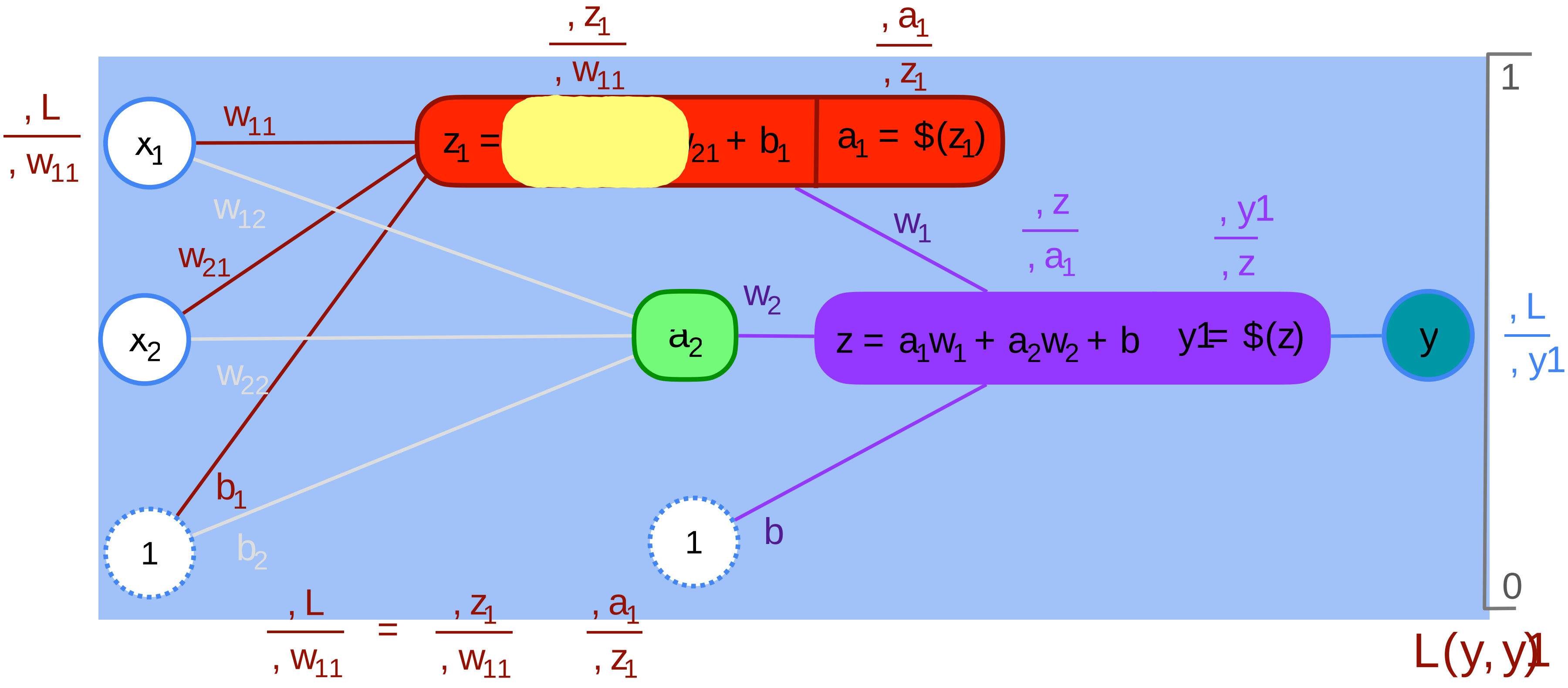
2,2,1 Neural Network



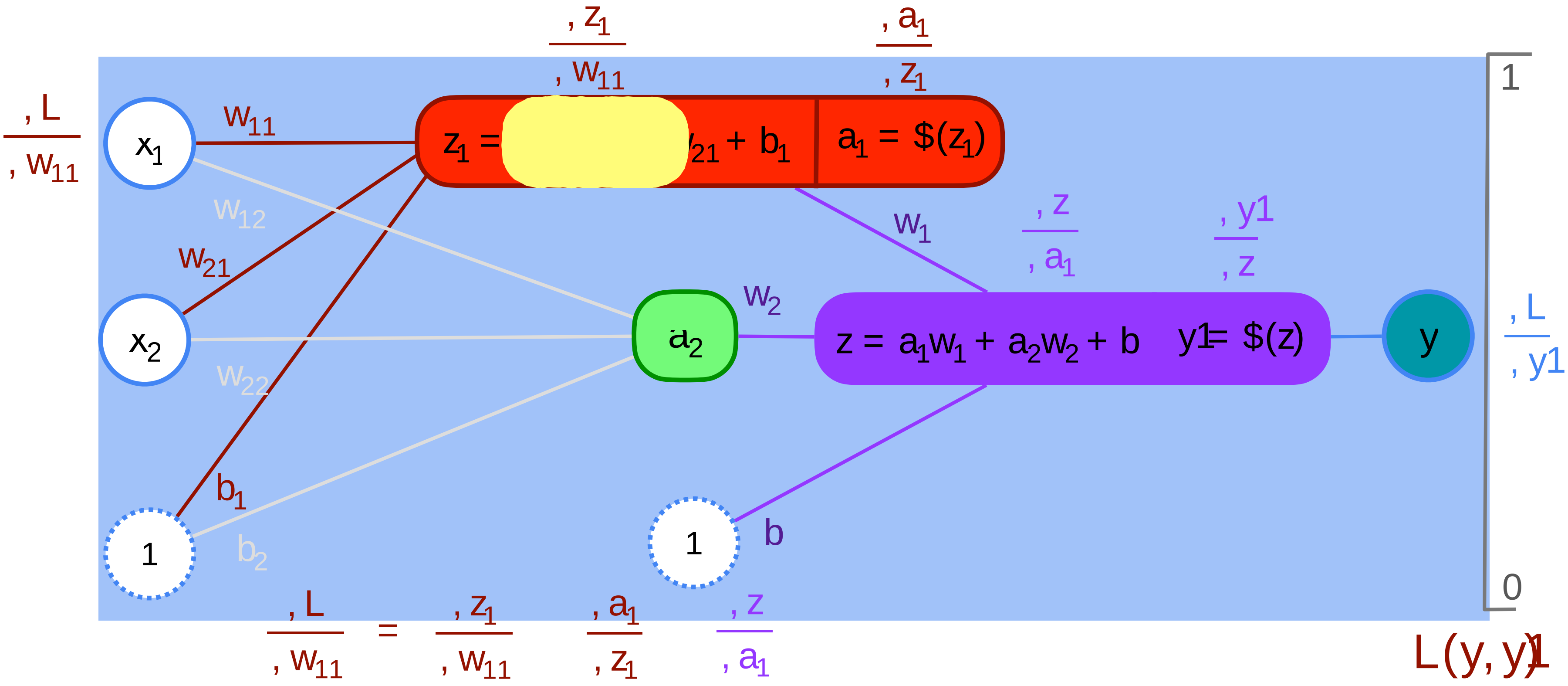
2,2,1 Neural Network



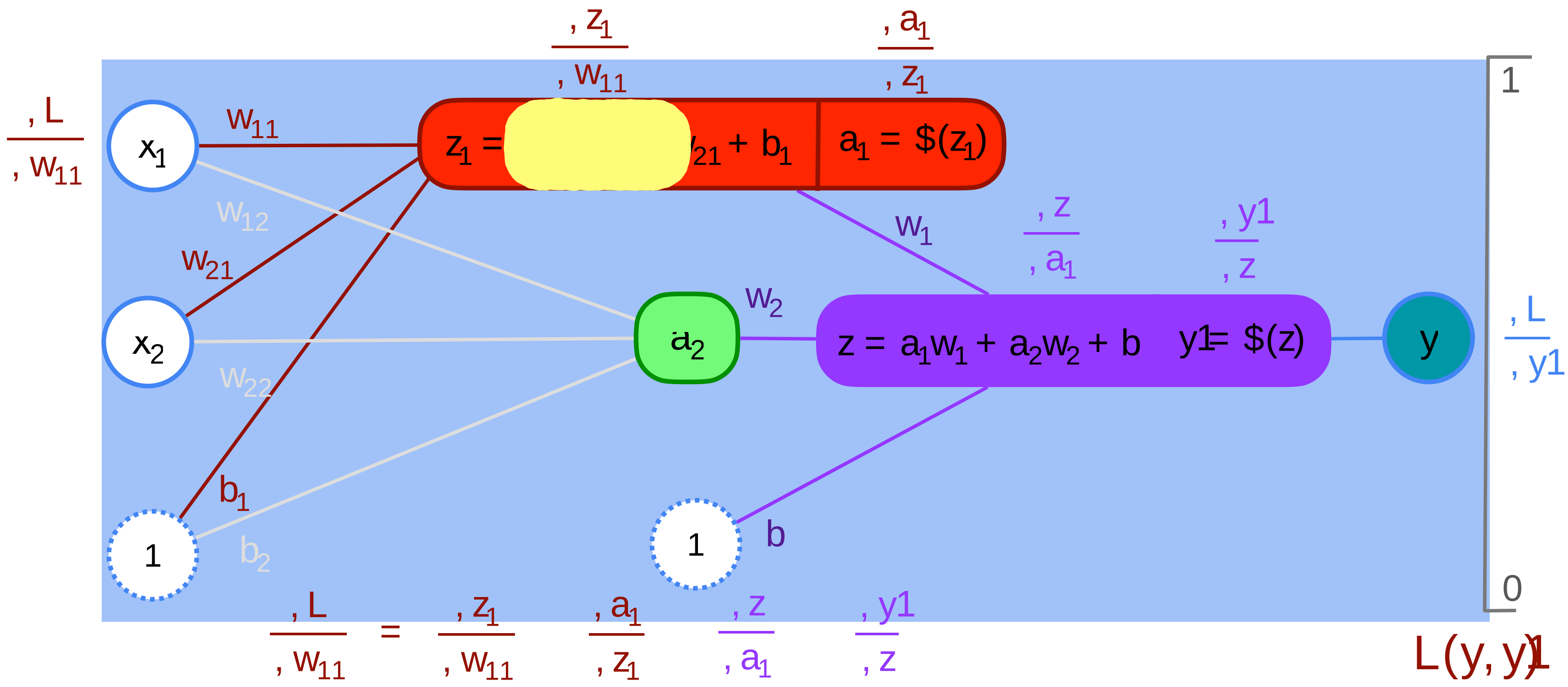
2,2,1 Neural Network



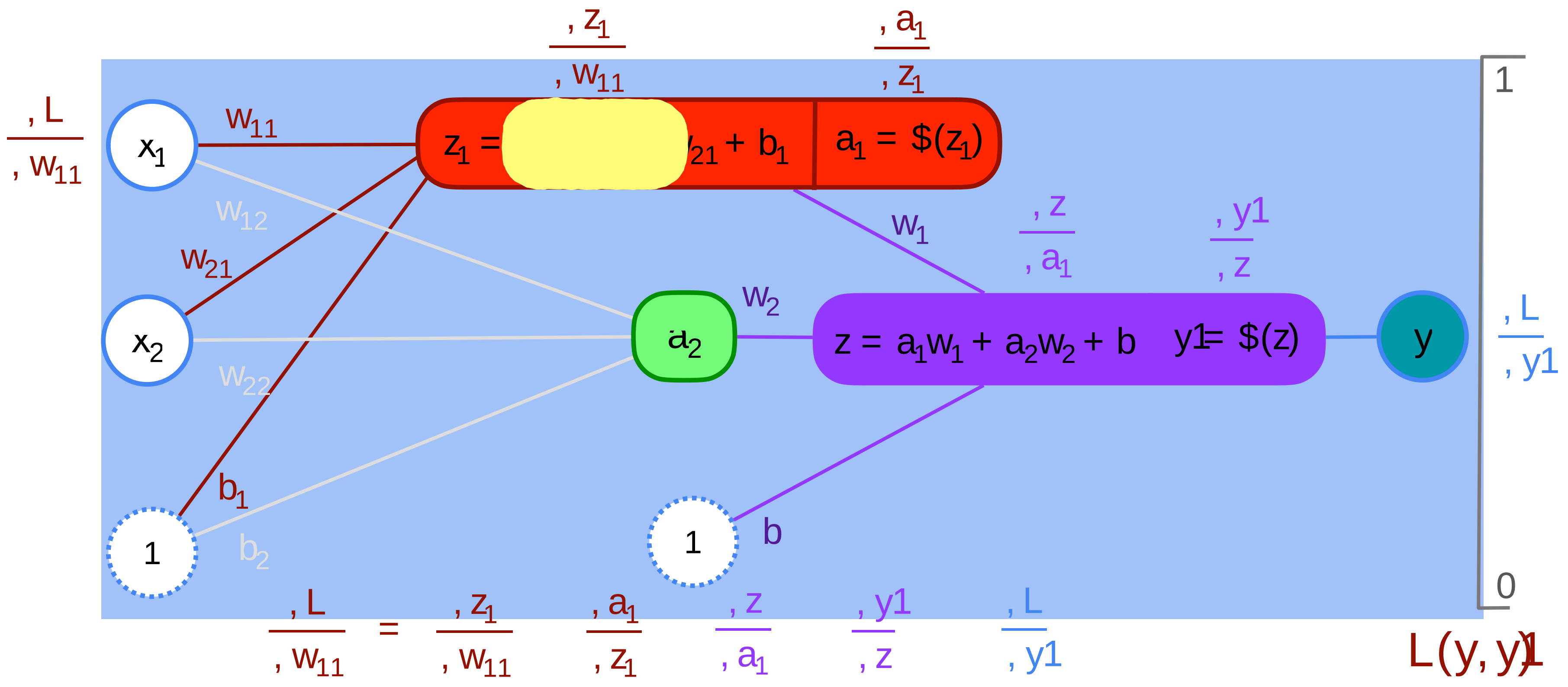
2,2,1 Neural Network



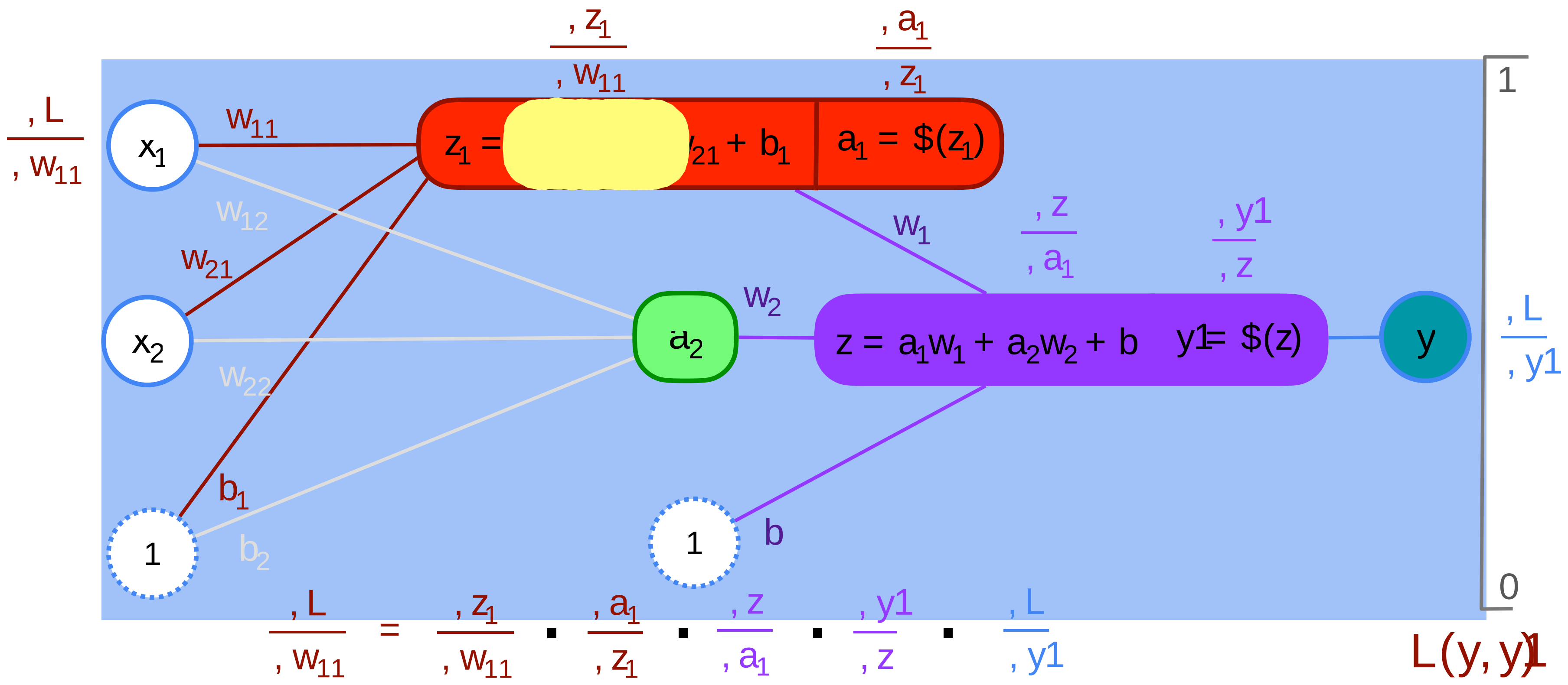
2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

2,2,1 Neural Network

$$L(y, y1) = -y\log(y1) " (1 " y)\log(1 " y1)$$

$$y1 = \$(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \$(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}}$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1 " (1" y)log(1" y1

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,w_{11}} =$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1\ "\ \ (1"\ y)\log(1"\ y1$

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{11}} = \quad \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,w_{11}} =$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) \text{ " } (1 \text{ " } y)\log(1 \text{ " } y1)$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$$\frac{, L}{, w_{11}} = \quad \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}} =$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$$\frac{, L}{, w_{11}} = \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}} = x_1$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1)\;"\;(1"\;y)\log(1"\;y1)$

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \quad \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}} = x_1$$

2,2,1 Neural Network

L(y, y1) = -y log(y1) " (1" y)log(1" y1)

y1 = $(z)

$z = a_1 w_1 + a_2 w_2 + b$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \quad \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}} = x_1$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) \text{ " } (1 \text{ " } y)\log(1 \text{ " } y1)$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \quad \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,w_{11}} = x_1 \quad a_1(1 \text{ " } a_1)$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1 " (1" y)log(1" y1

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,\ L}{,\ w_{11}} = \frac{,\ z_1}{,\ w_{11}} \cdot \frac{,\ a_1}{,\ z_1} \cdot \quad \frac{,\ y1}{,\ z} \cdot \frac{,\ L}{,\ y1}$$

$$\frac{,\ L}{,\ w_{11}} = x_1 \quad a_1(1 " \ a_1)$$

2,2,1 Neural Network

$L(y, y1) = -y \log(y1) - (1 - y)\log(1 - y1)$

$y1 = \sigma(z)$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \quad \frac{\partial y1}{\partial z} \cdot \frac{\partial L}{\partial y1}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1 - a_1)$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) \text{ " } (1 \text{ " } y)\log(1 \text{ " } y1)$

$y1 = \$(z)$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \frac{, a_1}{, z_1} \cdot \quad \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}} = x_1 \quad a_1(1 \text{ " } a_1) \quad w_1$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1 " (1" y)log(1" y1

y1 = $(z)

a_1 = $($z_1$)

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \quad \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}} = x_1 \quad a_1(1" \ a_1) \quad w_1$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1 \; " \; (1 " \; y)\log(1 " \; y1$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}} = x_1 \quad a_1(1 " \; a_1) \quad w_1$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) " (1" y)\log(1" y1)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \qquad \frac{, L}{, y1}$$

$$\frac{, L}{, w_{11}} = x_1 \quad a_1(1" a_1) \quad w_1 \quad y1(1" y1)$$

2,2,1 Neural Network

L(y, y1) = -y log(y1 " (1" y) log(1" y1

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \$(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{11}} = x_1 \quad a_1(1" \; a_1) \quad w_1 \quad y1(1" \; y1)$$

2,2,1 Neural Network

$$y1 = \$(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \$(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{11}} = x_1 \quad a_1(1\,"\;a_1) \quad w_1 \quad y1(1\,"\;y1)$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{11}} = x_1 \quad a_1(1 \text{ " } a_1) \quad w_1 \quad y1(1 \text{ " } y1) \quad \frac{-(y \text{ " } y1)}{y1(1 \text{ " } y1)}$$

2,2,1 Neural Network

$$y1 = \$(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \$(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{11}} = x_1 \cdot a_1(1 \text{ " } a_1) \cdot w_1 \cdot y1(1 \text{ " } y1) \cdot \frac{-(y \text{ " } y1)}{y1(1 \text{ " } y1)}$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{11}} = x_1 \cdot a_1(1 \text{"} \; a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"} \; y1)} \cdot \frac{-(y \text{"} \; y1)}{y1(1 \text{"} \; y1)}$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, w_{11}} = x_1 \cdot a_1(1 \text{"} \; a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"} \; y1)} \cdot \frac{-(y \text{"} \; y1)}{\cancel{y1(1 \text{"} \; y1)}}$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{11}} = x_1 \cdot a_1(1 \text{"}\ a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"}\ y1)} \cdot \frac{-(y \text{"}\ y1)}{\cancel{y1(1 \text{"}\ y1)}}$$

$$= -x_1 w_1 a_1 (1 \text{"}\ a_1)(y \text{"}\ y1)$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{11}} = \frac{, z_1}{, w_{11}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, w_{11}} = x_1 \cdot a_1(1 \text{"} \; a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"} \; y1)} \cdot \frac{-(y \text{"} \; y1)}{\cancel{y1(1 \text{"} \; y1)}}$$

$$= -x_1 w_1 a_1 (1 \text{"} \; a_1)(y \text{"} \; y1)$$

Perform gradient descent with

to find optimal value of w_{11} that gives the least error

2,2,1 Neural Network

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{11}} = x_1 \cdot a_1(1 \text{"} \ a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"} \ y1)} \cdot \frac{-(y \text{"} \ y1)}{\cancel{y1(1 \text{"} \ y1)}}$$

$$= -x_1w_1a_1(1 \text{"} \ a_1)(y \text{"} \ y1)$$

Perform gradient descent with

$$w_{11} \ \$ \ \ w_{11} \text{"} \ \# \frac{,L}{,w_{11}}$$

to find optimal value of w_{11} that gives the least error

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{11}} = \frac{,z_1}{,w_{11}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{11}} = x_1 \cdot a_1(1 " \ a_1) \cdot w_1 \cdot \cancel{y1(1 " \ y1)} \cdot \frac{-(y " \ y1)}{\cancel{y1(1 " \ y1)}}$$

$$= -x_1w_1a_1(1 " \ a_1)(y " \ y1)$$

Perform gradient descent with

$w_{11} \ \$ \ w_{11} " \ \#$

to find optimal value of w_{11} that gives the least error

2,2,1 Neural Network

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

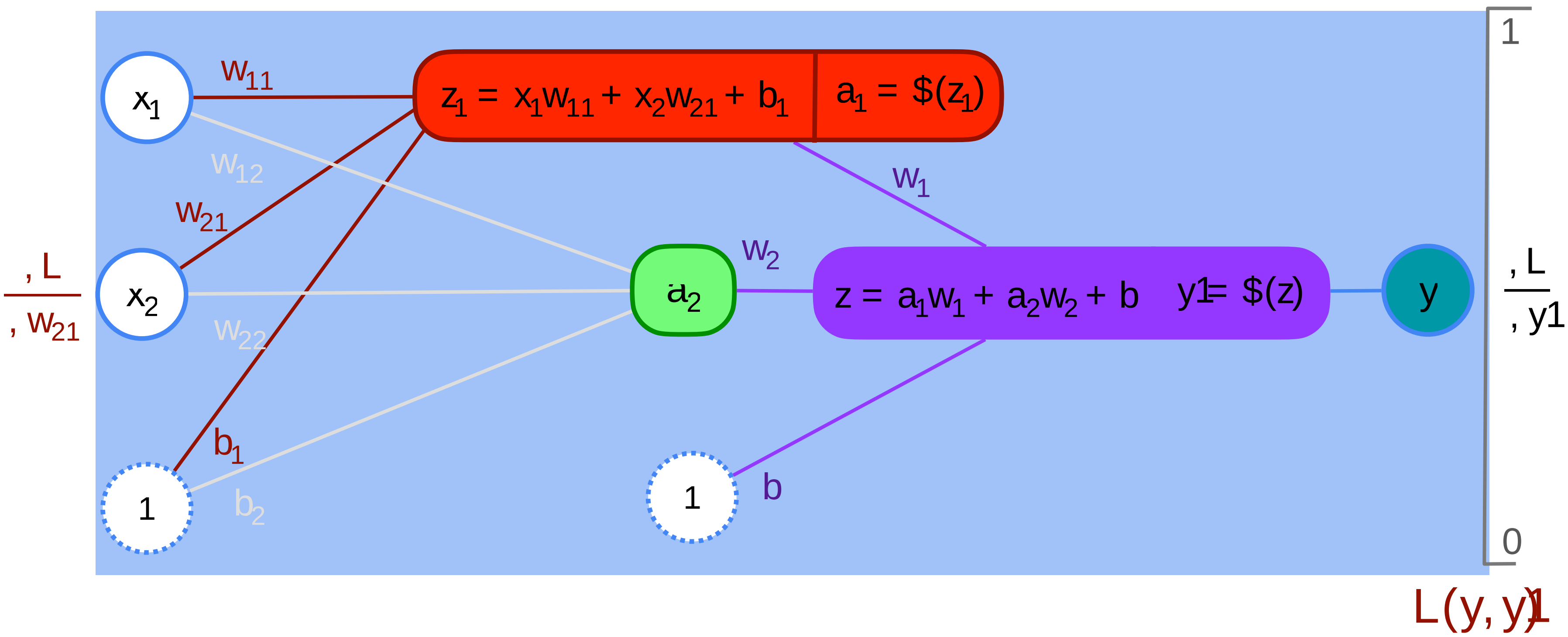
$$= -x_1 w_1 a_1 (1 - a_1)(y - \hat{y})$$

Perform gradient descent with

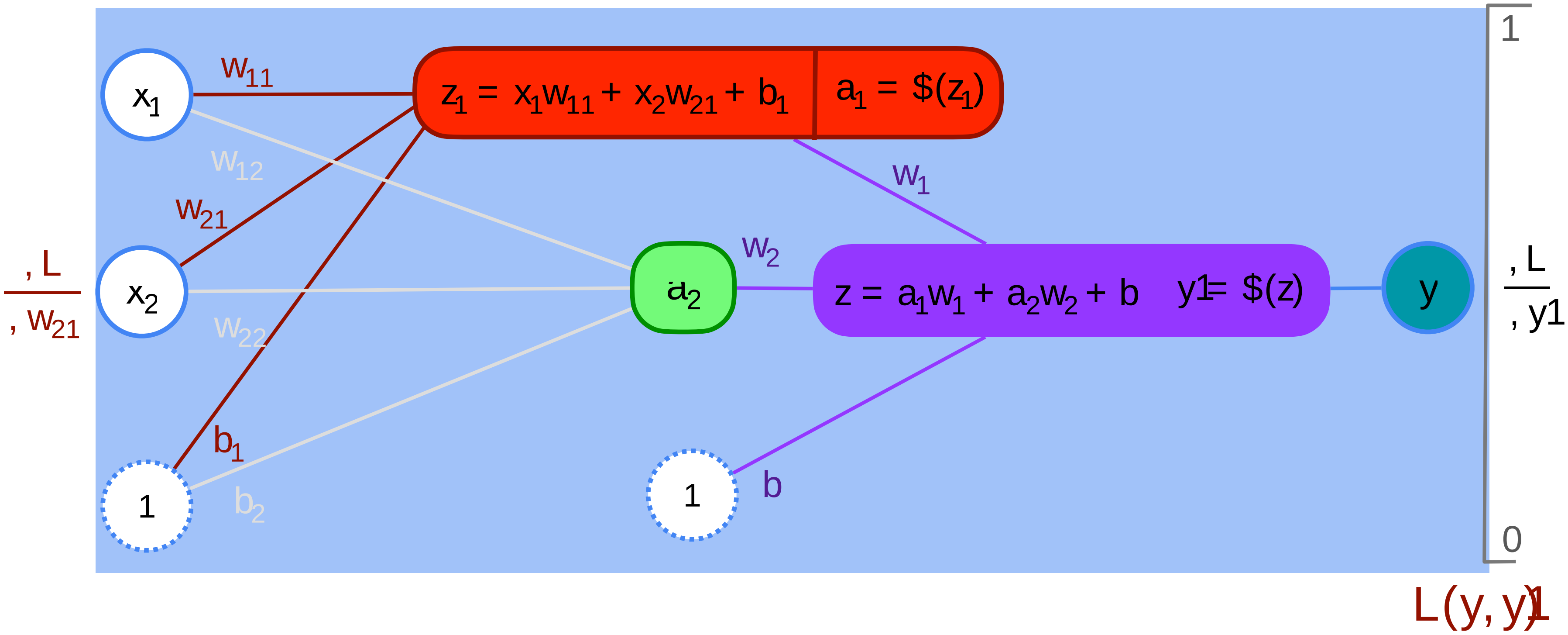
$$w_{11} \leftarrow w_{11} - \eta\, {-x_1 w_1 a_1 (1 - a_1)(y - \hat{y})}$$

to find optimal value of w_{11} that gives the least error

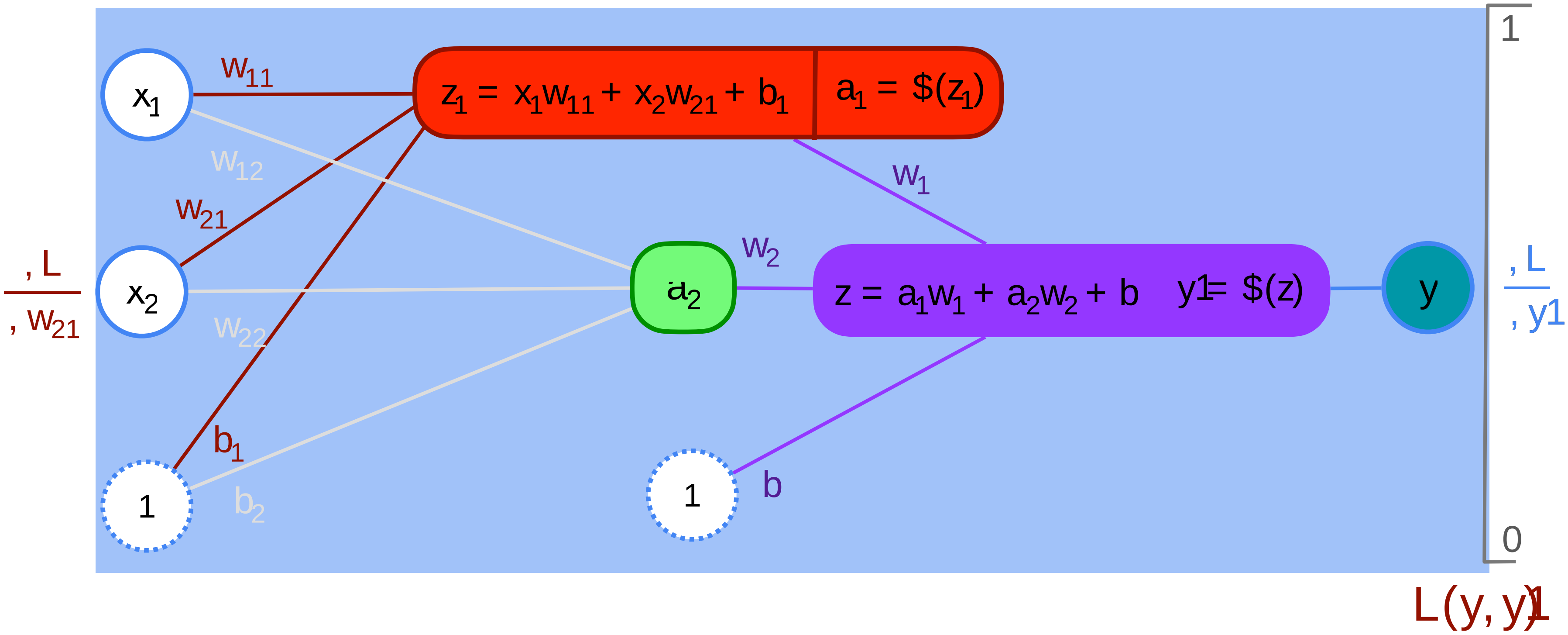
2,2,1 Neural Network



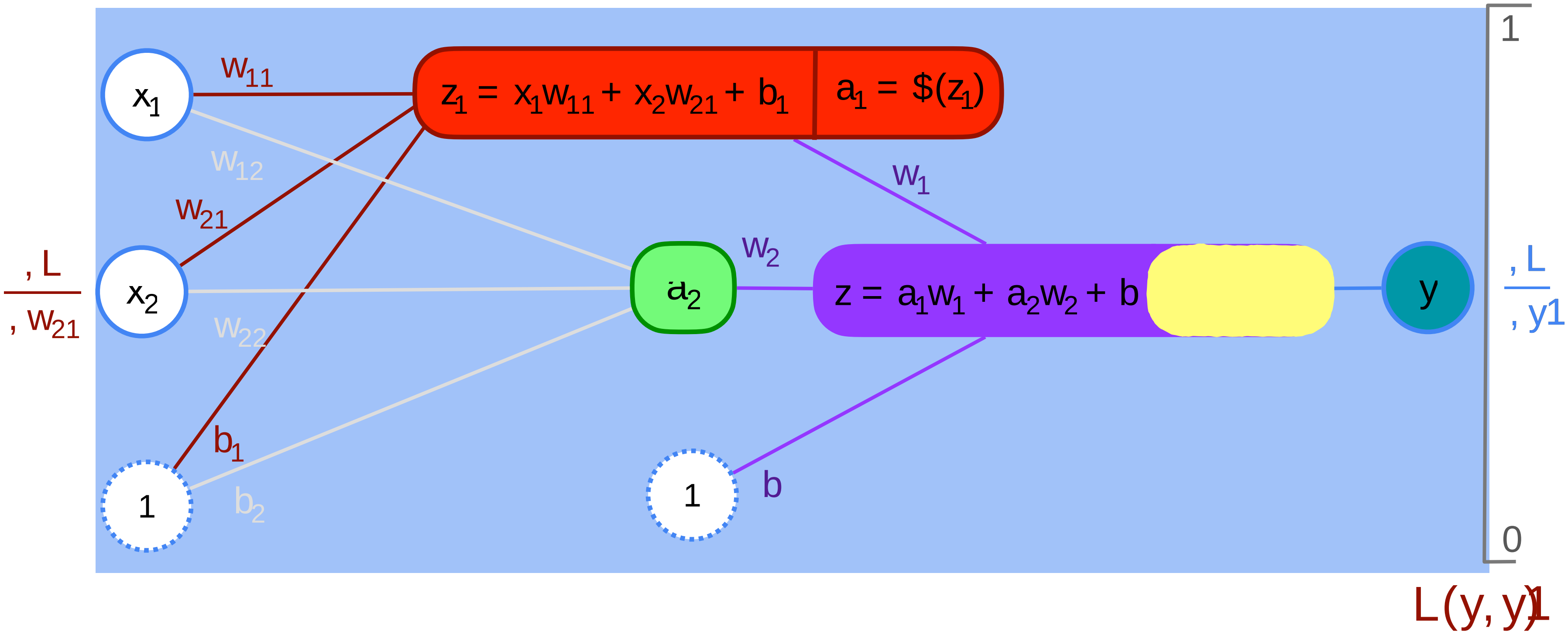
2,2,1 Neural Network



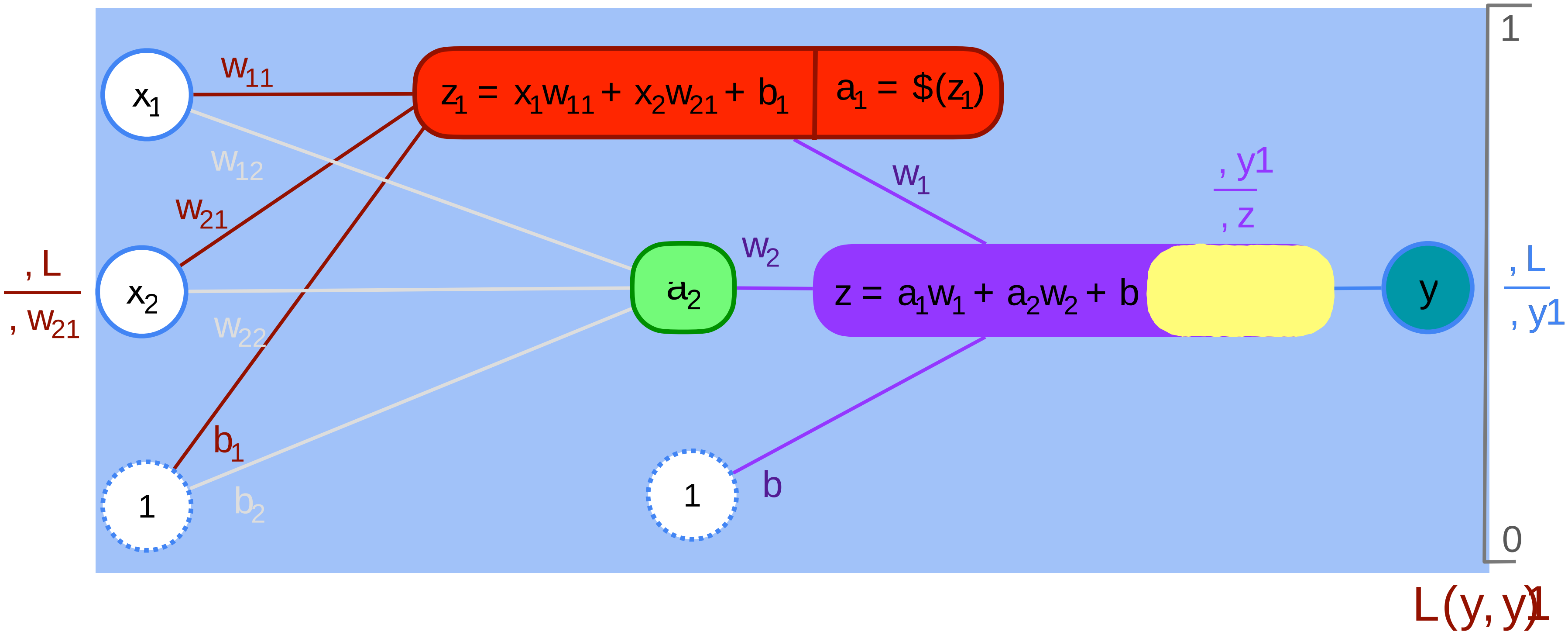
2,2,1 Neural Network



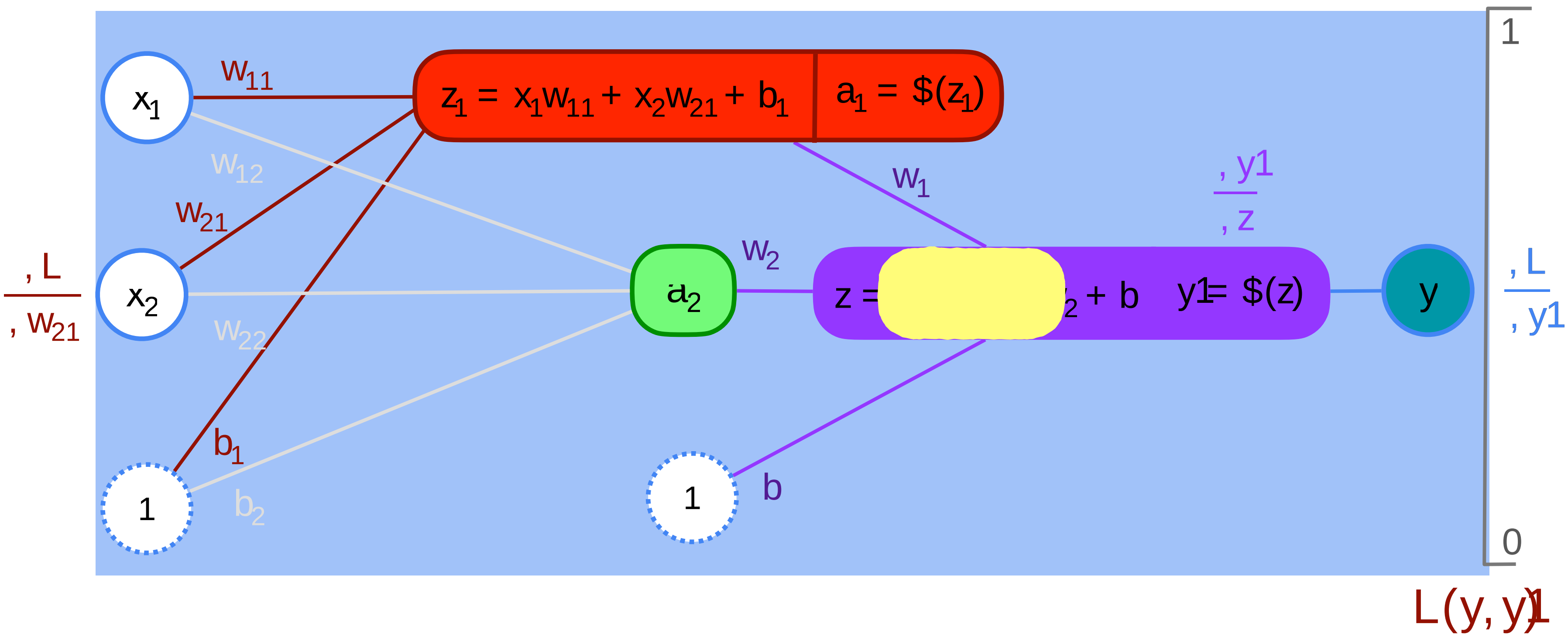
2,2,1 Neural Network



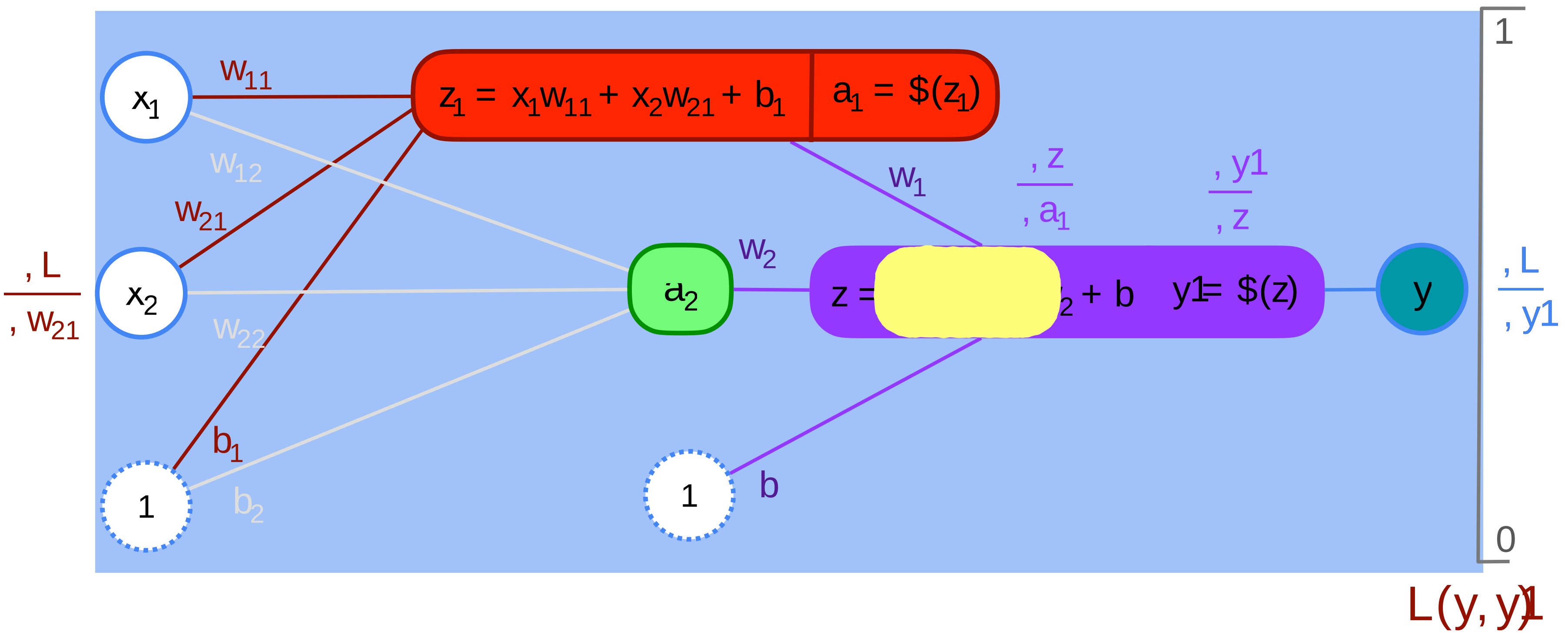
2,2,1 Neural Network



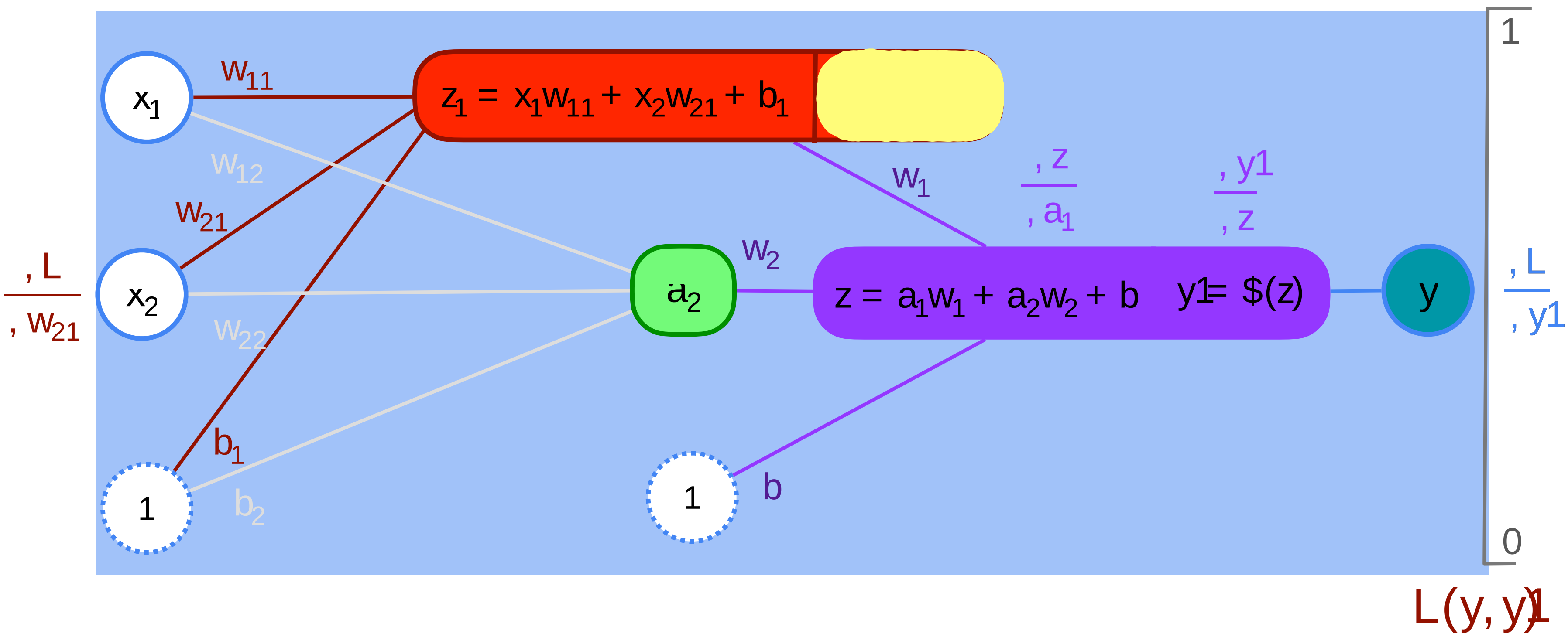
2,2,1 Neural Network



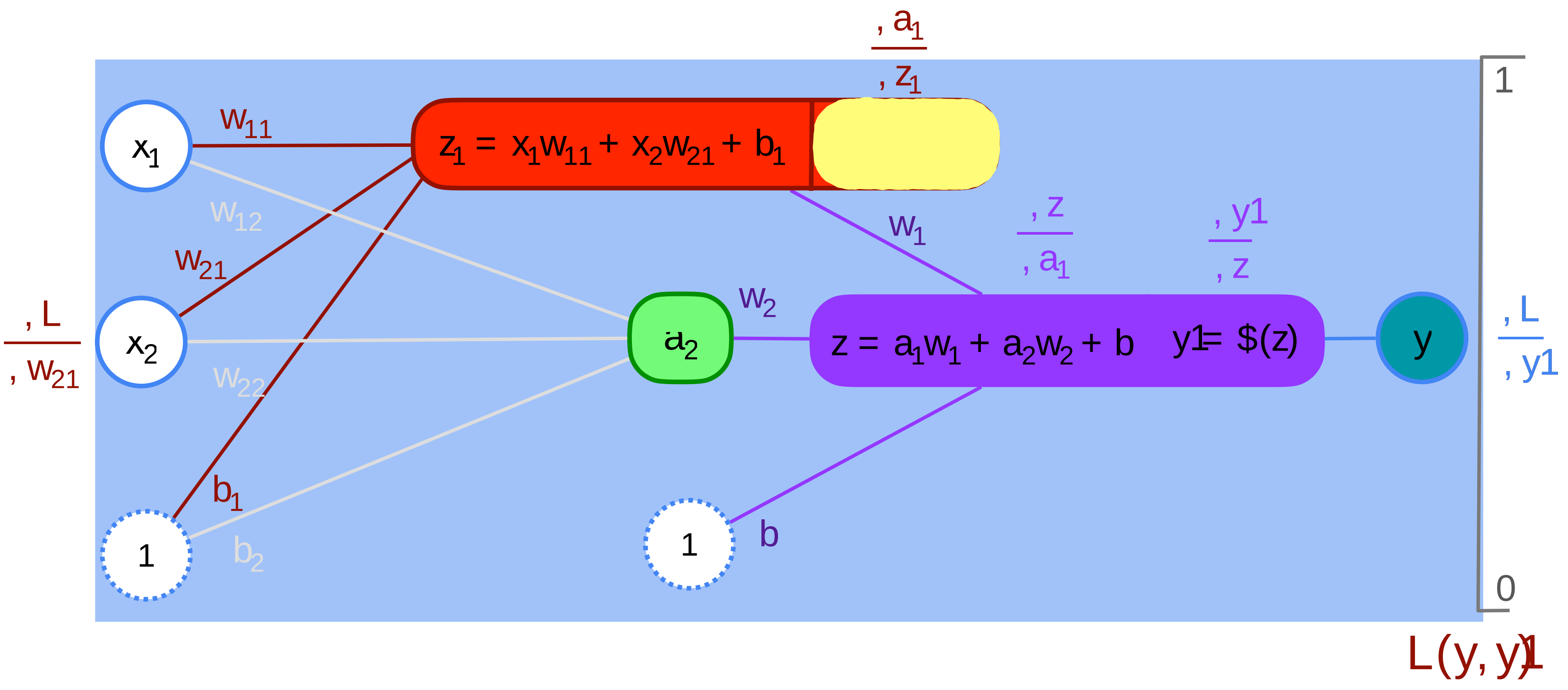
2,2,1 Neural Network



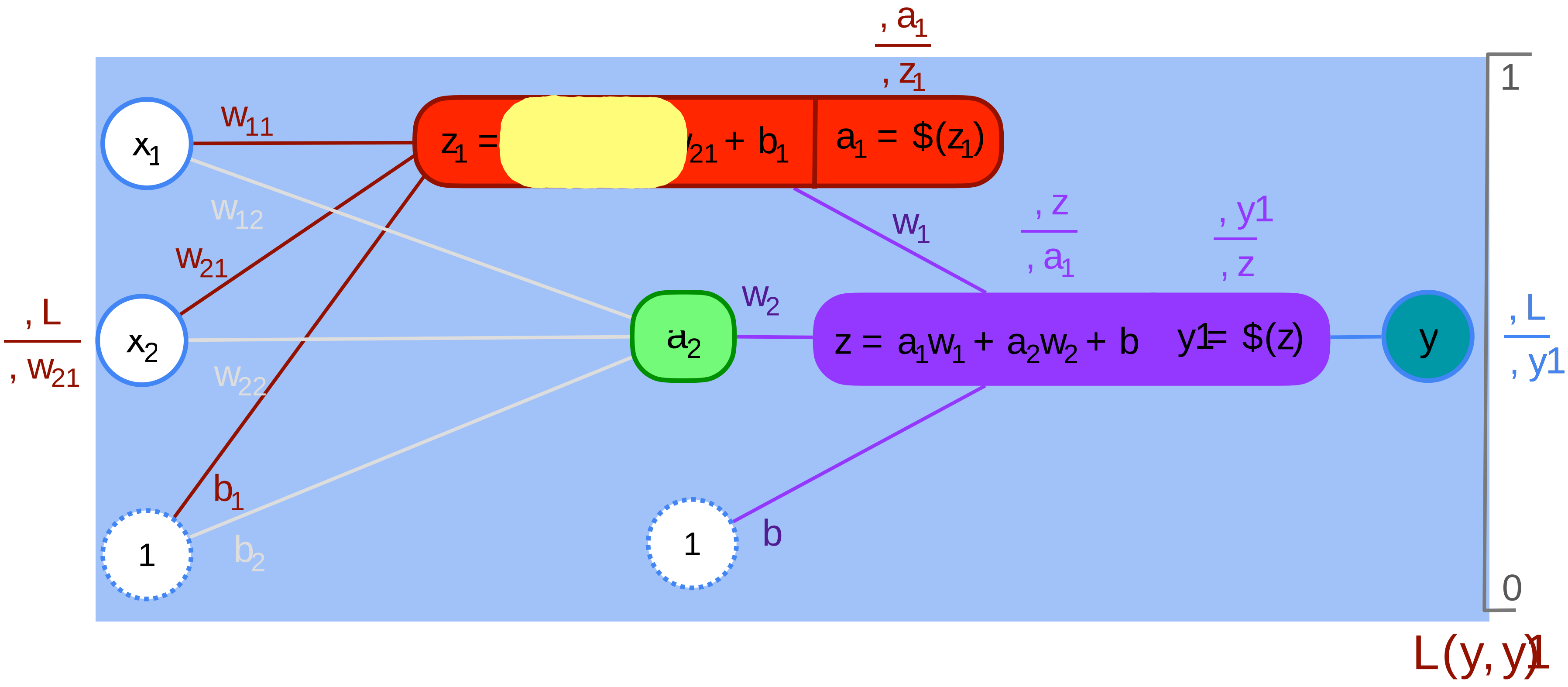
2,2,1 Neural Network



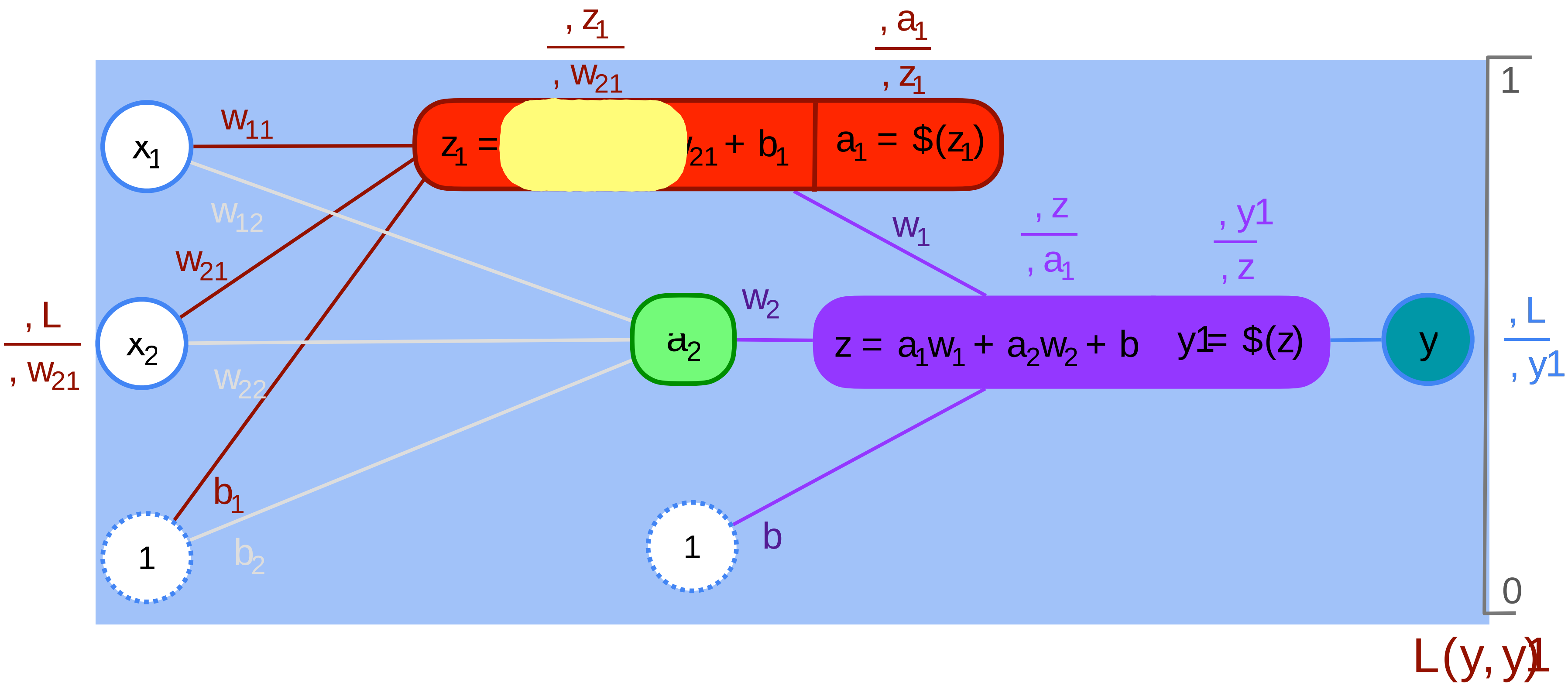
2,2,1 Neural Network



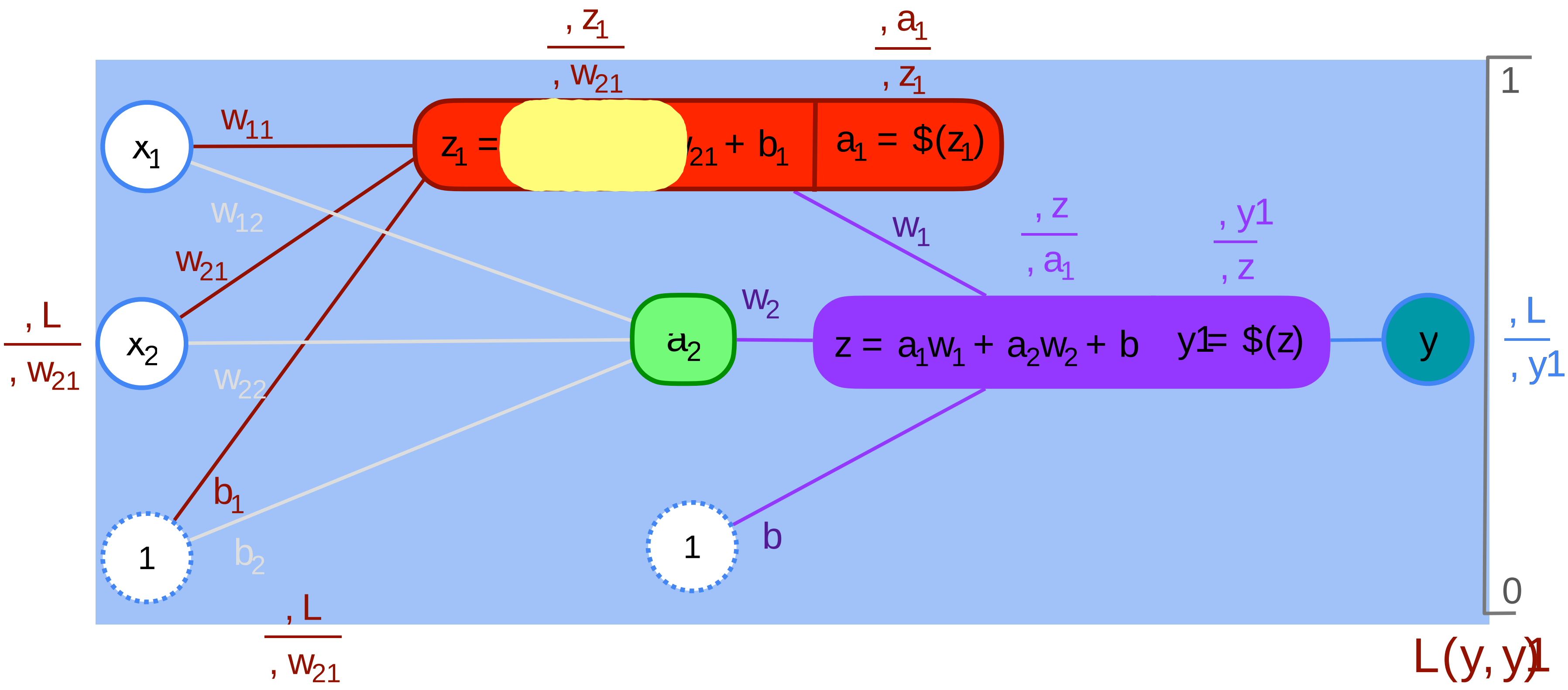
2,2,1 Neural Network



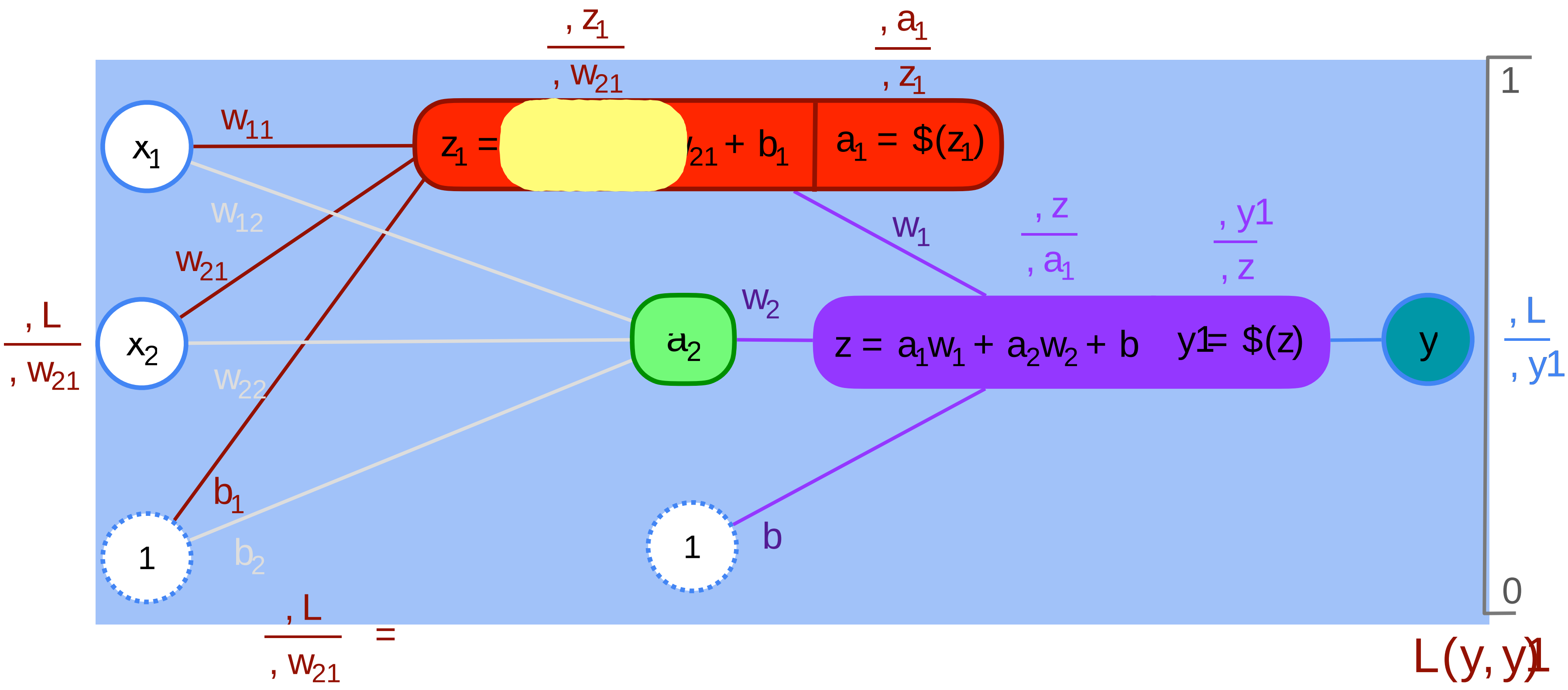
2,2,1 Neural Network



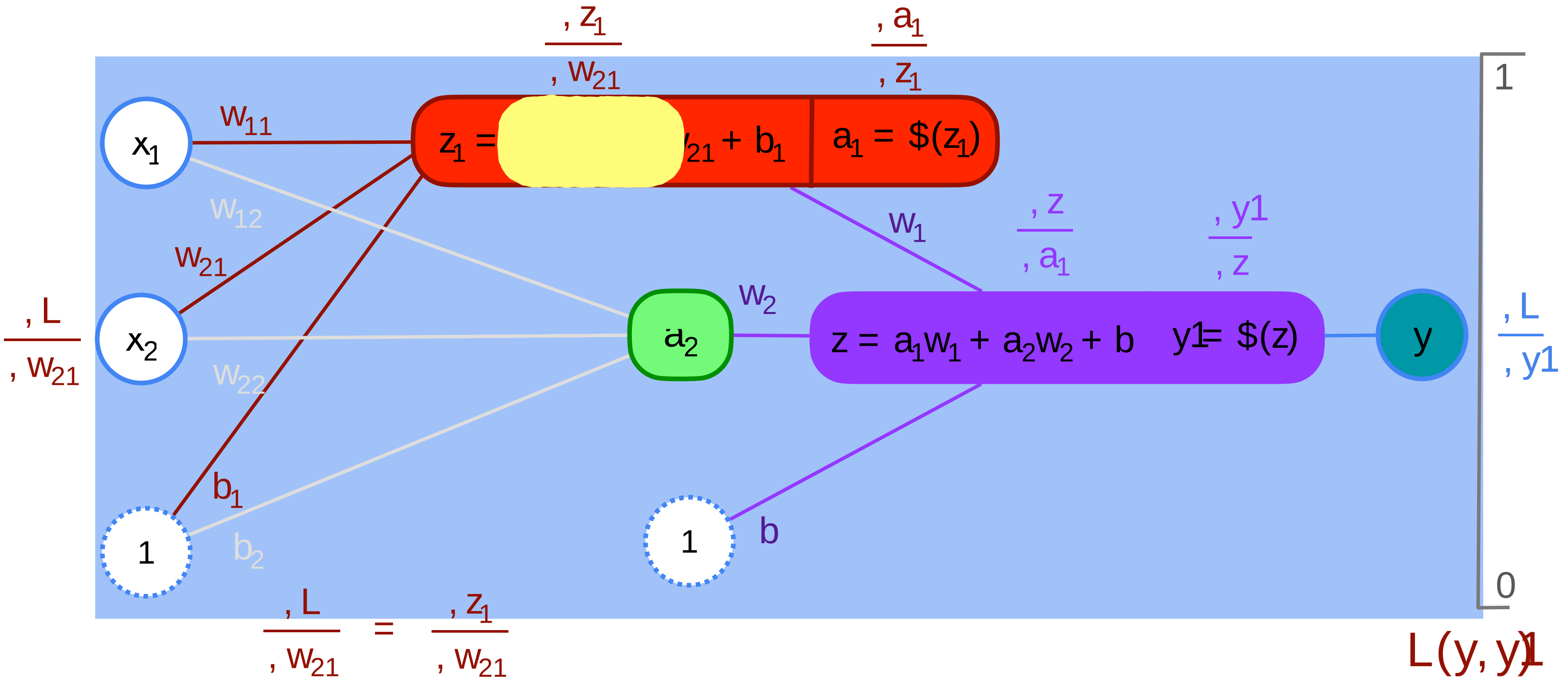
2,2,1 Neural Network



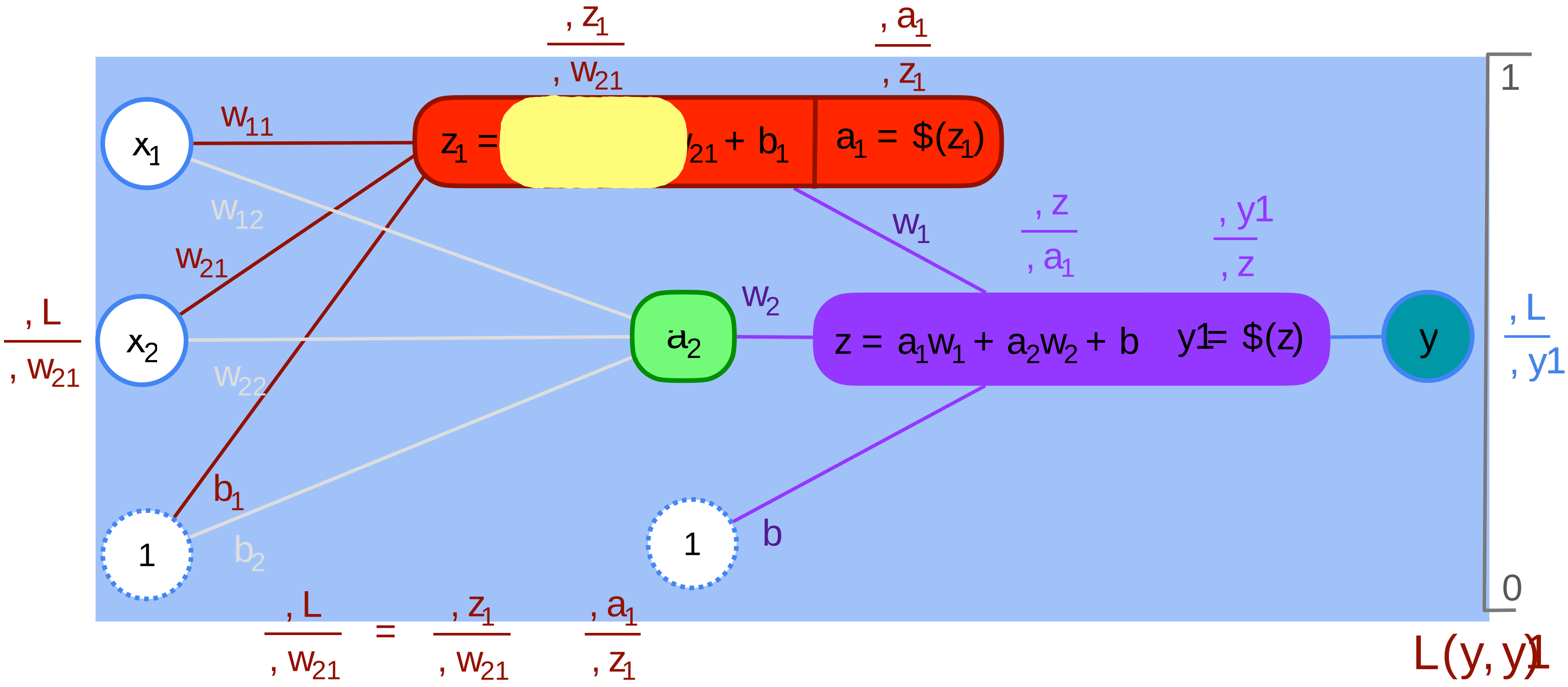
2,2,1 Neural Network



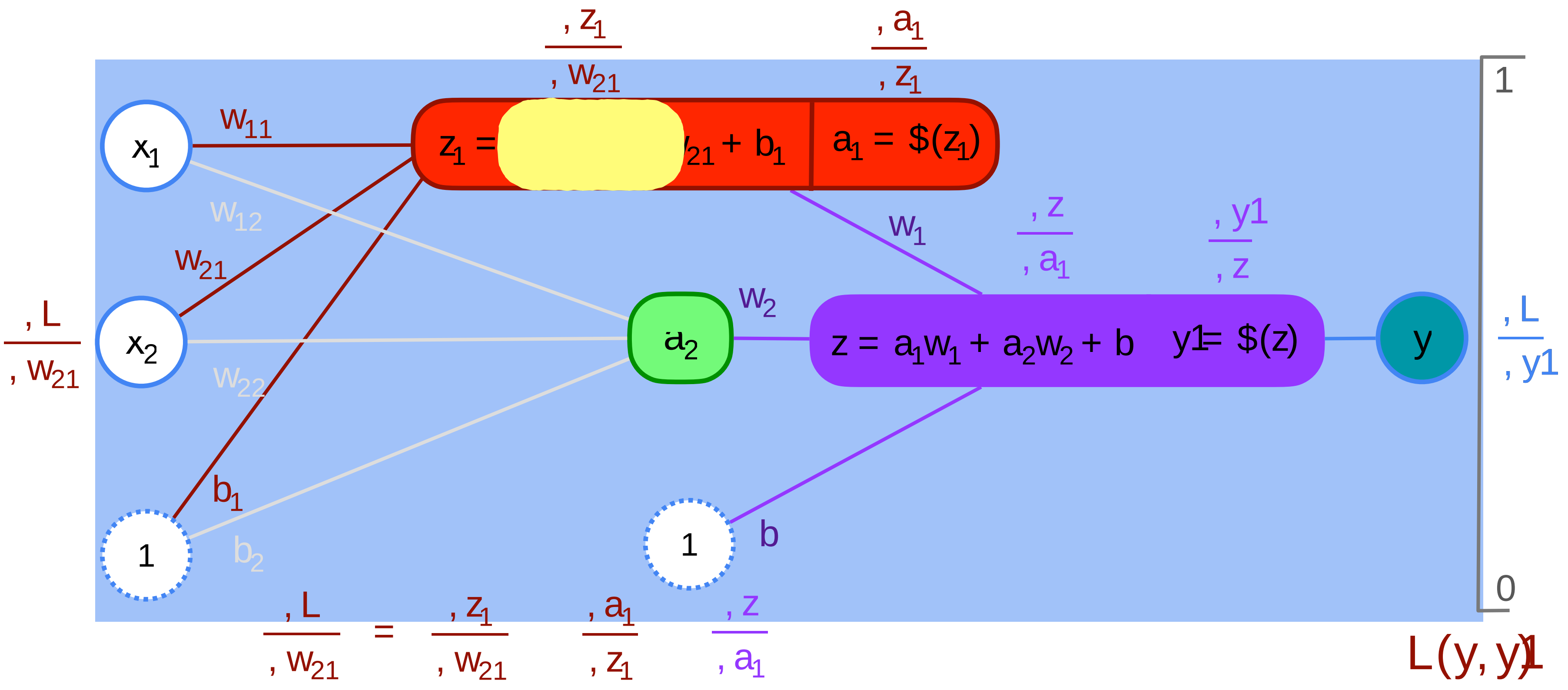
2,2,1 Neural Network



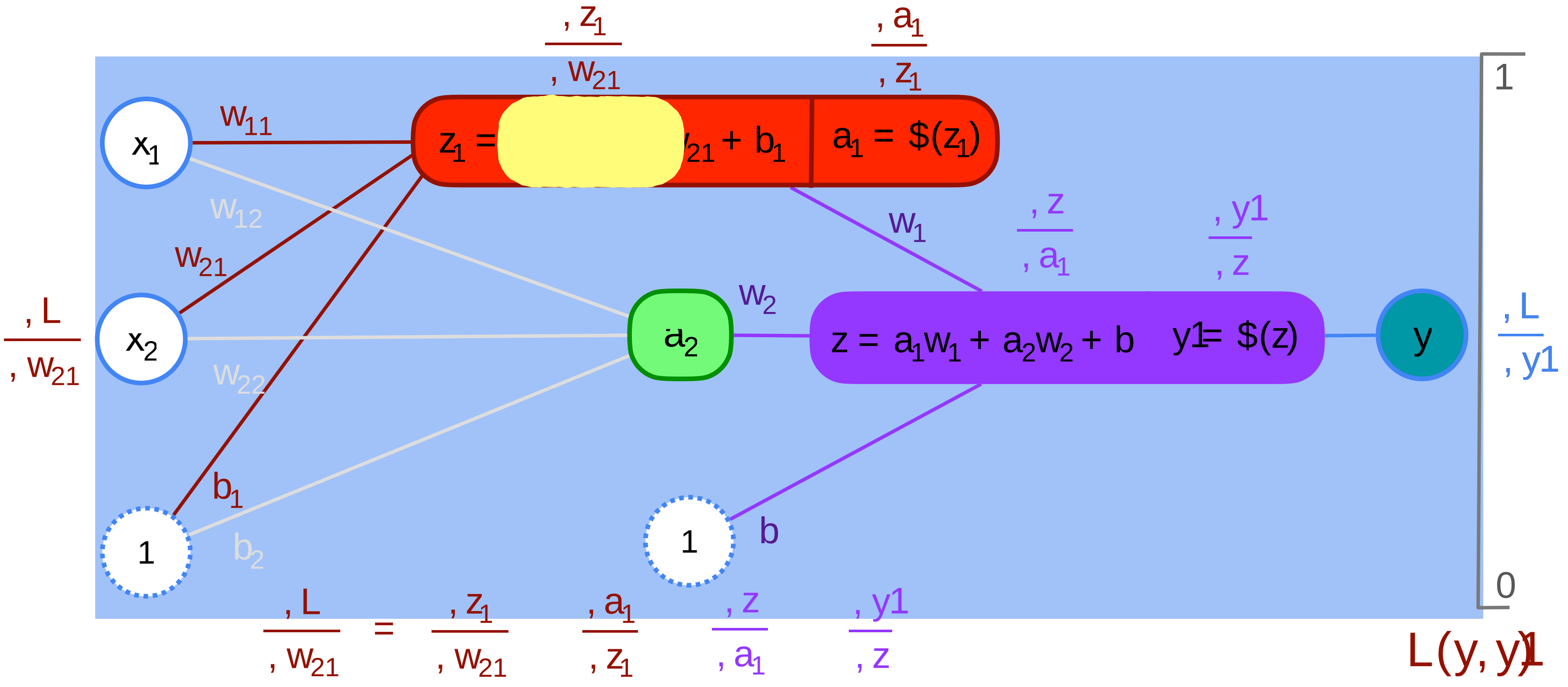
2,2,1 Neural Network



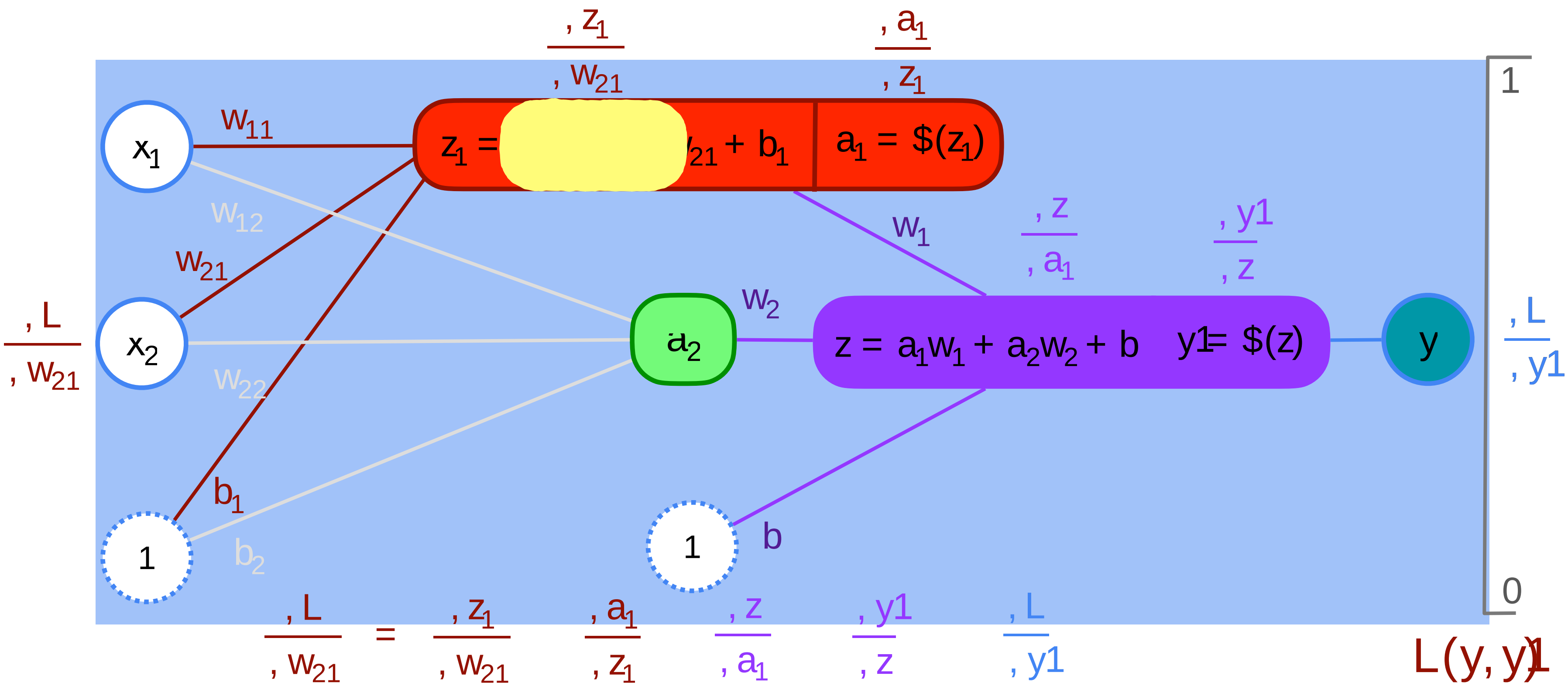
2,2,1 Neural Network



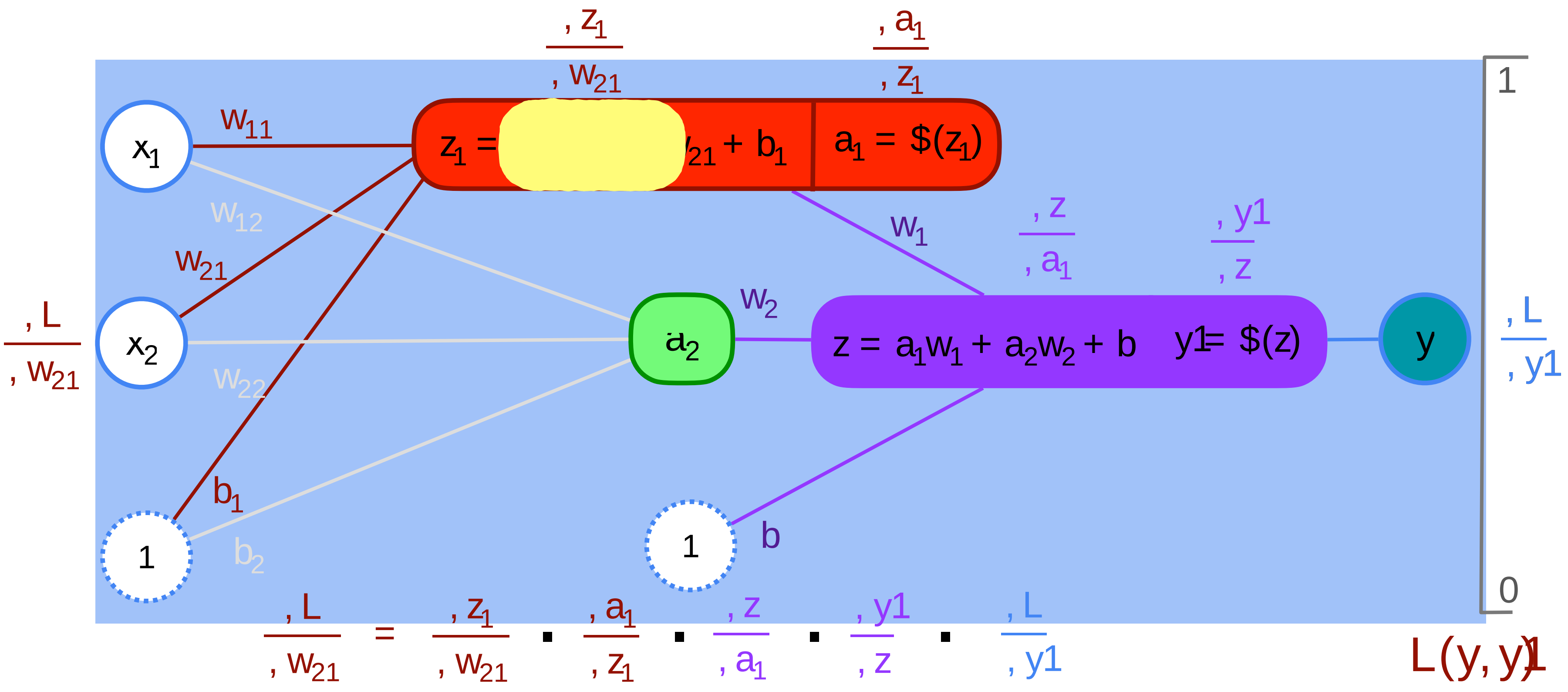
2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

2,2,1 Neural Network

L(y, y)1 = -ylog(y)1 " (1" y)log(1" y)1

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{21}} = \frac{, z_1}{, w_{21}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

2,2,1 Neural Network

$$L(y,\hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}}$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1 - y)\log(1 - \hat{y})$$

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} =$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) \;"\; (1" y)\log(1" y1)$

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{, L}{, w_{21}} = \quad \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{21}} =$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1 = $(z)

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$$\frac{,L}{,w_{21}} = \quad \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,w_{21}} =$$

2,2,1 Neural Network

L (y, y1) = -ylog(y1) " (1" y)log(1" y1)

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$$\frac{, L}{, w_{21}} = \quad \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{21}} = x_2$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \quad \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \quad \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,w_{21}} = x_2$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

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

$z = a_1 w_1 + a_2 w_2 + b$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \quad \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1-a_1)$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1 " (1" y)log(1" y1

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{, L}{, w_{21}} = \frac{, z_1}{, w_{21}} \cdot \frac{, a_1}{, z_1} \cdot \quad \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{21}} = x_2 \quad a_1(1" \ a_1)$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1)$ " $(1$ " $y)\log(1$ " $y1)$

$y1 = \$(z)$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{, L}{, w_{21}} = \frac{, z_1}{, w_{21}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_{21}} = x_2 \quad a_1(1 \text{"} \quad a_1)$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) \text{ " } (1\text{ " } y)\log(1\text{ " } y1)$

$y1 = \$(z)$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \qquad \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,w_{21}} = x_2 \quad a_1(1\text{ " } a_1) \quad w_1$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1)\ "\ (1"\ y)\log(1"\ y1)$

$y1 = \$(z)$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \qquad \frac{,L}{,y1}$$

$$\frac{,L}{,w_{21}} = x_2 \quad a_1(1"\ a_1) \quad w_1$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{21}} = \frac{, z_1}{, w_{21}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \qquad \frac{, L}{, y1}$$

$$\frac{, L}{, w_{21}} = x_2 \quad a_1(1" \ a_1) \quad w_1$$

2,2,1 Neural Network

$L(y, y1 = -y\log(y1\ \text{"}\ (1\text{"}\ y)\log(1\text{"}\ y1$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \qquad \frac{,L}{,y1}$$

$$\frac{,L}{,w_{21}} = x_2 \quad a_1(1\text{"}\ a_1) \quad w_1 \quad y(1\text{"}\ y)1$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) \;" \; (1" \; y)\log(1" \; y1)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{21}} = x_2 \quad a_1(1" \; a_1) \quad w_1 \quad y1(1" \; y1)$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{21}} = x_2 \quad a_1(1 \text{"} \; a_1) \quad w_1 \quad y1(1 \text{"} \; y1)$$

2,2,1 Neural Network

$$\frac{, L}{, w_{21}} = \frac{, z_1}{, w_{21}} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot$$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, w_{21}} = x_2 \quad a_1(1 \text{"} \ a_1) \quad w_1 \quad y1(1 \text{"} \ y1) \quad \frac{-(y \text{"} \ y1)}{y1(1 \text{"} \ y1)}$$

2,2,1 Neural Network

$\hat{y}1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \blacksquare \frac{,a_1}{,z_1} \blacksquare \frac{,z}{,a_1} \blacksquare \frac{,y1}{,z} \blacksquare$$

$$\frac{,L}{,w_{21}} = x_2 \blacksquare a_1(1 \text{"} \ a_1) \blacksquare w_1 \blacksquare y1(1 \text{"} \ y1) \blacksquare \frac{-(y \text{"} \ y1)}{y1(1 \text{"} \ y1)}$$

2,2,1 Neural Network

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

2,2,1 Neural Network

$y1 = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial y1}{\partial z} \cdot$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{y1(1 - y1)} \cdot \frac{-(y - y1)}{\cancel{y1(1 - y1)}}$$

$$= -x_2 w_1 a_1 (1 - a_1)(y - y1)$$

2,2,1 Neural Network

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

$$= -x_2 w_1 a_1 (1 - a_1)(y - \hat{y})$$

Perform gradient descent with

to find optimal value of w_{21} that gives the least error

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{21}} = x_2 \cdot a_1(1 " a_1) \cdot w_1 \cdot \cancel{y1(1 " y1)} \cdot \frac{-(y " y1)}{\cancel{y1(1 " y1)}}$$

$$= -x_2w_1a_1(1 " a_1)(y " y1)$$

Perform gradient descent with

$$w_{21} \$ w_{21} " \# \frac{,L}{,w_{21}}$$

to find optimal value of w_{21} that gives the least error

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{21}} = x_2 \cdot a_1(1 \text{"} \; a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"} \; y1)} \cdot \frac{-(y \text{"} \; y1)}{\cancel{y1(1 \text{"} \; y1)}}$$

$$= -x_2 w_1 a_1 (1 \text{"} \; a_1)(y \text{"} \; y1)$$

Perform gradient descent with

$w_{21} \; \$ \; w_{21} \text{"} \; \#$

to find optimal value of w_{21} that gives the least error

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,w_{21}} = \frac{,z_1}{,w_{21}} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_{21}} = x_2 \cdot a_1(1 \text{"} a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"} y)1} \cdot \frac{-(y \text{"} y)1}{\cancel{y1(1 \text{"} y)1}}$$

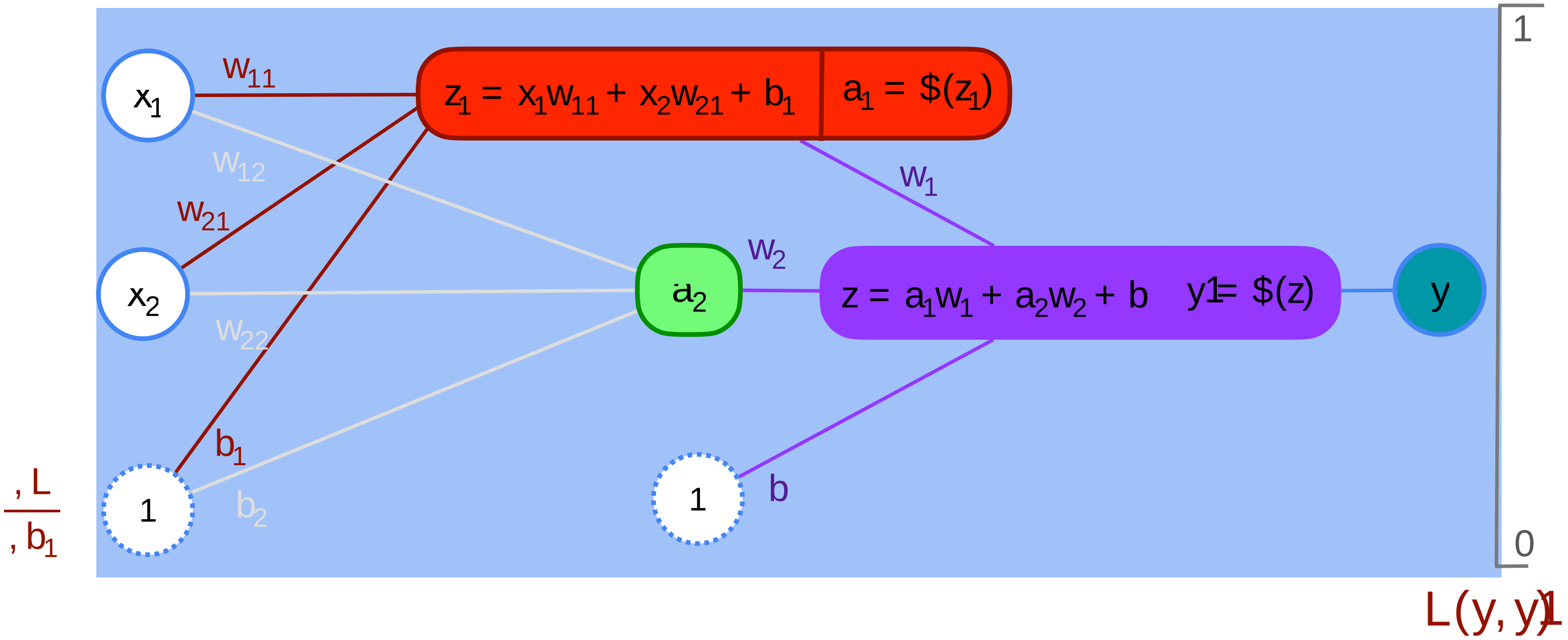
$$= -x_2w_1a_1(1 \text{"} a_1)(y \text{"} y)1$$

Perform gradient descent with

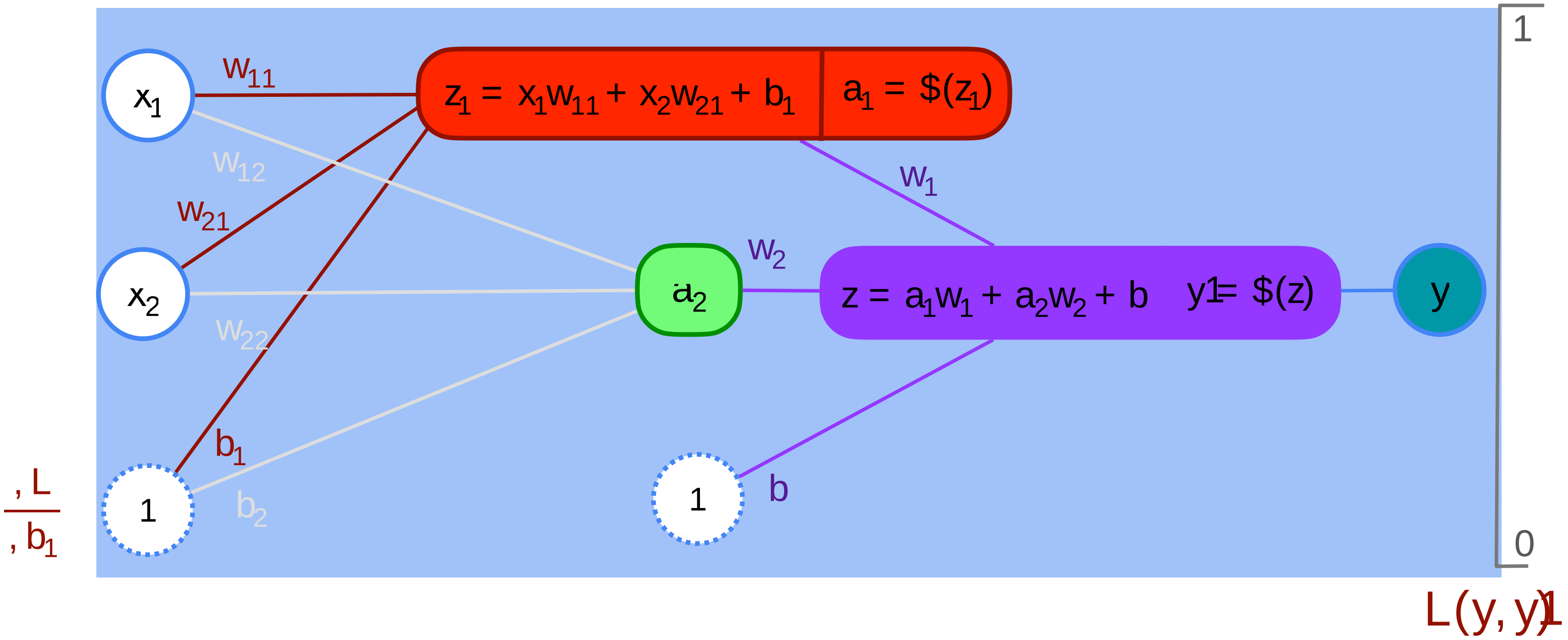
$$w_{21} \$ w_{21} \text{"} \#-x_2w_1a_1(1 \text{"} a_1)(y \text{"} y)1$$

to find optimal value of w_{21} that gives the least error

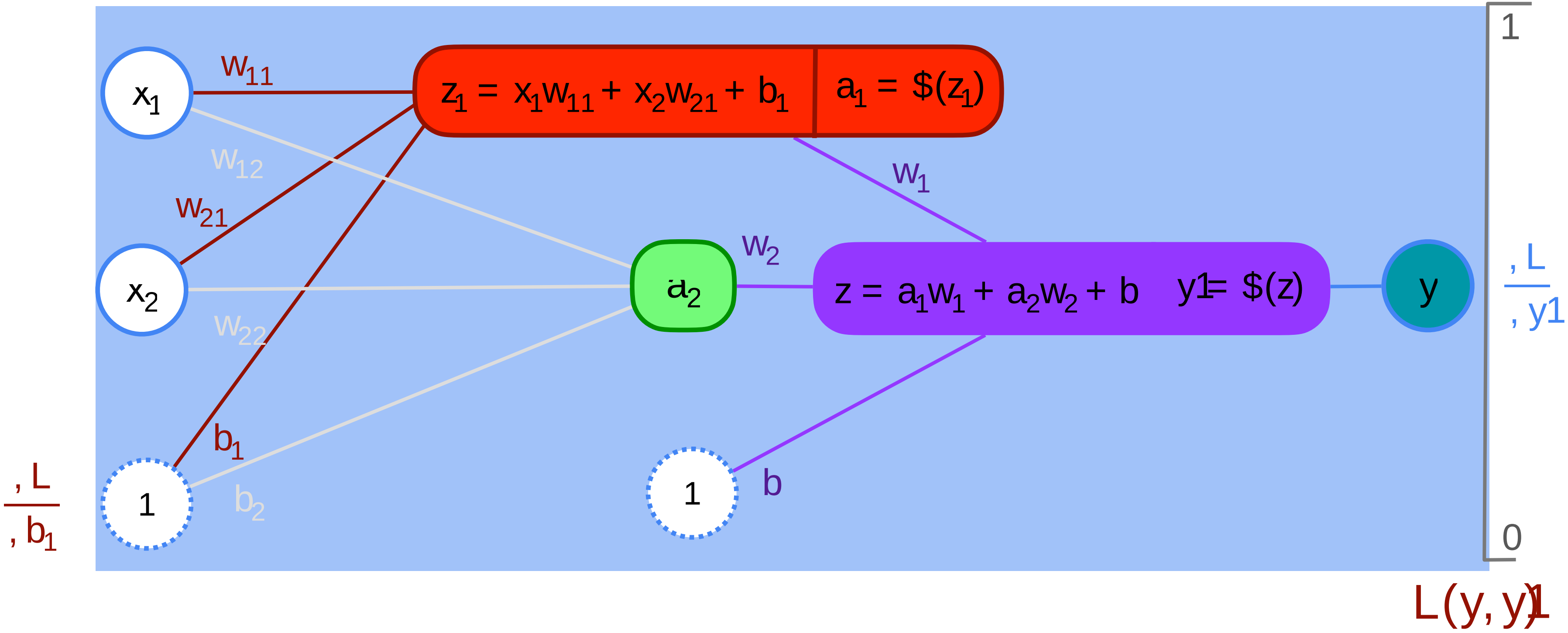
2,2,1 Neural Network



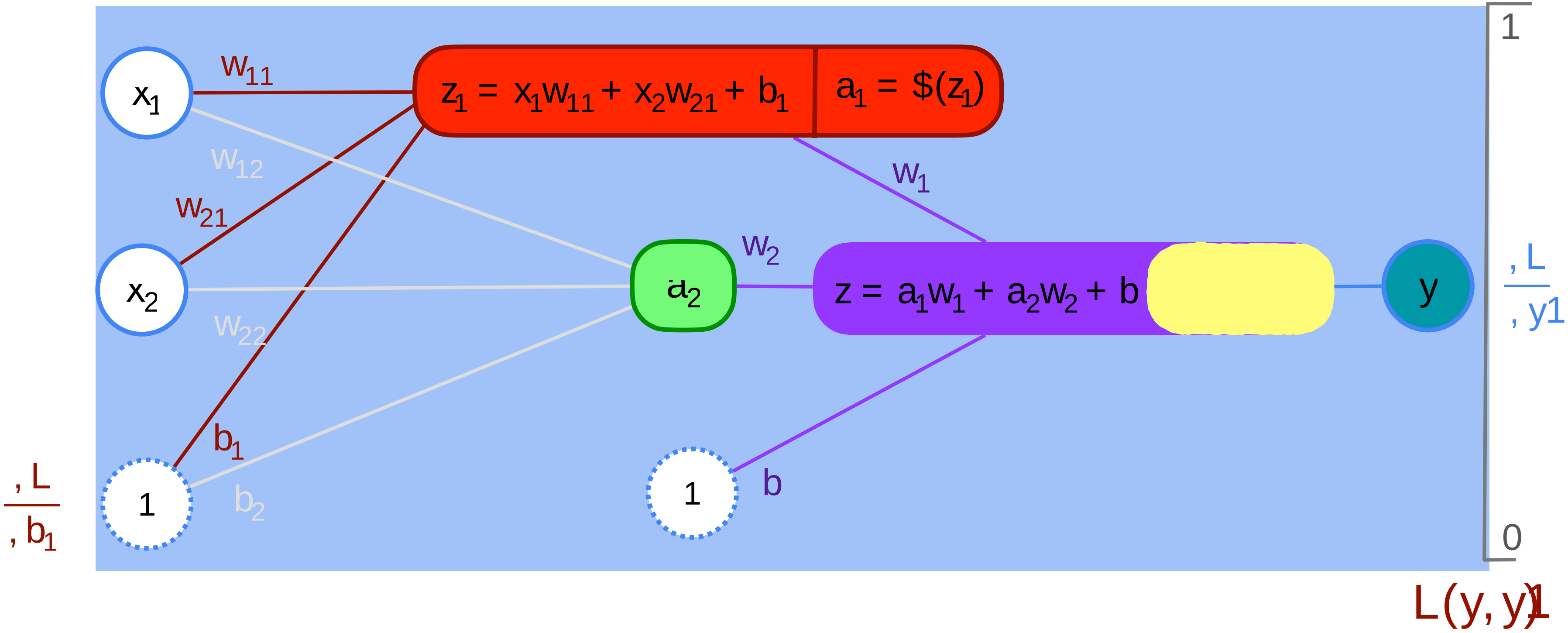
2,2,1 Neural Network



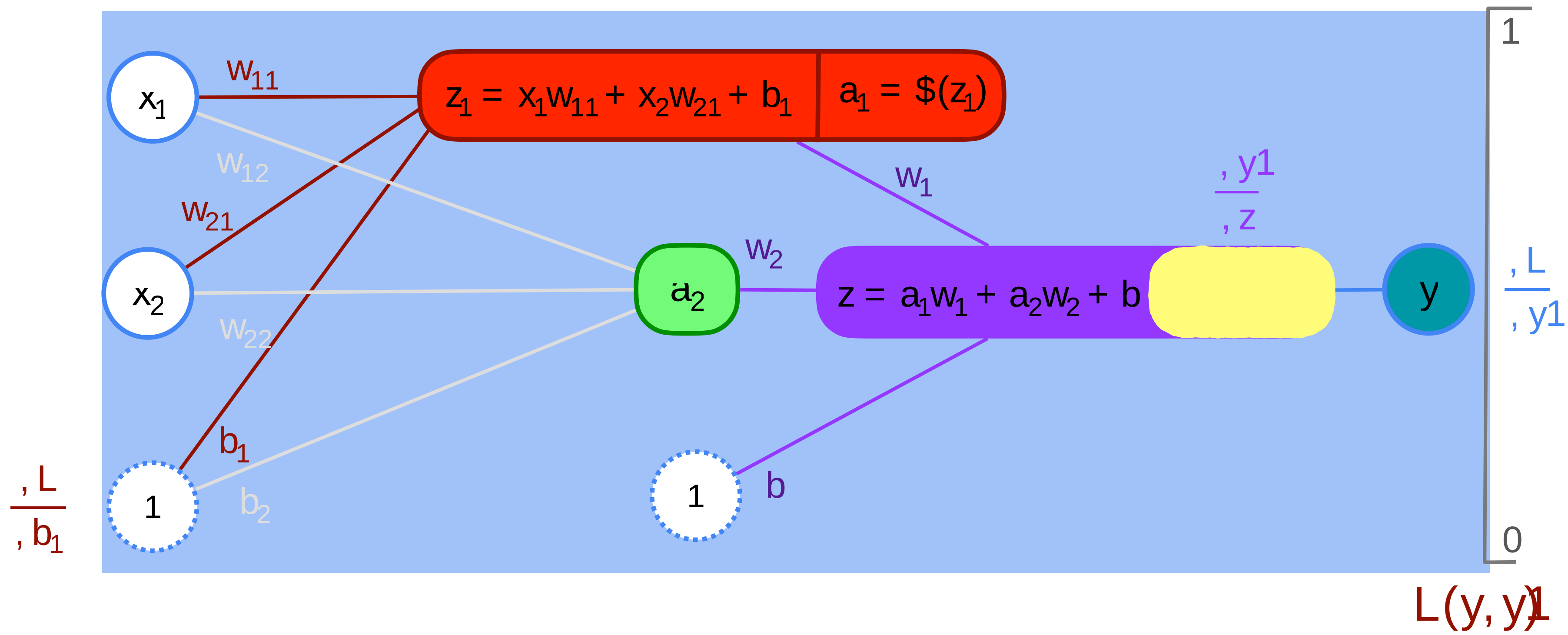
2,2,1 Neural Network



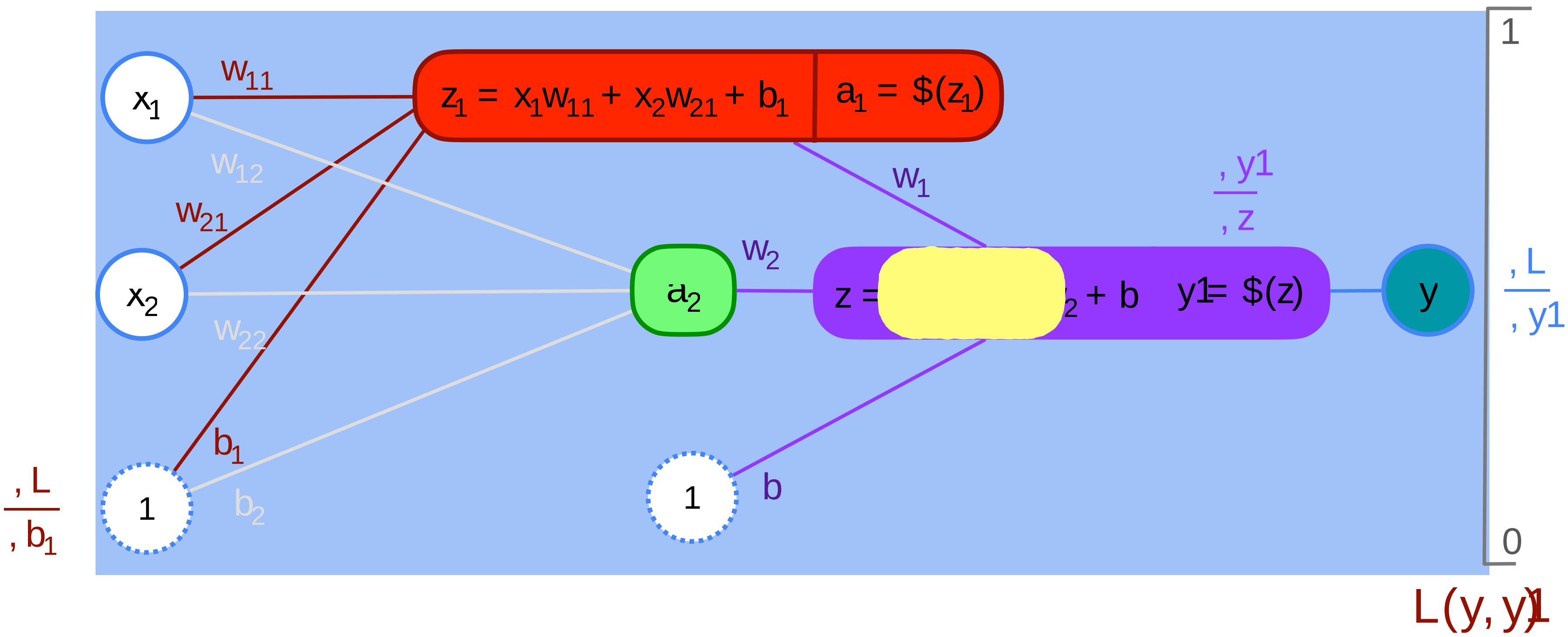
2,2,1 Neural Network



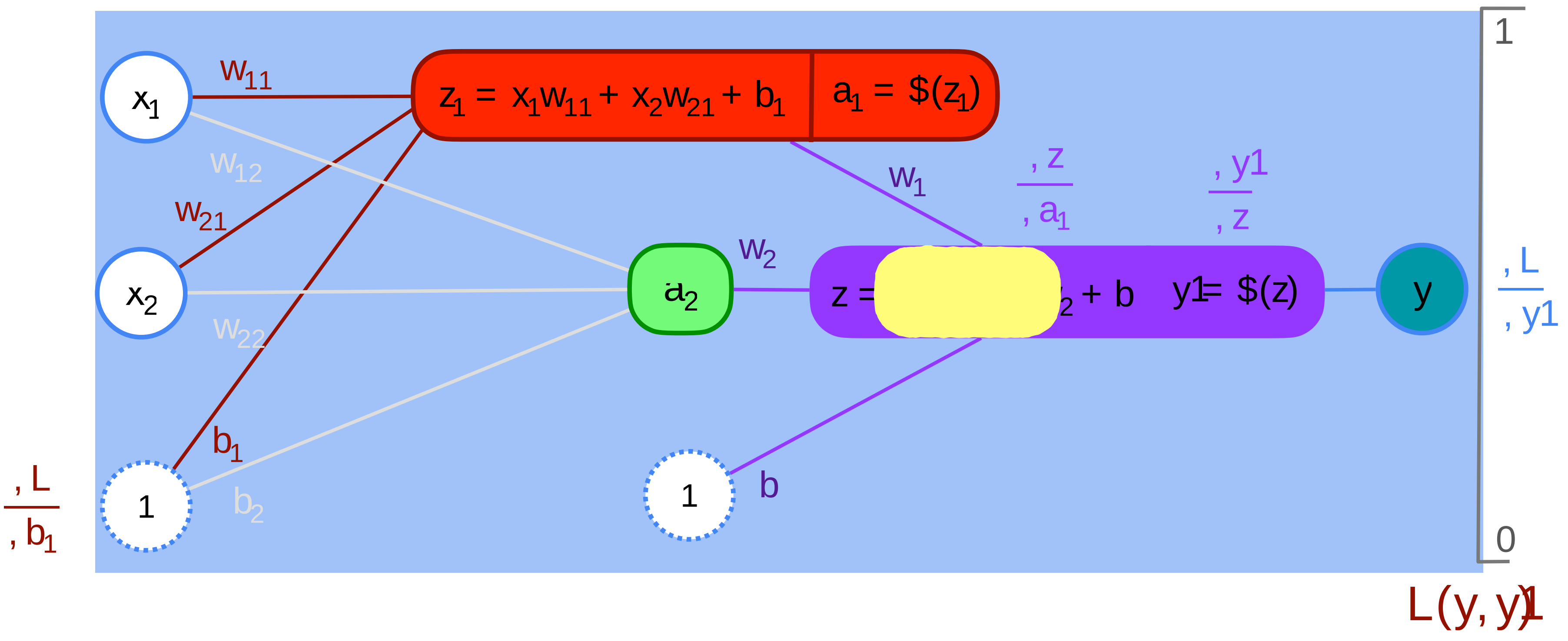
2,2,1 Neural Network



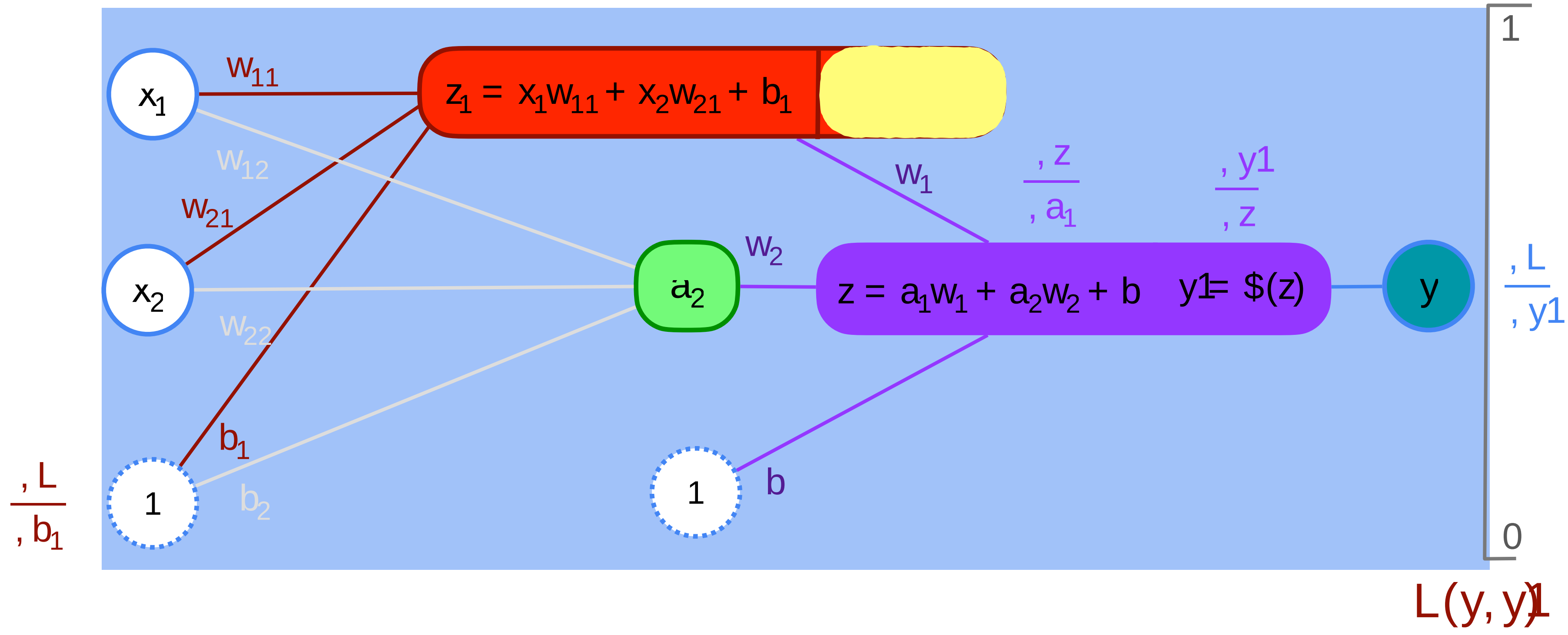
2,2,1 Neural Network



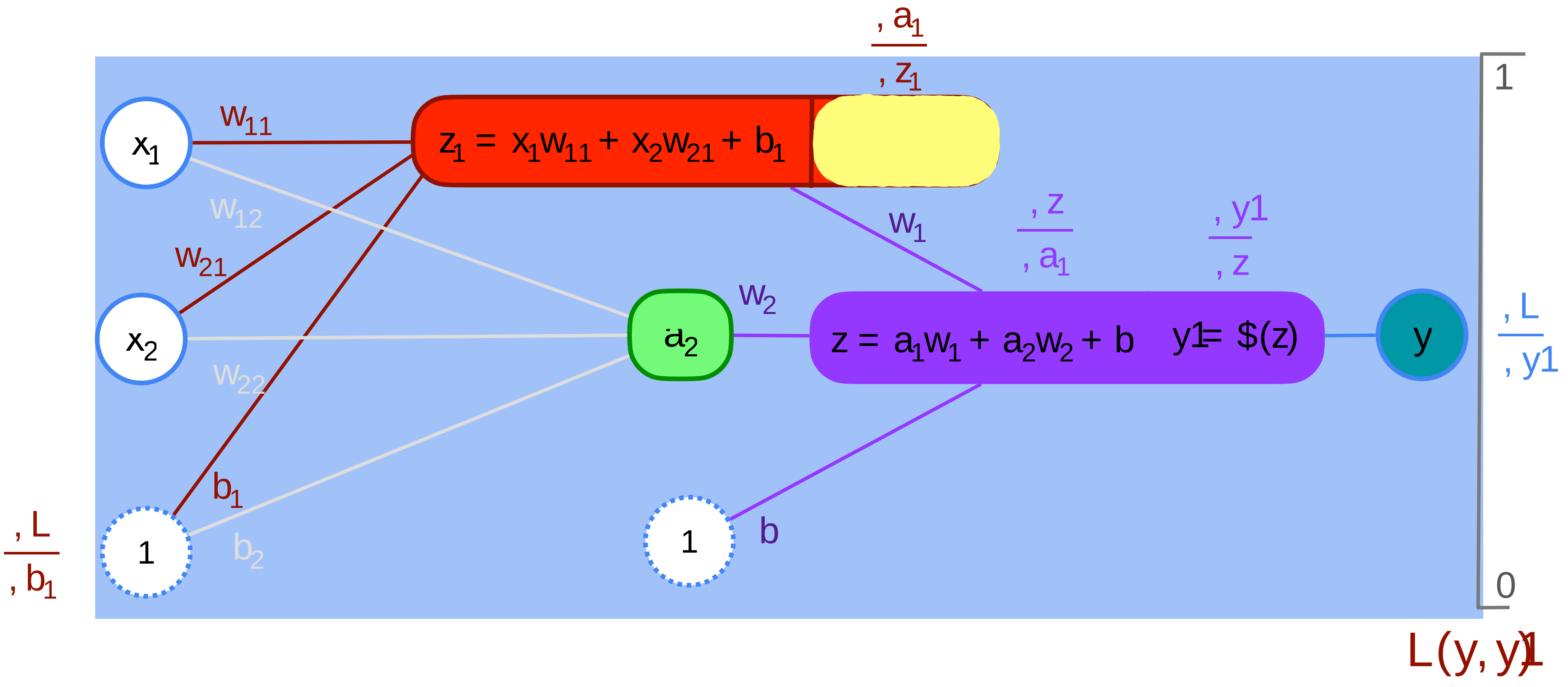
2,2,1 Neural Network



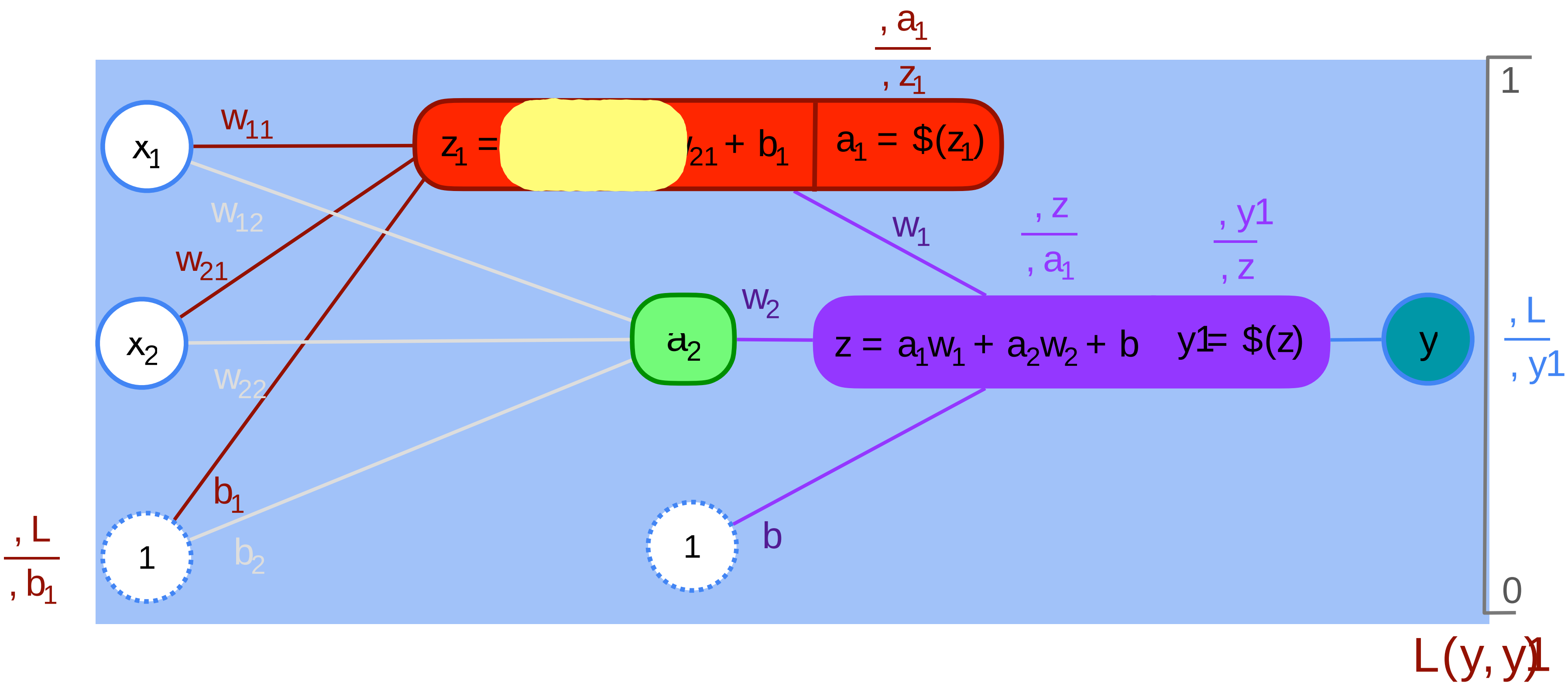
2,2,1 Neural Network



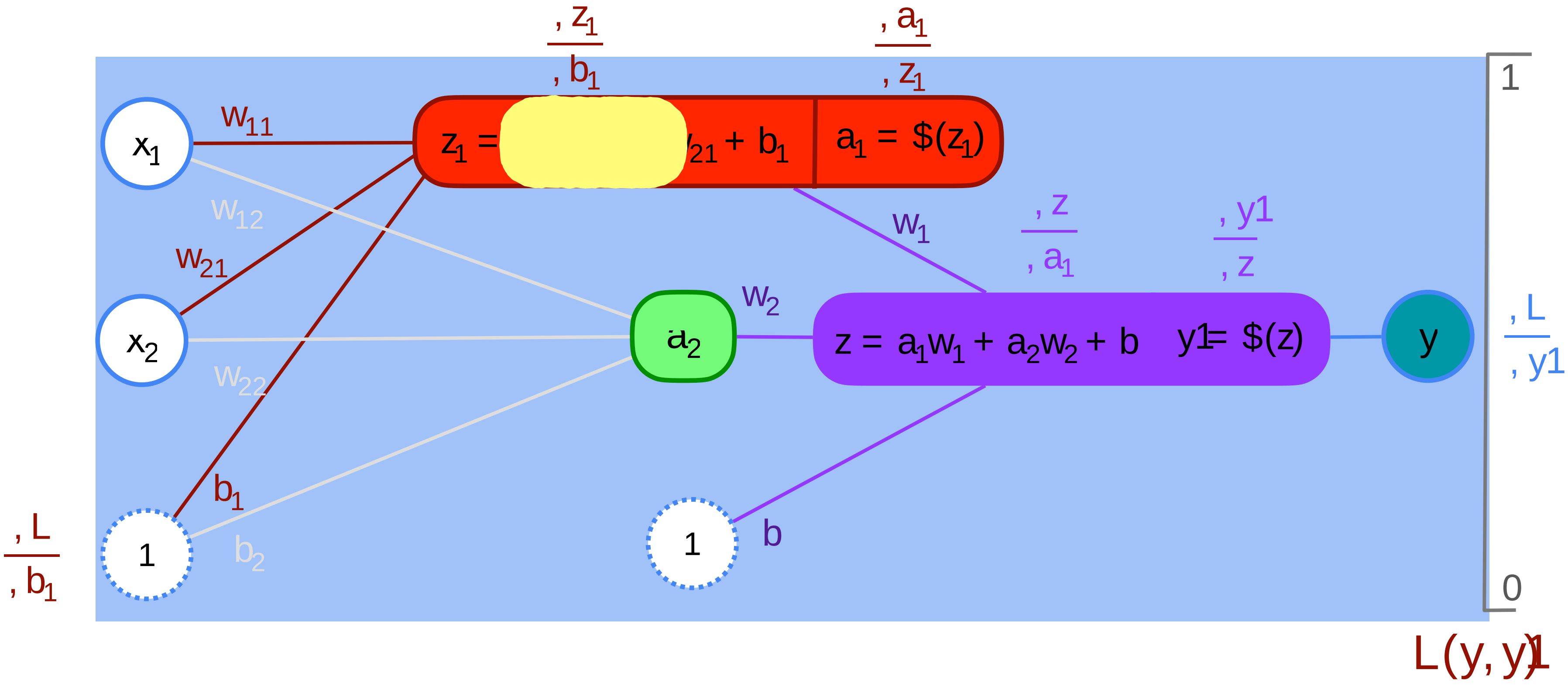
2,2,1 Neural Network



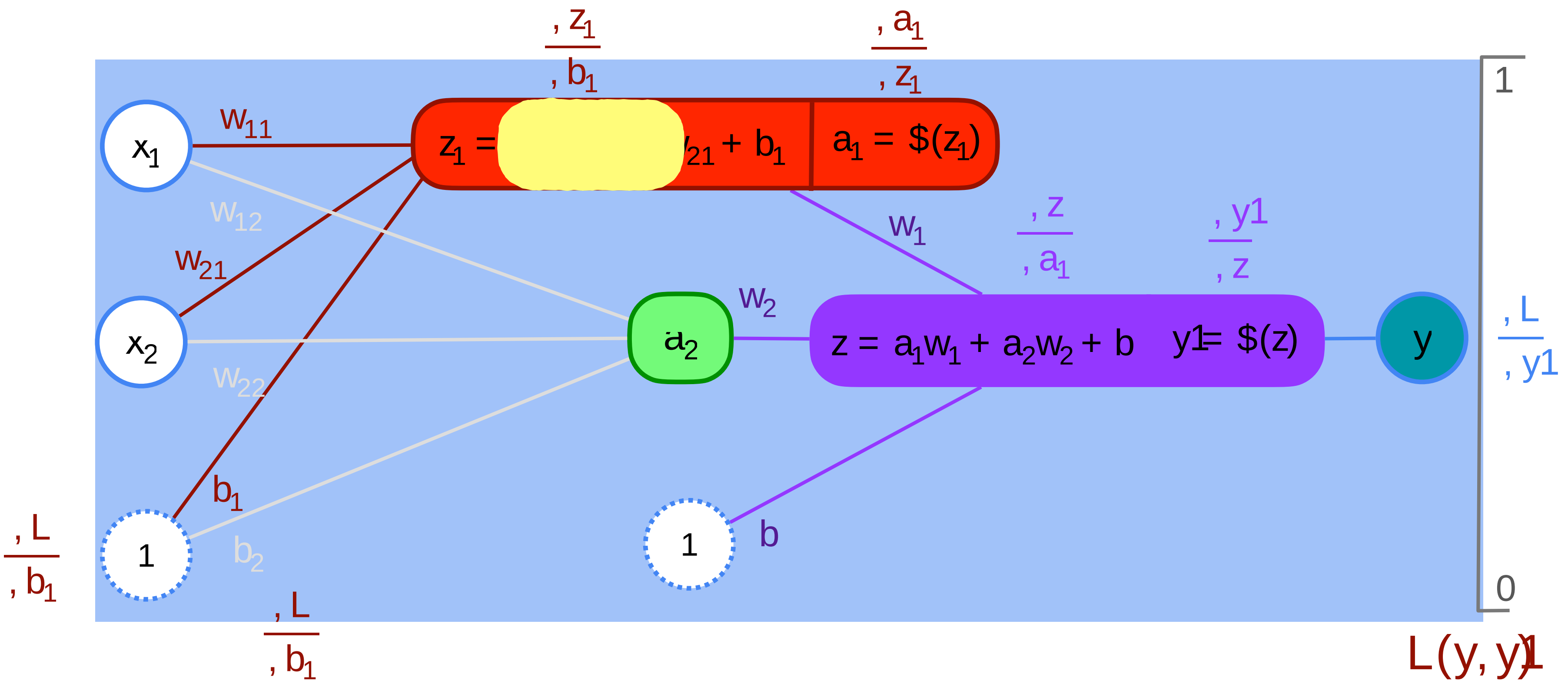
2,2,1 Neural Network



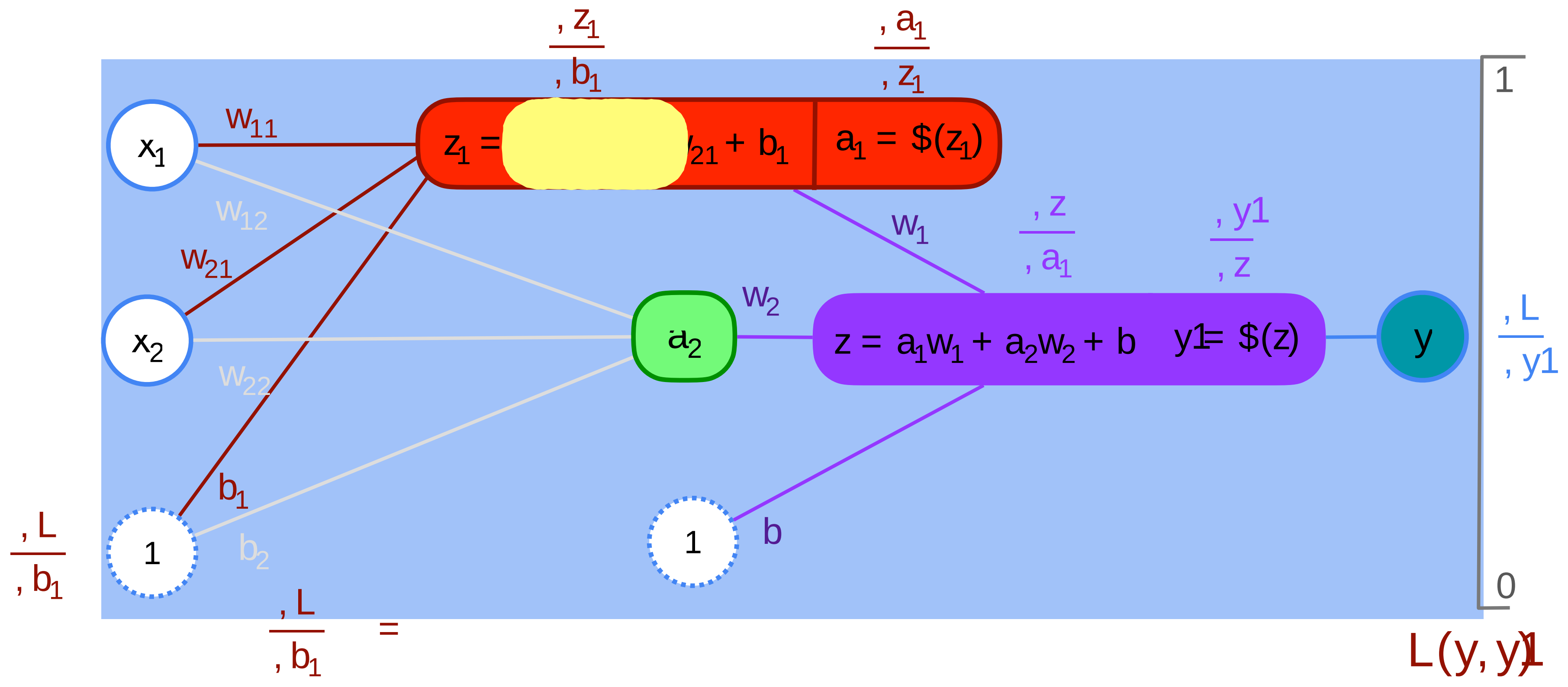
2,2,1 Neural Network



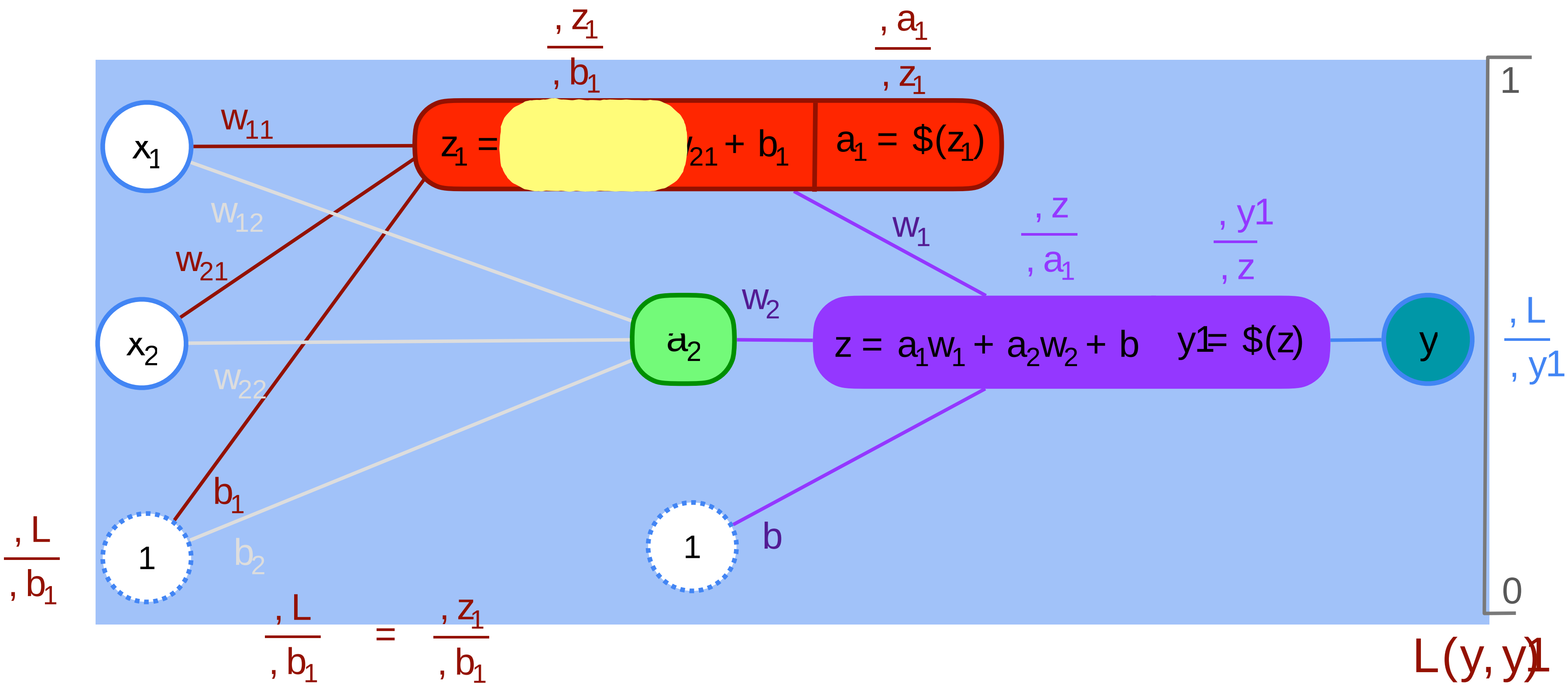
2,2,1 Neural Network



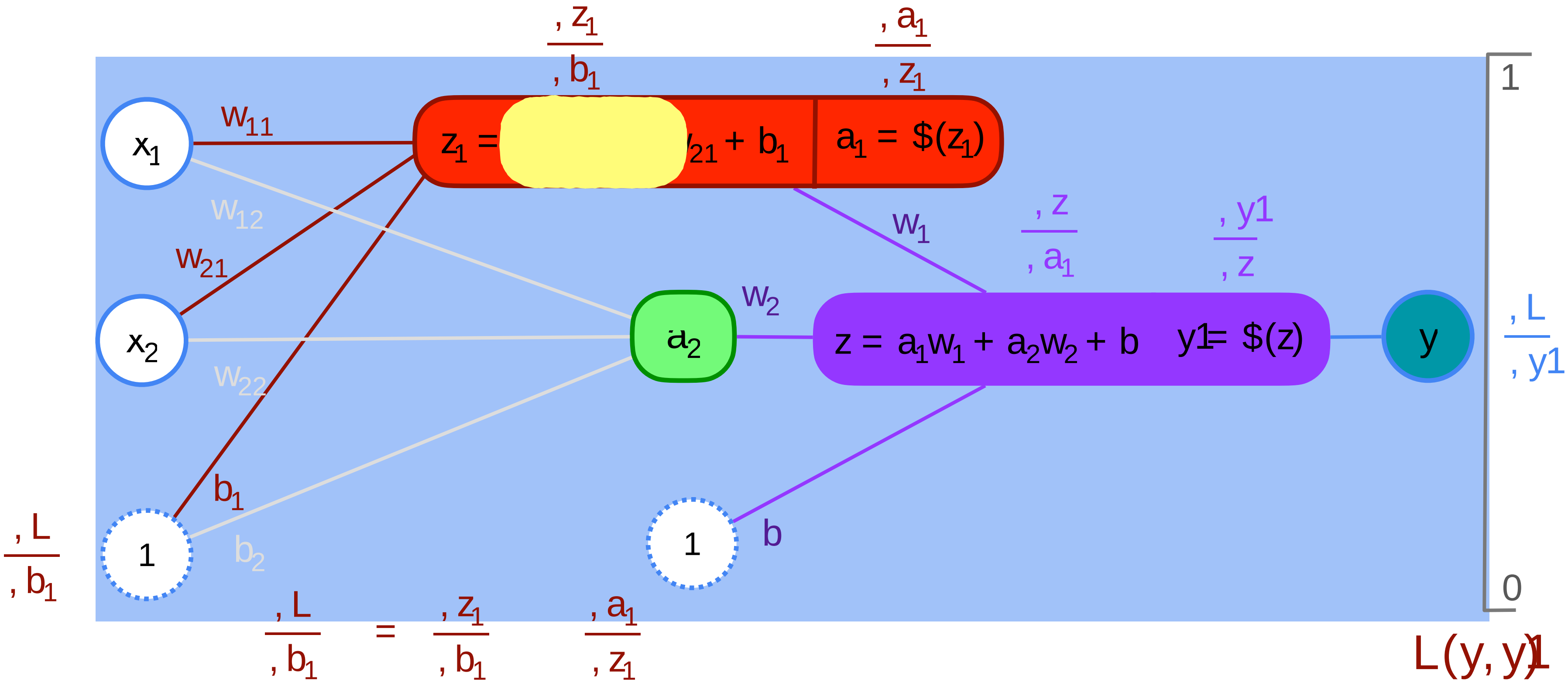
2,2,1 Neural Network



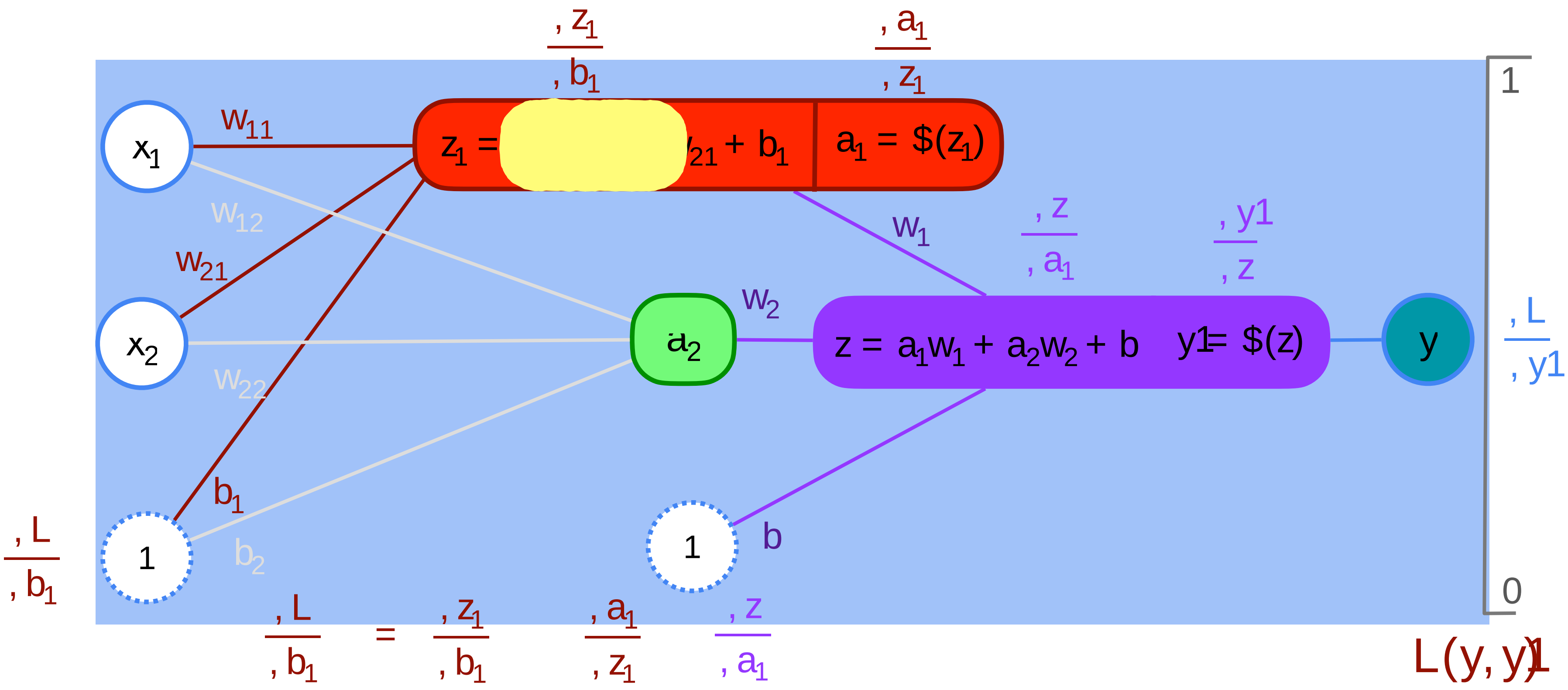
2,2,1 Neural Network



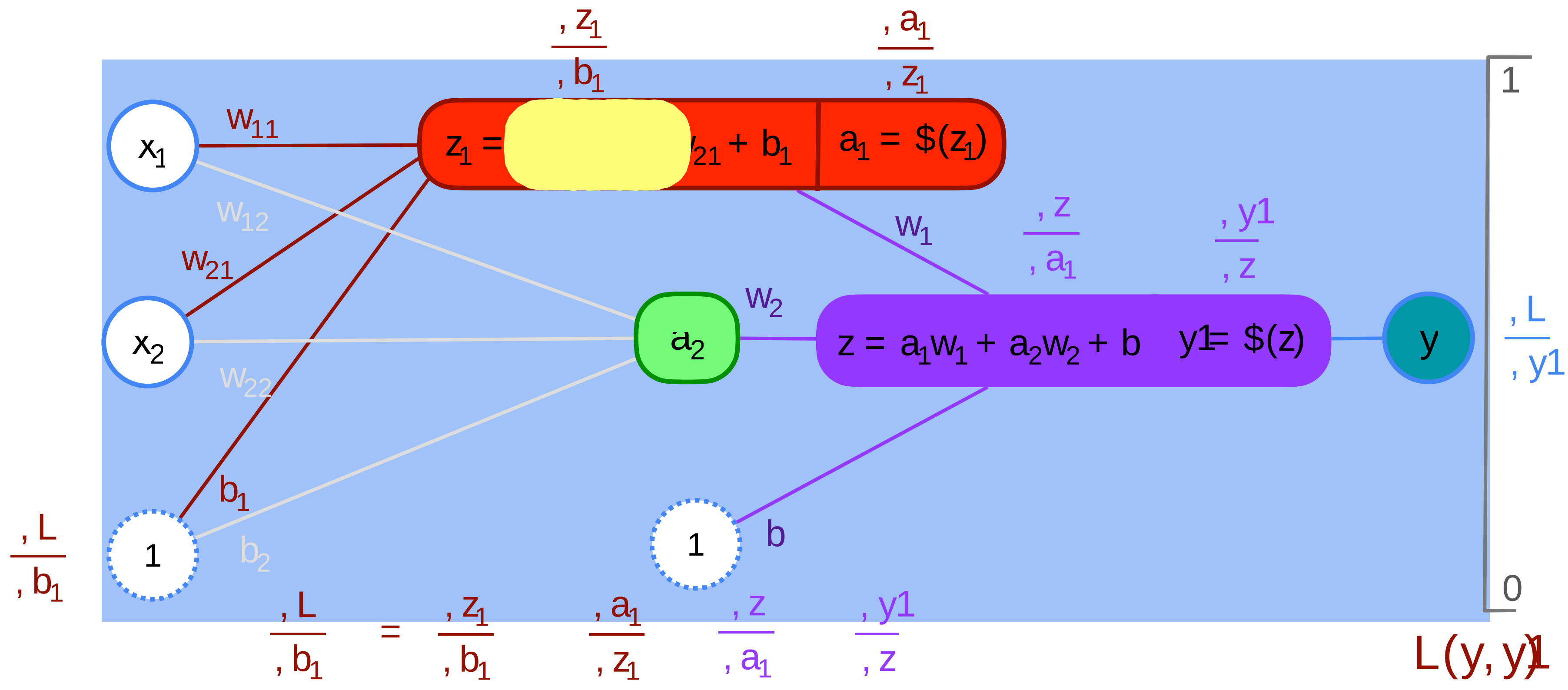
2,2,1 Neural Network



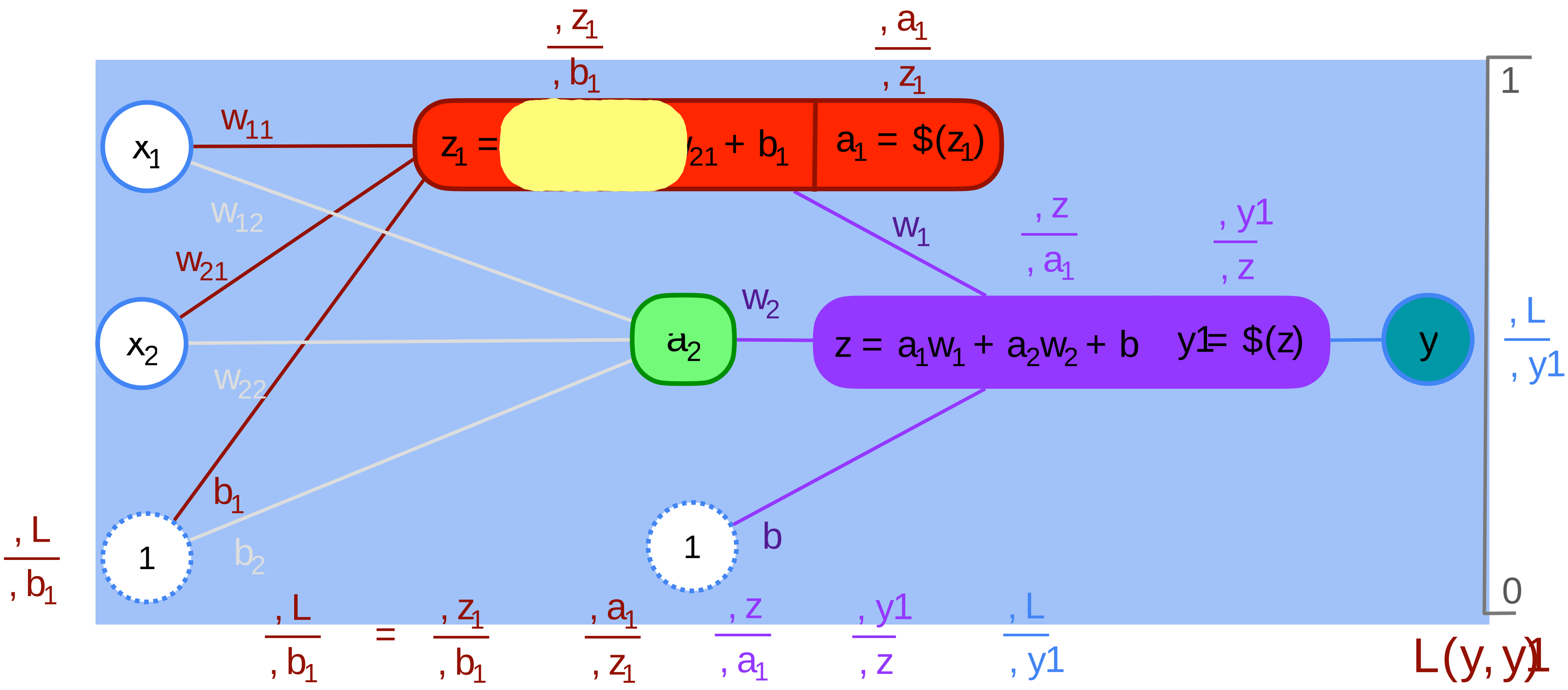
2,2,1 Neural Network



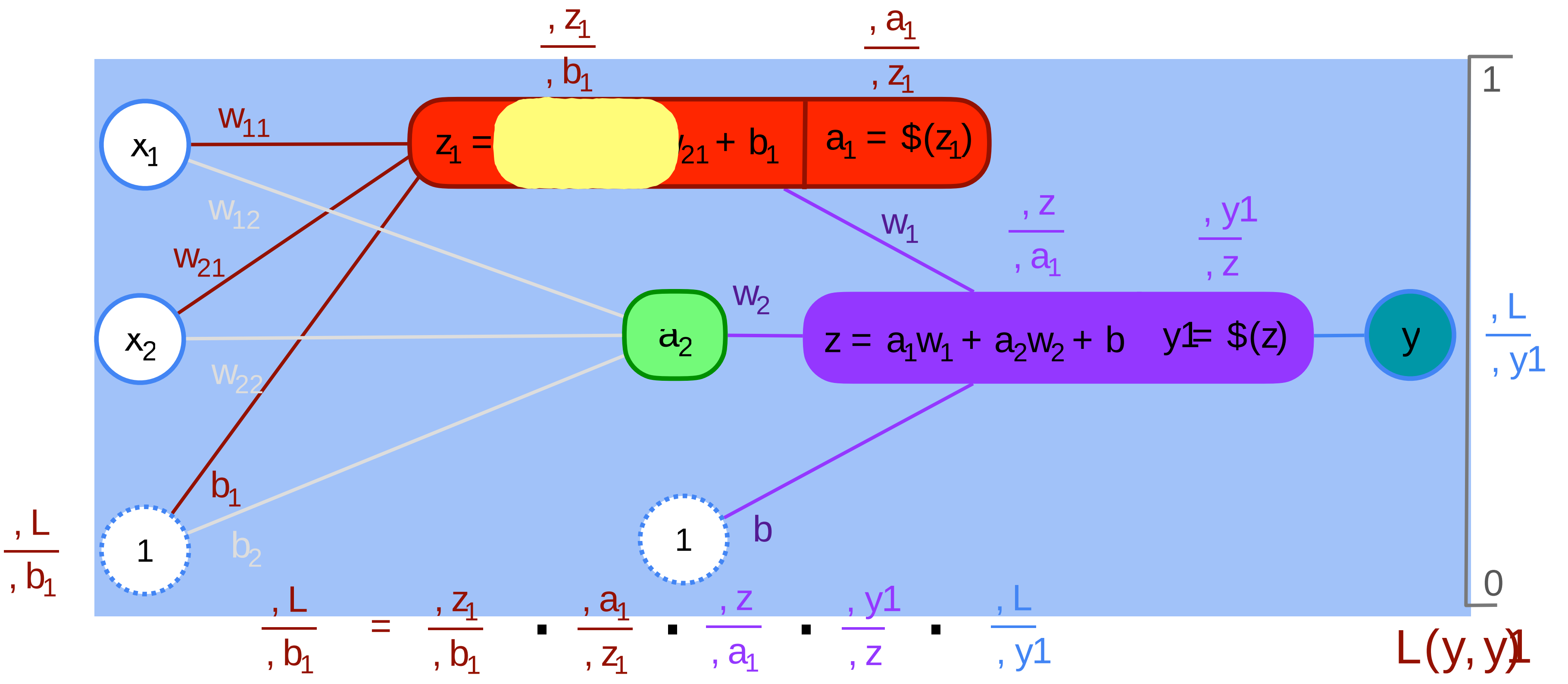
2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1)\ "\ (1"\ y)\log(1"\ y1)$

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1)\text{ " }(1\text{ " }y)\log(1\text{ " }y1)$

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,b_1}$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

a_1 = $($z_1$)

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,\ L}{,\ b_1} = \frac{,\ z_1}{,\ b_1} \cdot \frac{,\ a_1}{,\ z_1} \cdot \frac{,\ z}{,\ a_1} \cdot \frac{,\ y1}{,\ z} \cdot \frac{,\ L}{,\ y1}$$

$$\frac{,\ L}{,\ b_1} =$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

a_1 = $($z_1$)

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$\frac{, L}{, b_1} = \quad \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$

$\frac{, L}{, b_1} =$

2,2,1 Neural Network

L(y, y1)= -ylog(y1) " (1" y)log(1" y1)

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

a_1 = $($z_1$)

$$\frac{, L}{, b_1} = \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, b_1} =$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1)\ " \ (1" y)\log(1" y1)$

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$$\frac{, L}{, b_1} = \quad \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, b_1} = 1$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

a_1 = $($z_1$)

$$\frac{, L}{, b_1} = \frac{, z_1}{, b_1} \cdot \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, b_1} = 1$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1 \text{"} (1 \text{"} y)\log(1 \text{"} y1$

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,b_1} = 1$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1 - y)\log(1 - \hat{y})$$

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \quad \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1 - a_1)$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,b_1} = 1 \quad a_1(1 " \ a_1)$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1\ \text{"}\ (1\text{"}\ y)\log(1\text{"}\ y1$

$y1 = \$(z)$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, b_1} = \frac{, z_1}{, b_1} \cdot \frac{, a_1}{, z_1} \cdot \qquad \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, b_1} = 1 \quad a_1(1\text{"}\ a_1)$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) \text{ " } (1\text{"} y)\log(1\text{"} y1)$

$y1 = \$(z)$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \quad \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,b_1} = 1 \quad a_1(1\text{"} \; a_1) \quad w_1$$

2,2,1 Neural Network

L(y, y1= -ylog(y1 " (1" y)log(1" y1

y1= $(z)

a_1 = $(z_1)

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \quad \frac{,L}{,y1}$$

$$\frac{,L}{,b_1} = 1 \quad a_1(1 " \ a_1) \quad w_1$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1 " (1" y)log(1" y1

$$\frac{, L}{, b_1} = \frac{, z_1}{, b_1} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, b_1} = 1 \quad a_1(1 \text{"} \; a_1) \quad w_1$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \$(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

2,2,1 Neural Network

L(y, y1)= -ylog(y1 " (1" y)log(1" y1

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \quad \frac{,L}{,y1}$$

$$\frac{,L}{,b_1} = 1 \quad a_1(1" \; a_1) \quad w_1 \quad y1(1" \; y1)$$

2,2,1 Neural Network

L(y, y1)= -ylog(y1 " (1" y)log(1" y1

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,b_1} = 1 \quad a_1(1" \; a_1) \quad w_1 \quad y1(1" \; y1)$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,b_1} = 1 \quad a_1(1\text{"}\ a_1) \quad w_1 \quad y1(1\text{"}\ y1)$$

2,2,1 Neural Network

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1 - a_1) \quad w_1 \quad \hat{y}(1 - \hat{y}) \quad \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

y1 = \$(z)

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, b_1} = \frac{, z_1}{, b_1} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, b_1} = 1 \cdot a_1(1 \text{"} a_1) \cdot w_1 \cdot y1(1 \text{"} y1) \cdot \frac{-(y \text{"} y1)}{y1(1 \text{"} y1)}$$

2,2,1 Neural Network

$y1 = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial y1}{\partial z} \cdot$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{y1(1 - y1)} \cdot \frac{-(y - y1)}{y1(1 - y1)}$$

2,2,1 Neural Network

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

2,2,1 Neural Network

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

$$= -w_1 a_1 (1 - a_1)(y - \hat{y})$$

2,2,1 Neural Network

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

a_1 = $($z_1$)

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{,\,L}{,\,b_1} = \frac{,\,z_1}{,\,b_1} \cdot \frac{,\,a_1}{,\,z_1} \cdot \frac{,\,z}{,\,a_1} \cdot \frac{,\,y1}{,\,z} \cdot$$

$$\frac{,\,L}{,\,b_1} = 1 \cdot a_1(1\text{ " }a_1) \cdot w_1 \cdot \cancel{y1(1\text{ " }y1)} \cdot \frac{-(y\text{ " }y1)}{\cancel{y1(1\text{ " }y1)}}$$

$$= -w_1a_1(1\text{ " }a_1)(y\text{ " }y1)$$

Perform gradient descent with

to find optimal value of b_1 that gives the least error

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{, L}{, b_1} = \frac{, z_1}{, b_1} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, b_1} = 1 \cdot a_1(1 \text{"} \; a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"} \; y1)} \cdot \frac{-(y \text{"} \; y1)}{\cancel{y1(1 \text{"} \; y1)}}$$

$$= -w_1 a_1 (1 \text{"} \; a_1)(y \text{"} \; y1)$$

Perform gradient descent with

$$b_1 \; \$ \; b_1 \text{"} \; \# \frac{, L}{, b_1}$$

to find optimal value of b_1 that gives the least error

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{,L}{,b_1} = \frac{,z_1}{,b_1} \cdot \frac{,a_1}{,z_1} \cdot \frac{,z}{,a_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,b_1} = 1 \cdot a_1(1 \text{"}\ a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"}\ y1)} \cdot \frac{-(y \text{"}\ y1)}{\cancel{y1(1 \text{"}\ y1)}}$$

$$= -w_1 a_1 (1 \text{"}\ a_1)(y \text{"}\ y1)$$

Perform gradient descent with

$$b_1 \ \$ \ b_1 \text{"}\ \#$$

to find optimal value of b_1 that gives the least error

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \$(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{, L}{, b_1} = \frac{, z_1}{, b_1} \cdot \frac{, a_1}{, z_1} \cdot \frac{, z}{, a_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, b_1} = 1 \cdot a_1(1 \text{"} \; a_1) \cdot w_1 \cdot \cancel{y1(1 \text{"} \; y1)} \cdot \frac{-(y \text{"} \; y1)}{\cancel{y1(1 \text{"} \; y1)}}$$

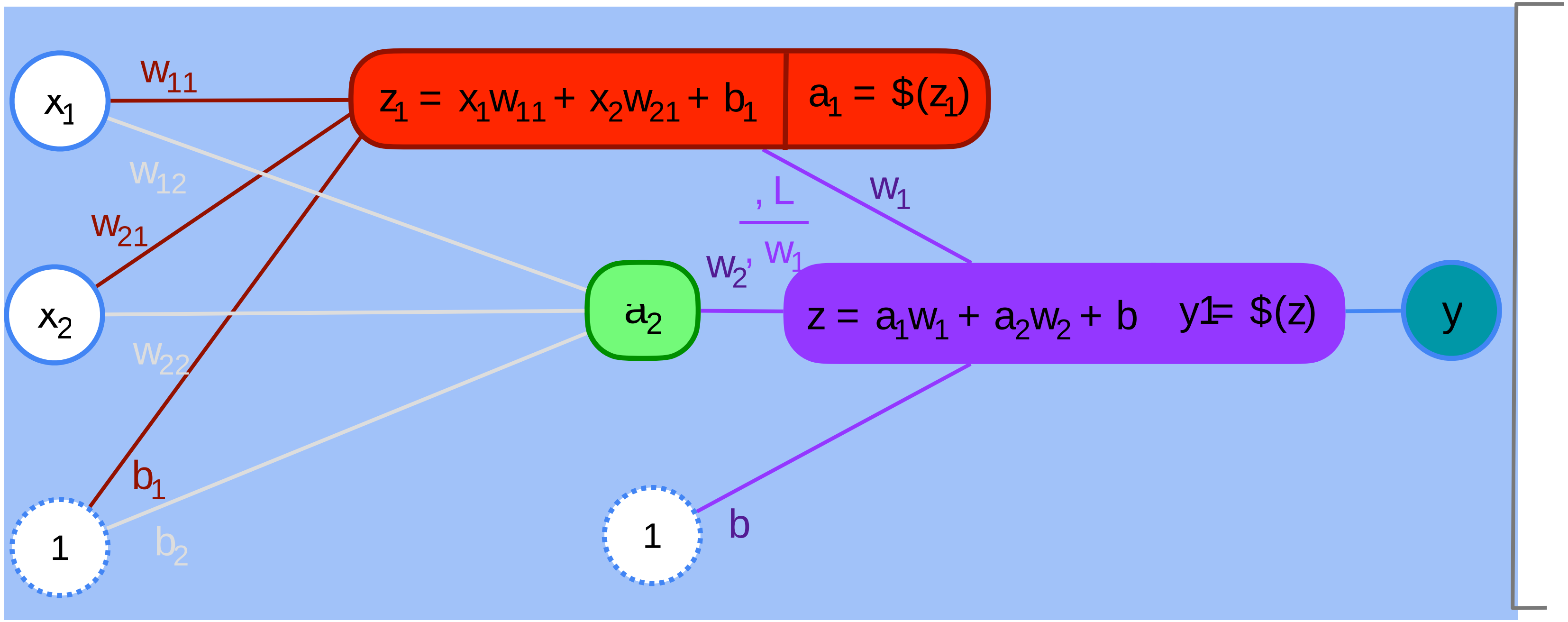
$$= -w_1a_1(1 \text{"} \; a_1)(y \text{"} \; y1)$$

Perform gradient descent with

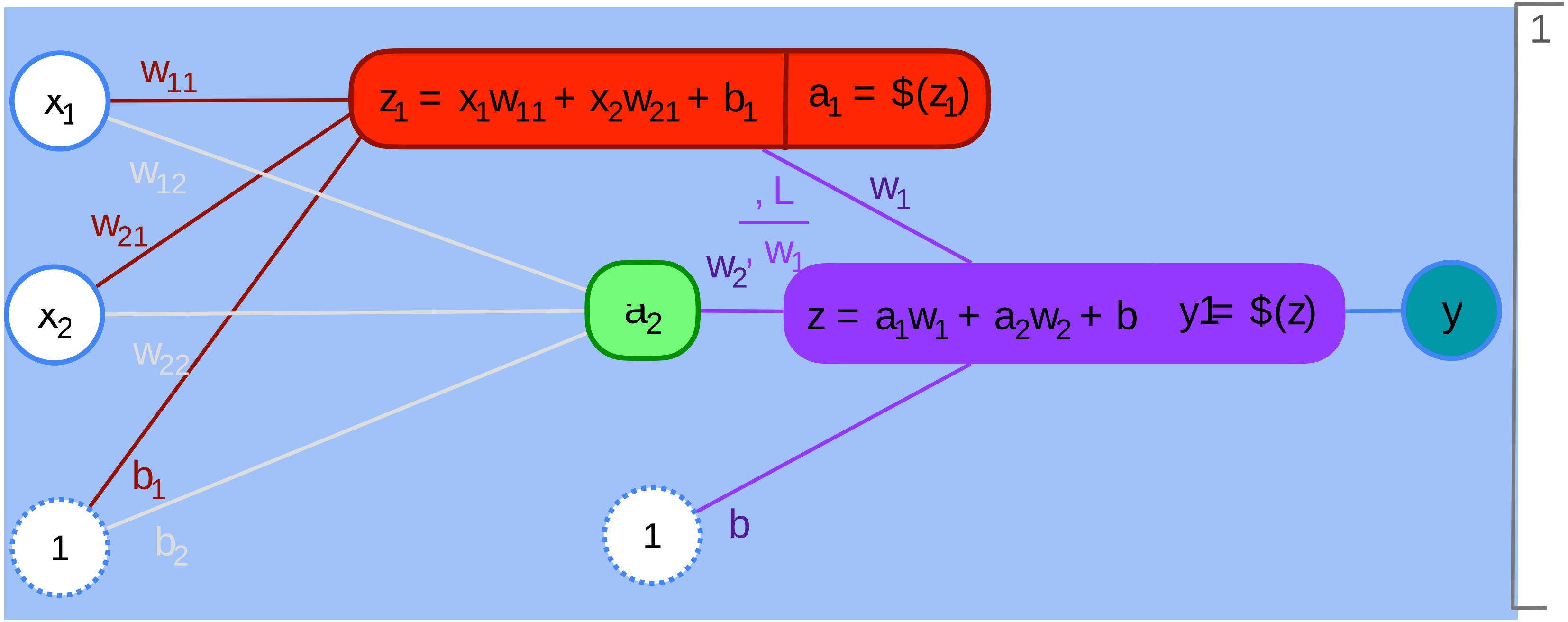
$$b_1 \; \$ \; b_1 \text{"} \; \#(-w_1a_1(1 \text{"} \; a_1)(y \text{"} \; y1))$$

to find optimal value of b_1 that gives the least error

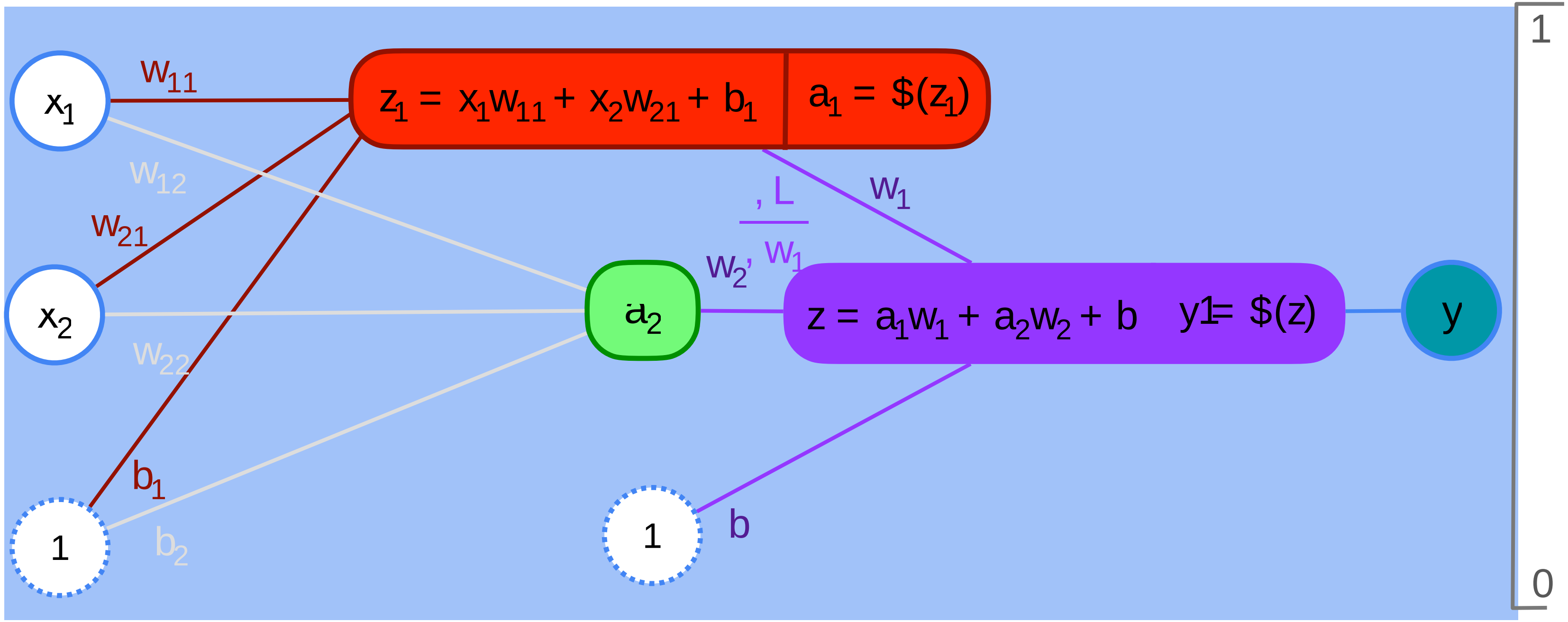
2,2,1 Neural Network



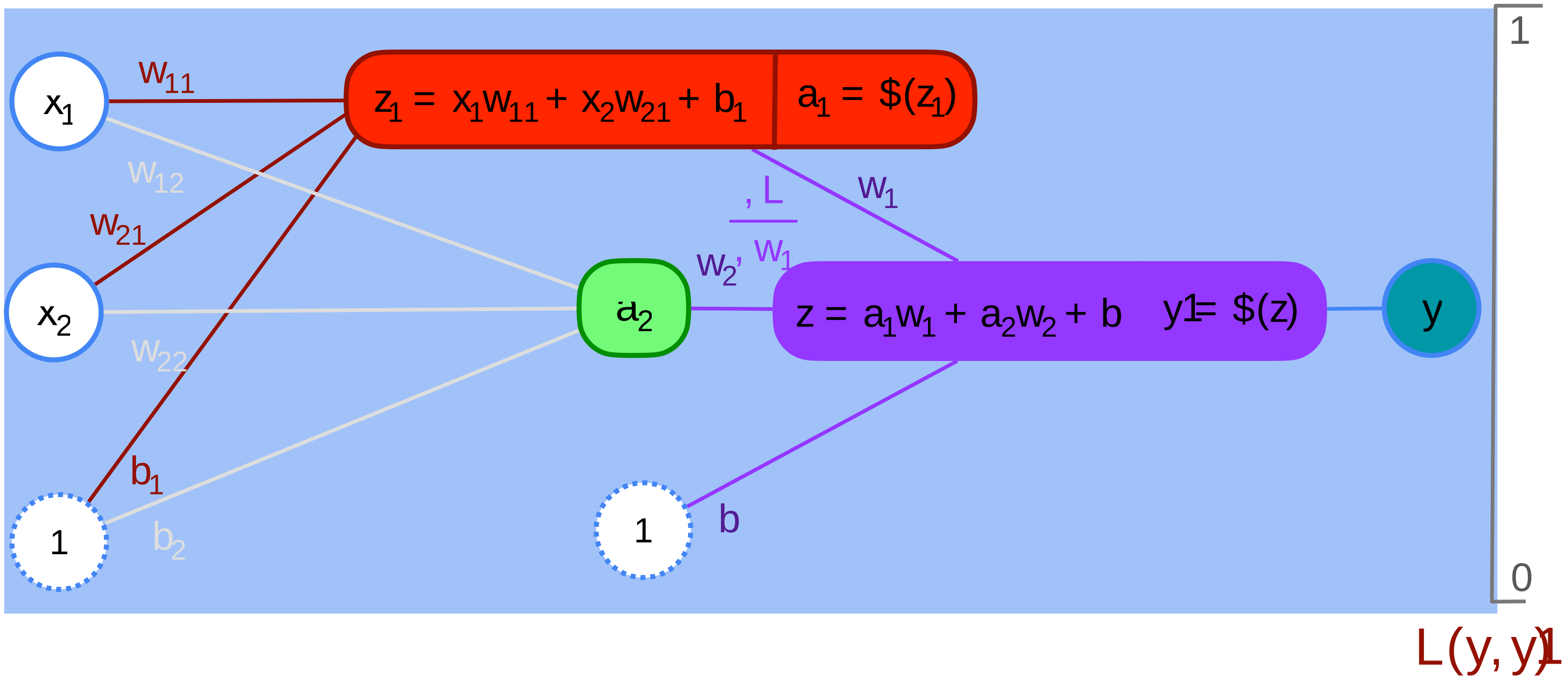
2,2,1 Neural Network



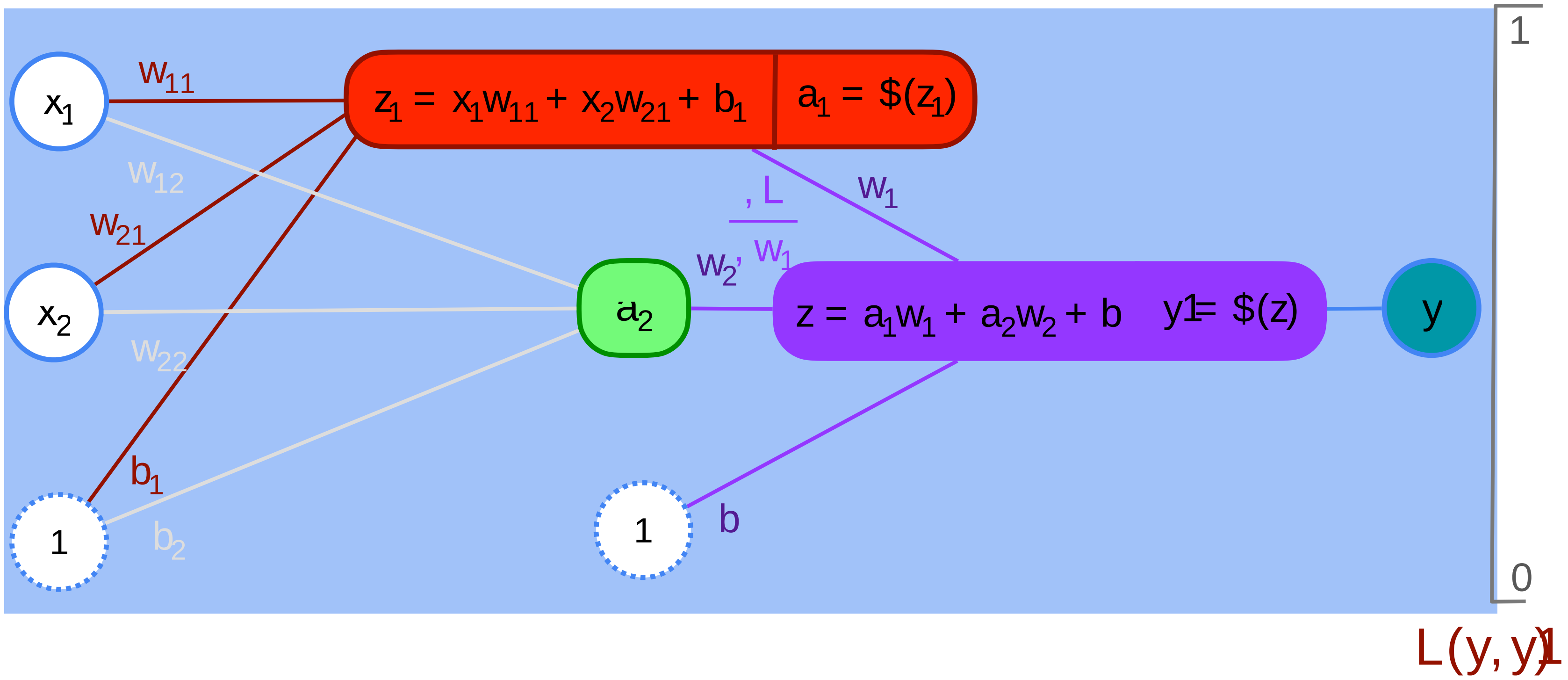
2,2,1 Neural Network



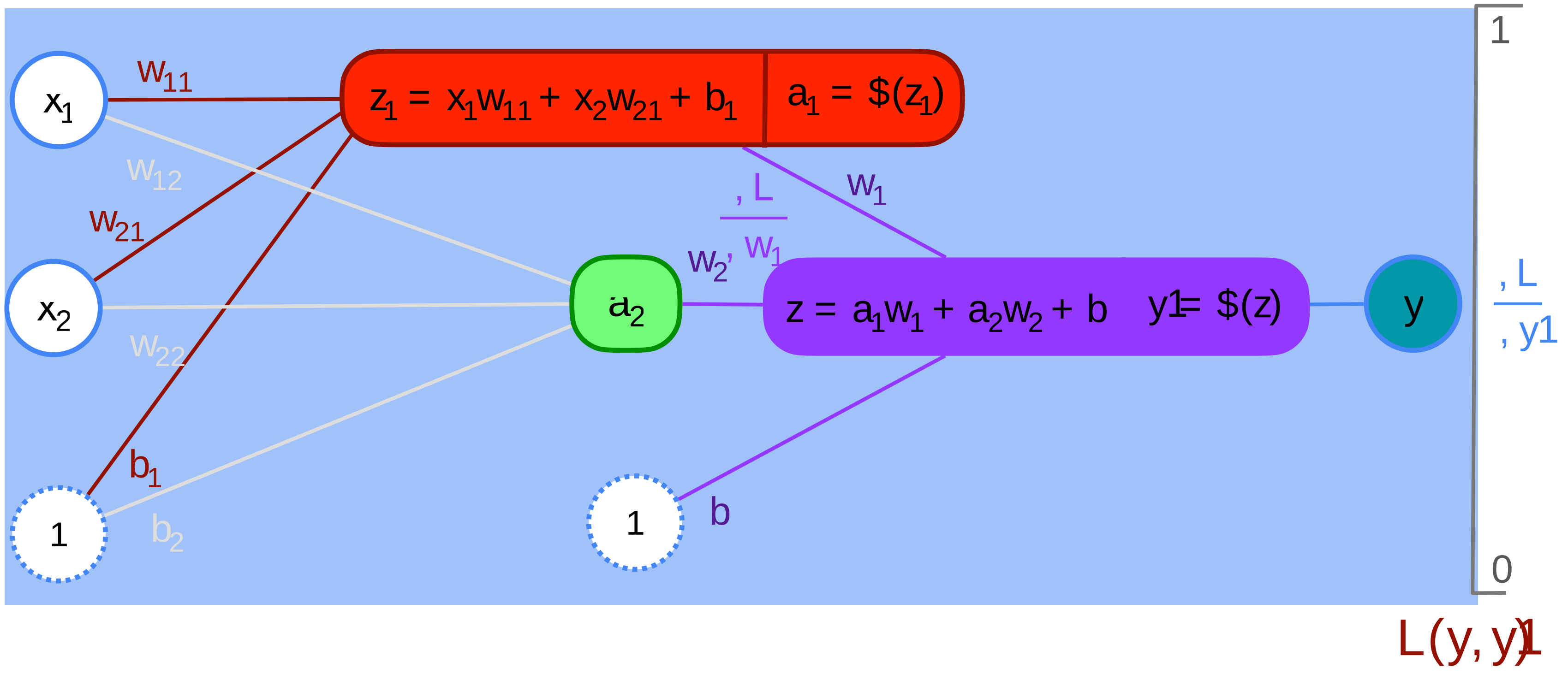
2,2,1 Neural Network



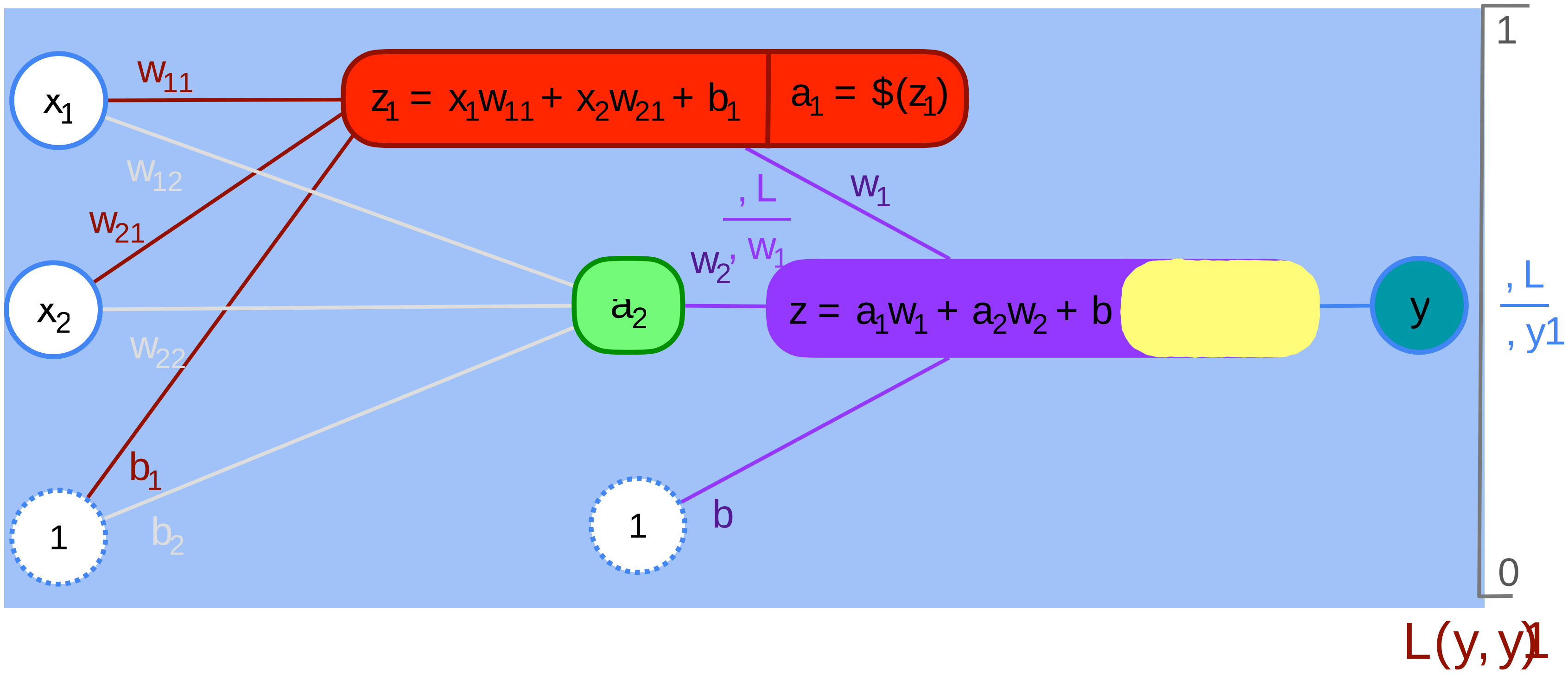
2,2,1 Neural Network



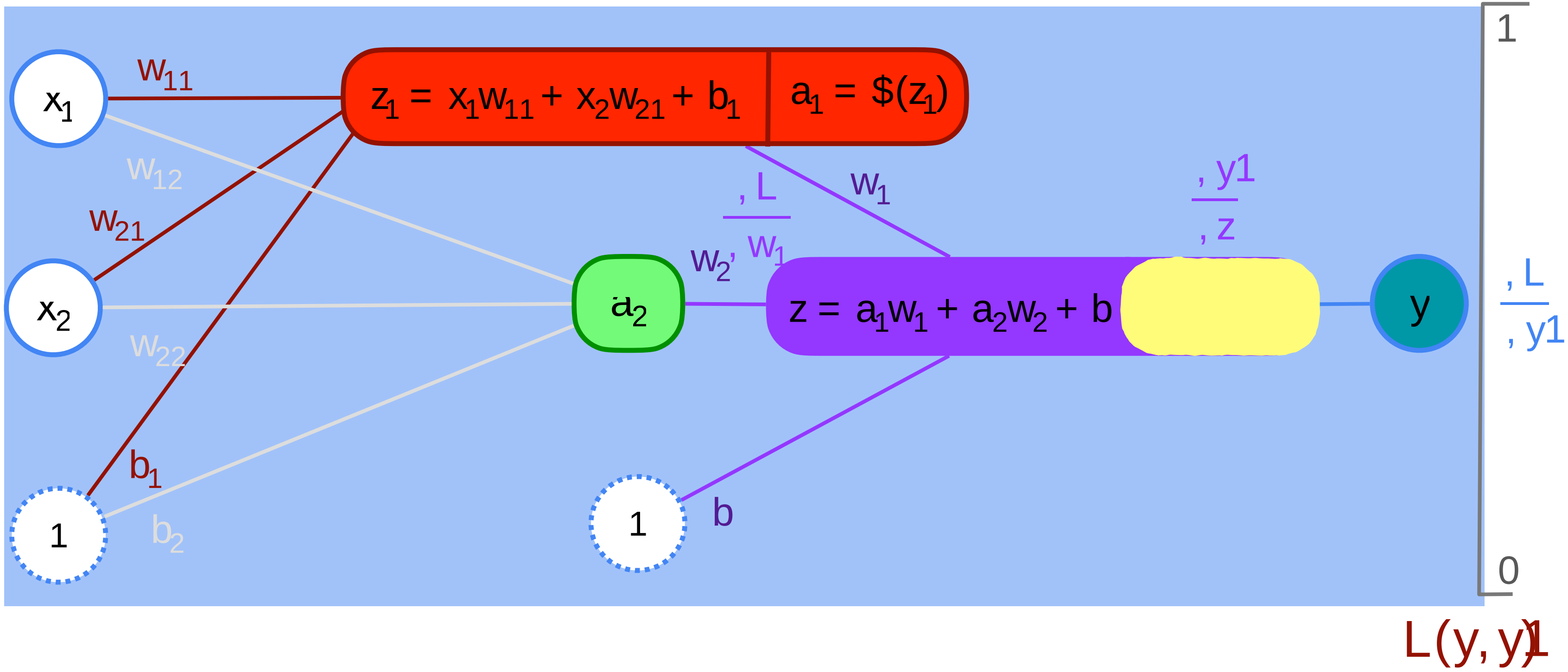
2,2,1 Neural Network



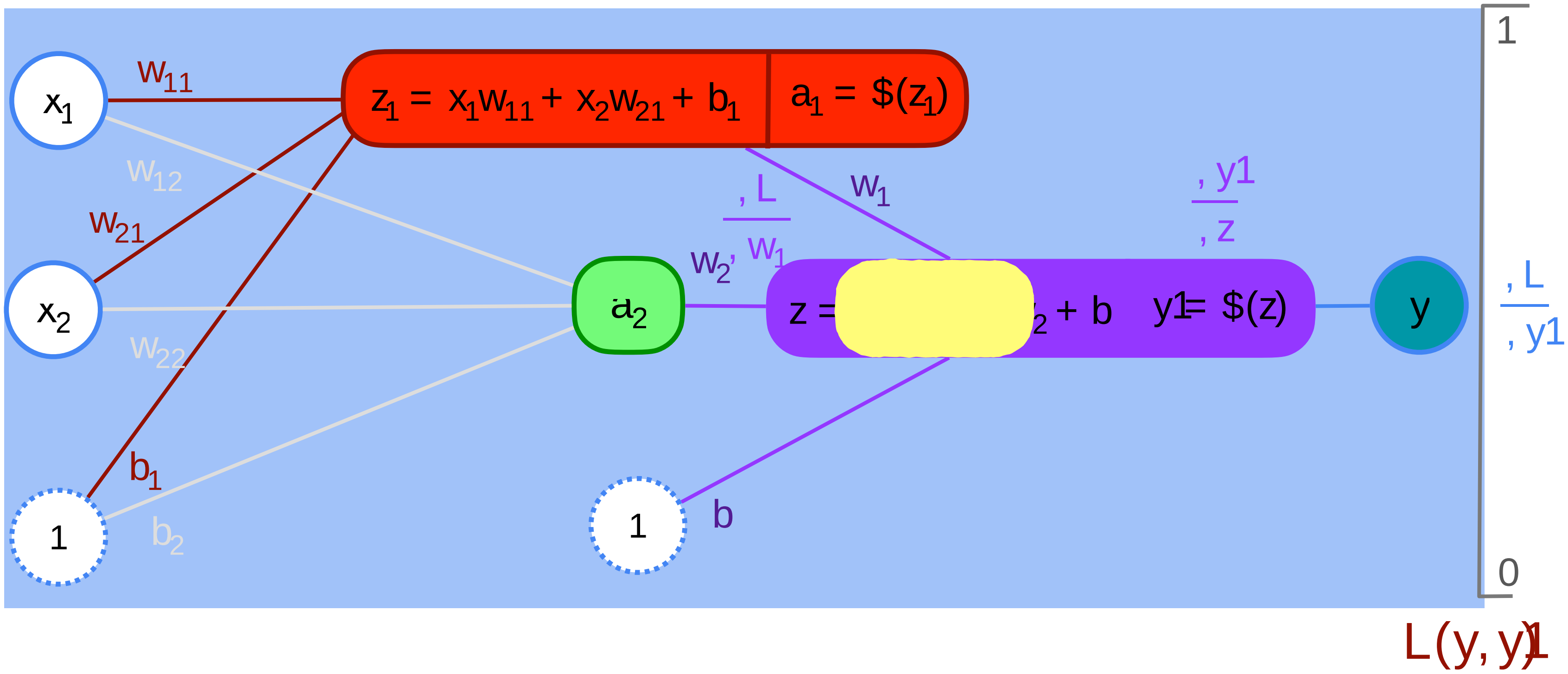
2,2,1 Neural Network



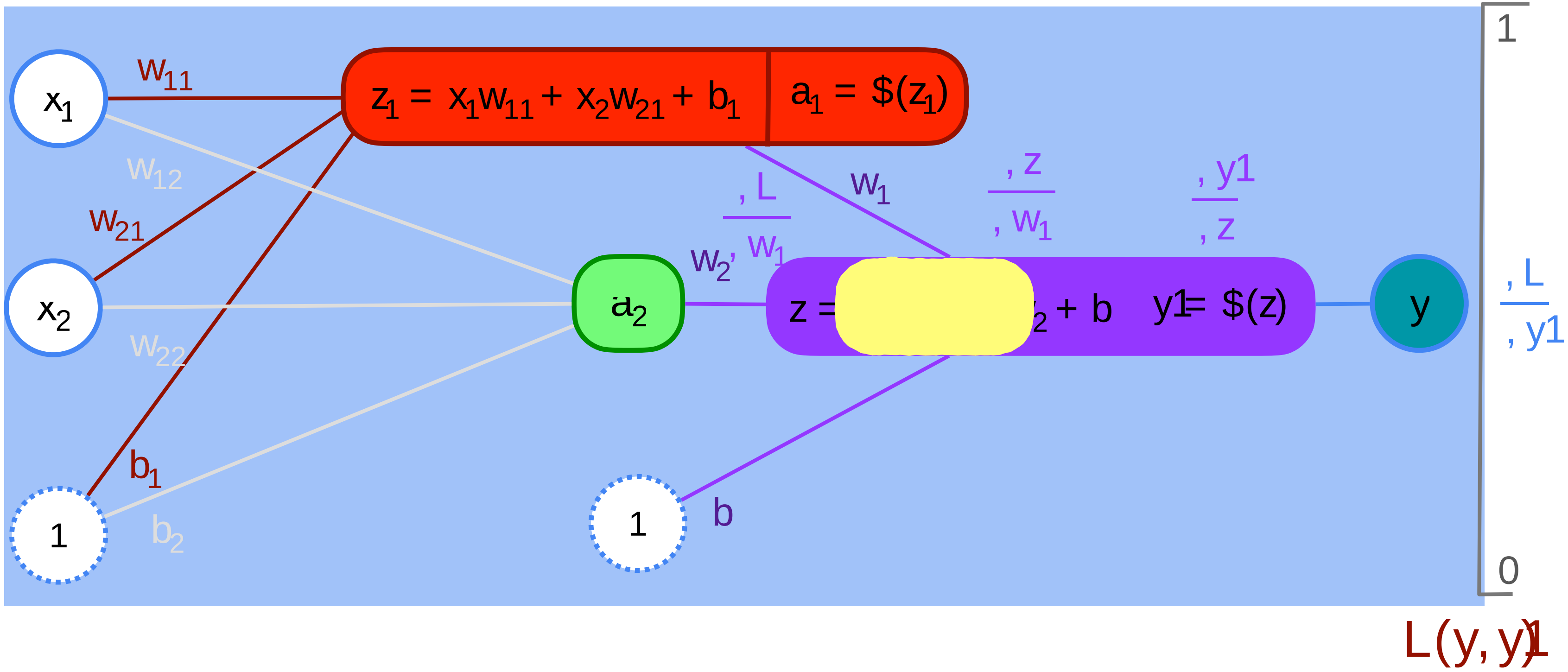
2,2,1 Neural Network



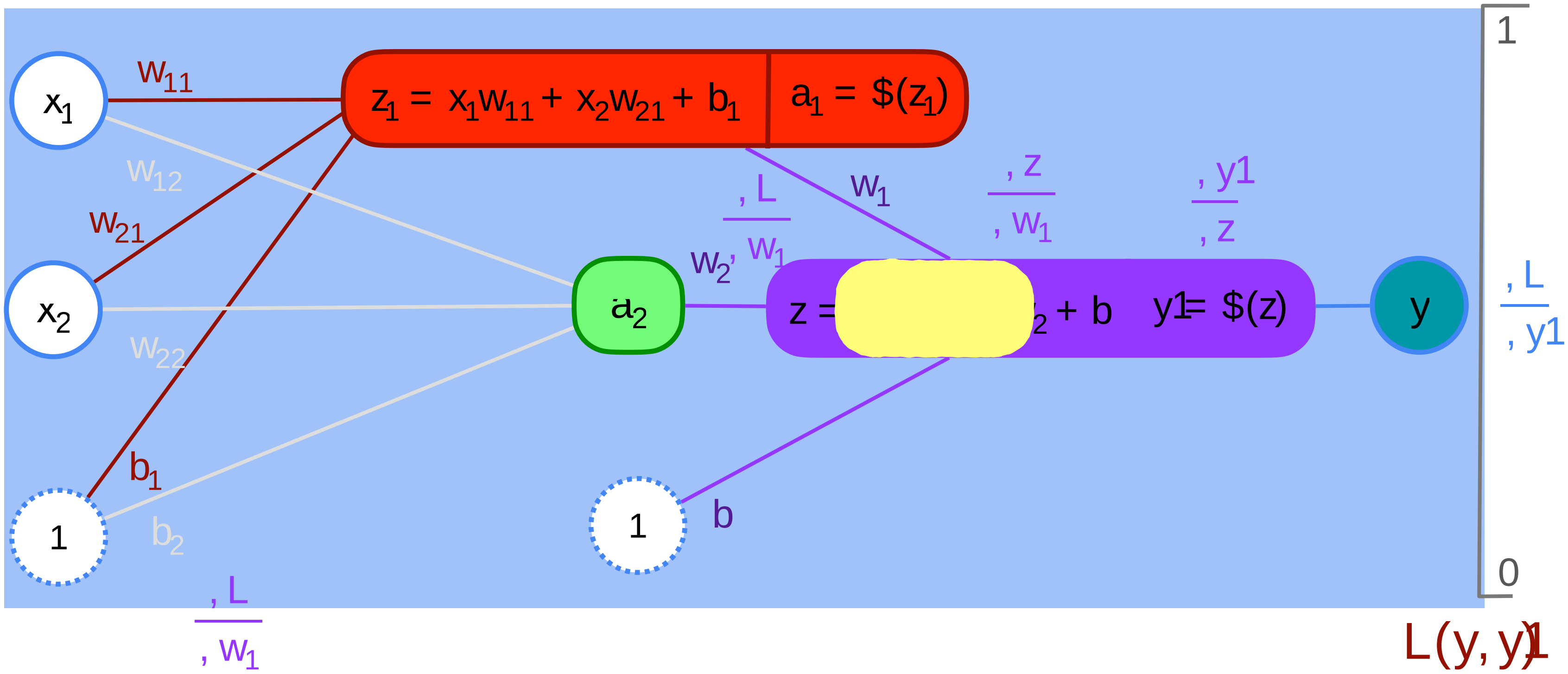
2,2,1 Neural Network



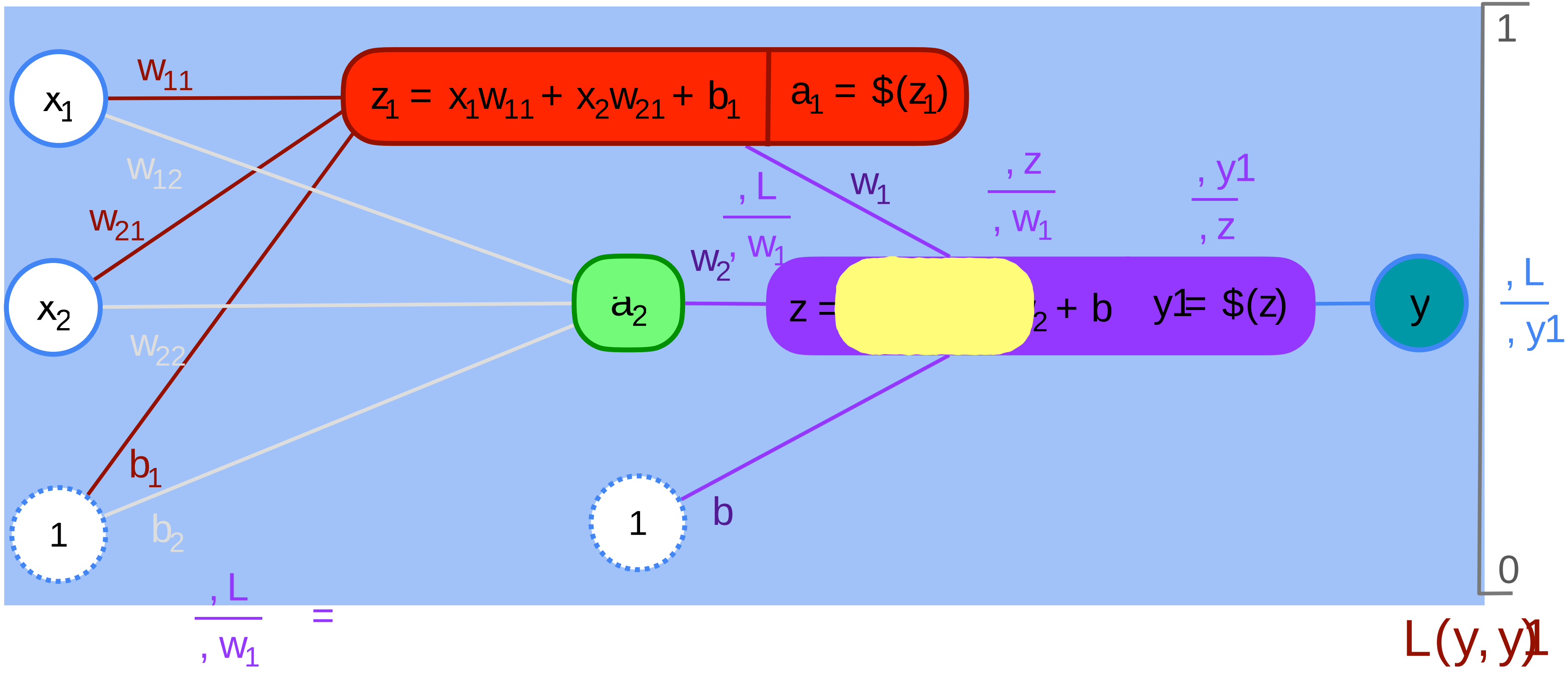
2,2,1 Neural Network



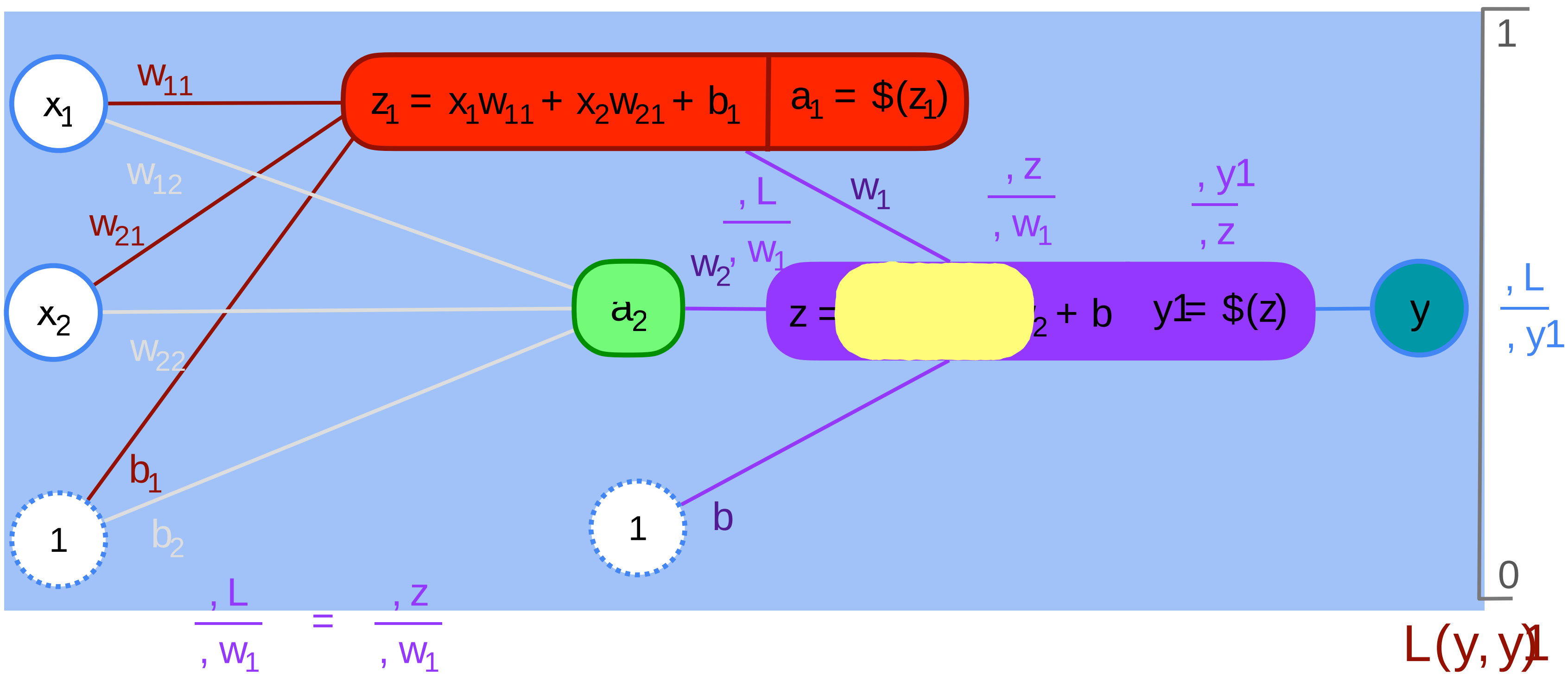
2,2,1 Neural Network



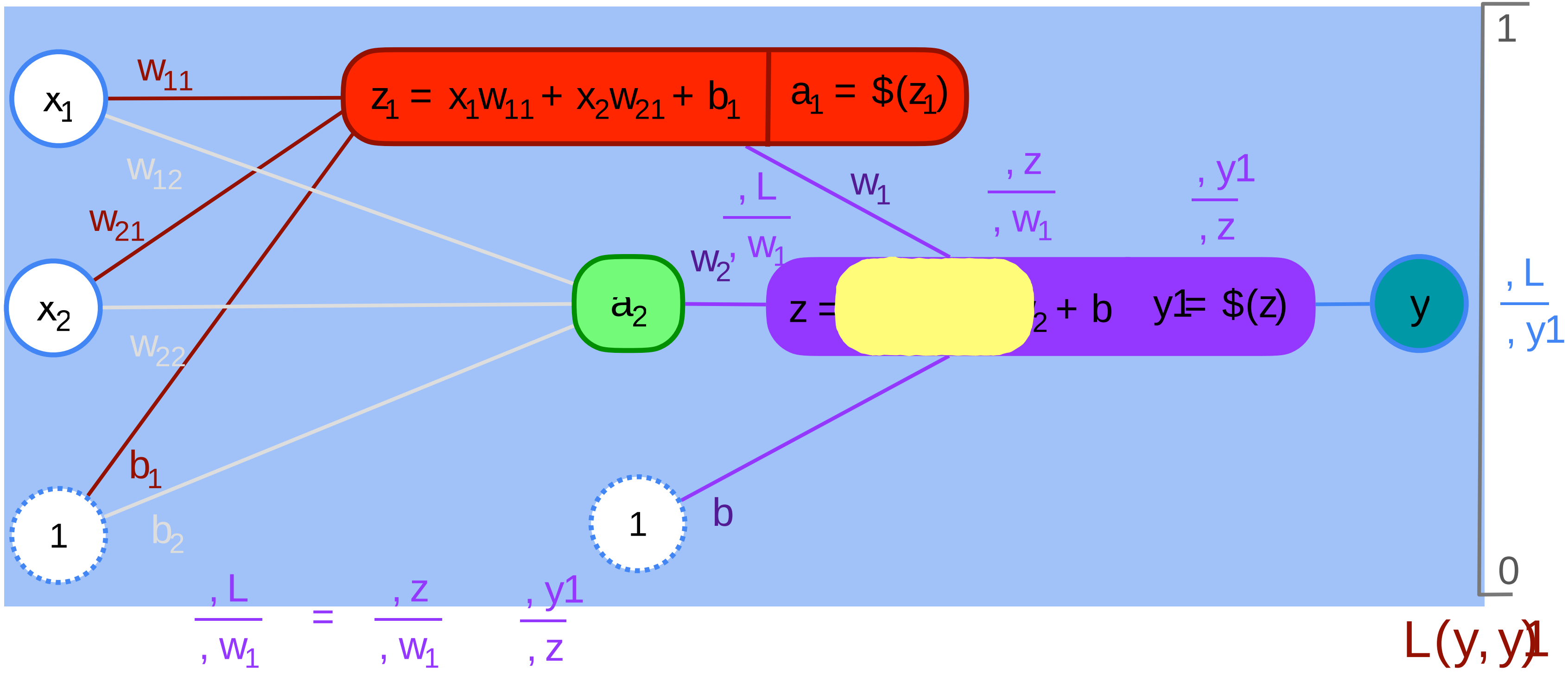
2,2,1 Neural Network



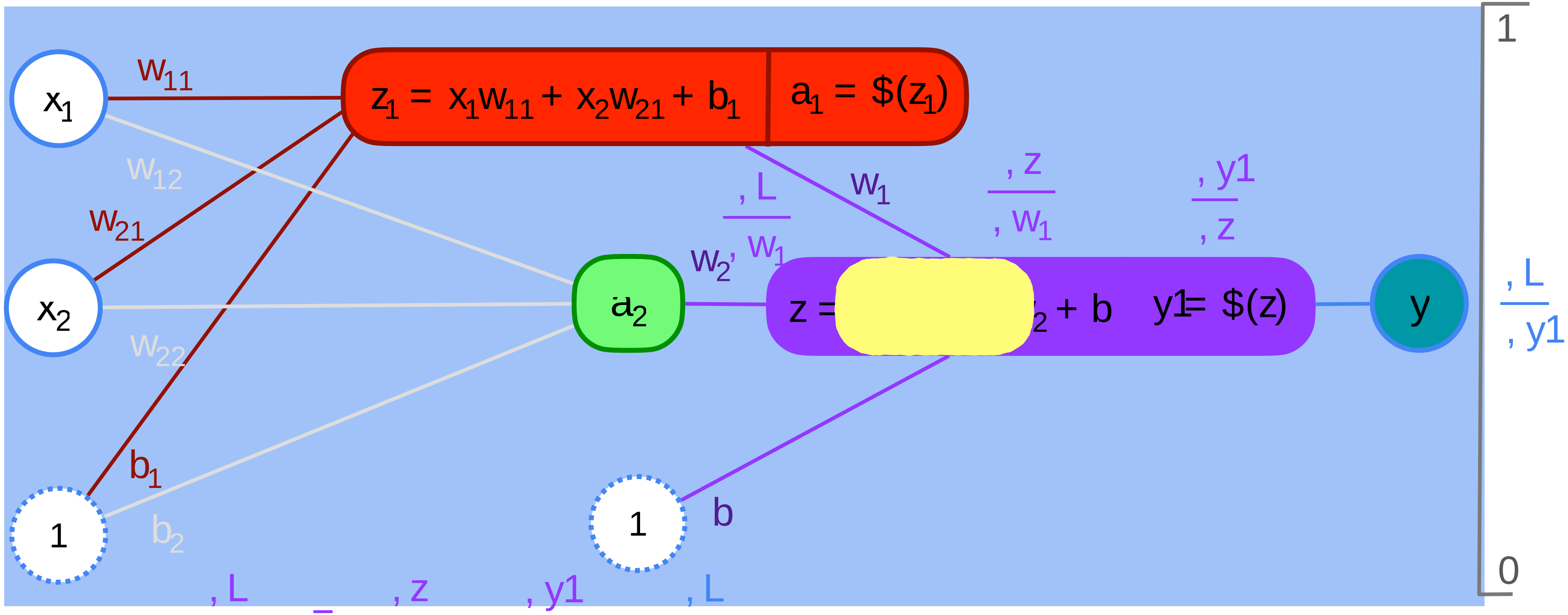
2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network



$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

2,2,1 Neural Network

$$\frac{,\, L}{,\, w_1} = \frac{,\, z}{,\, w_1} \cdot \frac{,\, y1}{,\, z} \cdot \frac{,\, L}{,\, y1}$$

y1= $(z)

$z = a_1 w_1 + a_2 w_2 + b$

2,2,1 Neural Network

L(y, y1)= -ylog(y1 " (1" y)log(1" y1

y1= $(z)

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

2,2,1 Neural Network

$$L(y,\hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

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

$$z = a_1w_1 + a_2w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1}$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1 - y)\log(1 - \hat{y})$

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

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} =$$

2,2,1 Neural Network

L(y, y1)= -ylog(y1) " (1" y)log(1" y1)

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{, L}{, w_1} = \qquad \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_1} =$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1= $(z)

$$\frac{, L}{, w_1} = \quad \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_1} =$$

2,2,1 Neural Network

$L(y, y1) = -y\log(y1) " (1" y)\log(1" y1)$

$y1 = \$(z)$

$$\frac{, L}{, w_1} = \quad \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_1} = a_1$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

$y1 = \$(z)$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \quad \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_1} = a_1$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \quad \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, w_1} = a_1$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \quad \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial w_1} = a_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, w_1} = a_1 \; y1(1 \,\text{"}\, y1)$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, w_1} = a_1 \quad y1(1 \text{"} \; y1) \quad \frac{-(y \text{"} \; y1)}{y1(1 \text{"} \; y1)}$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, w_1} = a_1 \cdot y1(1 \text{"} y1) \cdot \frac{-(y \text{"} y1)}{y1(1 \text{"} y1)}$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{,\ L}{,\ w_1} = \frac{,\ z}{,\ w_1} \cdot \frac{,\ y1}{,\ z} \cdot$$

$$\frac{,\ L}{,\ w_1} = a_1 \cdot \cancel{y1(1 \text{ " } y1)} \cdot \frac{-(y \text{ " } y1)}{y1(1 \text{ " } y1)}$$

2,2,1 Neural Network

$y1 = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial y1}{\partial z} \cdot$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{y1(1 - y1)} \cdot \frac{-(y - y1)}{\cancel{y1(1 - y1)}}$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, w_1} = a_1 \cdot \cancel{y1(1 \text{"} \; y1)} \cdot \frac{-(y \text{"} \; y1)}{\cancel{y1(1 \text{"} \; y1)}}$$

$$= -a_1(y \text{"} \; y1)$$

2,2,1 Neural Network

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,L}{,w_1} = \frac{,z}{,w_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_1} = a_1 \cdot \cancel{y1(1 \text{"} y1)} \cdot \frac{-(y \text{"} y1)}{\cancel{y1(1 \text{"} y1)}}$$

$$= -a_1(y \text{"} y1)$$

to find optimal value of w_1 that gives the least error

2,2,1 Neural Network

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,L}{,w_1} = \frac{,z}{,w_1} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,w_1} = a_1 \cdot \cancel{y1(1 \text{"} y1)} \cdot \frac{-(y \text{ " } y1)}{\cancel{y1(1 \text{"} y1)}}$$

$$= -a_1(y \text{ " } y1)$$

Perform gradient descent with

to find optimal value of w_1 that gives the least error

2,2,1 Neural Network

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, w_1} = a_1 \cdot \cancel{y1(1 \text{ " } y1)} \cdot \frac{-(y \text{ " } y1)}{\cancel{y1(1 \text{ " } y1)}}$$

$$= -a_1(y \text{ " } y1)$$

Perform gradient descent with

$$w_1 \ \$ \ w_1 \text{ " } \# \frac{, L}{, w_1}$$

to find optimal value of w_1 that gives the least error

2,2,1 Neural Network

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,\ L}{,\ w_1} = \frac{,\ z}{,\ w_1} \cdot \frac{,\ y1}{,\ z} \cdot$$

$$\frac{,\ L}{,\ w_1} = a_1 \cdot \cancel{y(1\text{"}\ y1)} \cdot \frac{-(y\ \text{"}\ \ y1)}{\cancel{y(1\text{"}\ y1)}}$$

$$= -a_1(y\ \text{"}\ \ y1)$$

Perform gradient descent with

$w_1\ \$\ \ w_1\ \text{"}\ \#$

to find optimal value of w_1 that gives the least error

2,2,1 Neural Network

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{, L}{, w_1} = \frac{, z}{, w_1} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, w_1} = a_1 \cdot \cancel{y1(1 \text{"} y1)} \cdot \frac{-(y \text{"} y1)}{\cancel{y1(1 \text{"} y1)}}$$

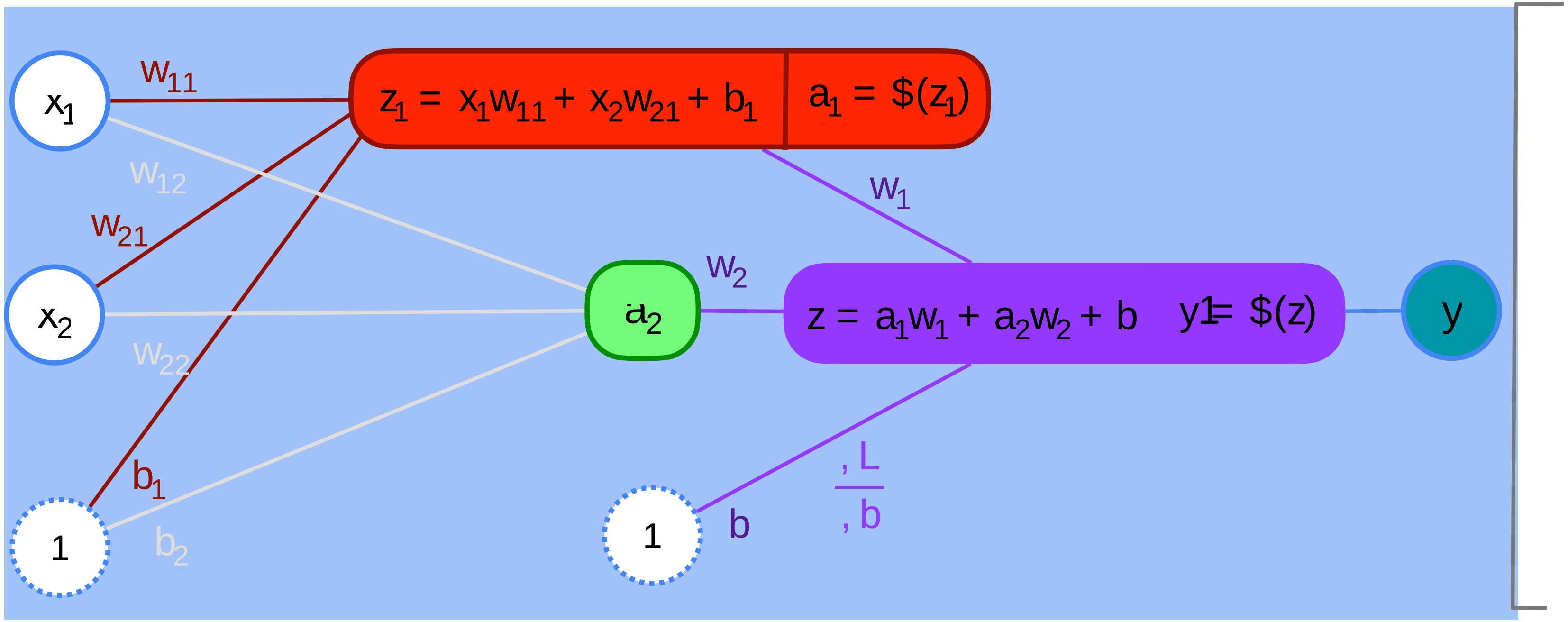
$$= -a_1(y \text{"} y1)$$

Perform gradient descent with

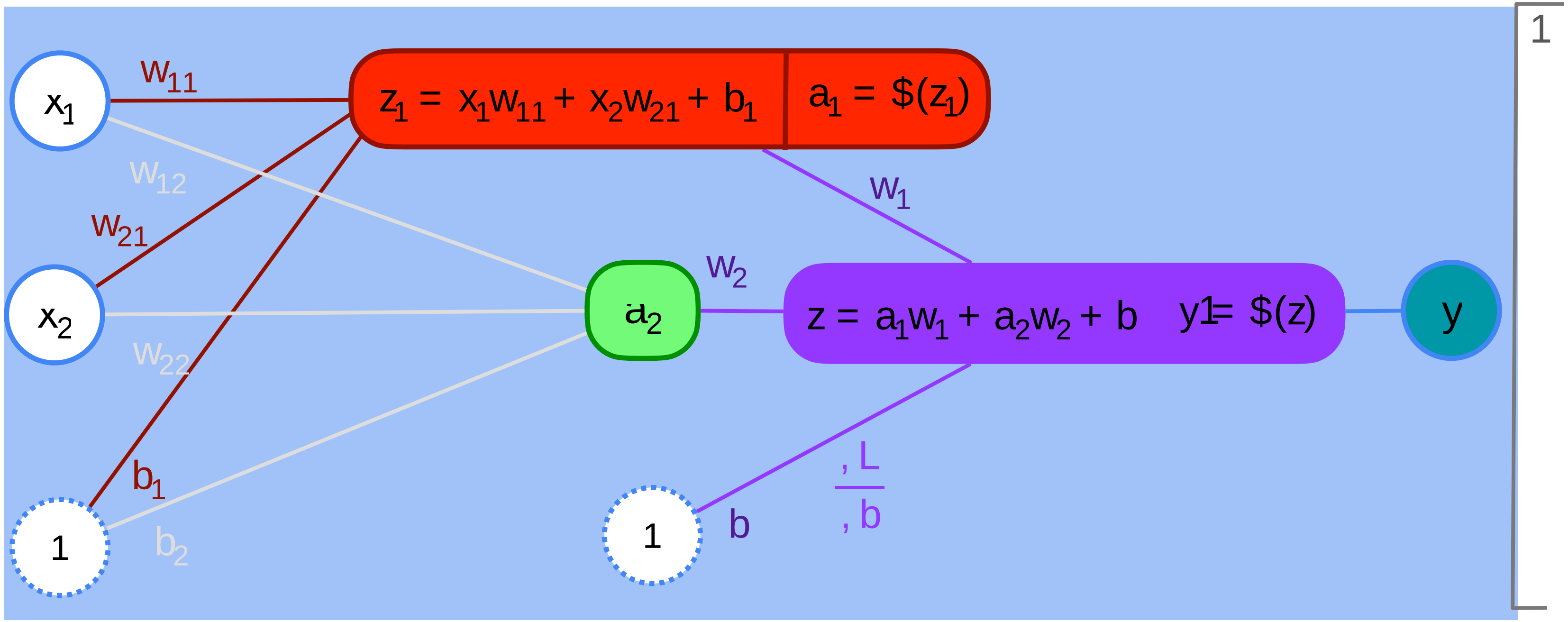
$$w_1 \ \$ \ w_1 \text{"} \#(-a_1(y \text{"} y1))$$

to find optimal value of w_1 that gives the least error

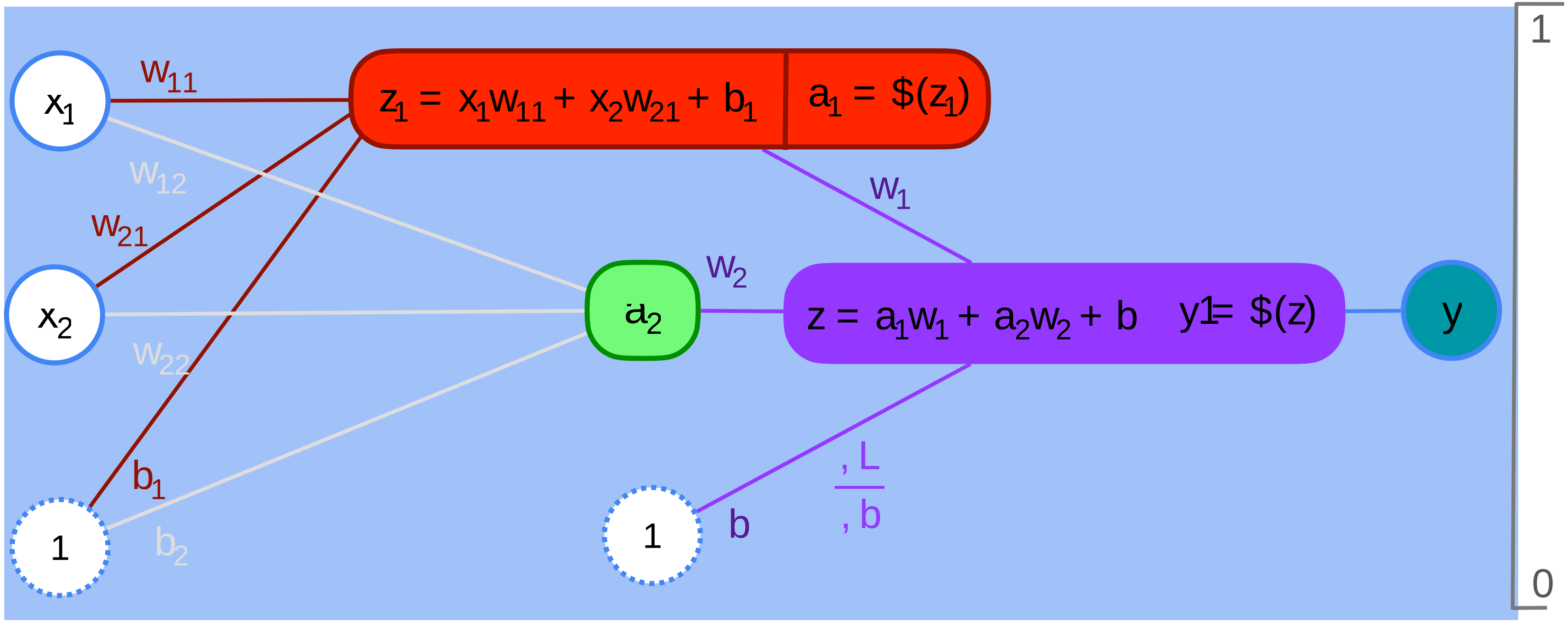
2,2,1 Neural Network



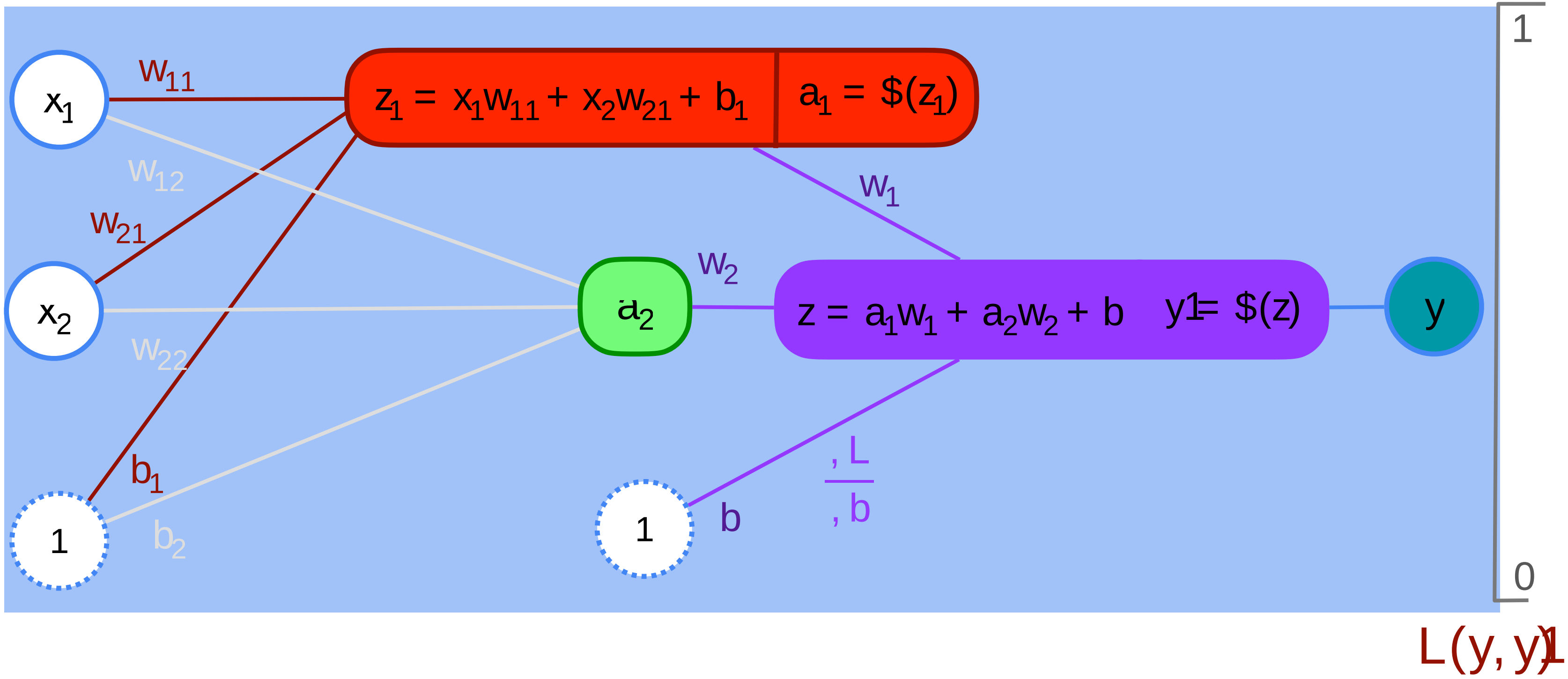
2,2,1 Neural Network



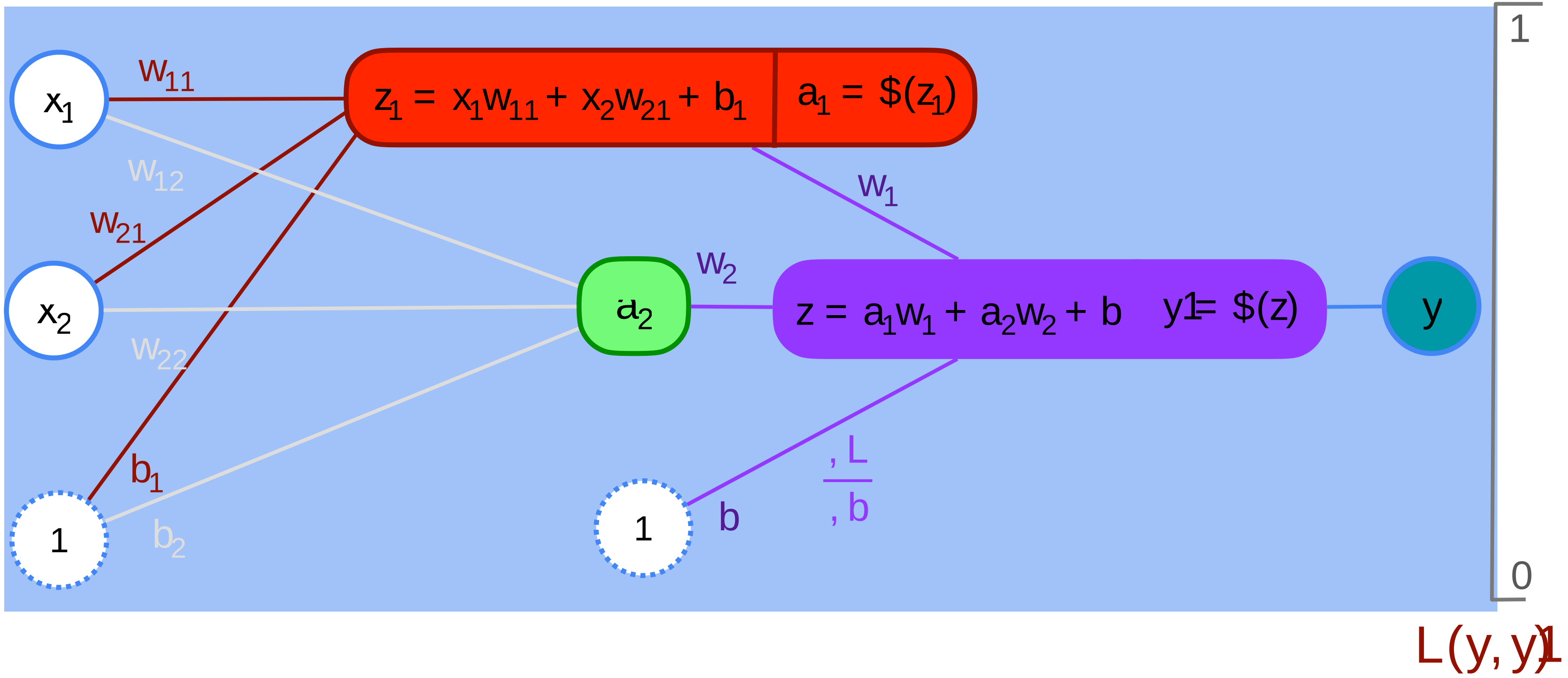
2,2,1 Neural Network



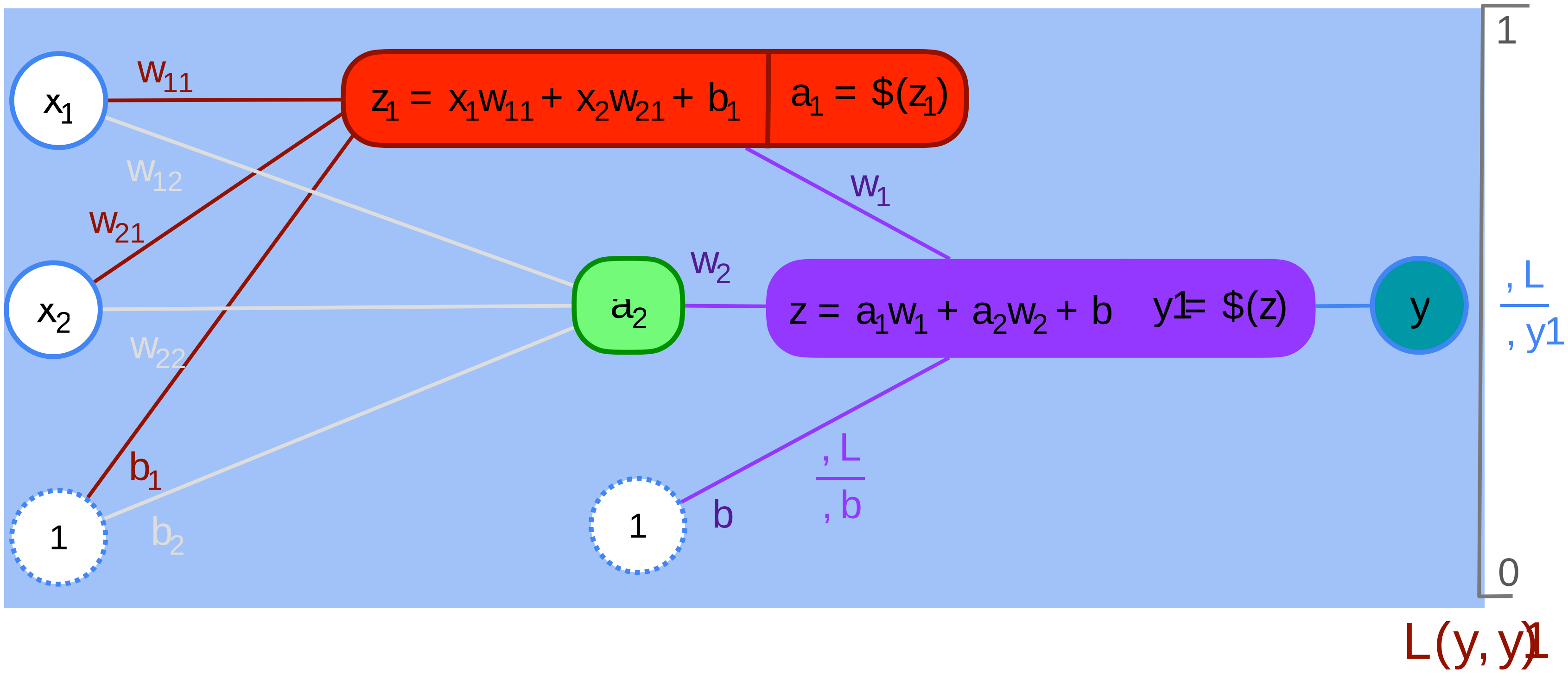
2,2,1 Neural Network



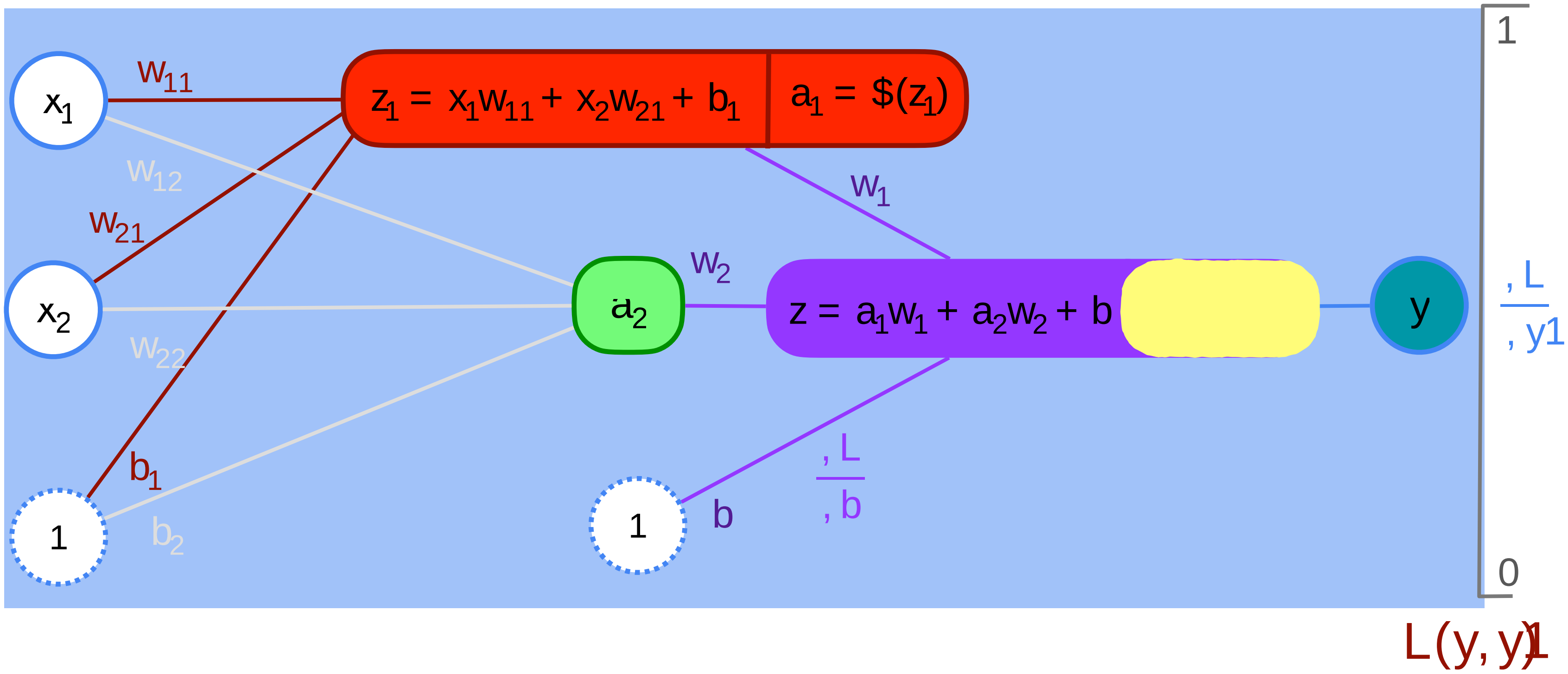
2,2,1 Neural Network



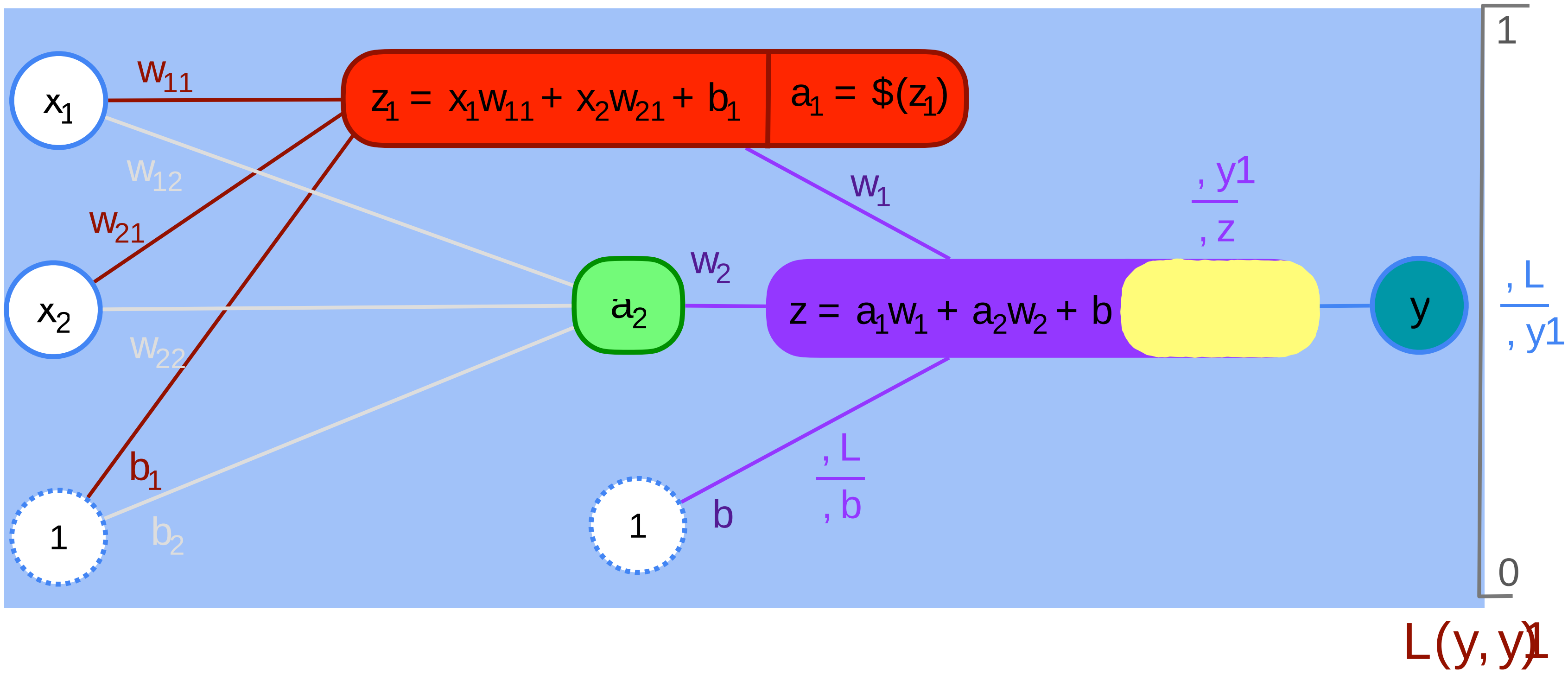
2,2,1 Neural Network



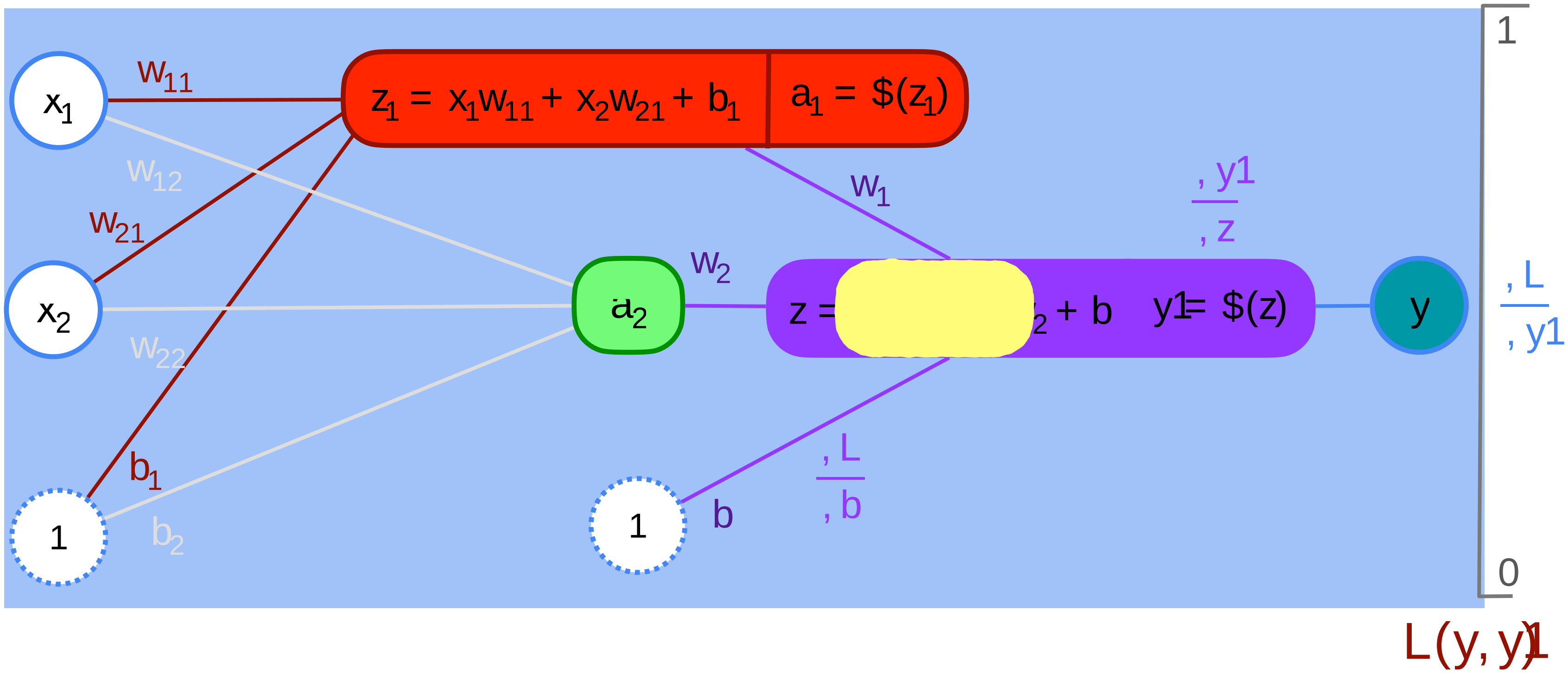
2,2,1 Neural Network



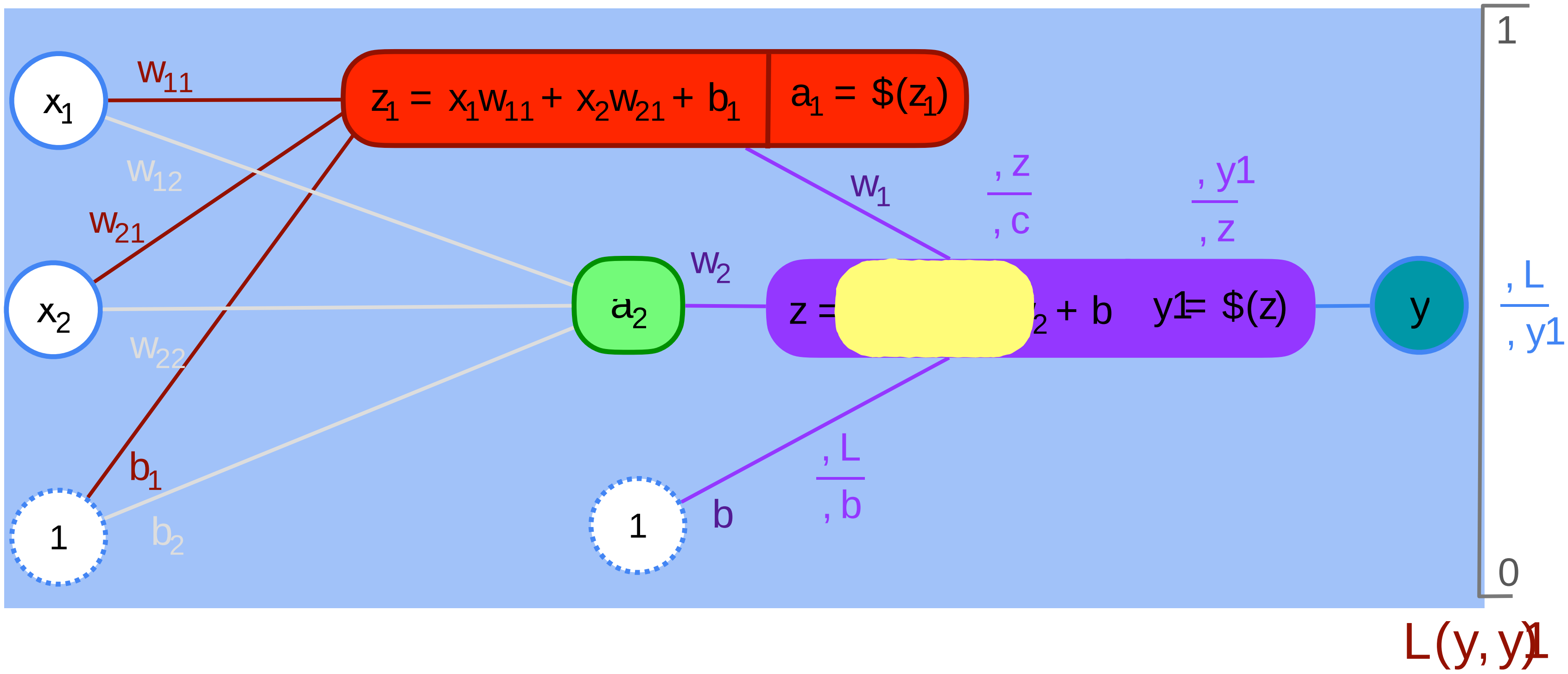
2,2,1 Neural Network



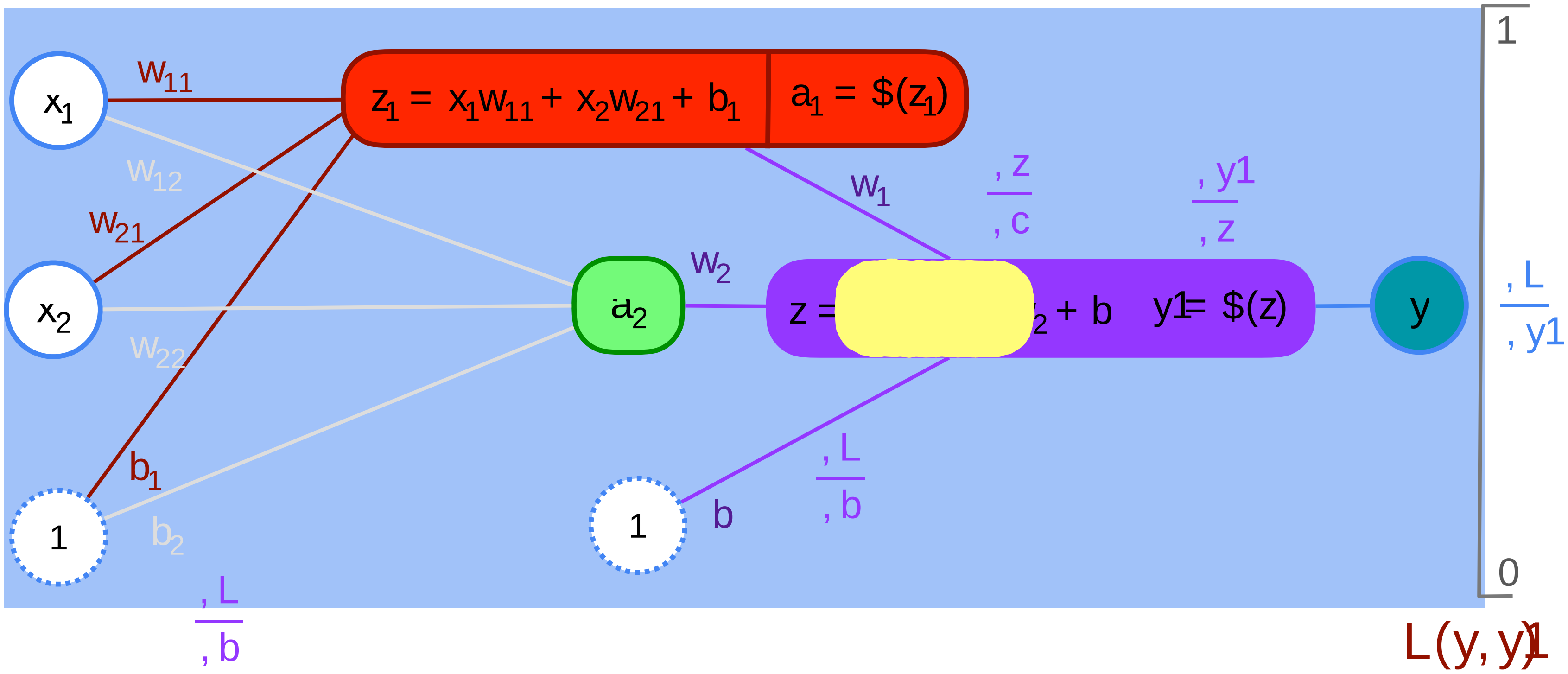
2,2,1 Neural Network



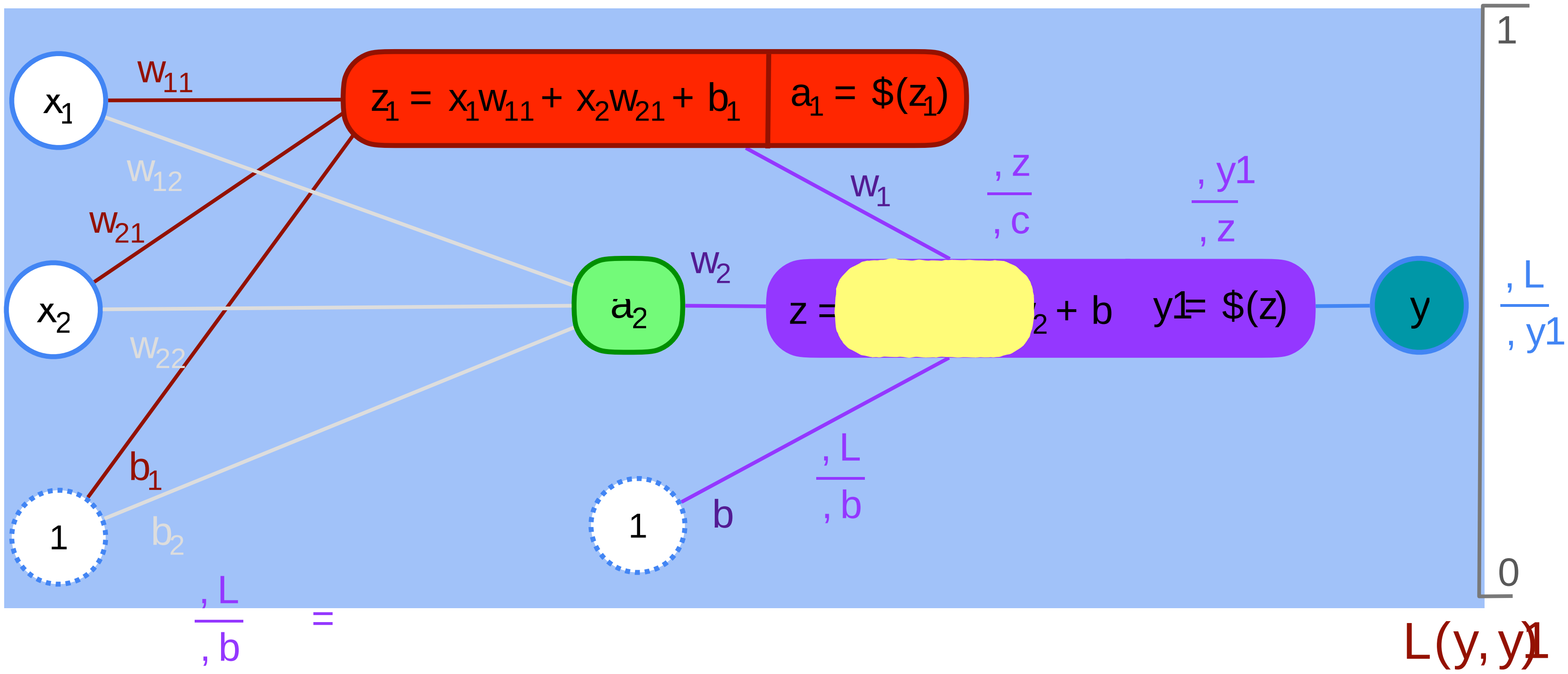
2,2,1 Neural Network



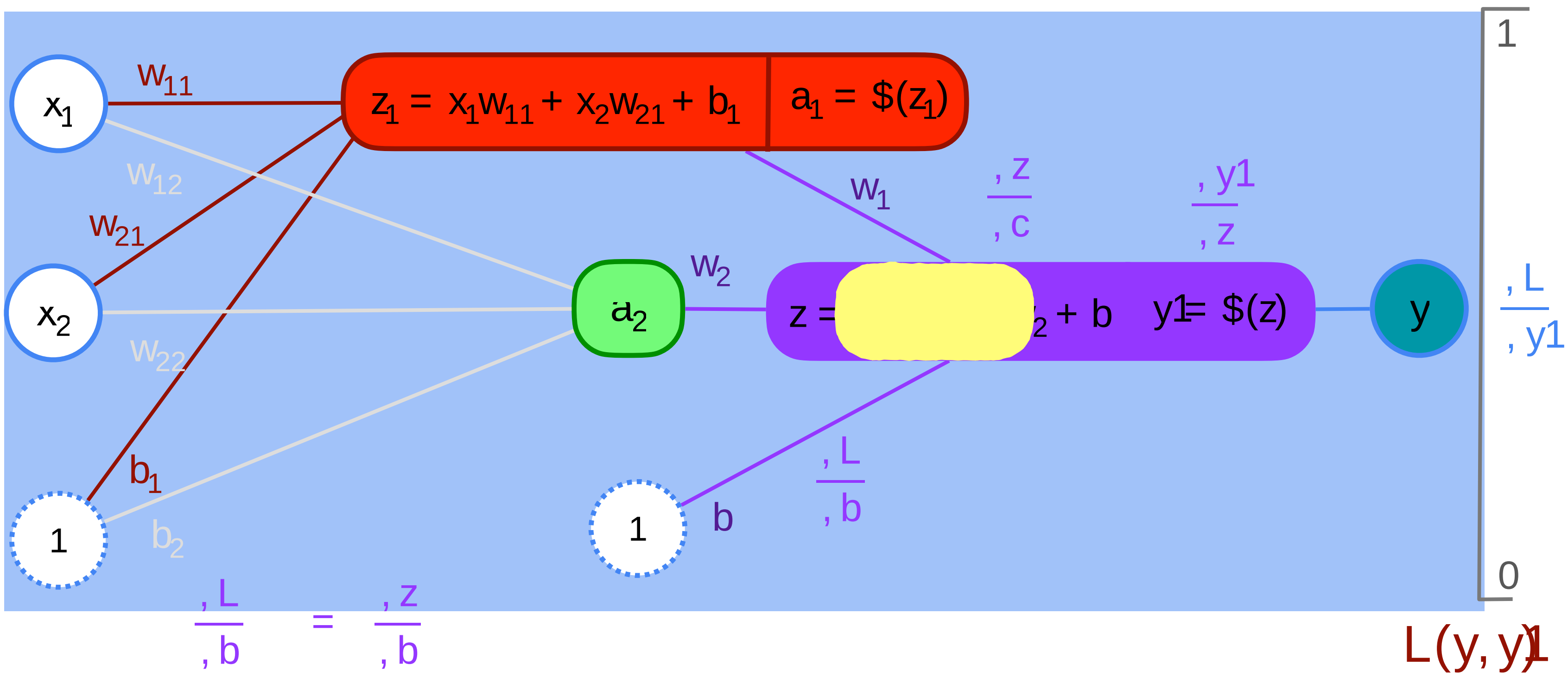
2,2,1 Neural Network



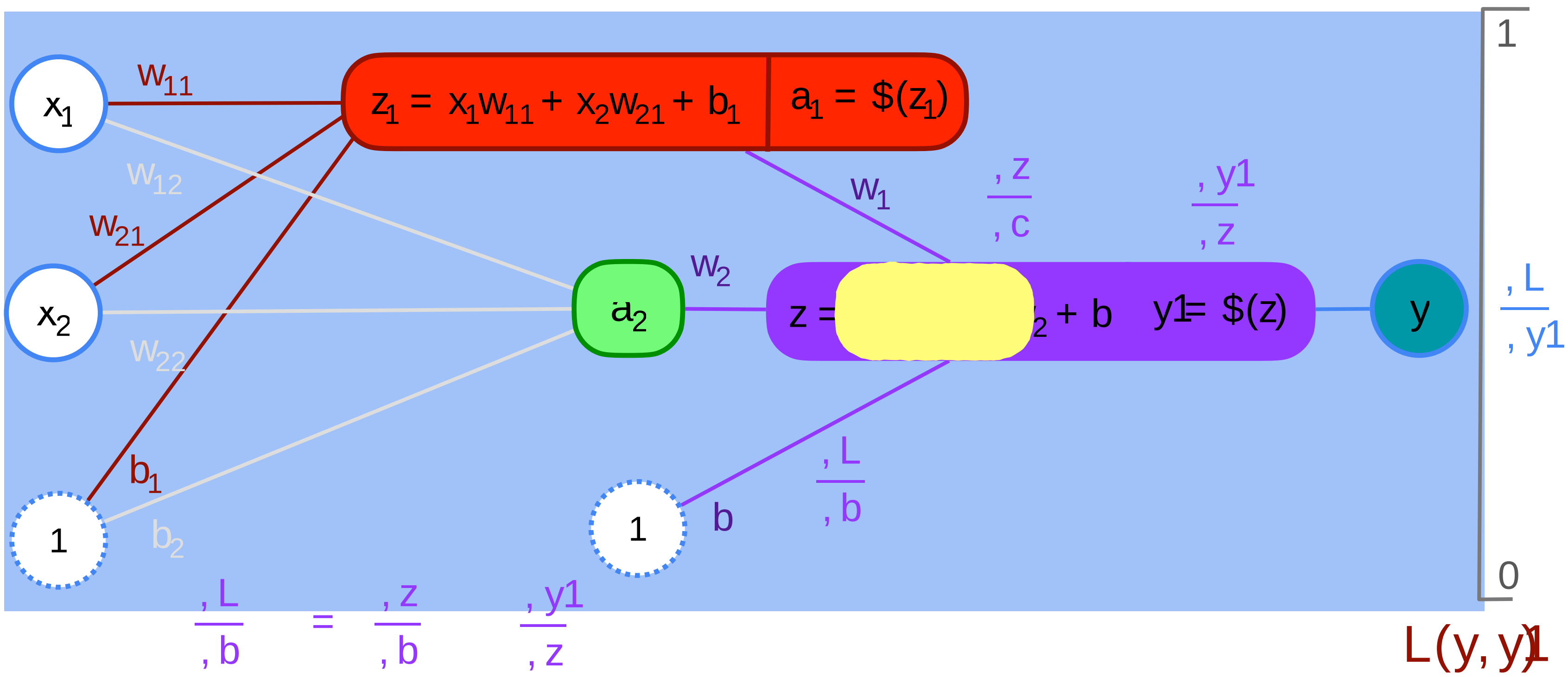
2,2,1 Neural Network



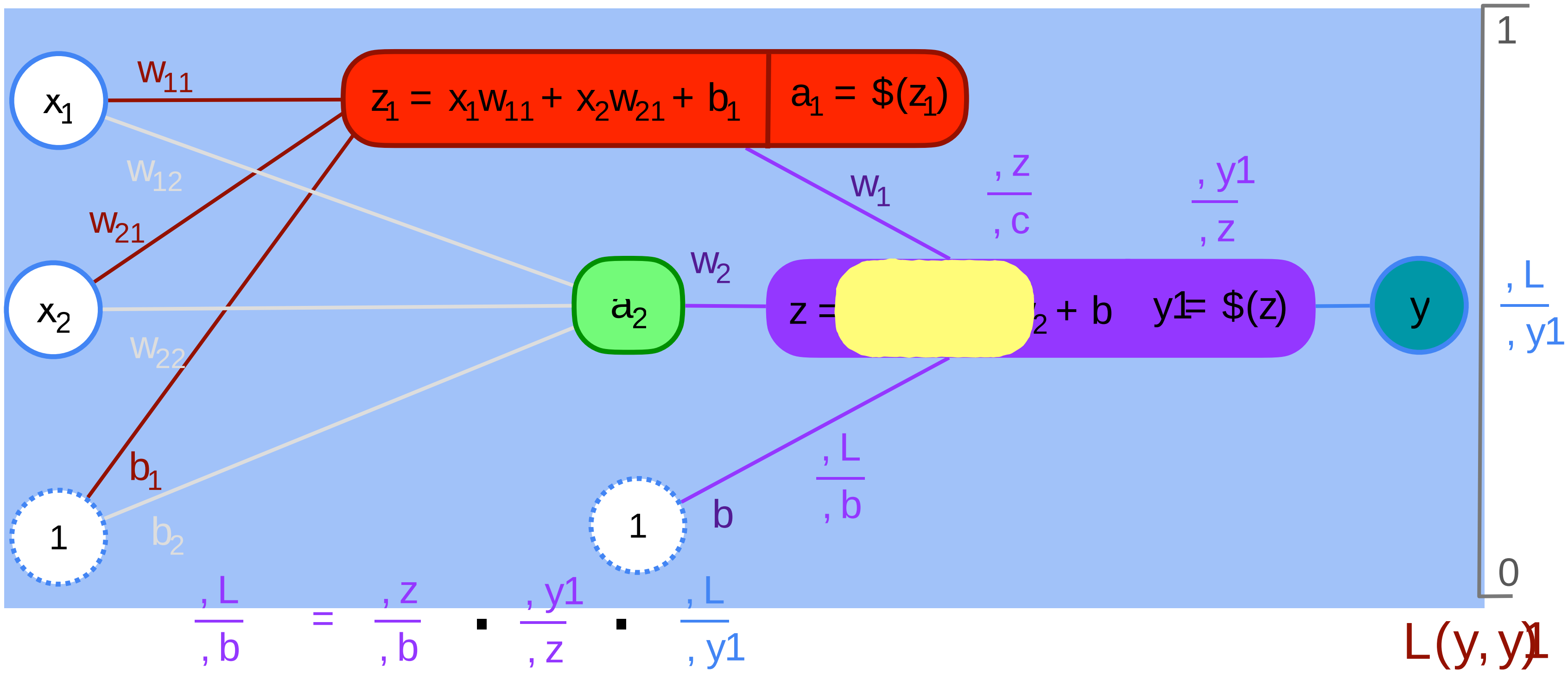
2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

2,2,1 Neural Network

L(y, y1)= -ylog(y1) " (1" y)log(1" y1)

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b}$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1 = $(z)

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, b} = \frac{, z}{, b} \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, b} =$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,L}{,b} = \cdot \frac{,y1}{,z} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,b} =$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

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

$$\frac{\partial L}{\partial b} = \quad \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} =$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

y1 = $(z)

$$\frac{, L}{, b} = \quad \cdot \frac{, y1}{, z} \cdot \frac{, L}{, y1}$$

$$\frac{, L}{, b} = 1$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1 " (1" y)log(1" y1

y1= $(z)

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \quad \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,b} = 1$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \frac{,L}{,y1}$$

$$\frac{,L}{,b} = 1$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

$z = a_1w_1 + a_2w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \quad \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

L(y, y1) = -ylog(y1) " (1" y)log(1" y1)

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, b} = \frac{, z}{, b} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, b} = 1 \quad y(1 \text{"} \; y)1$$

2,2,1 Neural Network

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,b} = 1 \quad y1(1 \text{ " } y1)$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,b} = 1 \quad y1(1 \text{"} \; y1) \quad \frac{-(y \text{"} \; y1)}{y1(1 \text{"} \; y1)}$$

2,2,1 Neural Network

y1 = $(z)

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, b} = \frac{, z}{, b} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, b} = 1 \cdot y1(1 \text{"} \ y1) \cdot \frac{-(y \text{"} \ y1)}{y1(1 \text{"} \ y1)}$$

2,2,1 Neural Network

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

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, b} = \frac{, z}{, b} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, b} = 1 \cdot \cancel{y1(1 \text{"} y1)} \cdot \frac{-(y \text{"} y1)}{\cancel{y1(1 \text{"} y1)}}$$

2,2,1 Neural Network

y1= $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,b} = 1 \cdot \cancel{y1(1 \text{"} y1)} \cdot \frac{-(y \text{"} y1)}{\cancel{y1(1 \text{"} y1)}}$$

$$= -(y \text{"} y1)$$

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,b} = 1 \cdot \cancel{y1(1 \text{"} y1)} \cdot \frac{-(y \text{"} y1)}{\cancel{y1(1 \text{"} y1)}}$$

$$= -(y \text{"} y1)$$

to find optimal
value of b that gives
the least error

2,2,1 Neural Network

y1 = $(z)

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,\ L}{,\ b} = \frac{,\ z}{,\ b} \cdot \frac{,\ y1}{,\ z} \cdot$$

$$\frac{,\ L}{,\ b} = 1 \cdot \cancel{y1(1\ "\ y1)} \cdot \frac{-(y\ "\ y1)}{\cancel{y1(1\ "\ y1)}}$$

$$= -(y\ "\ y1)$$

Perform gradient descent with

to find optimal value of b that gives the least error

2,2,1 Neural Network

$\hat{y}1 = \$(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{, L}{, b} = \frac{, z}{, b} \cdot \frac{, y1}{, z} \cdot$$

$$\frac{, L}{, b} = 1 \cdot \cancel{y1(1 " y1)} \cdot \frac{-(y " y1)}{\cancel{y1(1 " y1)}}$$

$$= -(y " y1)$$

Perform gradient descent with

$$b \ \$ \ b " \# \frac{, L}{, b}$$

to find optimal value of b that gives the least error

2,2,1 Neural Network

$y1 = \$(z)$

$z = a_1w_1 + a_2w_2 + b$

$$\frac{,L}{,b} = \frac{,z}{,b} \cdot \frac{,y1}{,z} \cdot$$

$$\frac{,L}{,b} = 1 \cdot \cancel{y1(1 \text{"} y1)} \cdot \frac{-(y \text{"} y1)}{\cancel{y1(1 \text{"} y1)}}$$

$$= -(y \text{"} y1)$$

Perform gradient descent with

$$b \ \$ \ b \text{"} \ \#$$

to find optimal value of b that gives the least error

2,2,1 Neural Network

y1= $(z)

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{,\ L}{,\ b} = \frac{,\ z}{,\ b} \cdot \frac{,\ y1}{,\ z} \cdot$$

$$\frac{,\ L}{,\ b} = 1 \cdot \cancel{y1(1\ "\ y1)} \cdot \frac{-(y\ "\ \ y1)}{\cancel{y1(1\ "\ y1)}}$$

$$= -(y\ "\ \ y1)$$

Perform gradient descent with

b $ b " #(-(y " y1))

to find optimal value of b that gives the least error

Optimization in Neural Networks and NewtonÕs Method

Gradient Descent and Backpropagation

Back Propagation Introduction

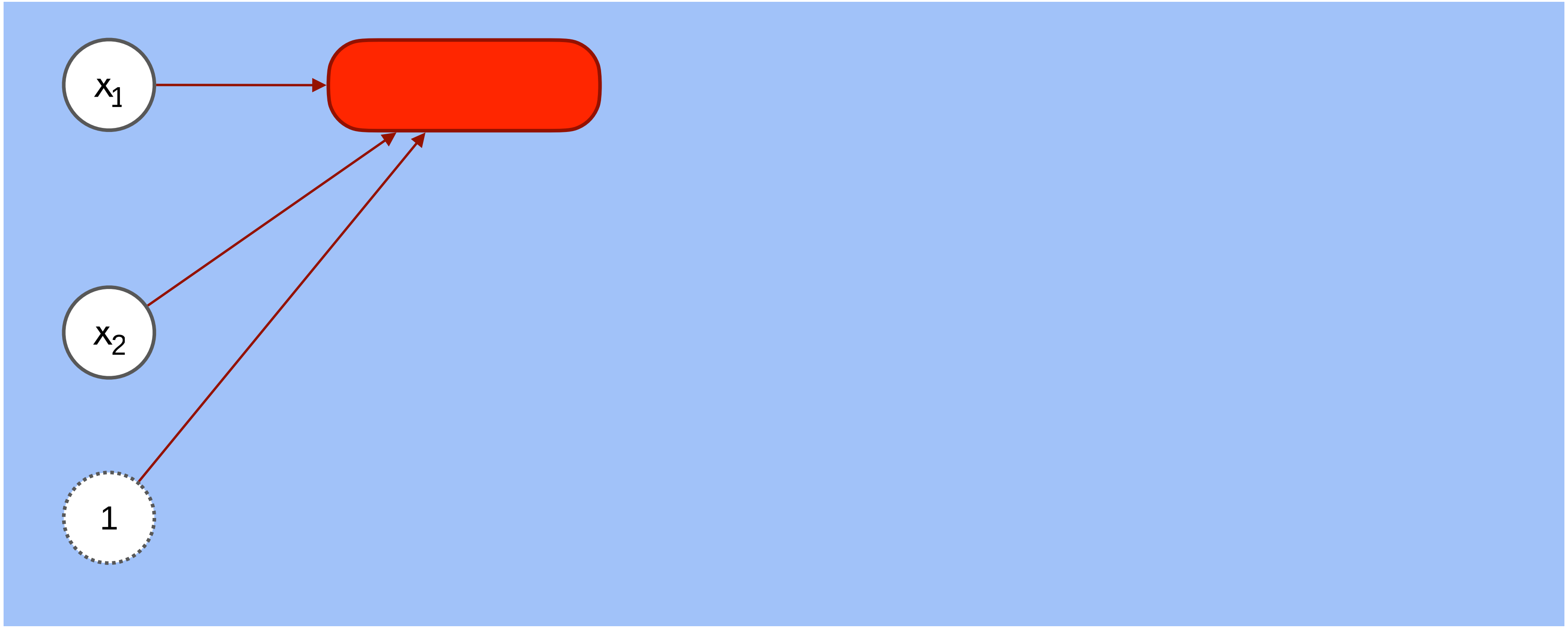
Back Propagation Introduction

Back Propagation Introduction

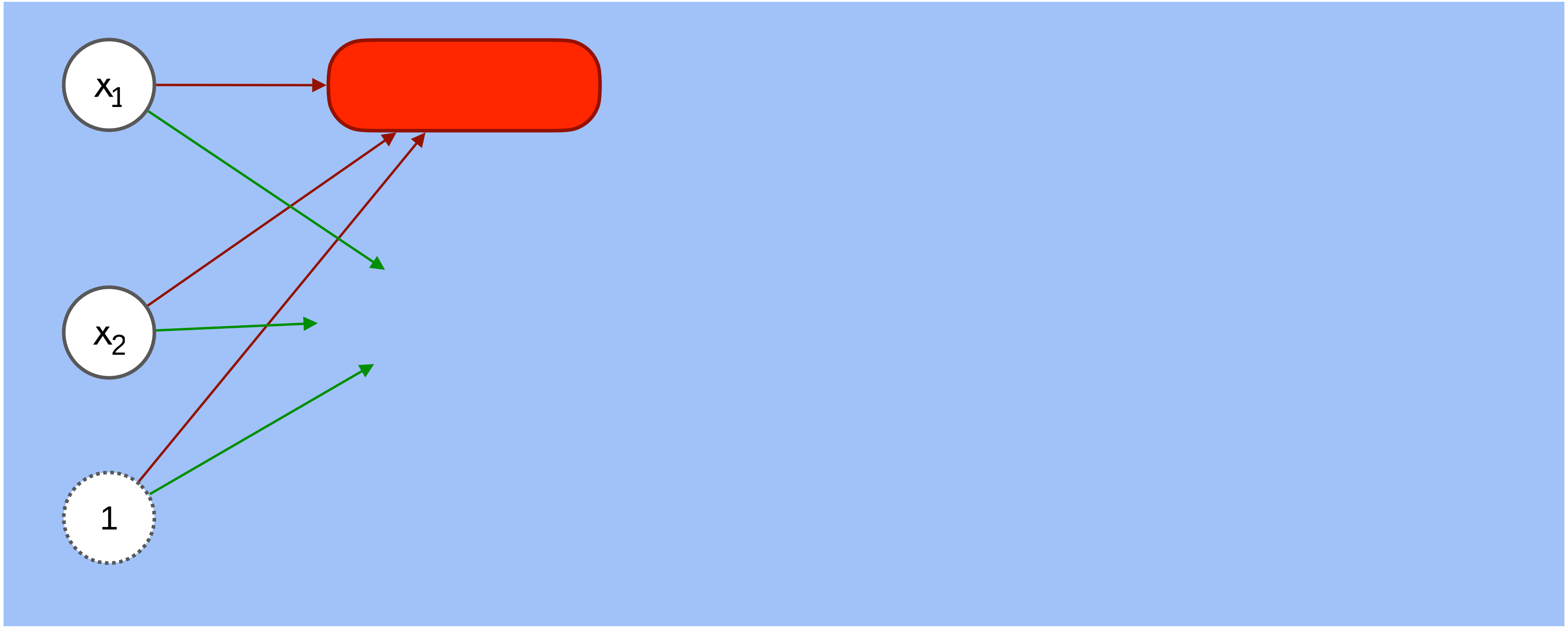
Back Propagation Introduction

Back Propagation Introduction

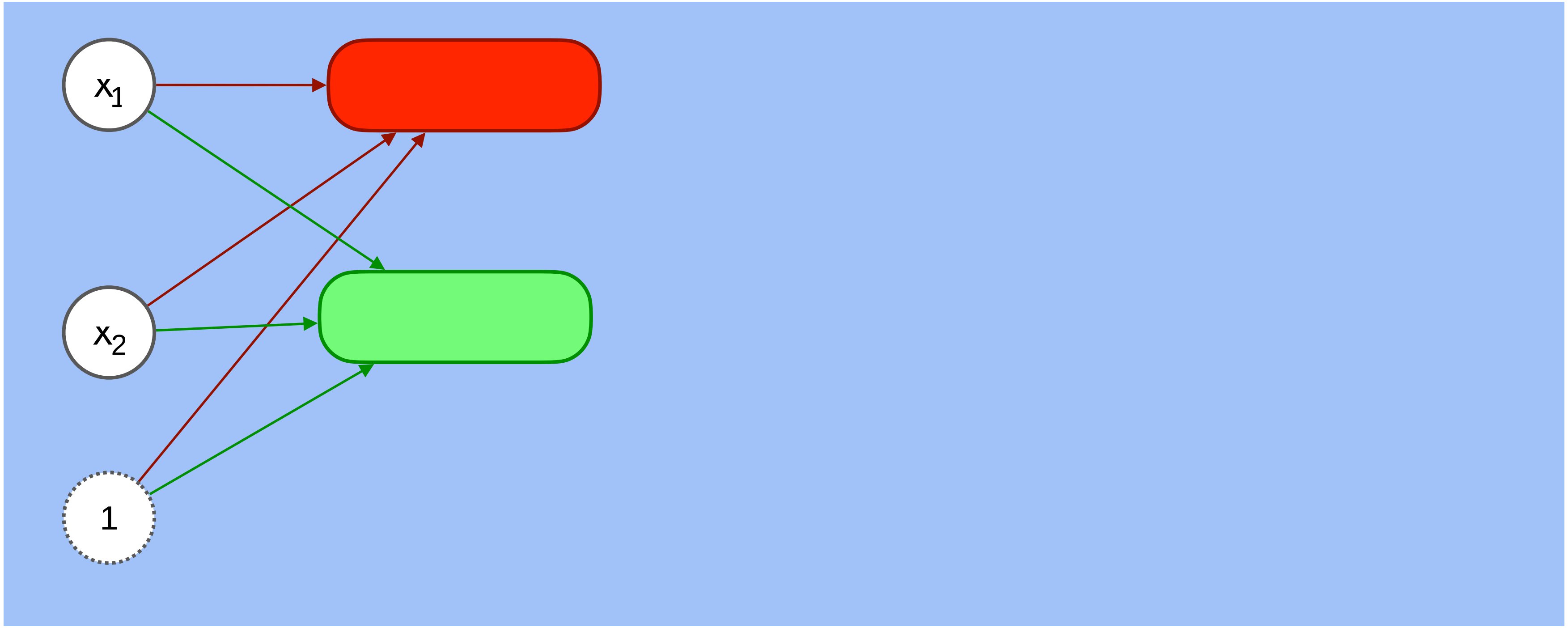
Back Propagation Introduction



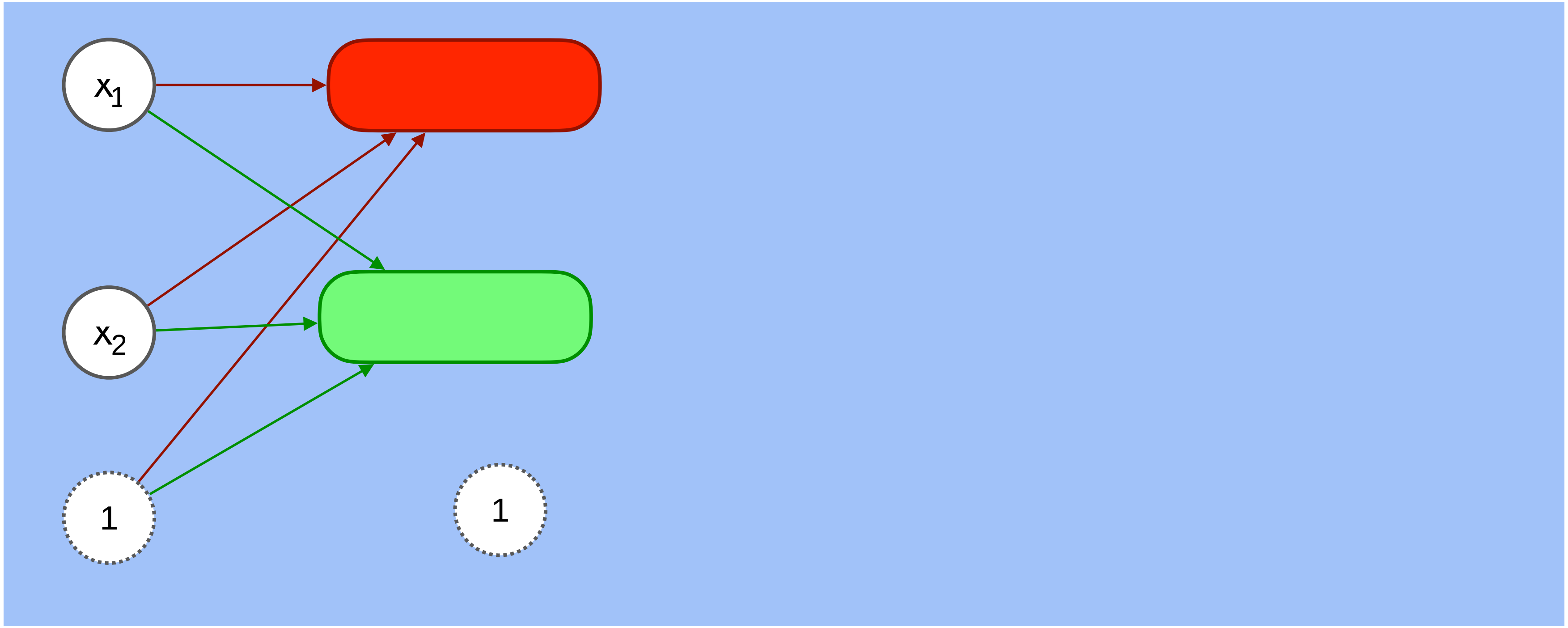
Back Propagation Introduction



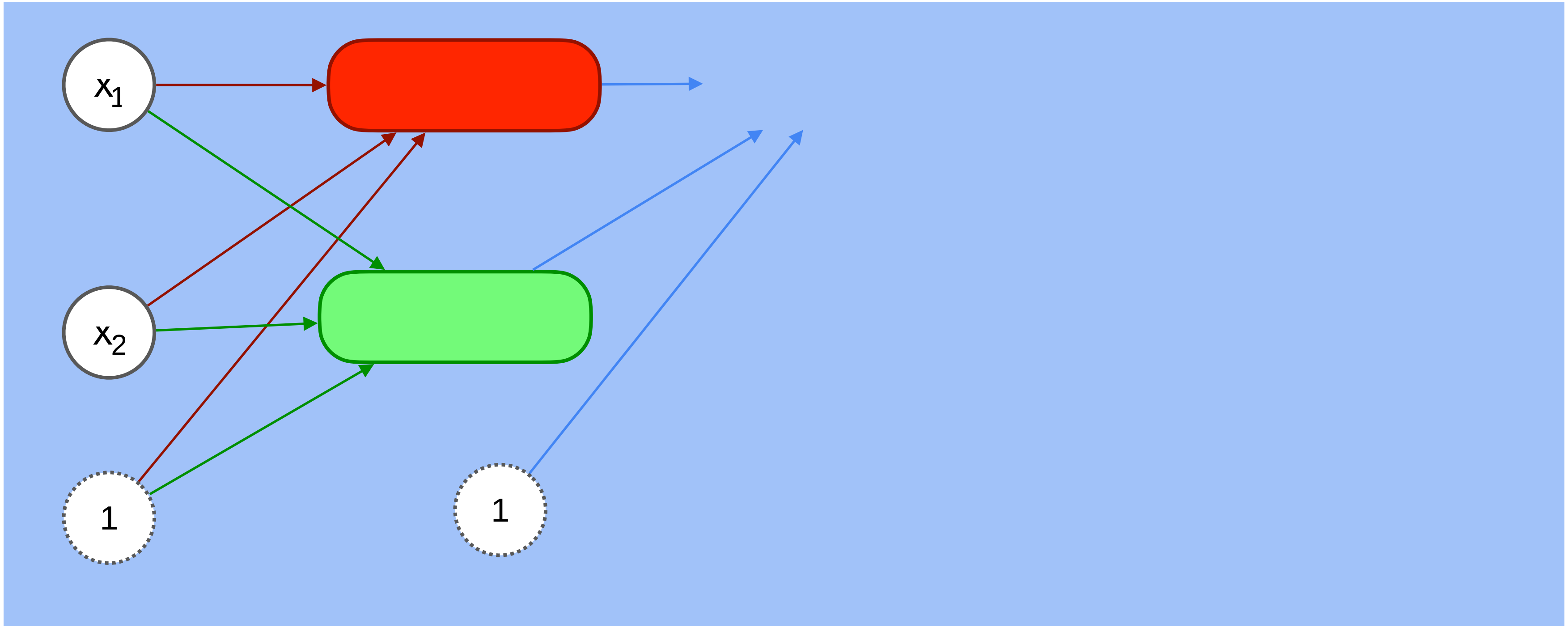
Back Propagation Introduction



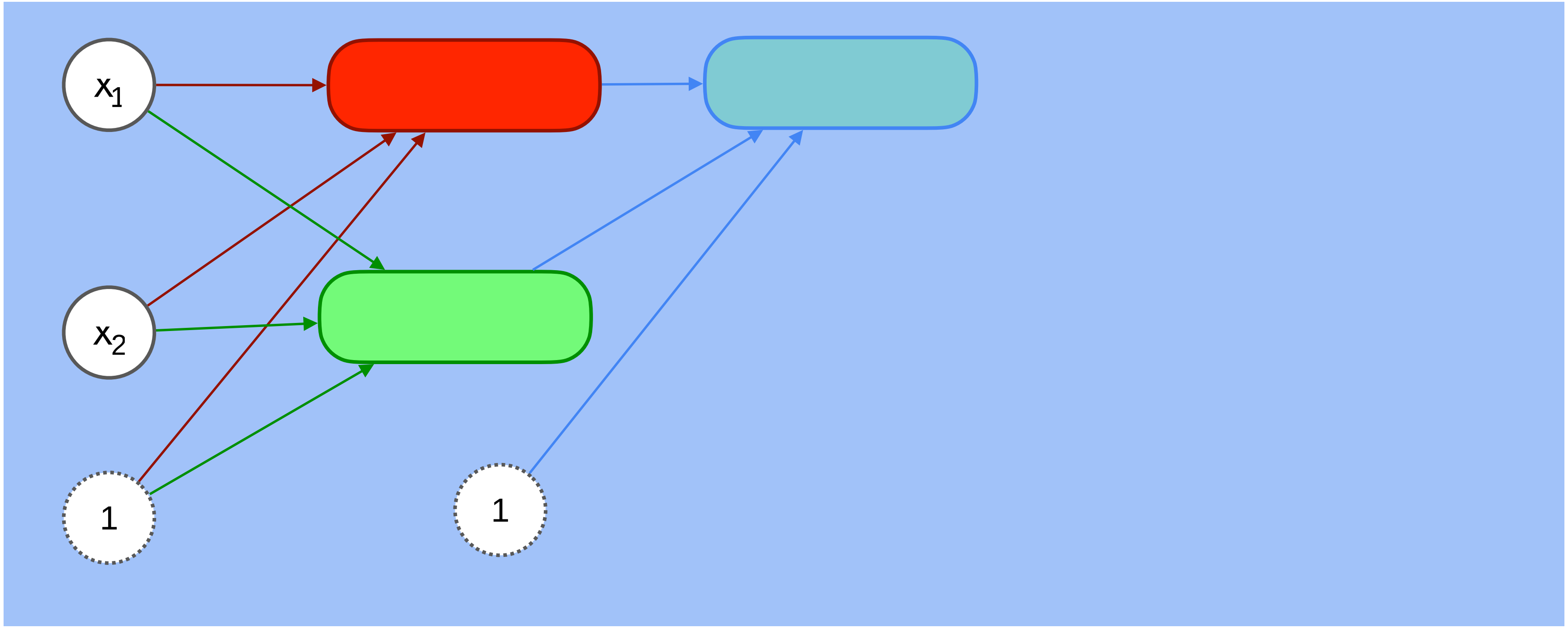
Back Propagation Introduction



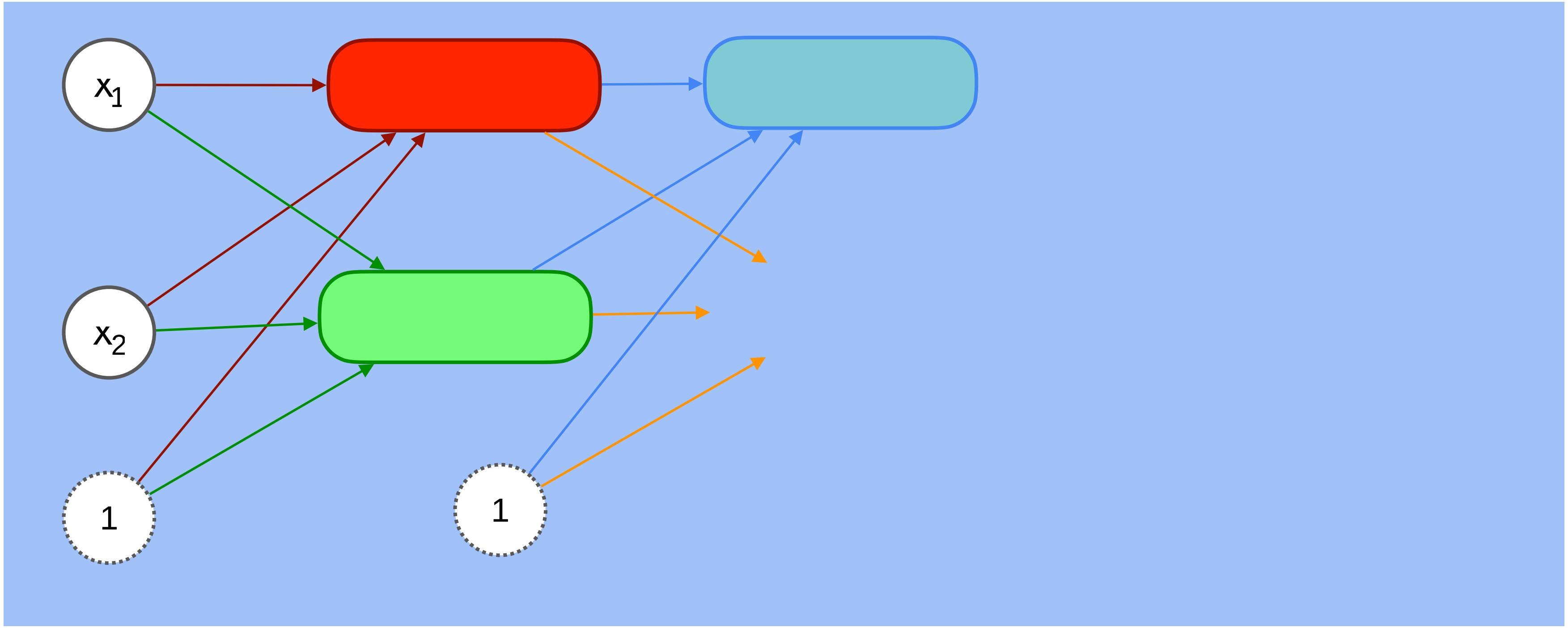
Back Propagation Introduction



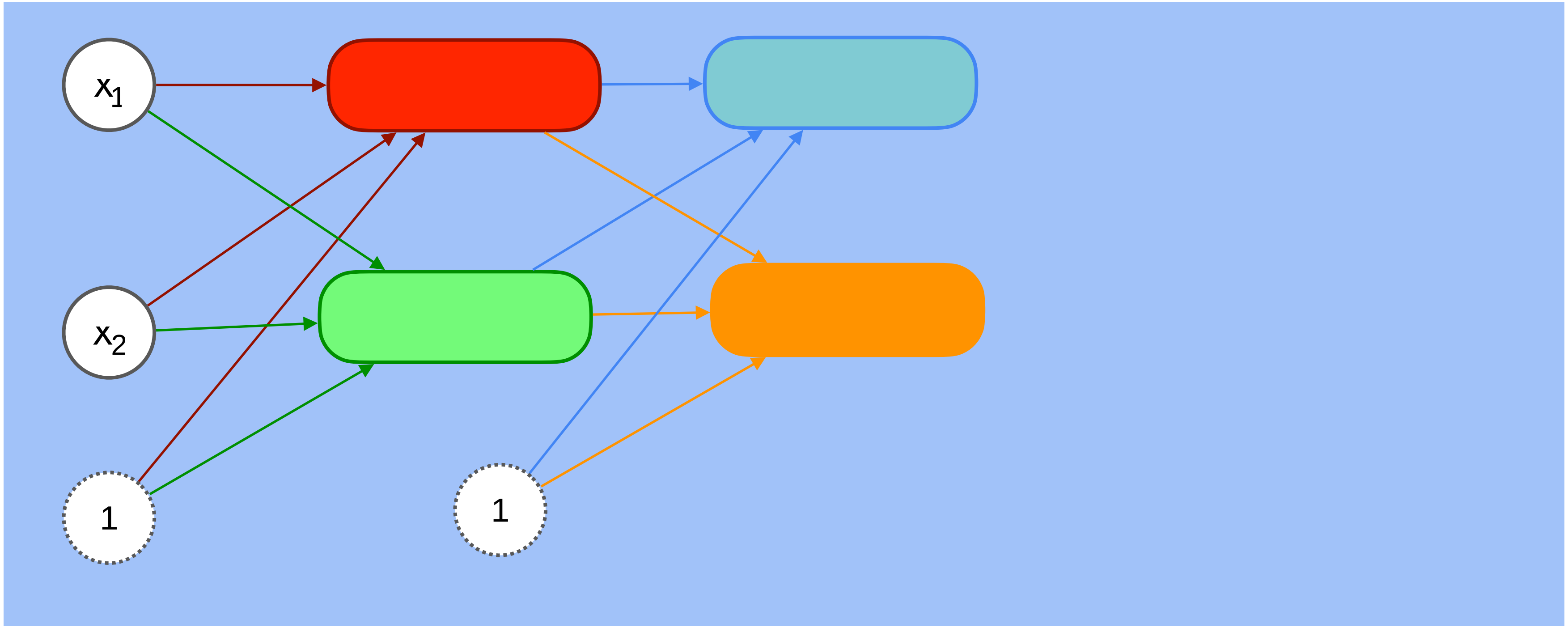
Back Propagation Introduction



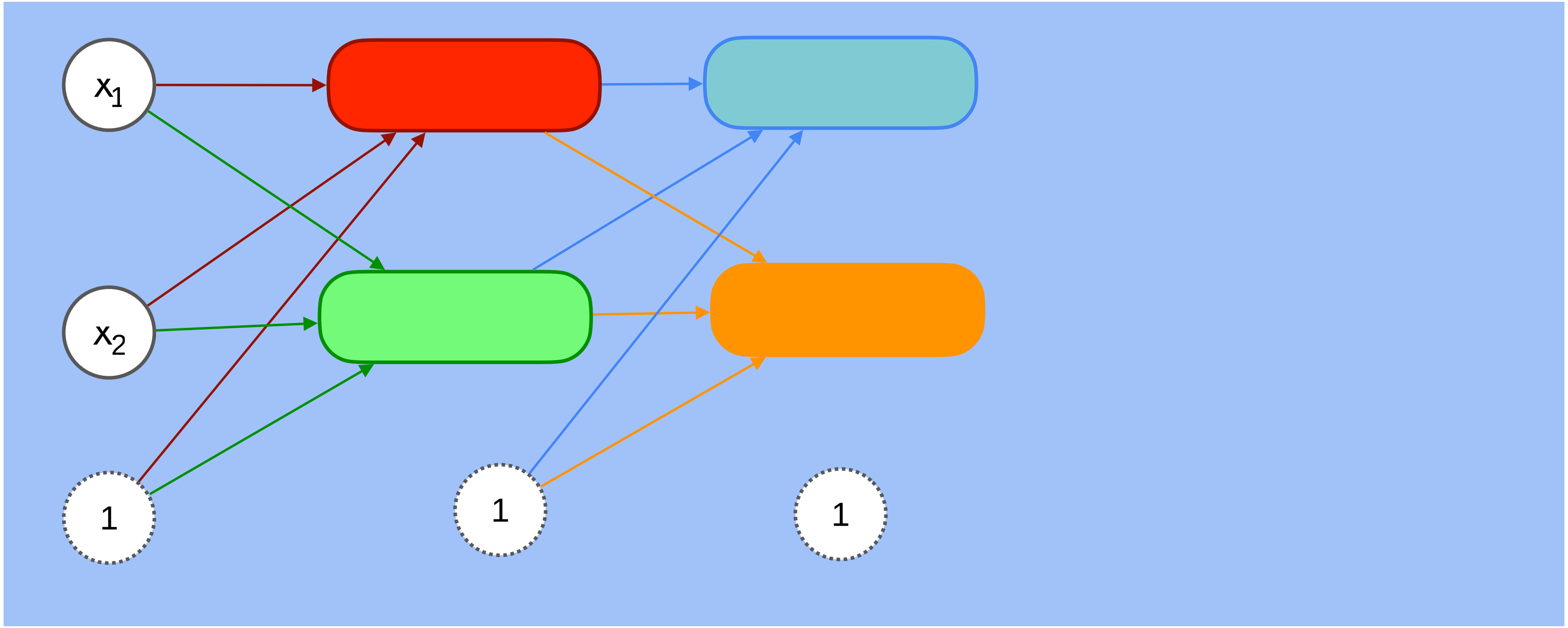
Back Propagation Introduction



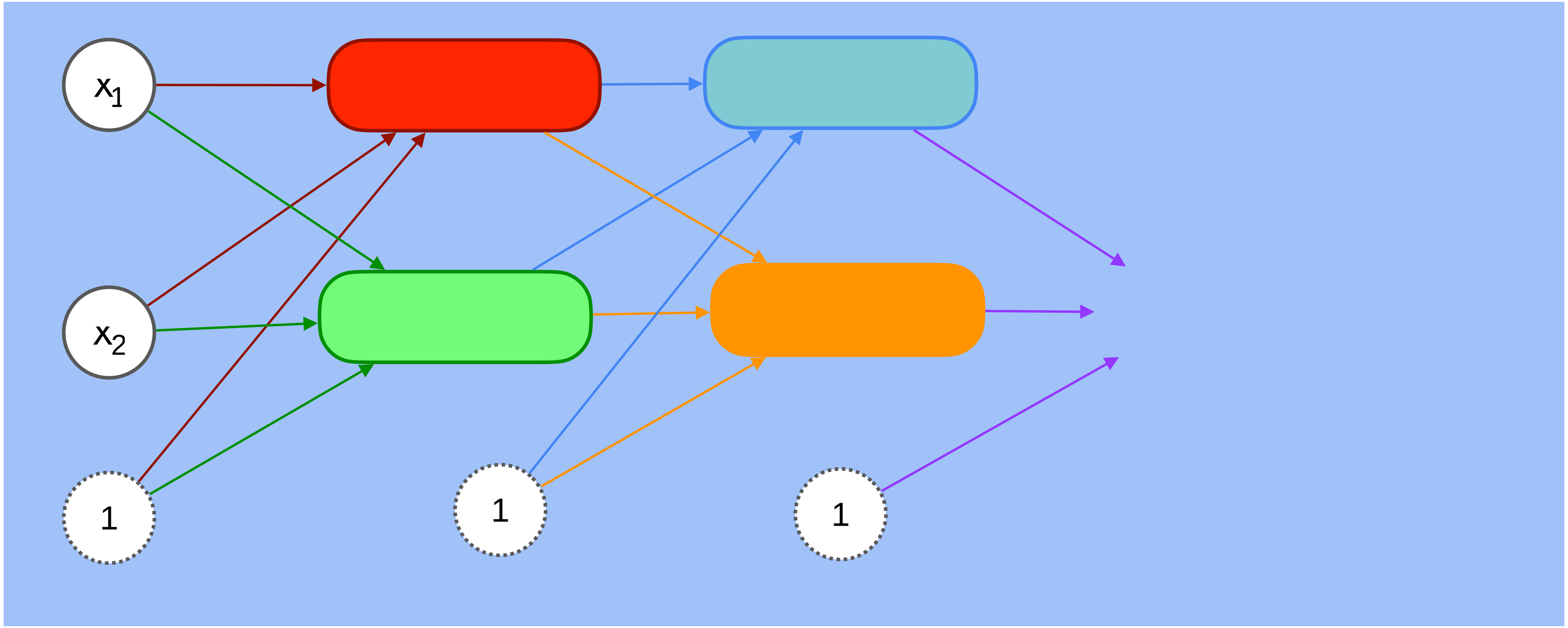
Back Propagation Introduction



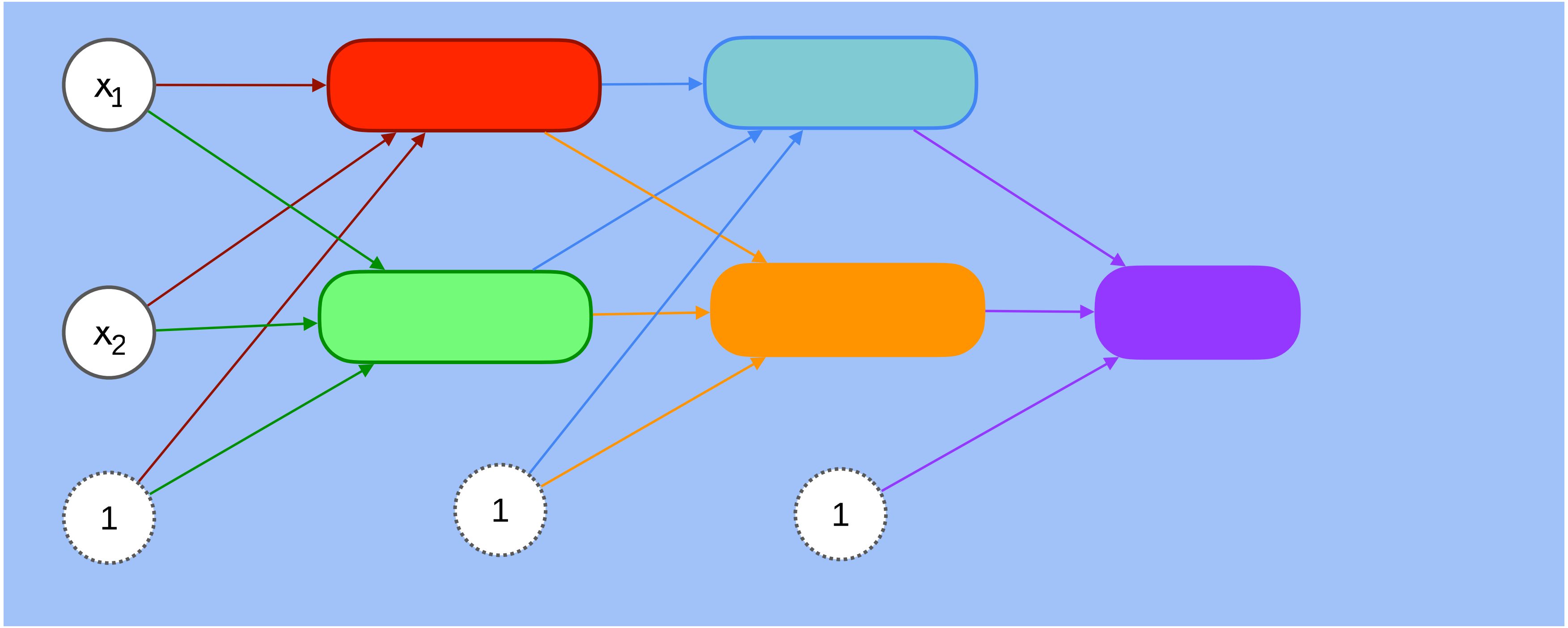
Back Propagation Introduction



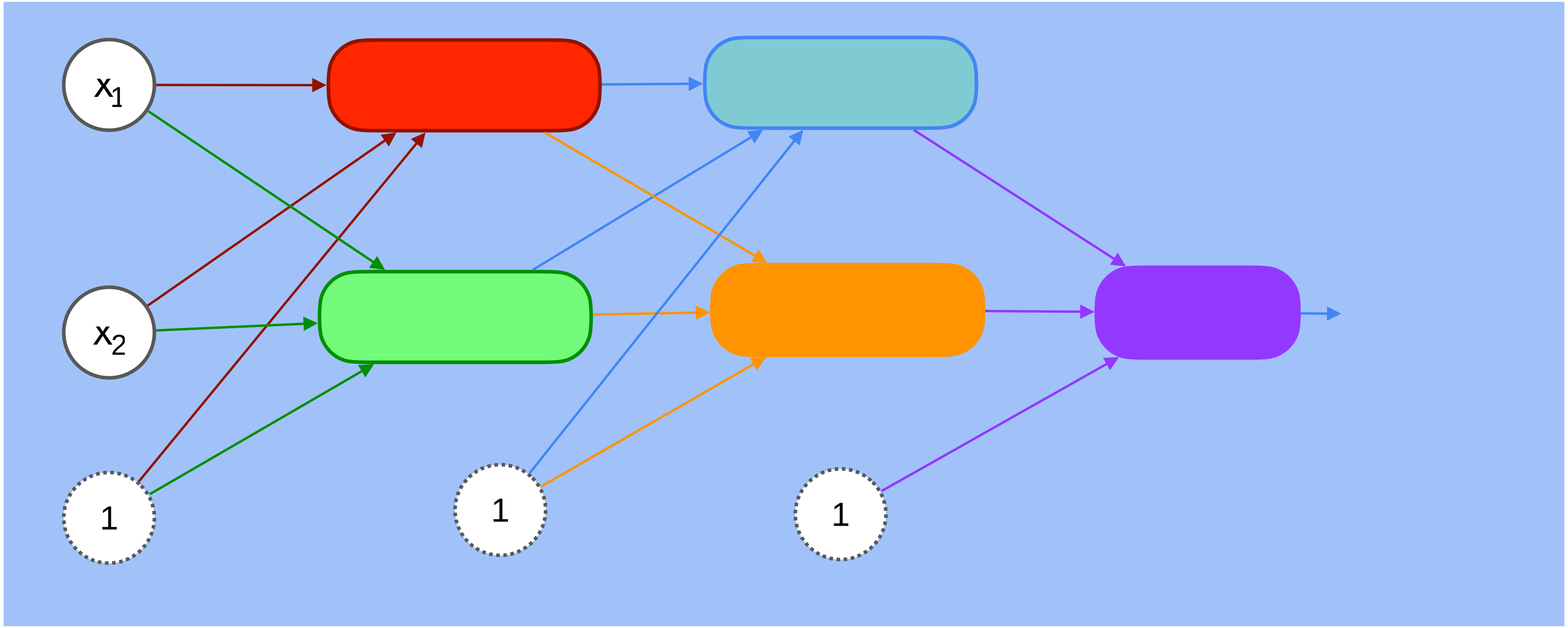
Back Propagation Introduction



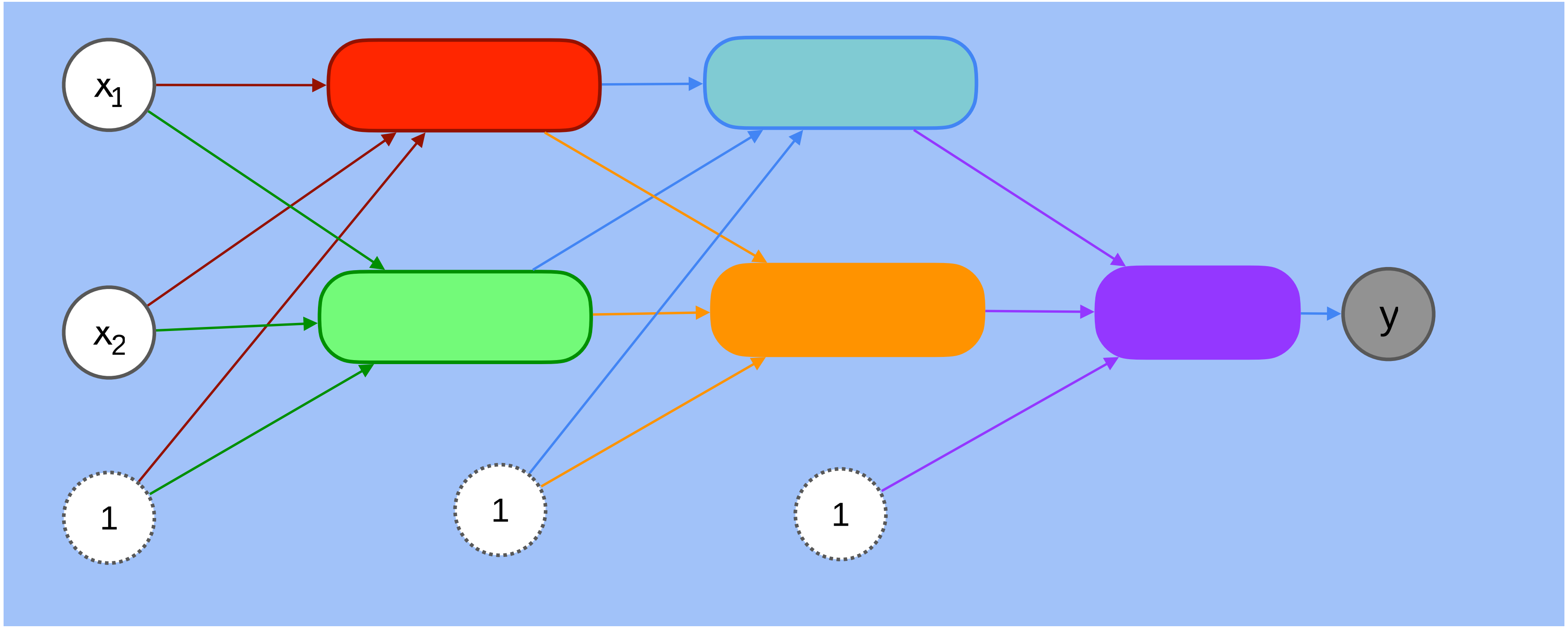
Back Propagation Introduction



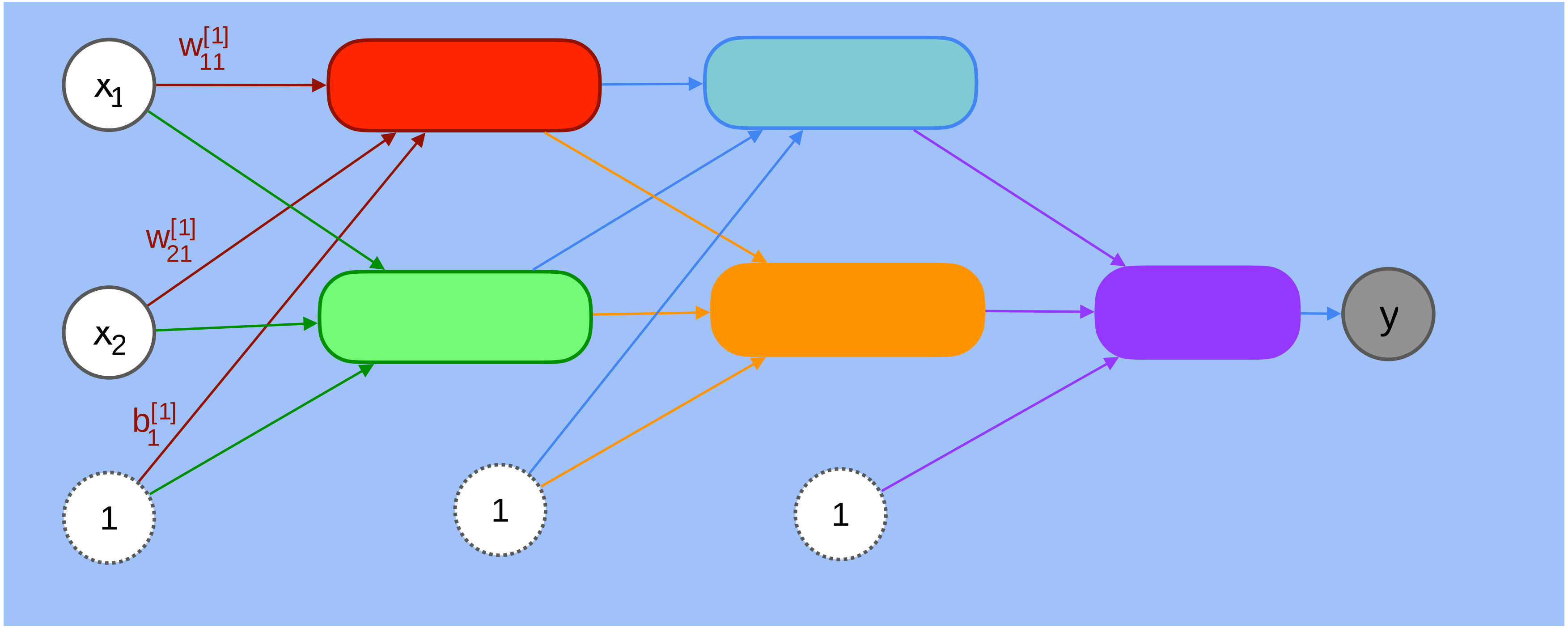
Back Propagation Introduction



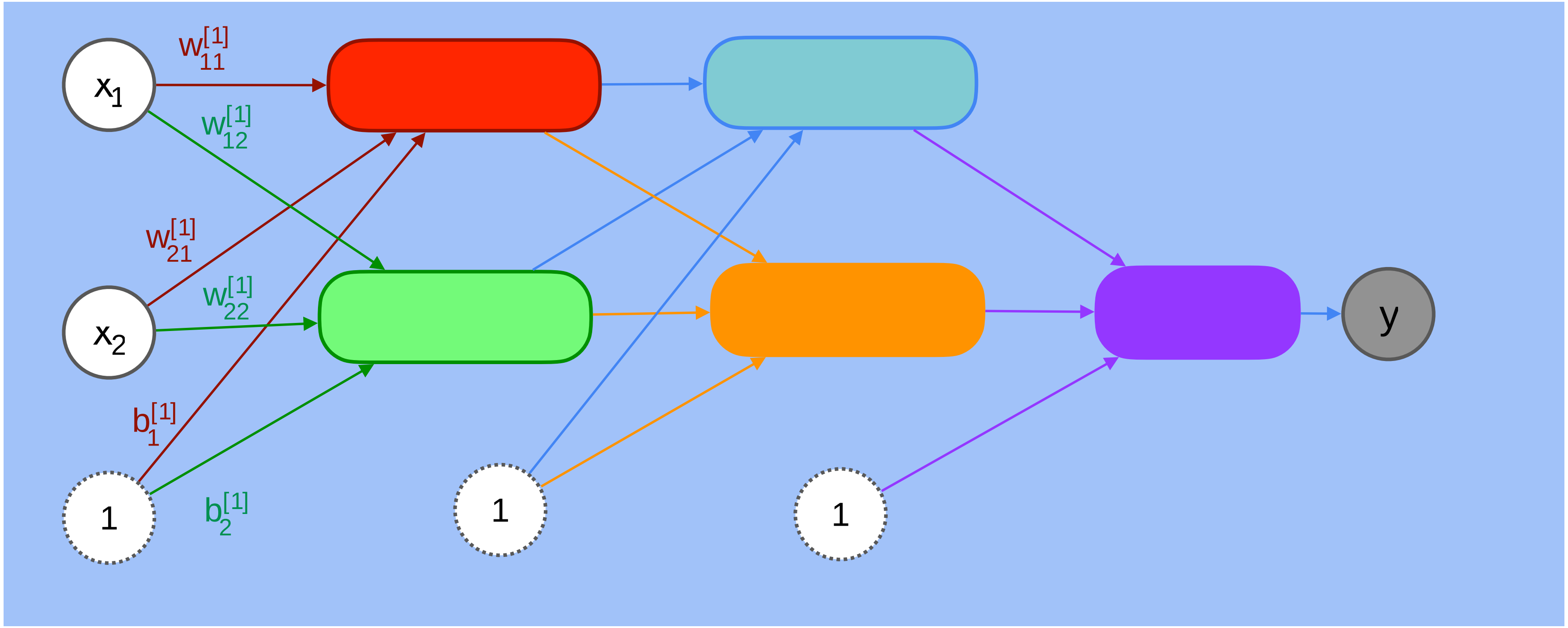
Back Propagation Introduction



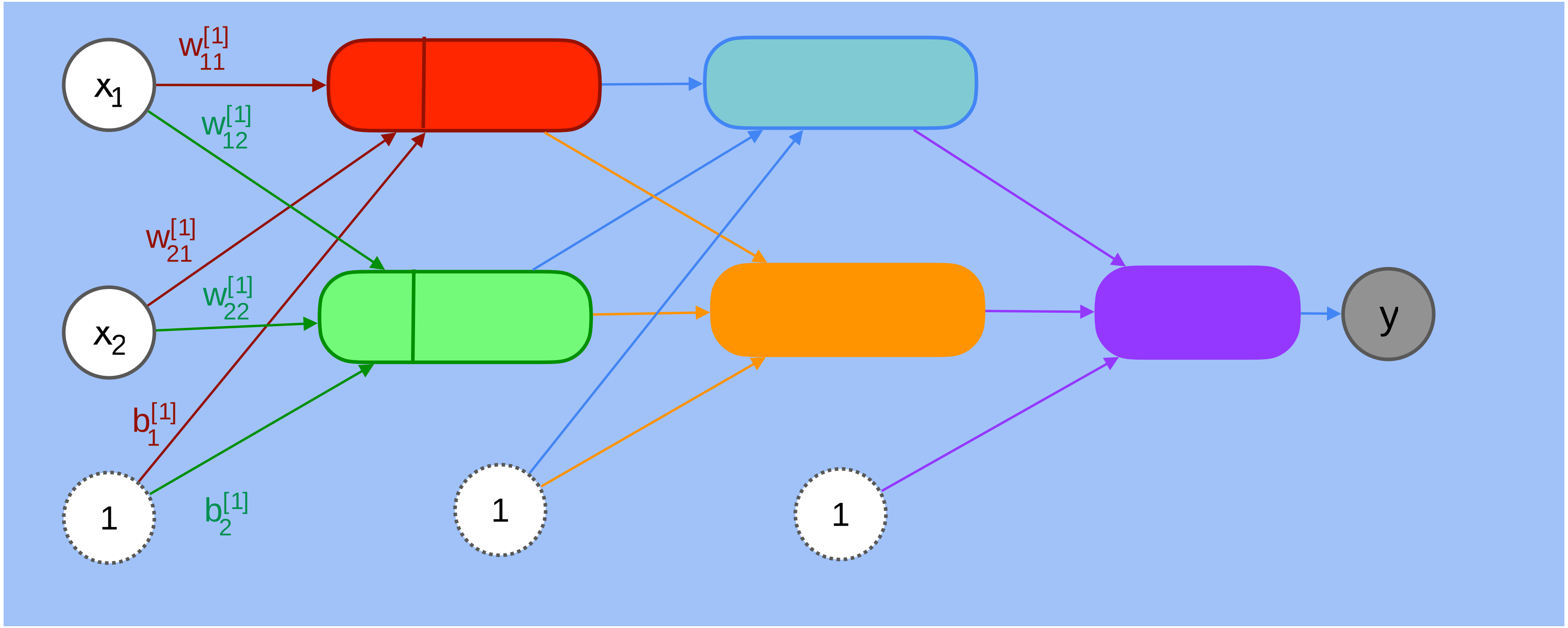
Back Propagation Introduction



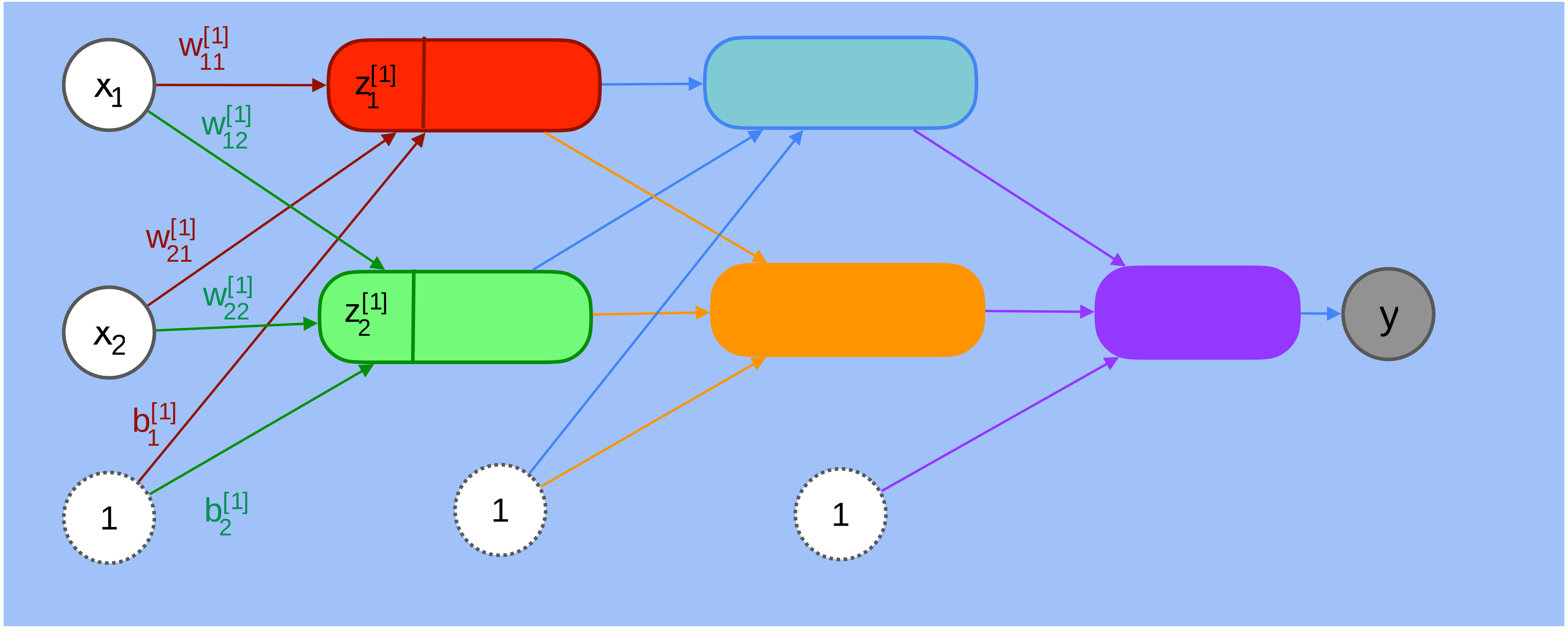
Back Propagation Introduction



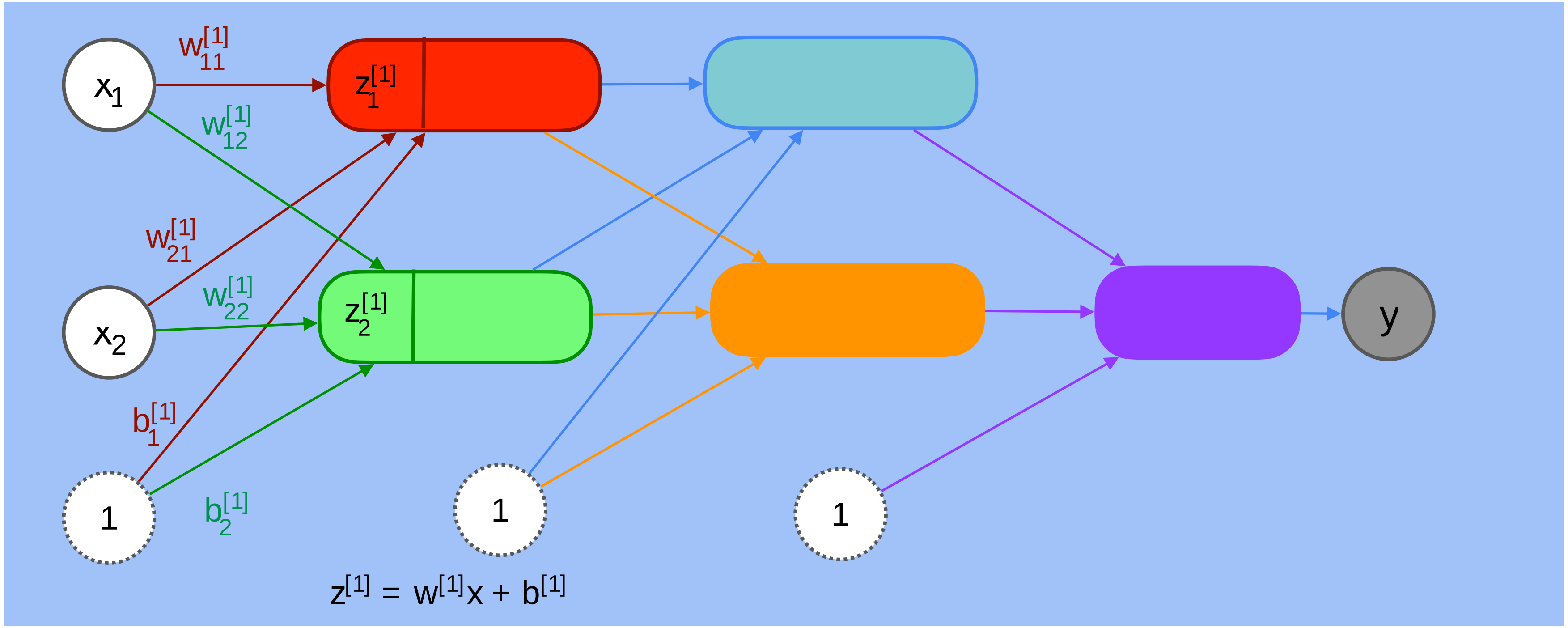
Back Propagation Introduction



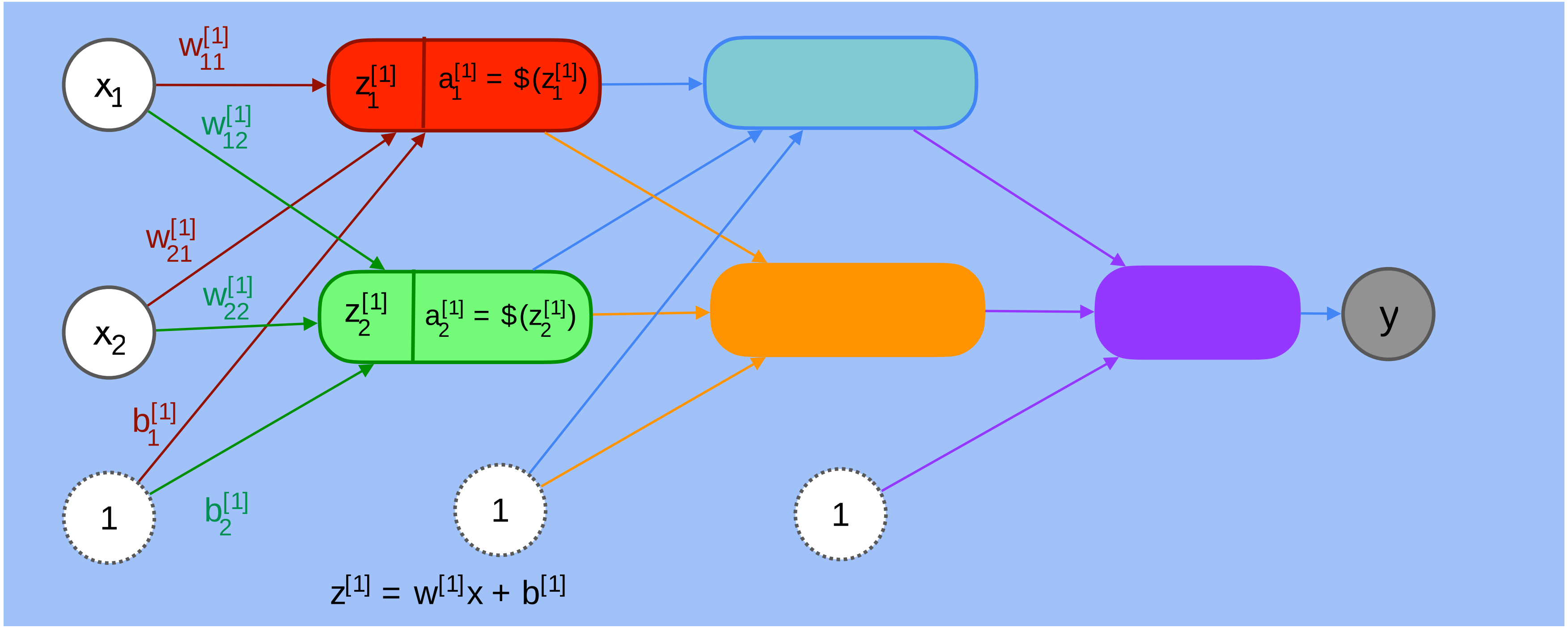
Back Propagation Introduction



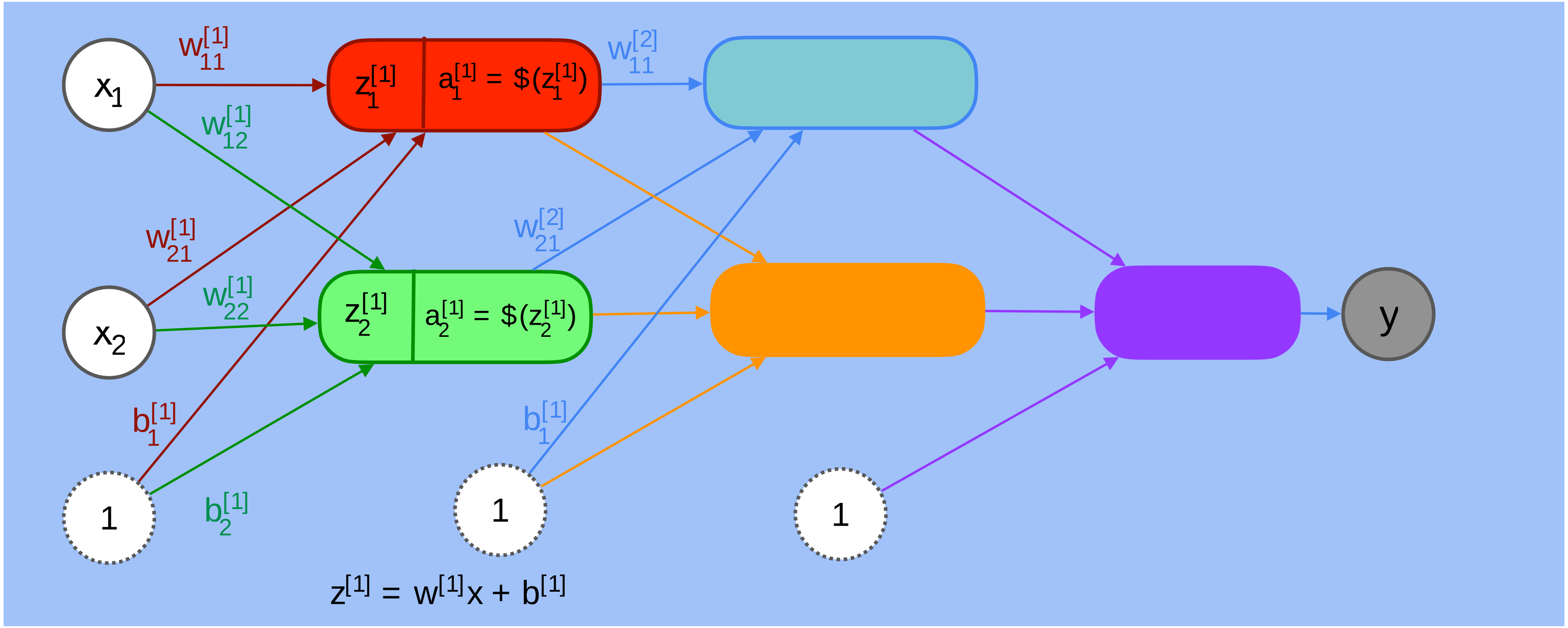
Back Propagation Introduction



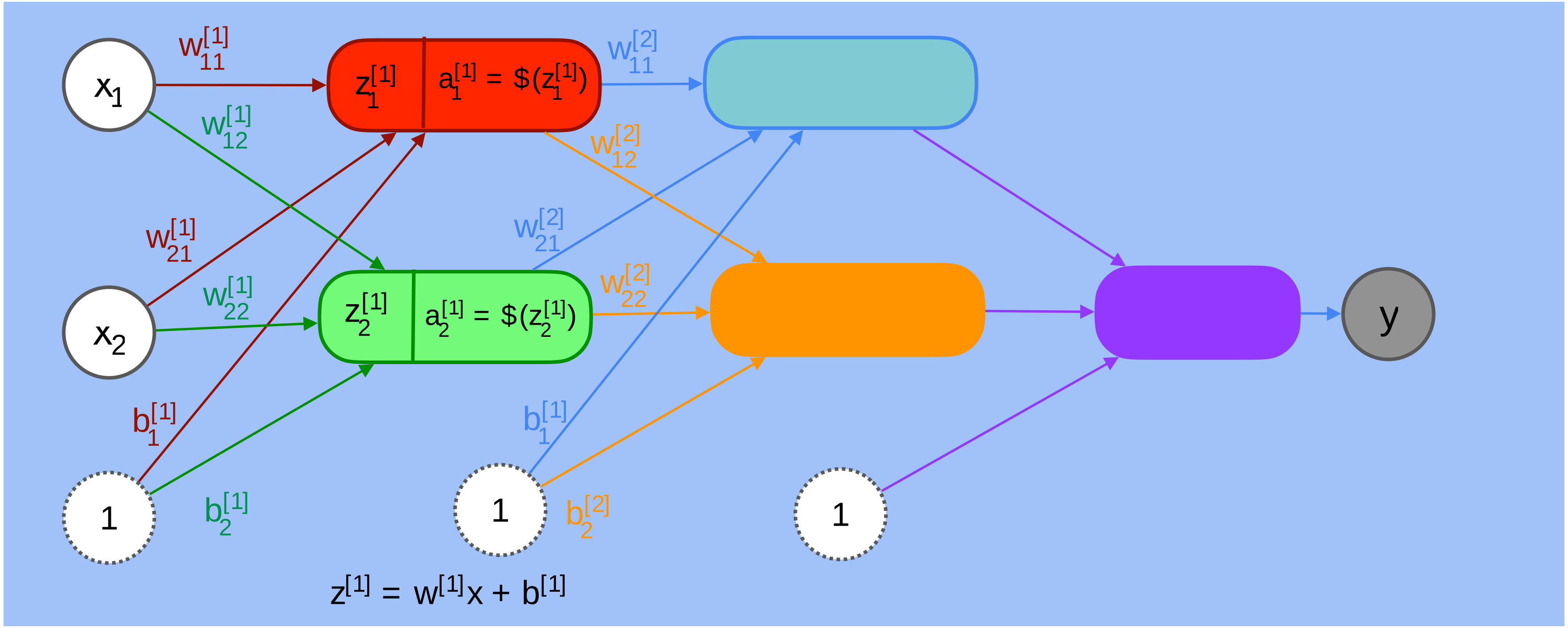
Back Propagation Introduction



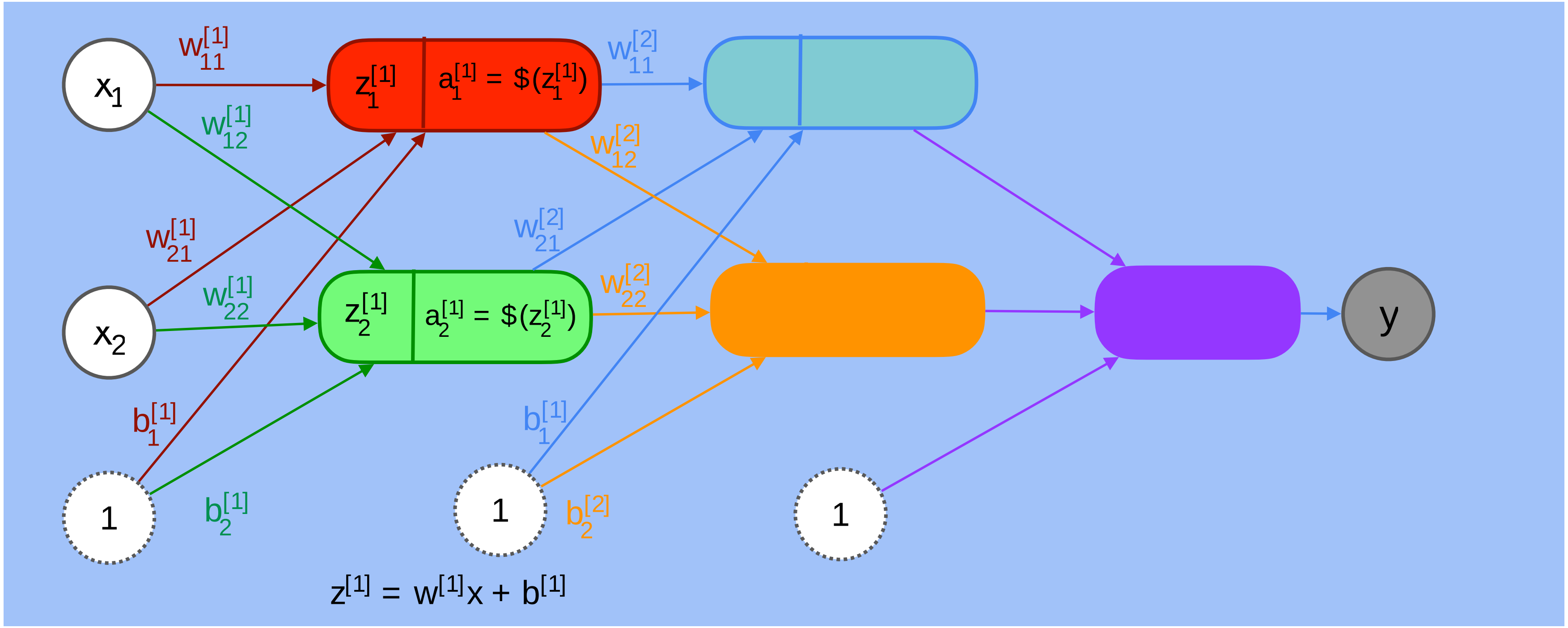
Back Propagation Introduction



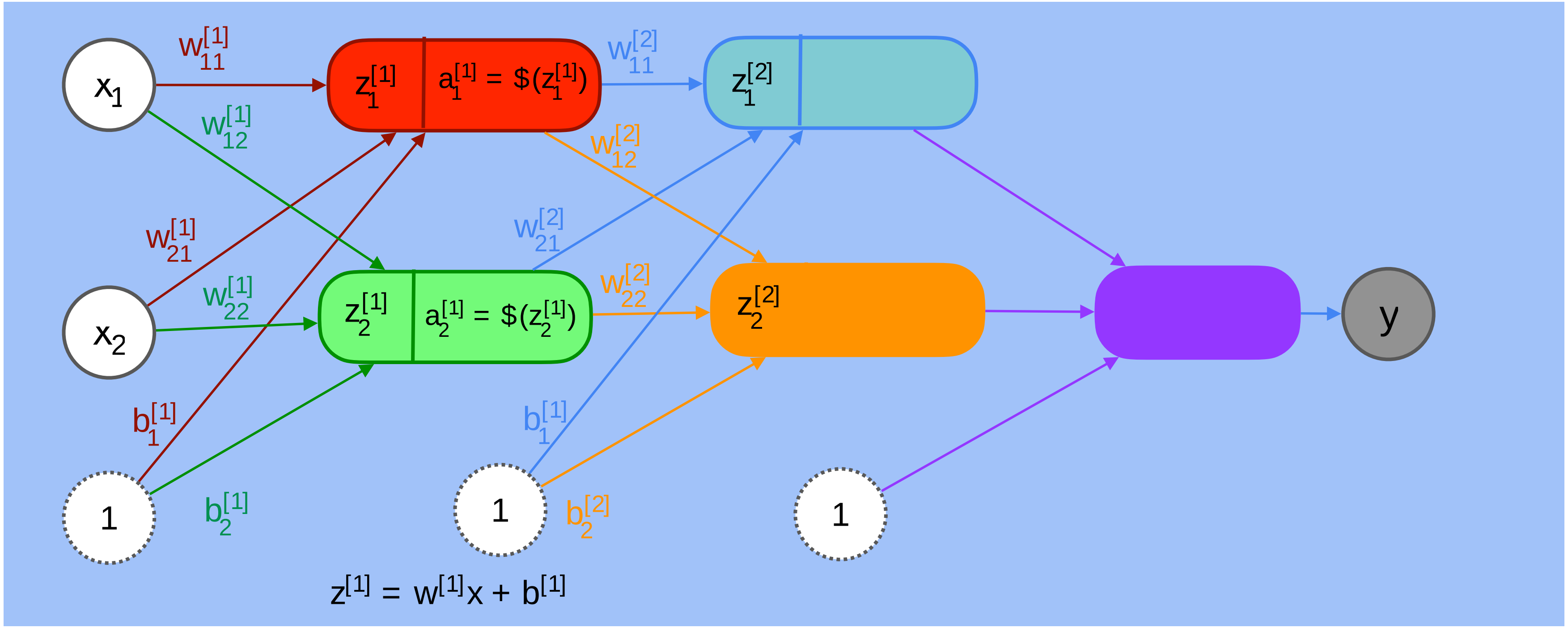
Back Propagation Introduction



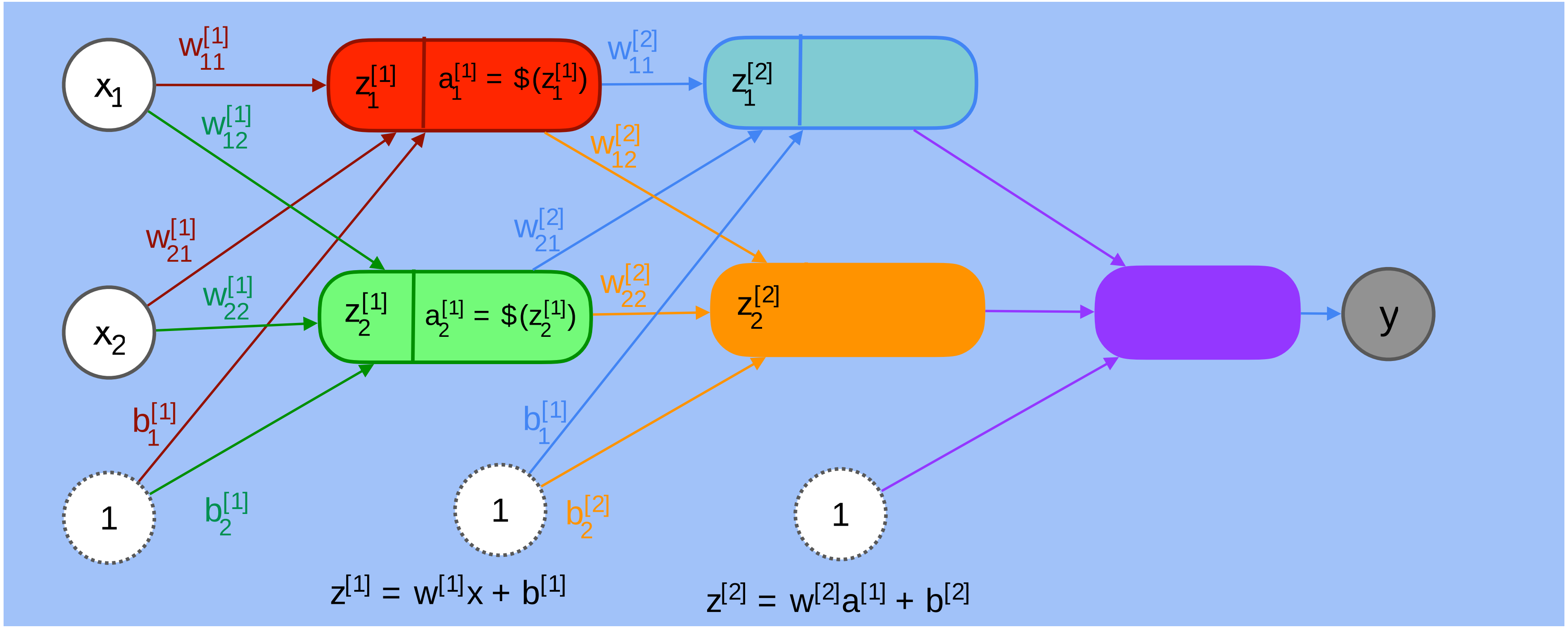
Back Propagation Introduction



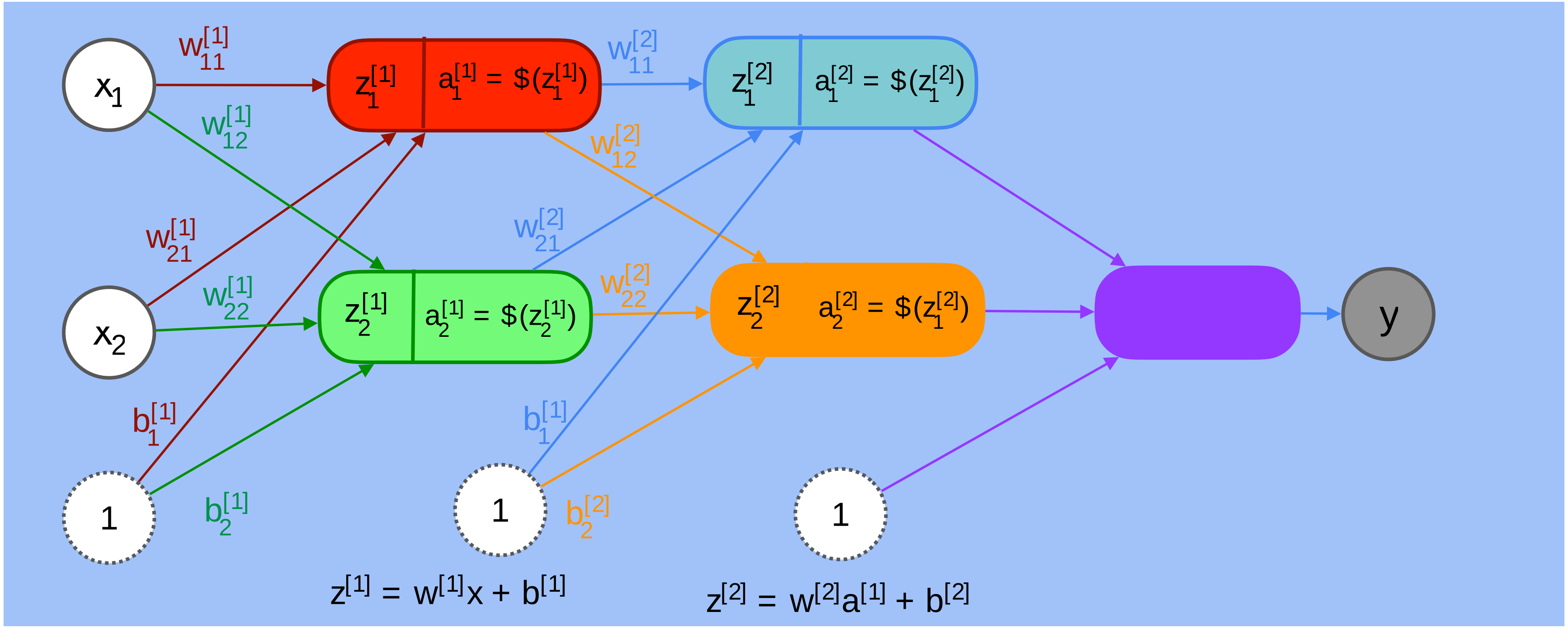
Back Propagation Introduction



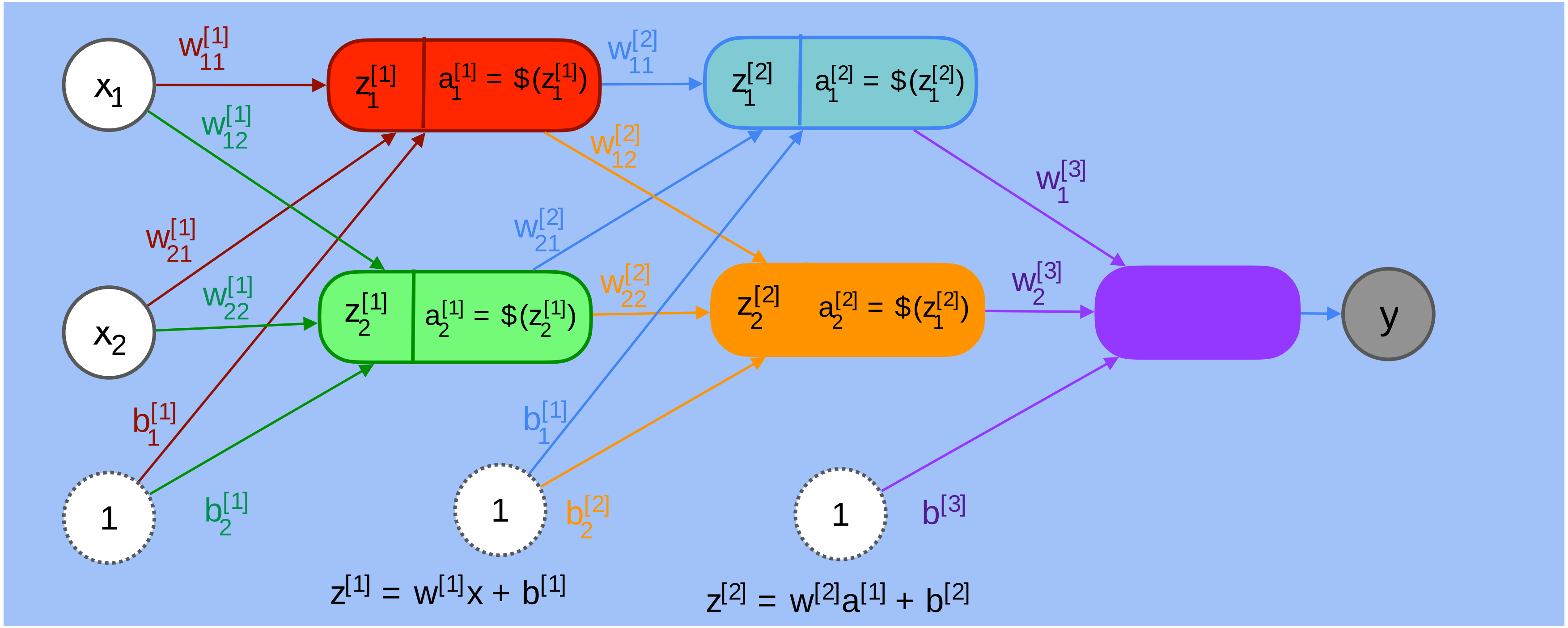
Back Propagation Introduction



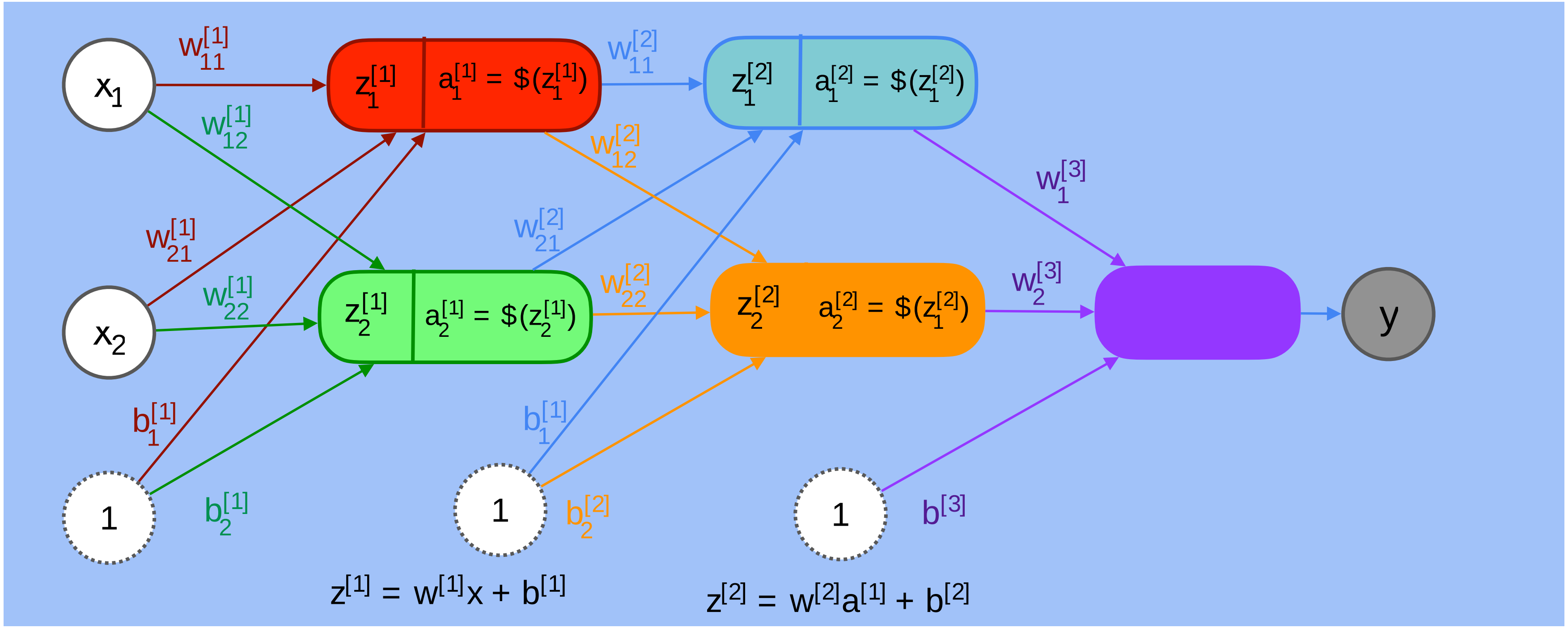
Back Propagation Introduction



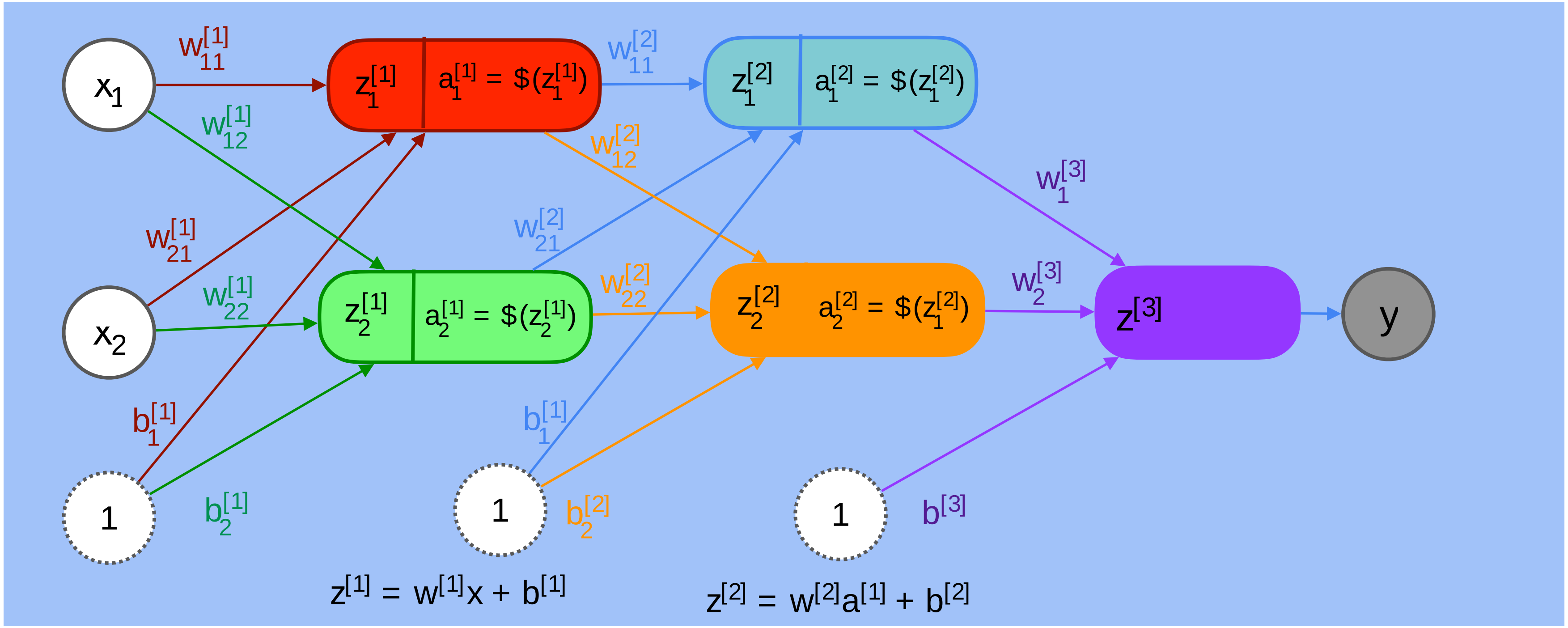
Back Propagation Introduction



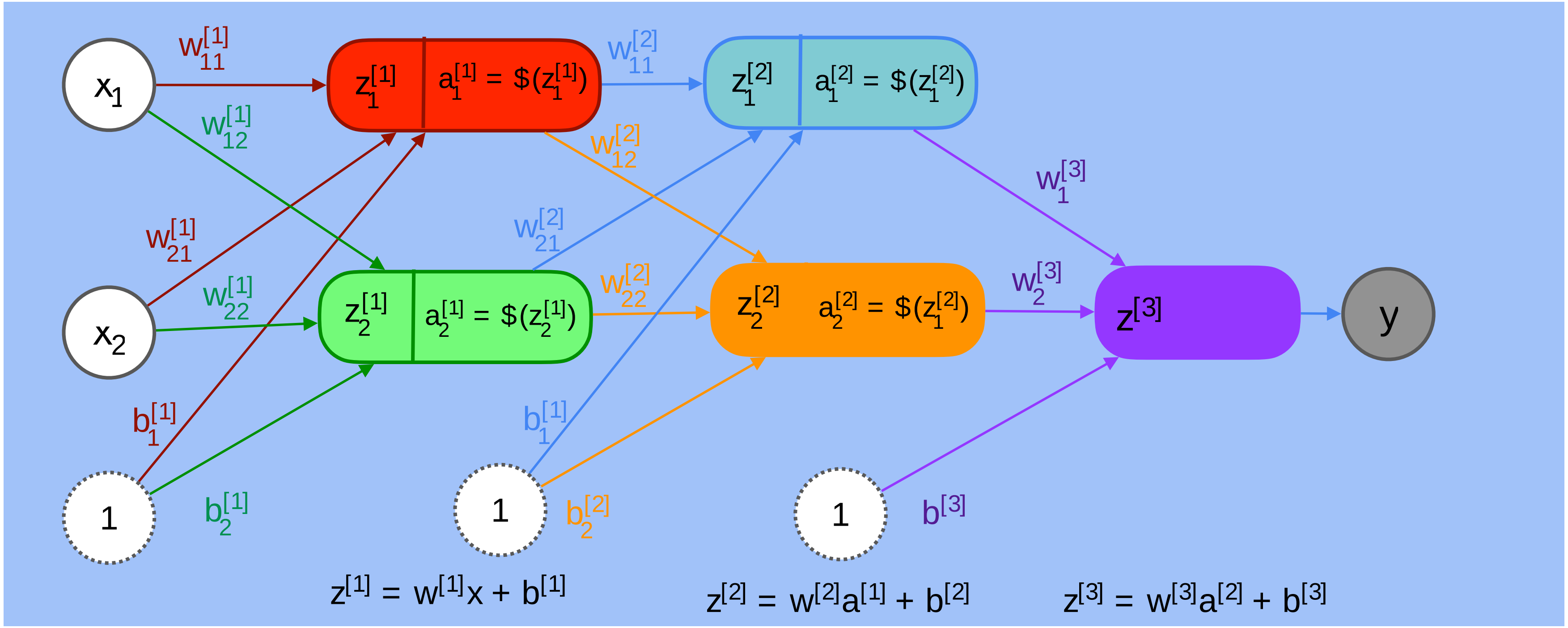
Back Propagation Introduction



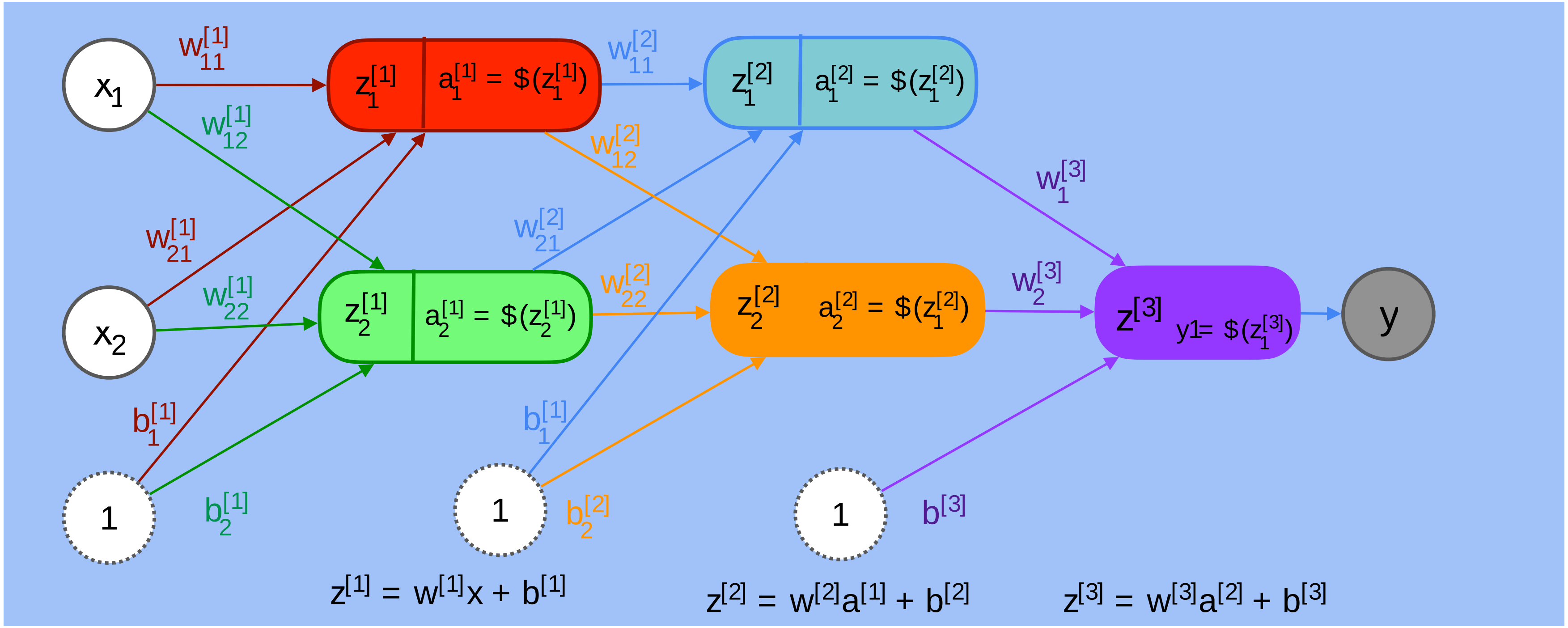
Back Propagation Introduction



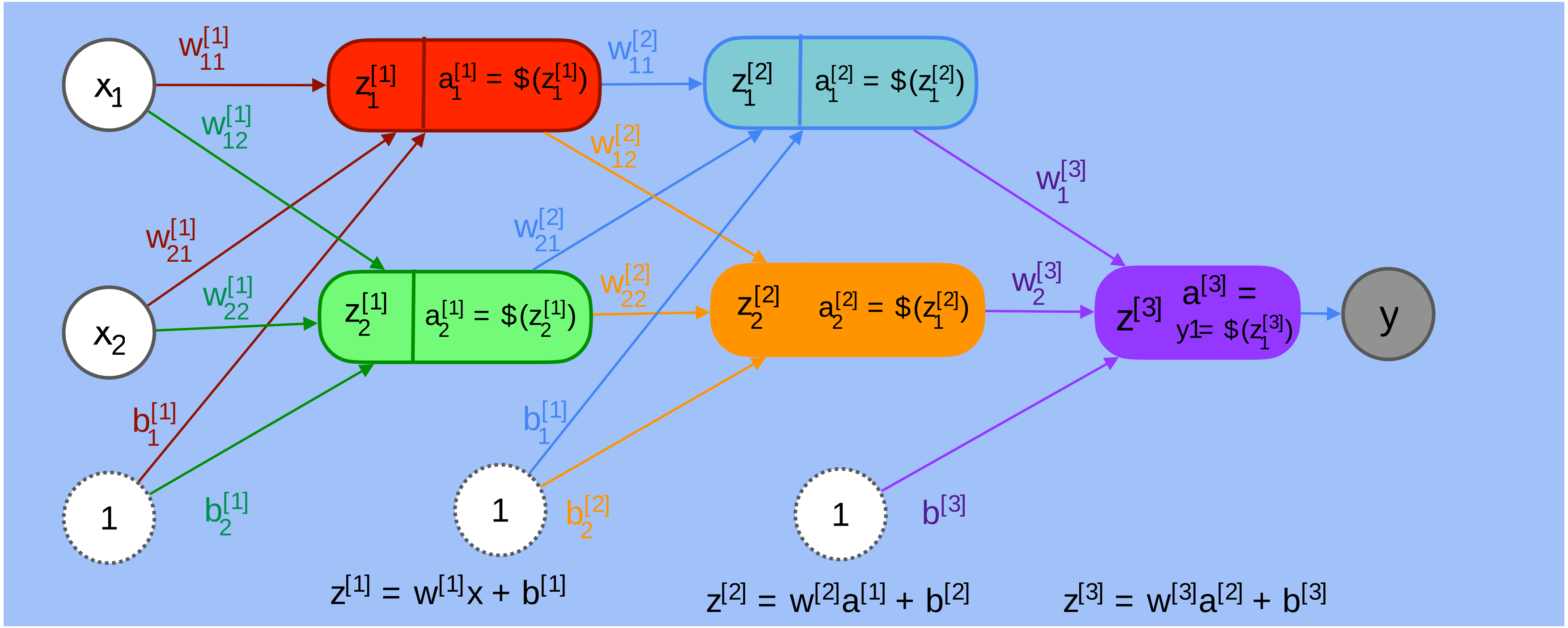
Back Propagation Introduction



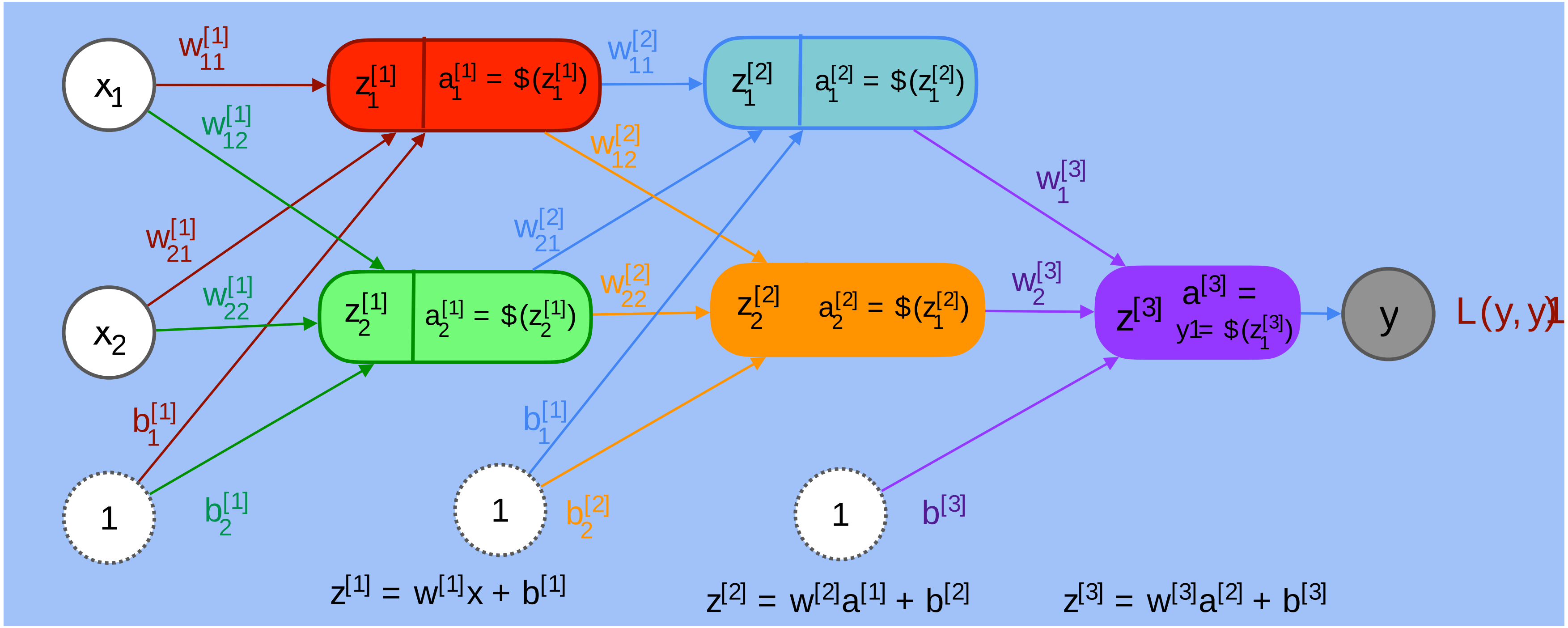
Back Propagation Introduction



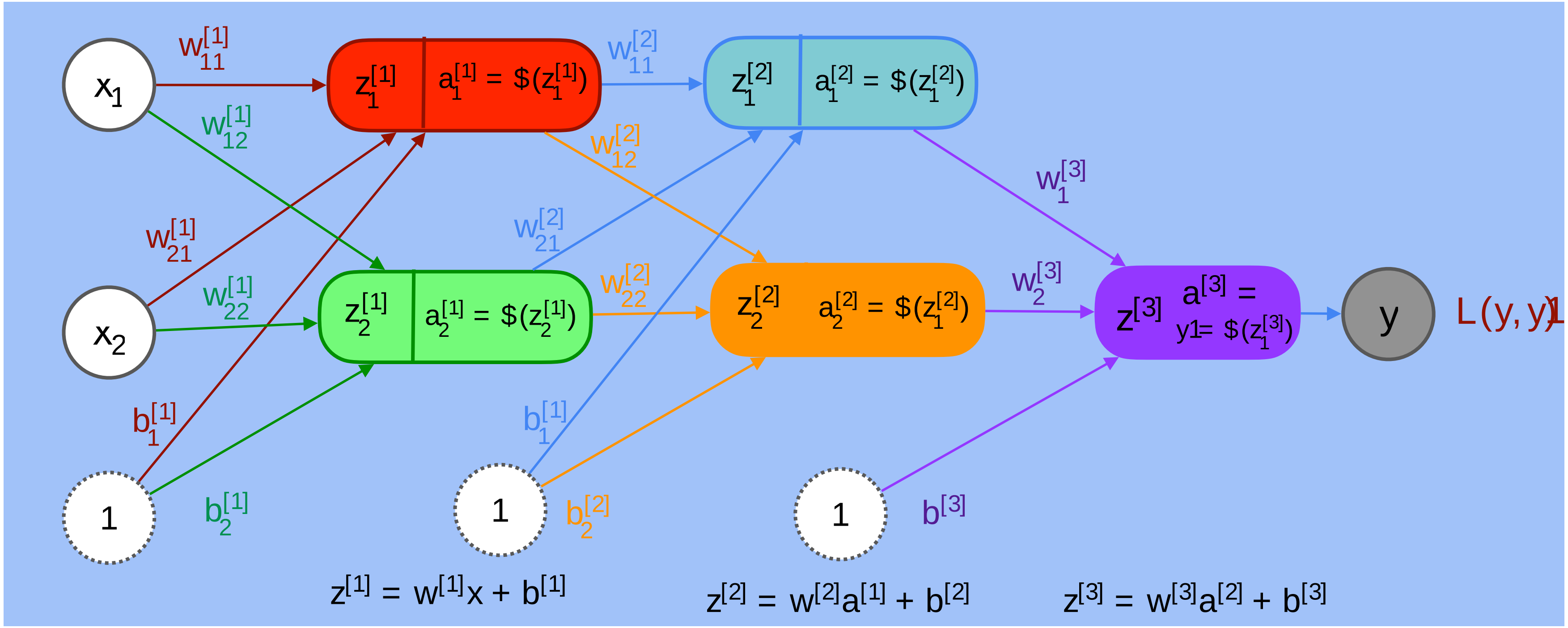
Back Propagation Introduction



Back Propagation Introduction

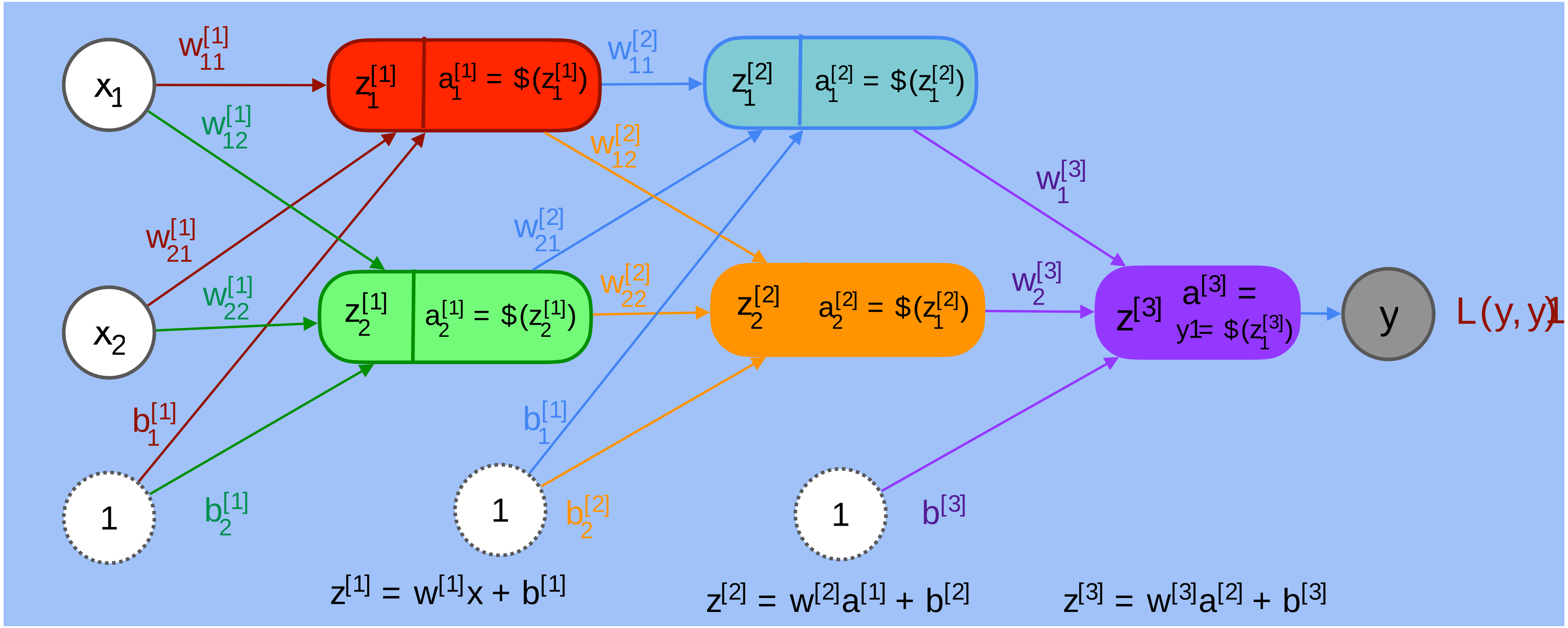


Back Propagation Introduction



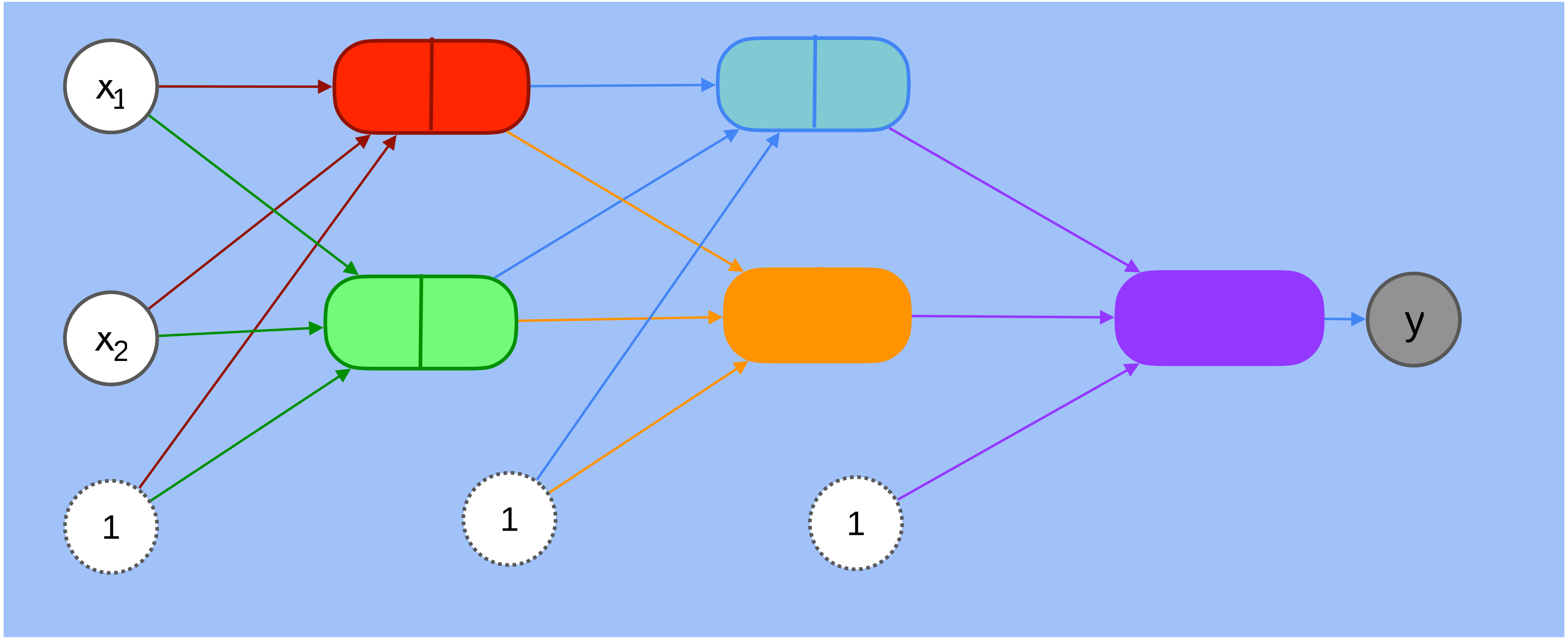
L(y, y1

Back Propagation Introduction

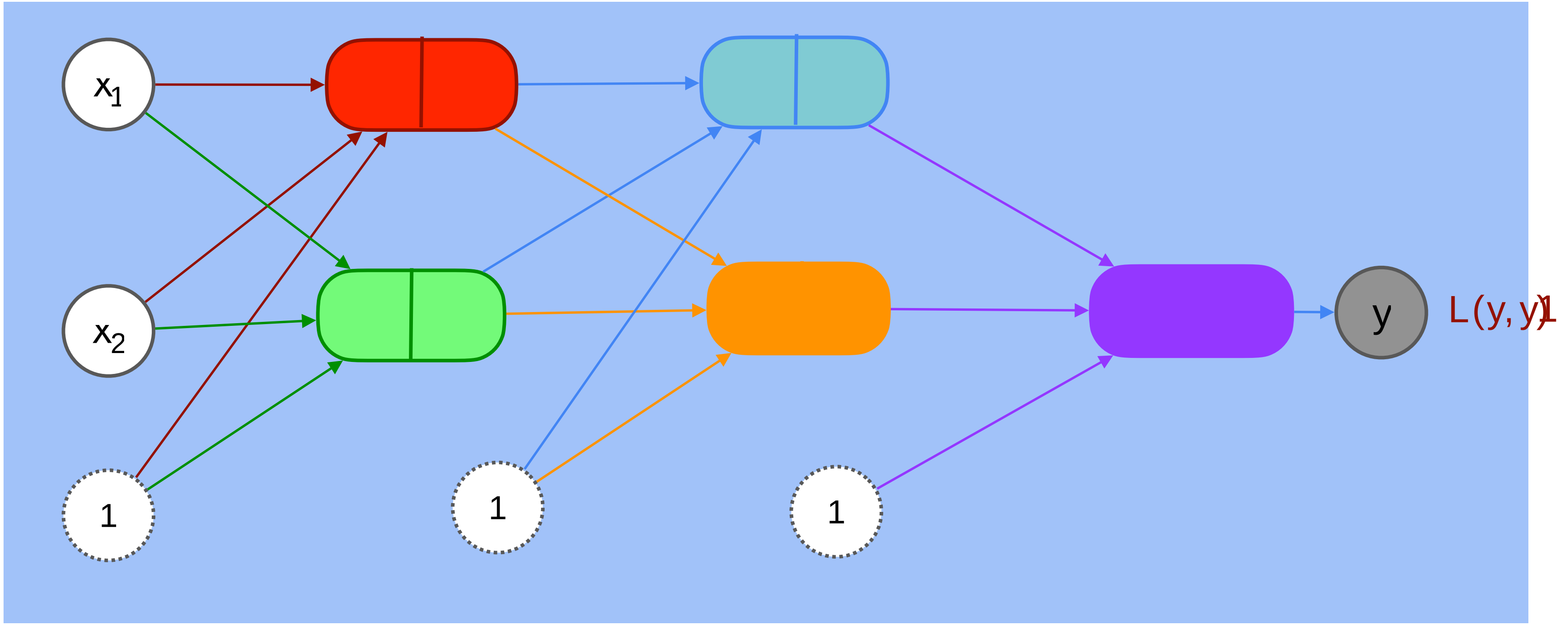


L(y, y1 = -ylog(y1 " (1 " y)log(1 " y1

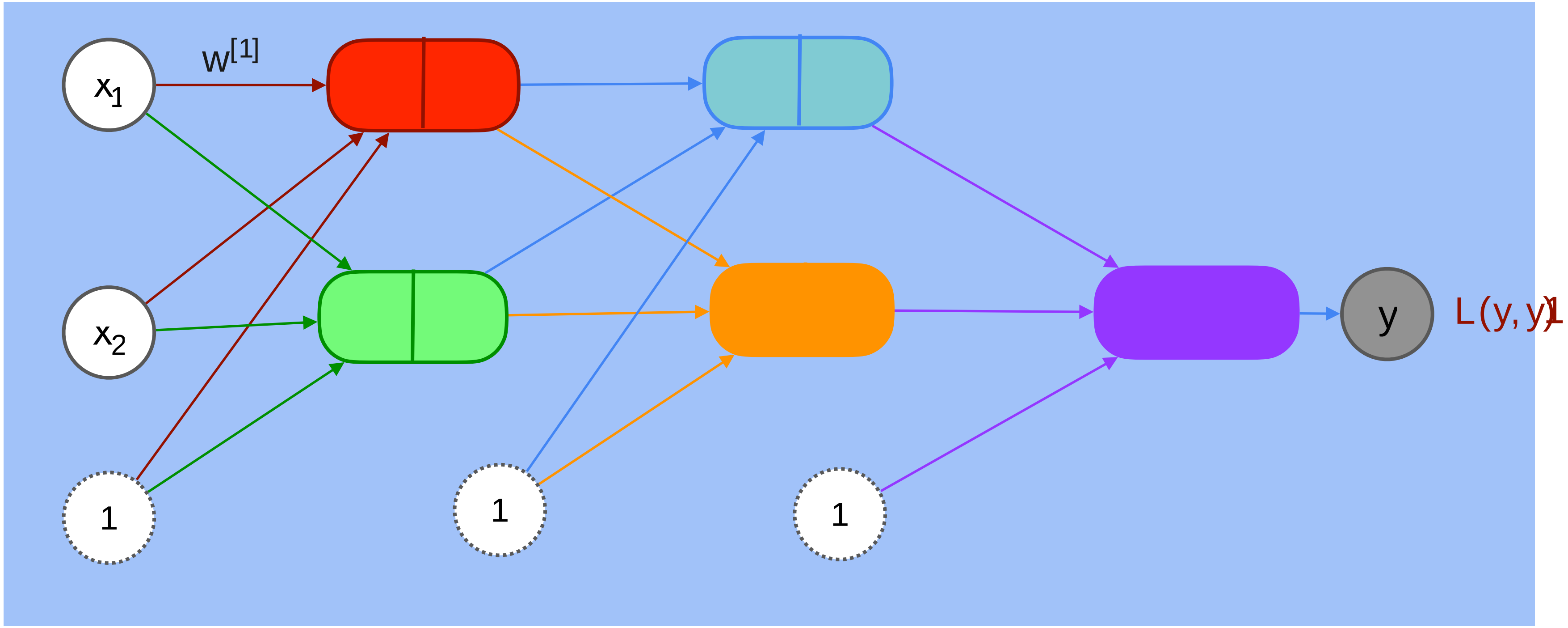
Back Propagation Introduction



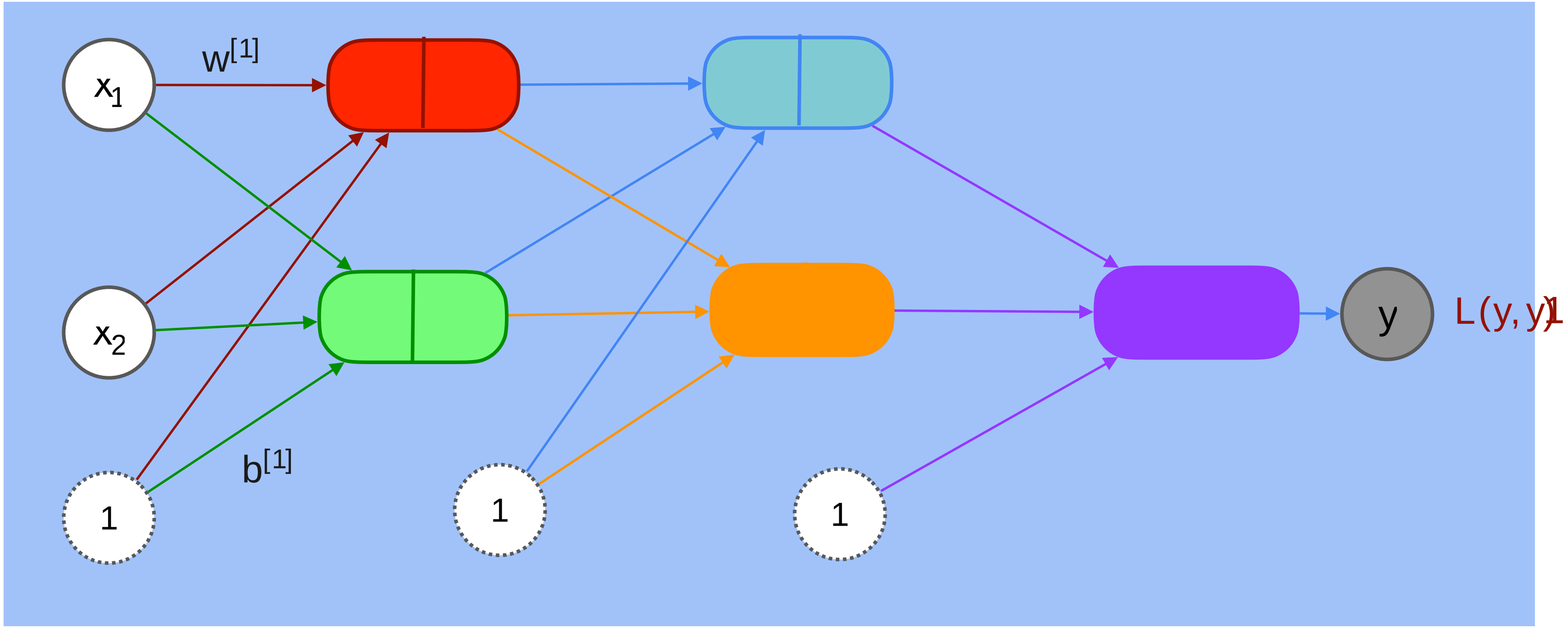
Back Propagation Introduction



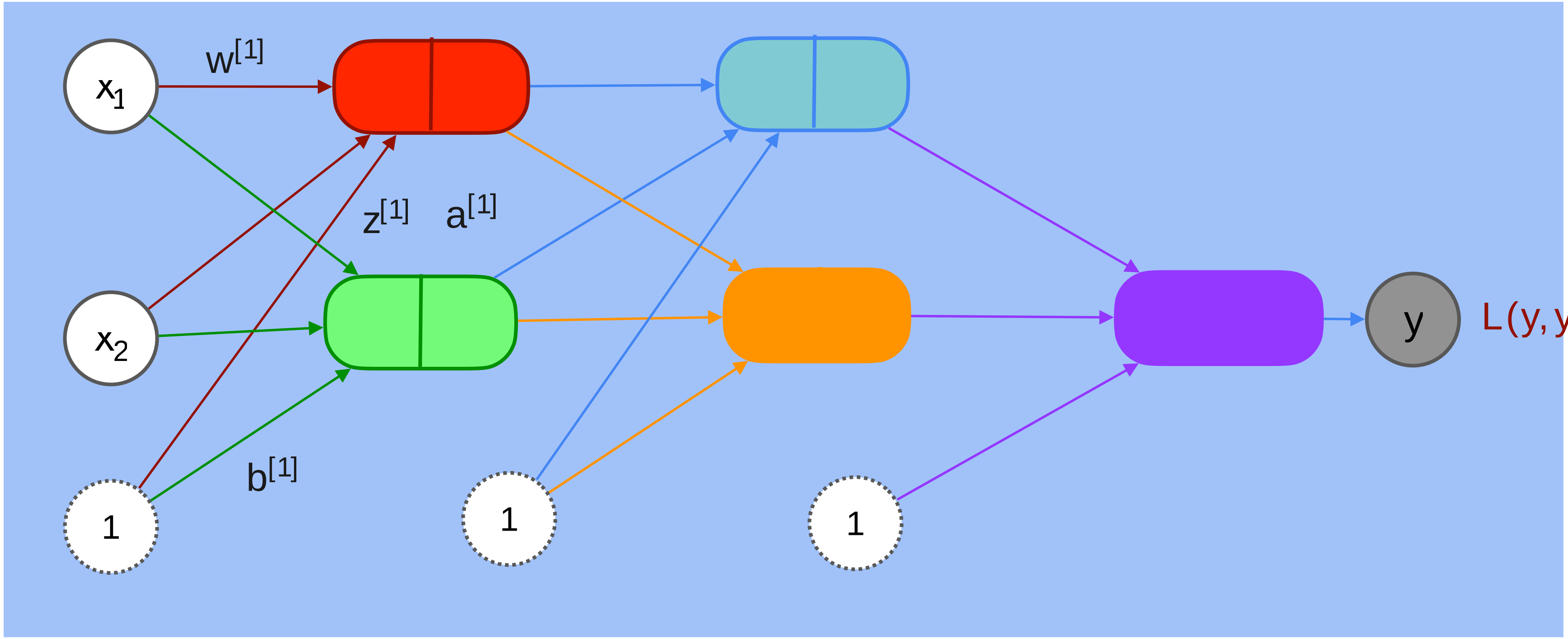
Back Propagation Introduction



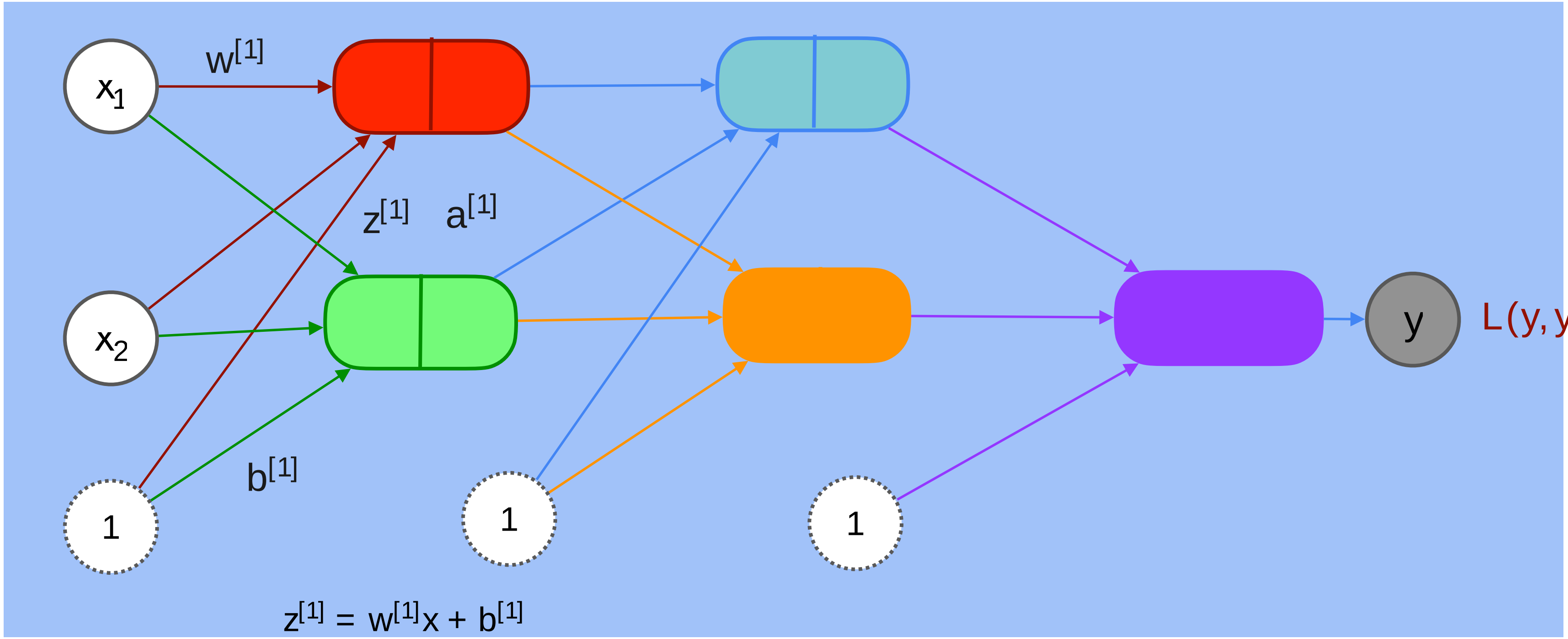
Back Propagation Introduction



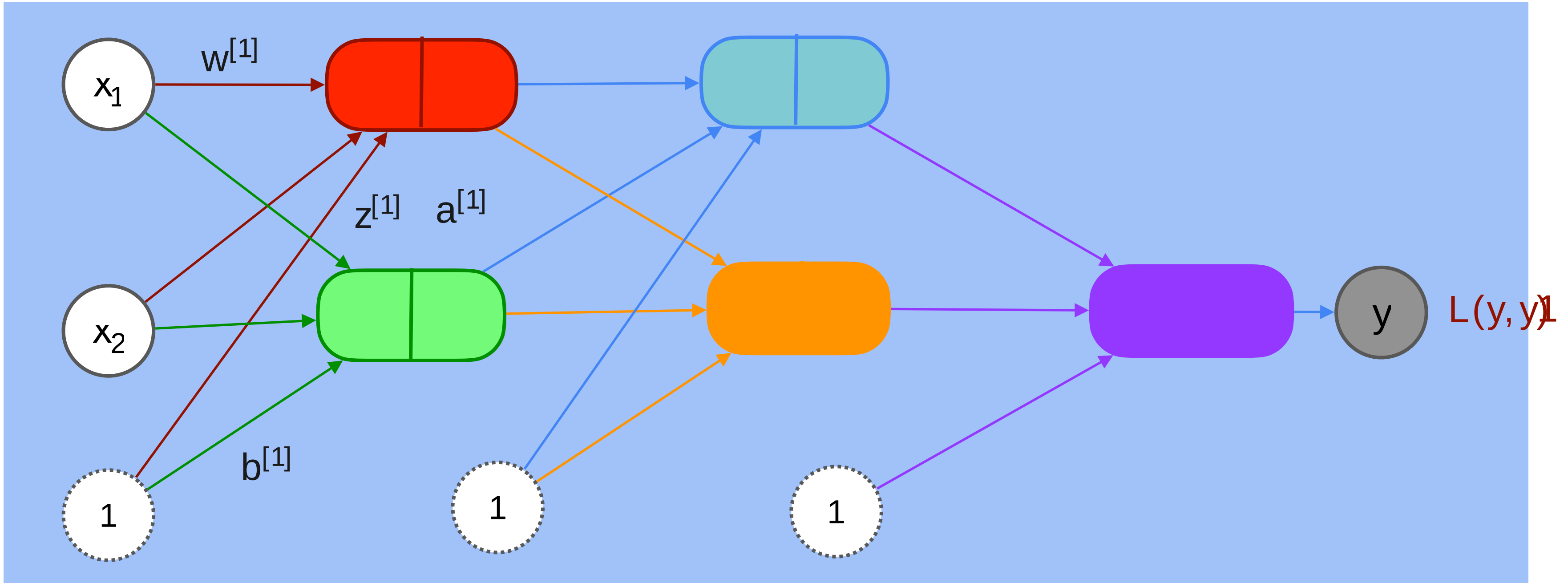
Back Propagation Introduction



Back Propagation Introduction

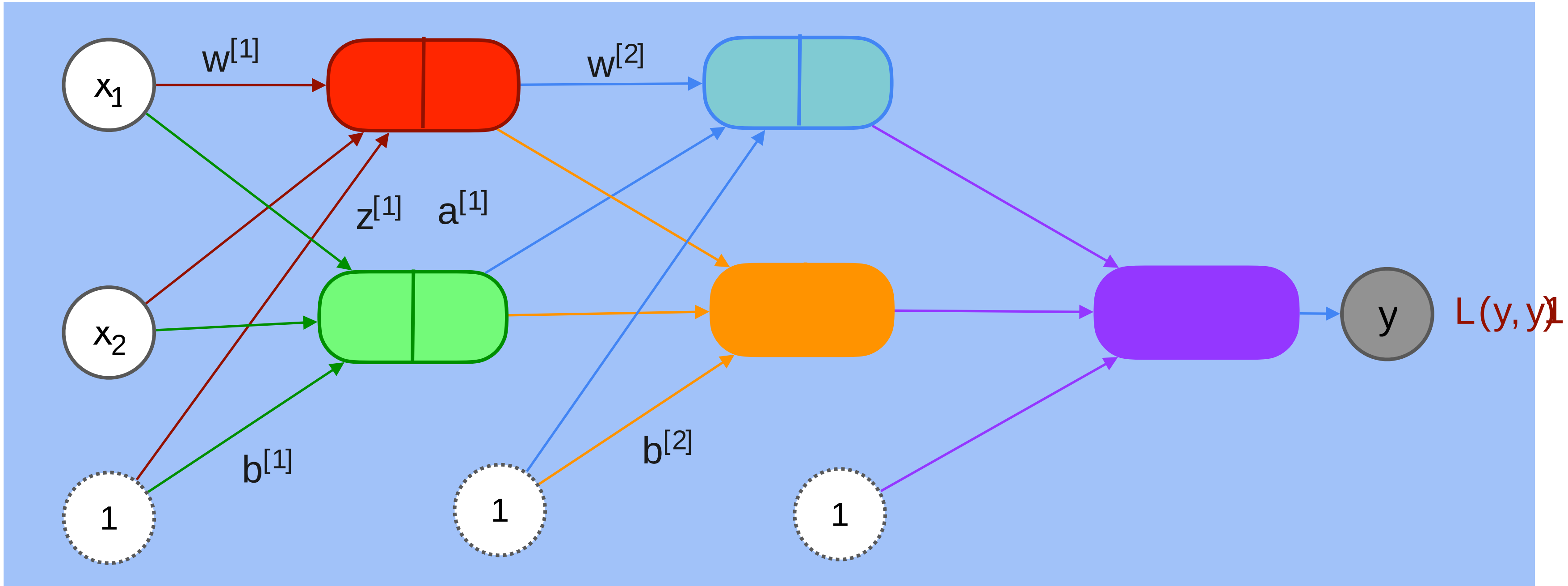


Back Propagation Introduction



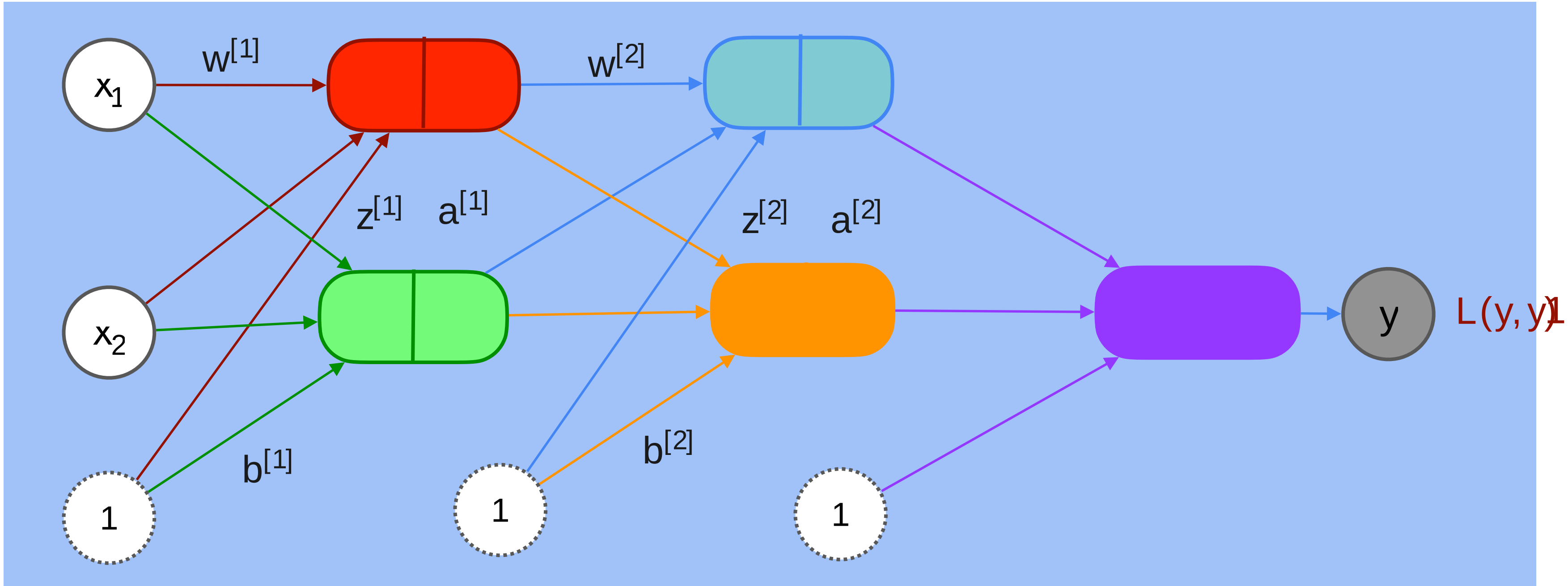
$z^{[1]} = w^{[1]}x + b^{[1]}$
$a^{[1]} = \$(z^{[1]})$

Back Propagation Introduction



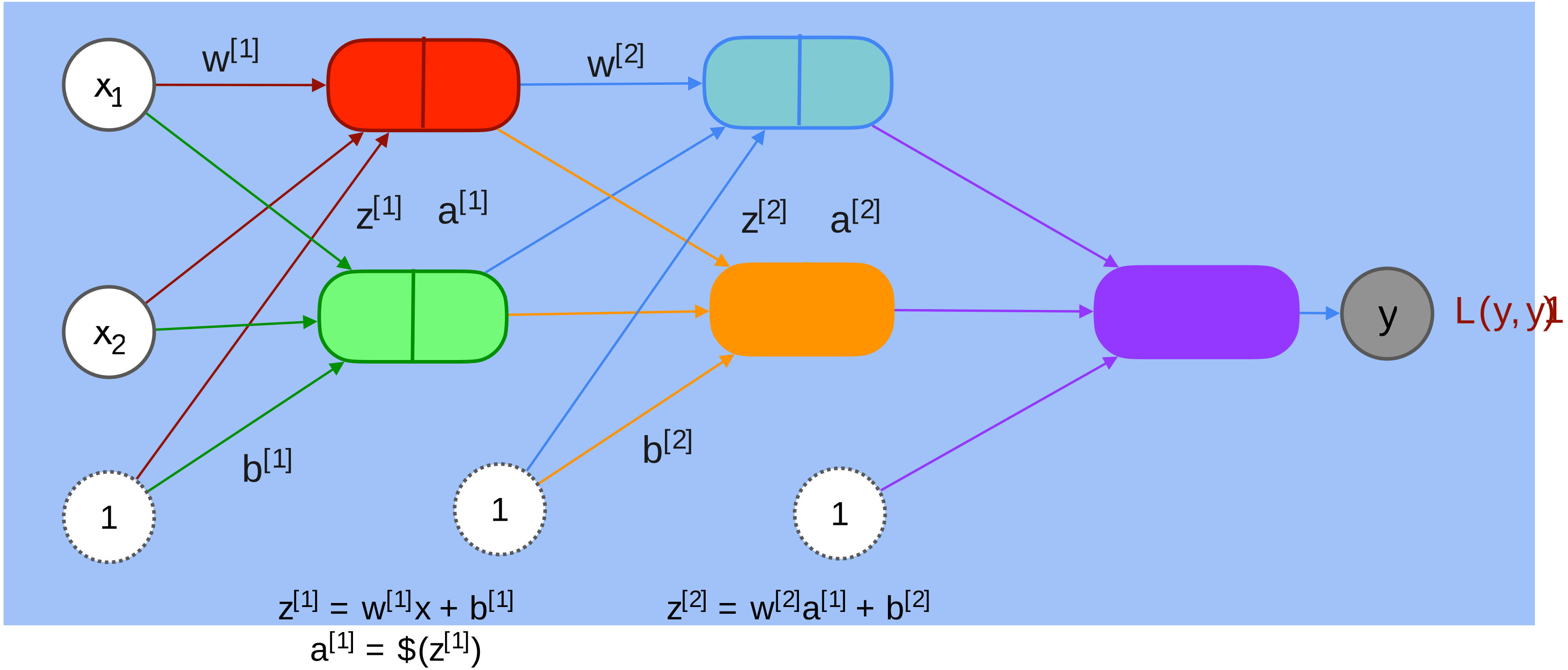
$$z^{[1]} = w^{[1]}x + b^{[1]}$$
$$a^{[1]} = \$(z^{[1]})$$

Back Propagation Introduction

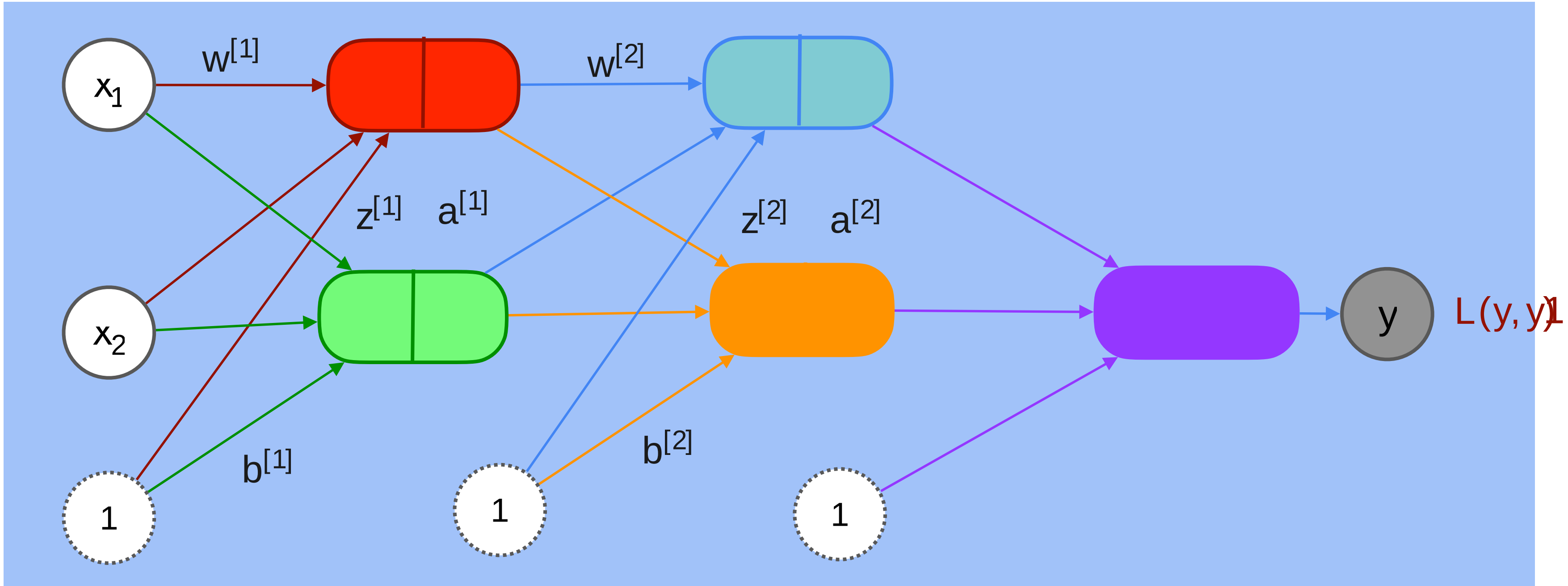


$$z^{[1]} = w^{[1]}x + b^{[1]}$$
$$a^{[1]} = \$(z^{[1]})$$

Back Propagation Introduction



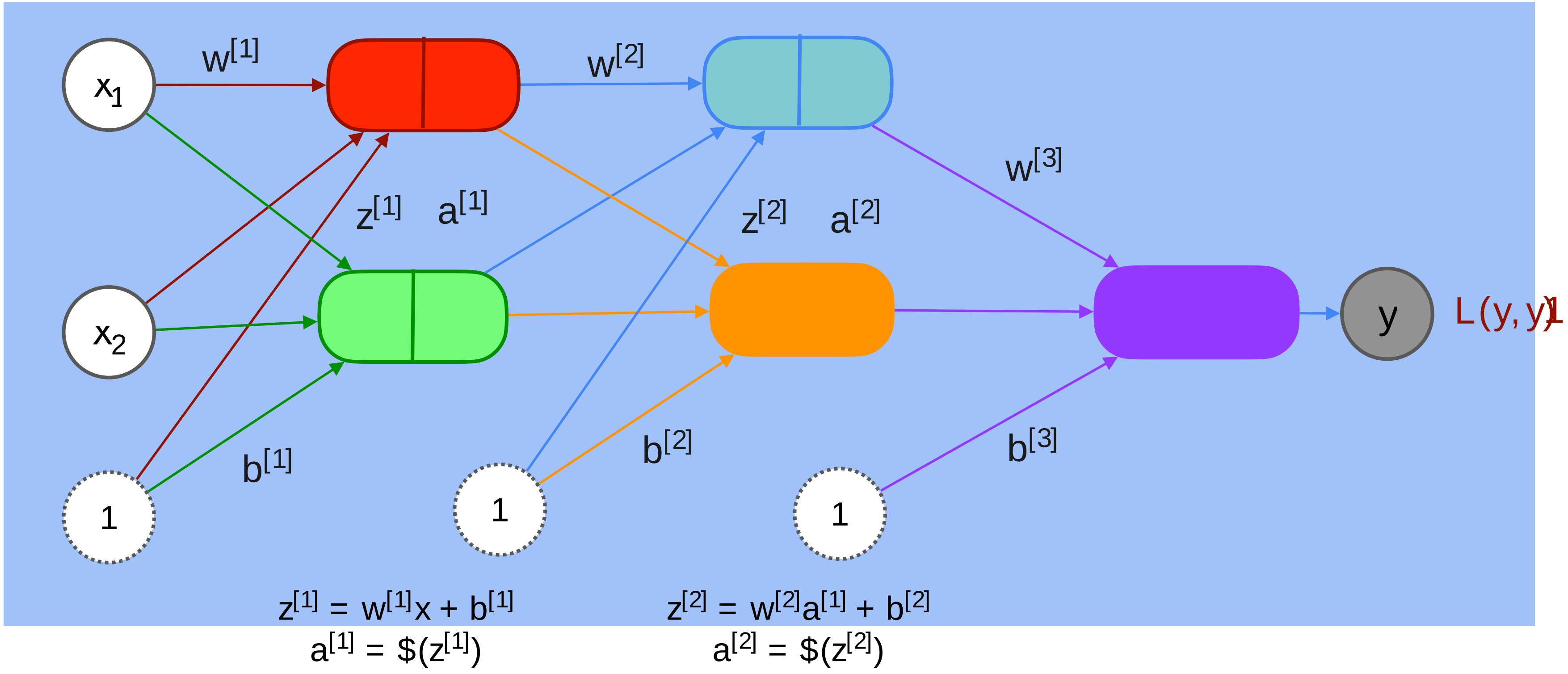
Back Propagation Introduction



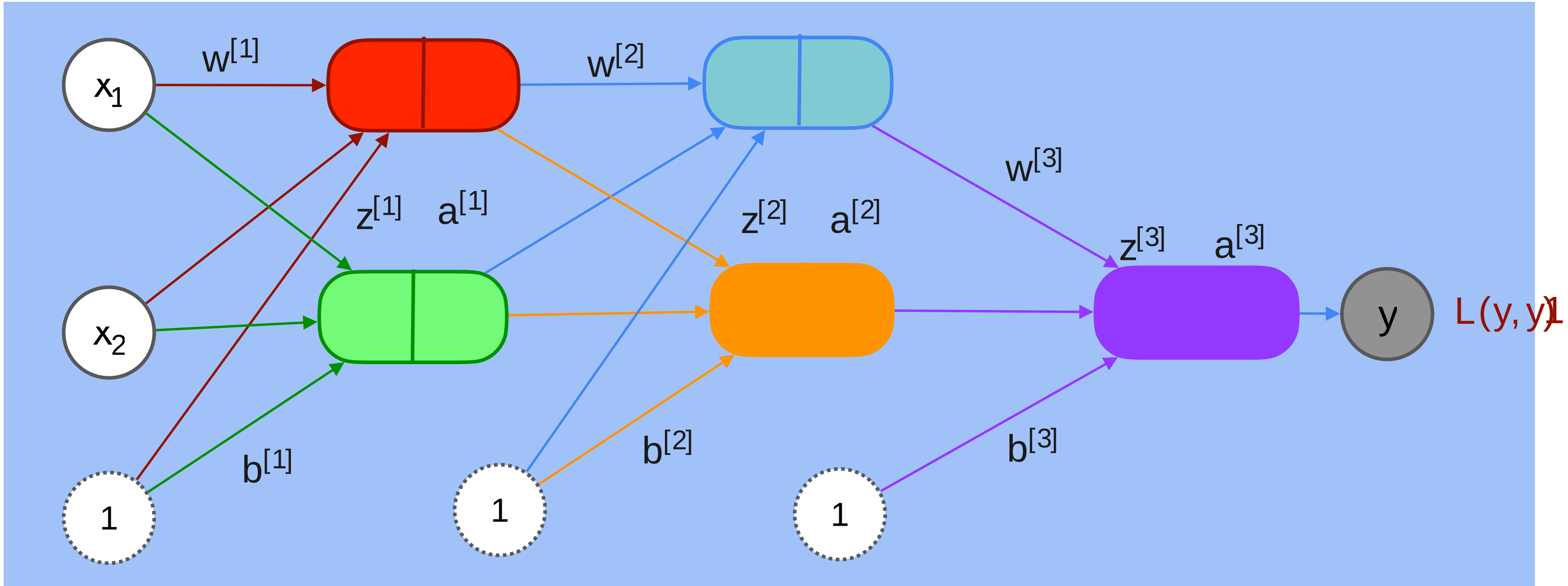
$z^{[1]} = w^{[1]}x + b^{[1]}$
$a^{[1]} = \$(z^{[1]})$

$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$
$a^{[2]} = \$(z^{[2]})$

Back Propagation Introduction



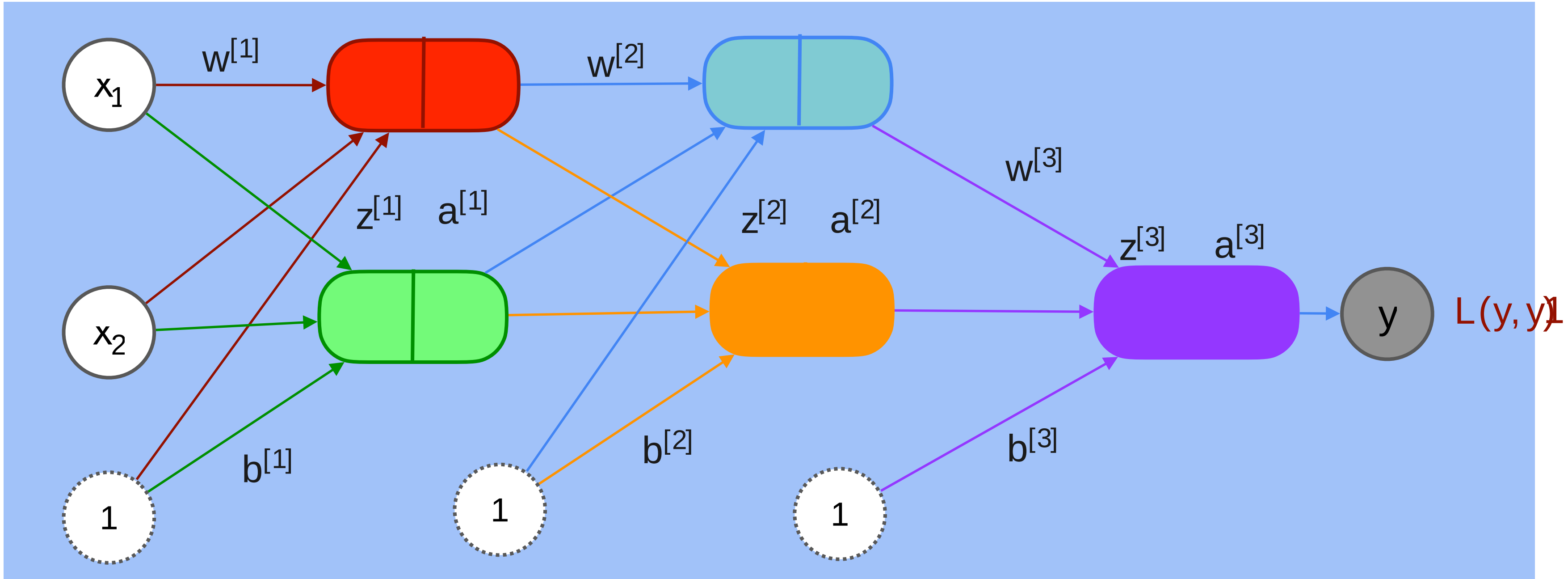
Back Propagation Introduction



$$z^{[1]} = w^{[1]}x + b^{[1]}$$
$$a^{[1]} = \$(z^{[1]})$$

$$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$$
$$a^{[2]} = \$(z^{[2]})$$

Back Propagation Introduction

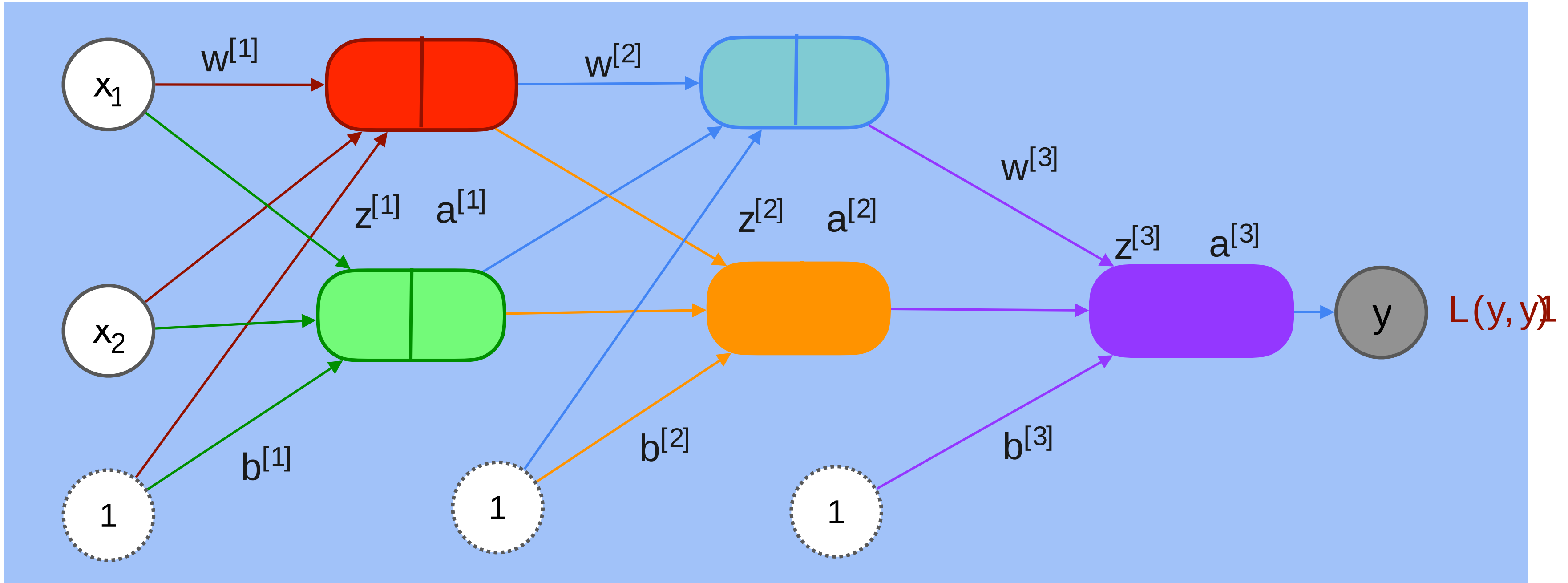


$z^{[1]} = w^{[1]}x + b^{[1]}$
$a^{[1]} = \$(z^{[1]})$

$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$
$a^{[2]} = \$(z^{[2]})$

$z^{[3]} = w^{[3]}a^{[2]} + b^{[3]}$

Back Propagation Introduction

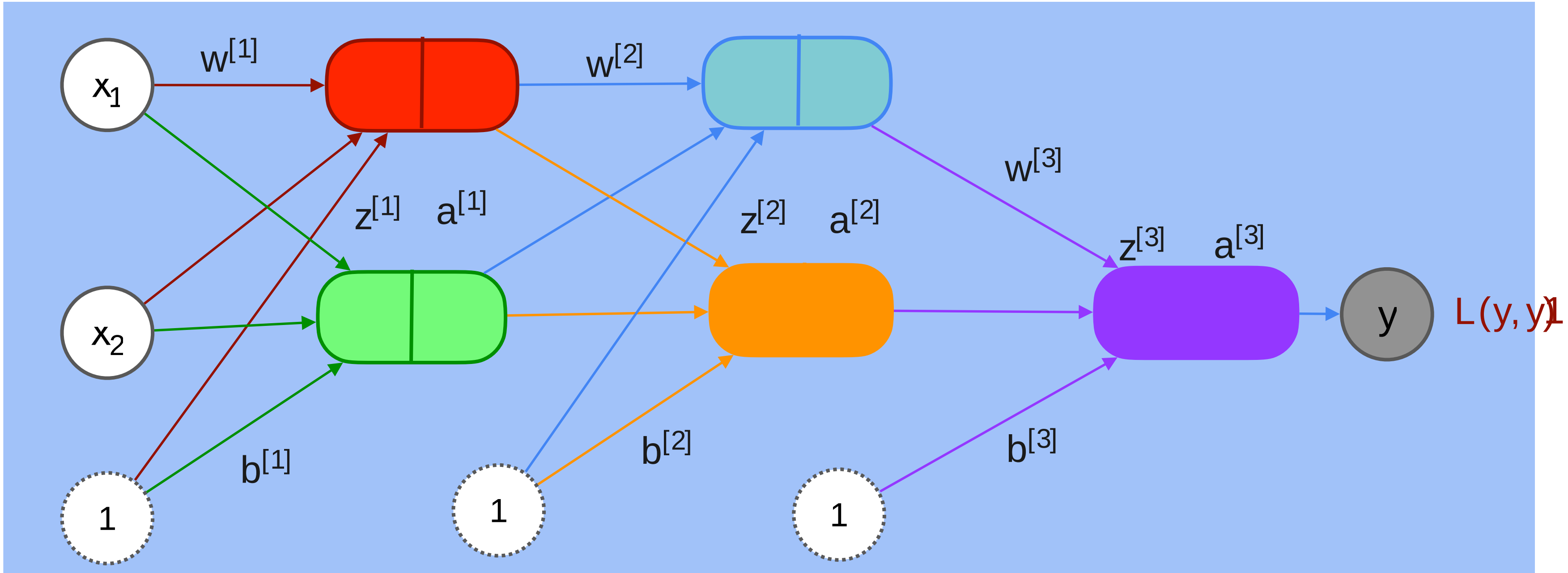


$z^{[1]} = w^{[1]}x + b^{[1]}$
$a^{[1]} = \$(z^{[1]})$

$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$
$a^{[2]} = \$(z^{[2]})$

$z^{[3]} = w^{[3]}a^{[2]} + b^{[3]}$
$a^{[3]} = \$(z^{[3]})$

Back Propagation Introduction

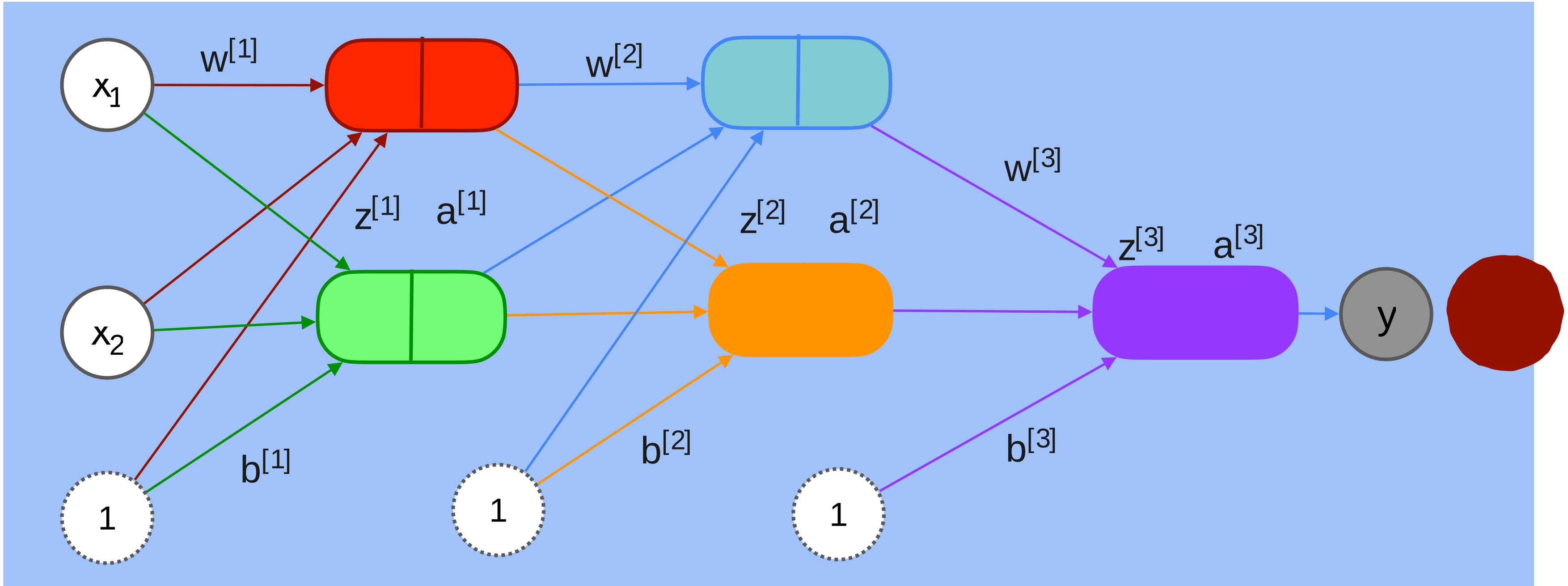


$z^{[1]} = w^{[1]}x + b^{[1]}$
$a^{[1]} = \$(z^{[1]})$

$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$
$a^{[2]} = \$(z^{[2]})$

$z^{[3]} = w^{[3]}a^{[2]} + b^{[3]}$
$a^{[3]} = \$(z^{[3]})$

Back Propagation Introduction

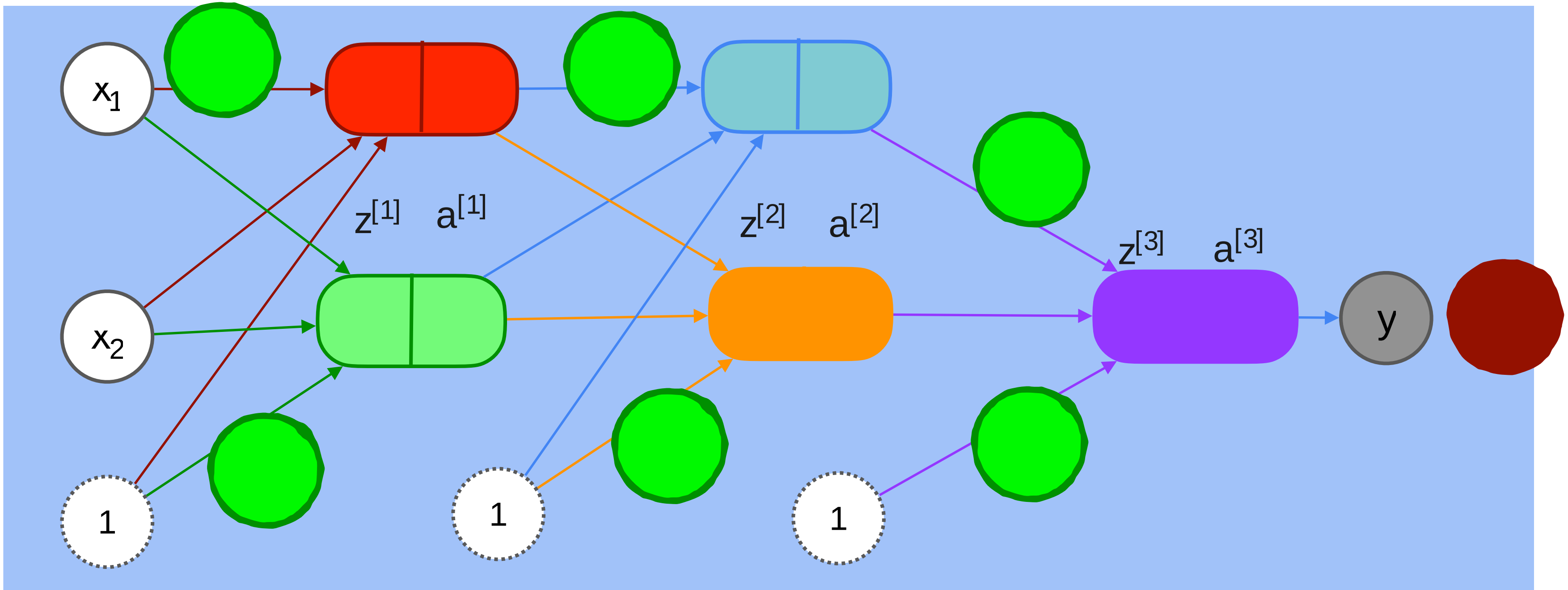


$z^{[1]} = w^{[1]}x + b^{[1]}$
$a^{[1]} = \$(z^{[1]})$

$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$
$a^{[2]} = \$(z^{[2]})$

$z^{[3]} = w^{[3]}a^{[2]} + b^{[3]}$
$a^{[3]} = \$(z^{[3]})$

Back Propagation Introduction

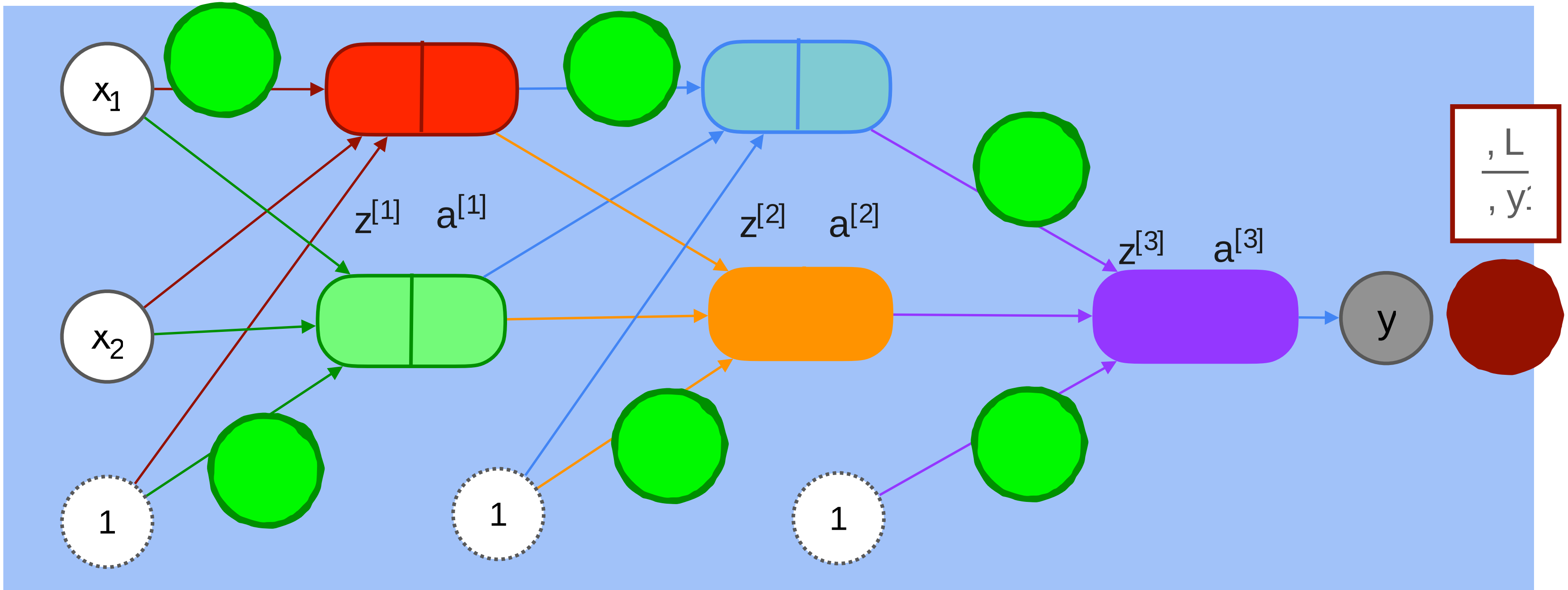


$z^{[1]} = w^{[1]}x + b^{[1]}$
$a^{[1]} = \$(z^{[1]})$

$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$
$a^{[2]} = \$(z^{[2]})$

$z^{[3]} = w^{[3]}a^{[2]} + b^{[3]}$
$a^{[3]} = \$(z^{[3]})$

Back Propagation Introduction

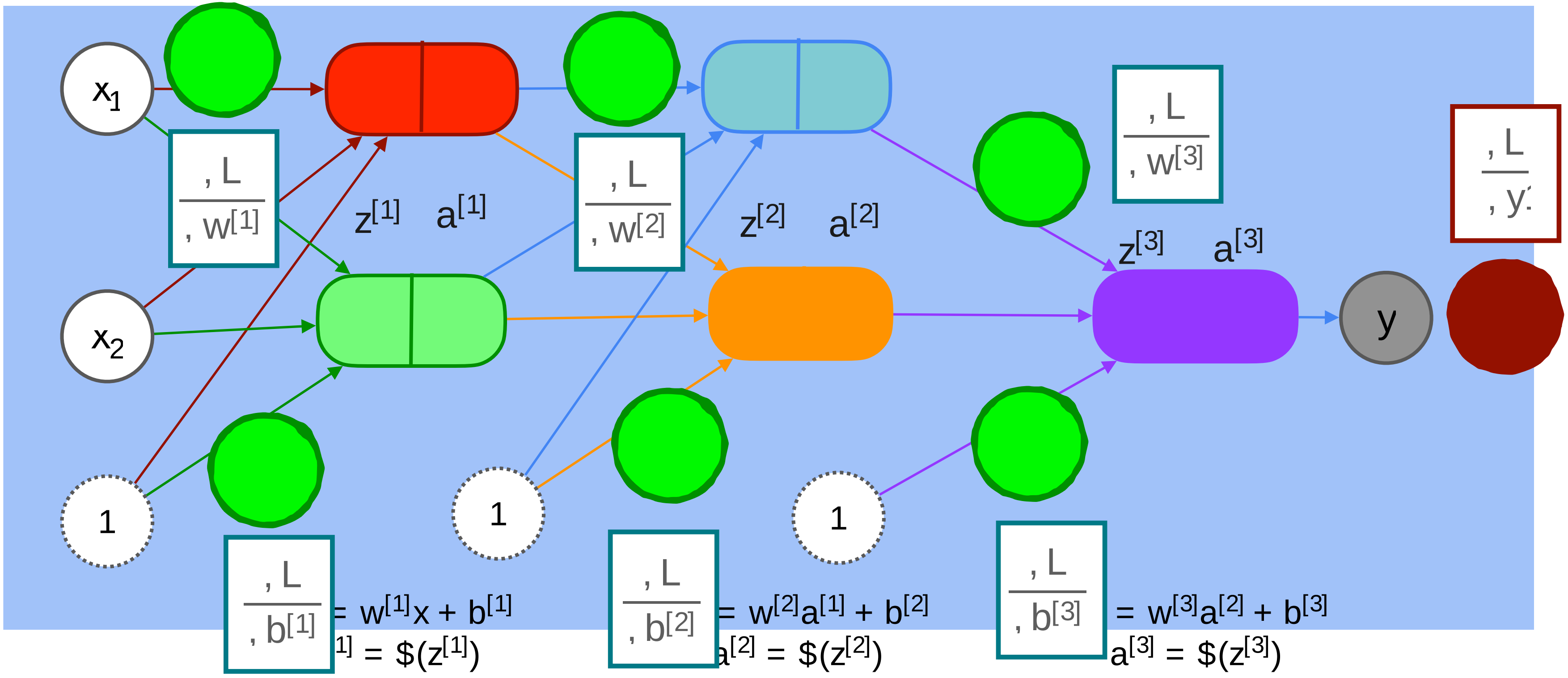


$z^{[1]} = w^{[1]}x + b^{[1]}$
$a^{[1]} = \$(z^{[1]})$

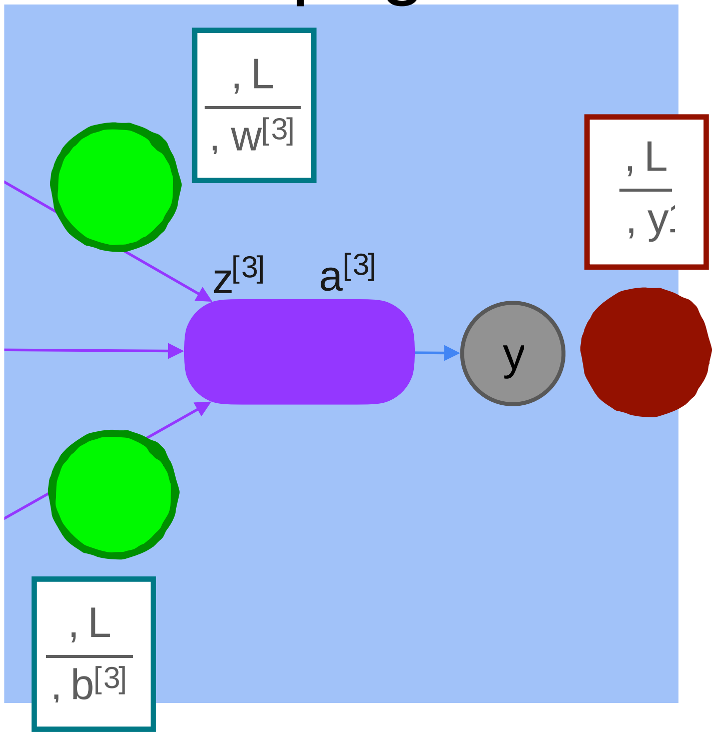
$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$
$a^{[2]} = \$(z^{[2]})$

$z^{[3]} = w^{[3]}a^{[2]} + b^{[3]}$
$a^{[3]} = \$(z^{[3]})$

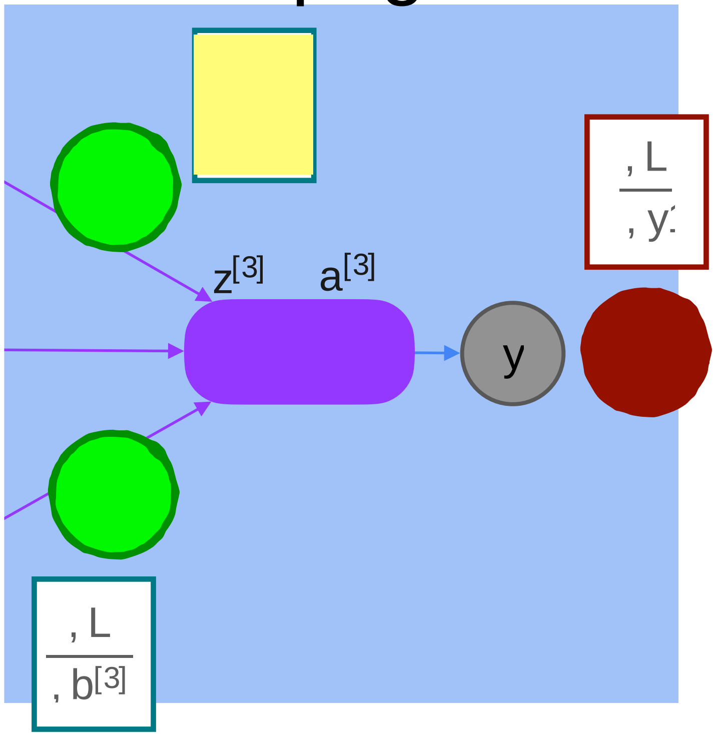
Back Propagation Introduction



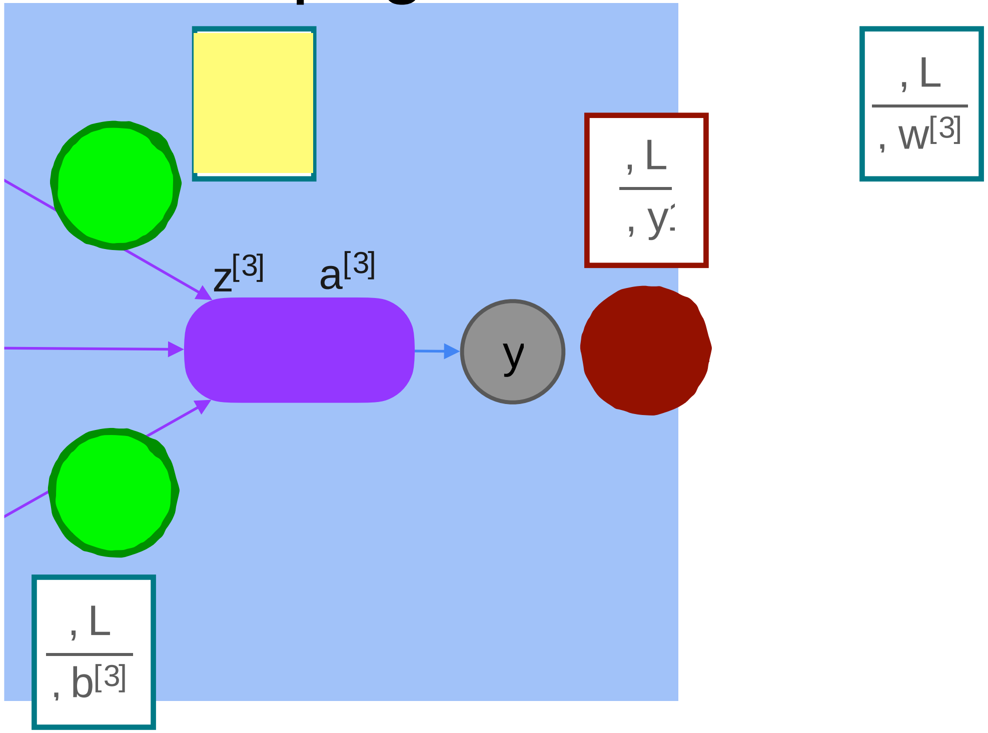
Back Propagation Introduction



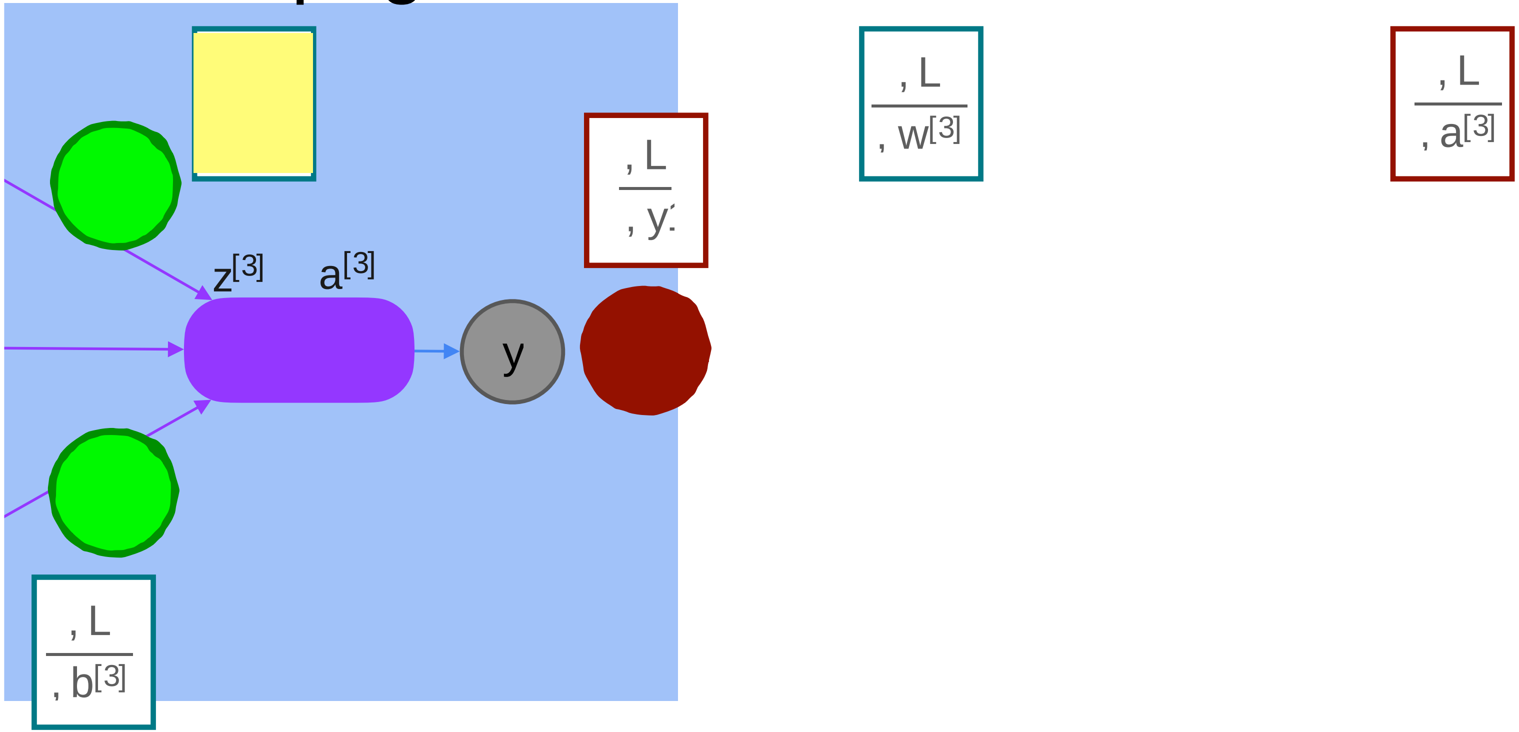
Back Propagation Introduction



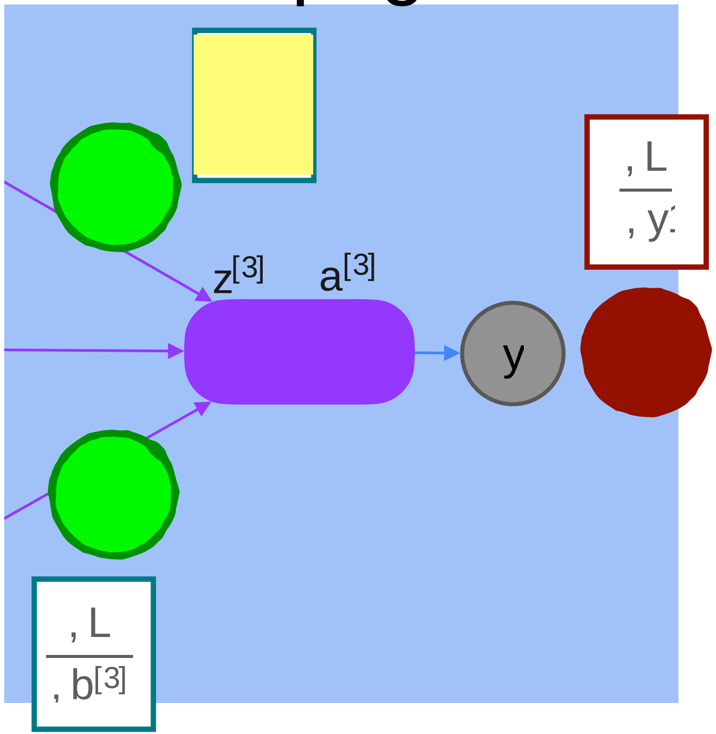
Back Propagation Introduction



Back Propagation Introduction



Back Propagation Introduction

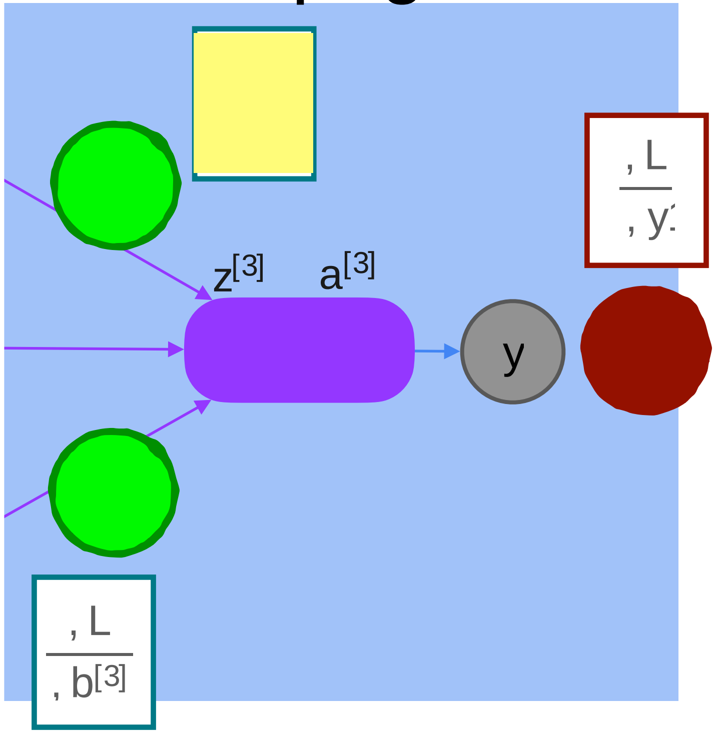


$\frac{, L}{, w^{[3]}}$

$\frac{, a^{[3]}}{, z^{[3]}}$

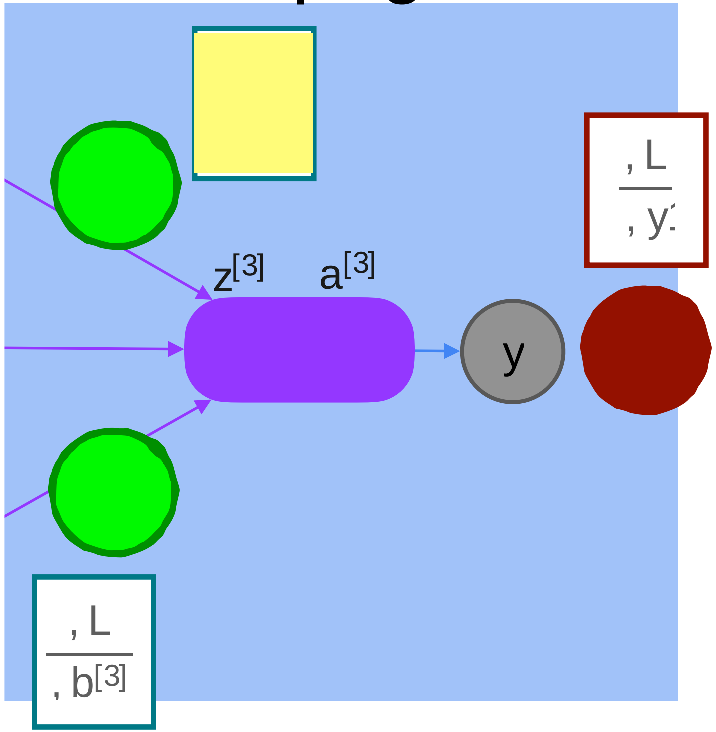
$\frac{, L}{, a^{[3]}}$

Back Propagation Introduction



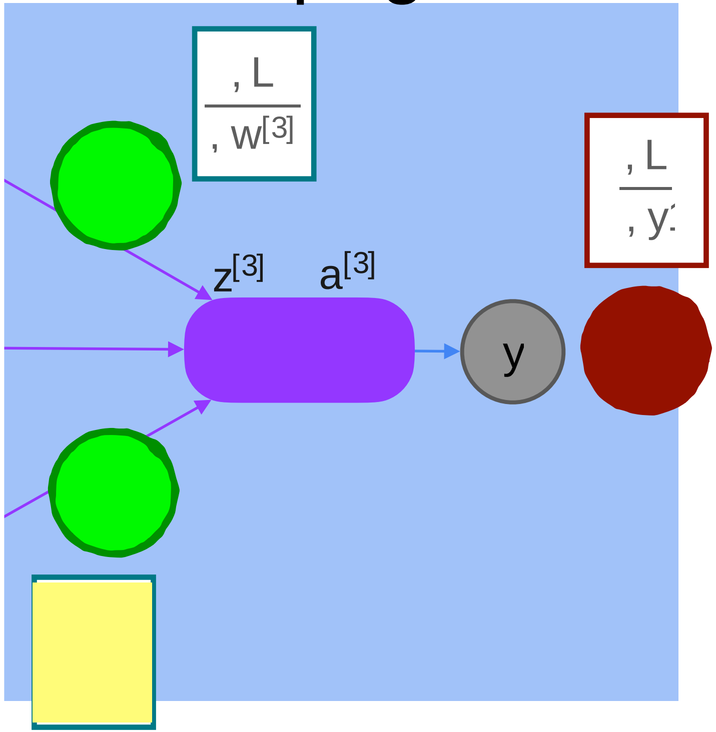
$\frac{, L}{, w^{[3]}}$ $\frac{, z^{[3]}}{, w^{[3]}}$ $\frac{, a^{[3]}}{, z^{[3]}}$ $\frac{, L}{, a^{[3]}}$

Back Propagation Introduction



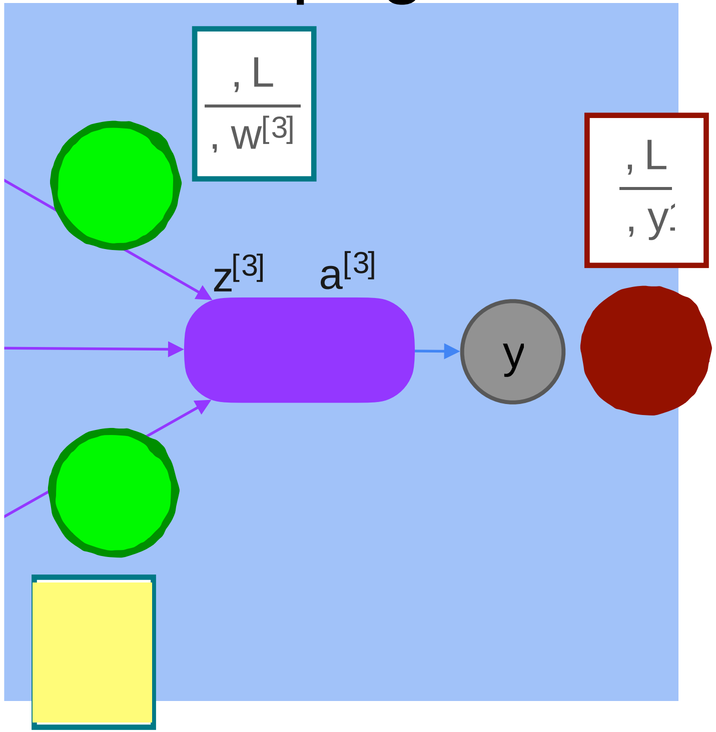
$$\boxed{\frac{,\ L}{,\ w^{[3]}}} = \boxed{\frac{,\ z^{[3]}}{,\ w^{[3]}}} \cdot \boxed{\frac{,\ a^{[3]}}{,\ z^{[3]}}} \cdot \boxed{\frac{,\ L}{,\ a^{[3]}}}$$

Back Propagation Introduction



$$\boxed{\frac{,\ L}{,\ w^{[3]}}} = \boxed{\frac{,\ z^{[3]}}{,\ w^{[3]}}} \cdot \boxed{\frac{,\ a^{[3]}}{,\ z^{[3]}}} \cdot \boxed{\frac{,\ L}{,\ a^{[3]}}}$$

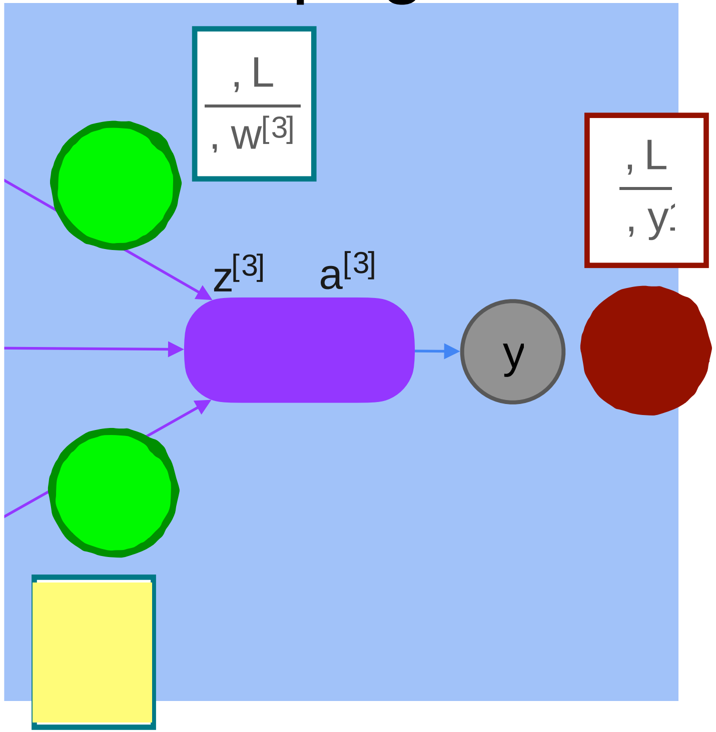
Back Propagation Introduction



$$\frac{, L}{, w^{[3]}} = \frac{, z^{[3]}}{, w^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, L}{, a^{[3]}}$$

$$\frac{, L}{, b^{[3]}}$$

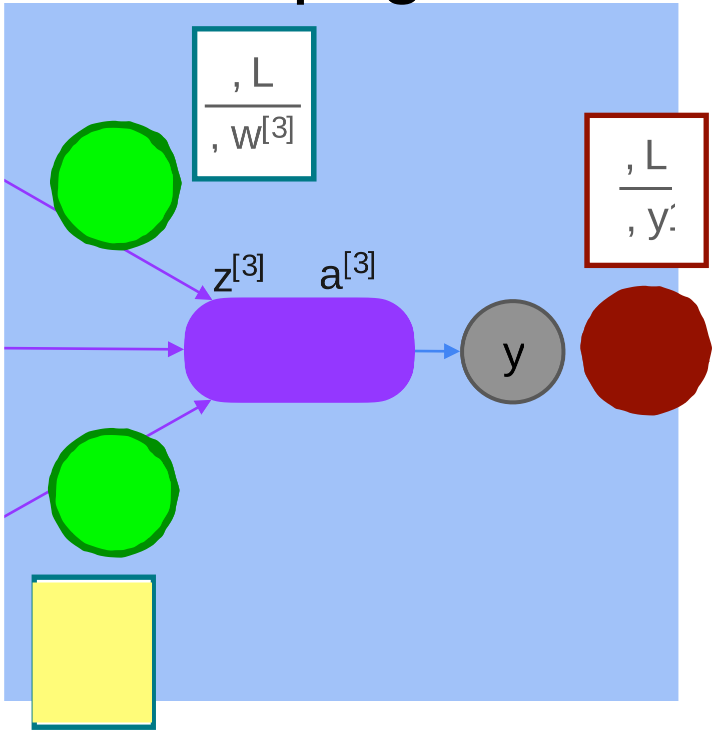
Back Propagation Introduction



$$\boxed{\frac{, L}{, w^{[3]}}} = \boxed{\frac{, z^{[3]}}{, w^{[3]}}} \cdot \boxed{\frac{, a^{[3]}}{, z^{[3]}}} \cdot \boxed{\frac{, L}{, a^{[3]}}}$$

$$\boxed{\frac{, L}{, b^{[3]}}} \quad \boxed{\frac{, z^{[3]}}{, b^{[3]}}} \quad \boxed{\frac{, a^{[3]}}{, z^{[3]}}} \quad \boxed{\frac{, L}{, a^{[3]}}}$$

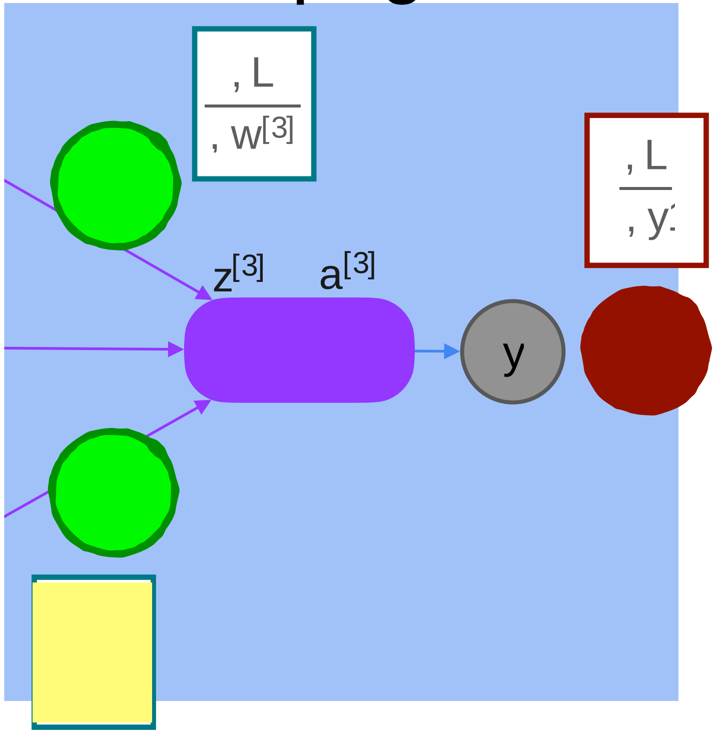
Back Propagation Introduction



$$\frac{, L}{, w^{[3]}} = \frac{, z^{[3]}}{, w^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, L}{, a^{[3]}}$$

$$\frac{, L}{, b^{[3]}} = \frac{, z^{[3]}}{, b^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, L}{, a^{[3]}}$$

Back Propagation Introduction

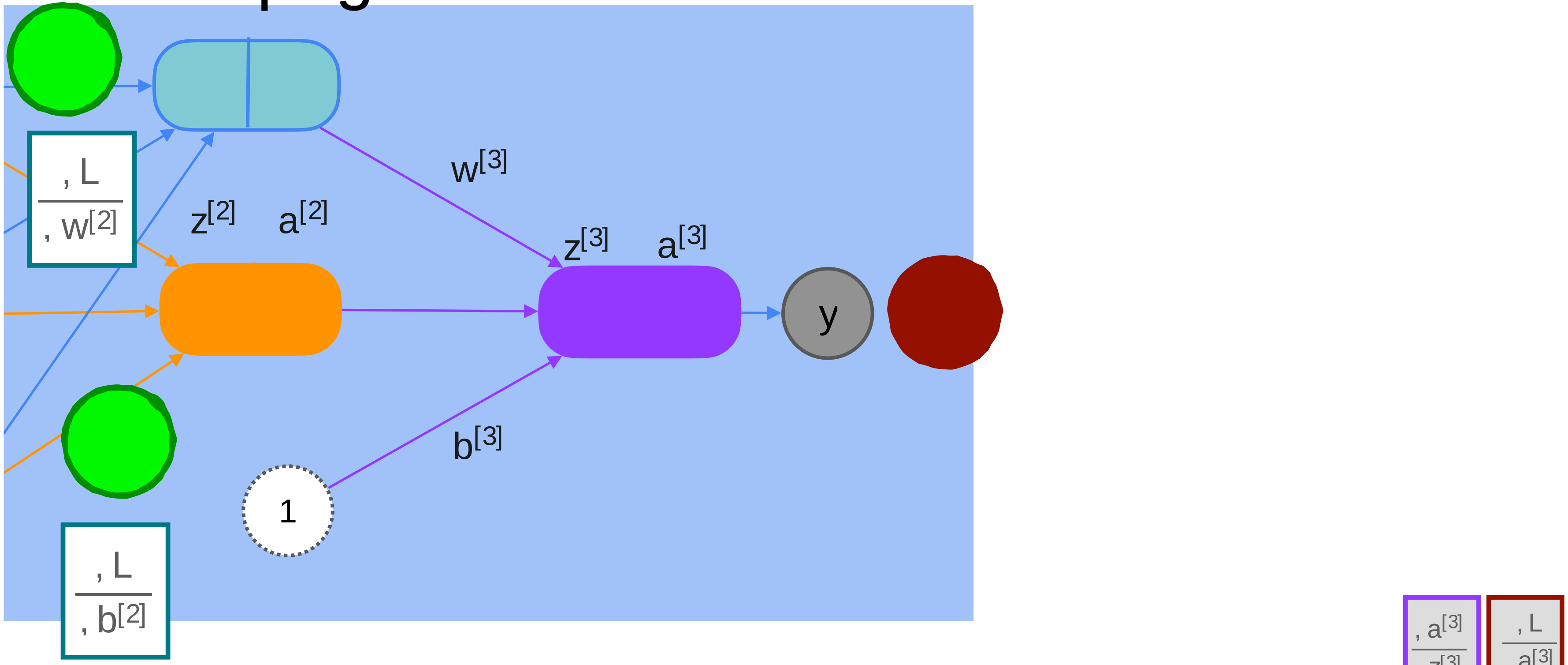


$$\frac{,\ L}{,\ w^{[3]}} = \frac{,\ z^{[3]}}{,\ w^{[3]}} \cdot \frac{,\ a^{[3]}}{,\ z^{[3]}} \cdot \frac{,\ L}{,\ a^{[3]}}$$

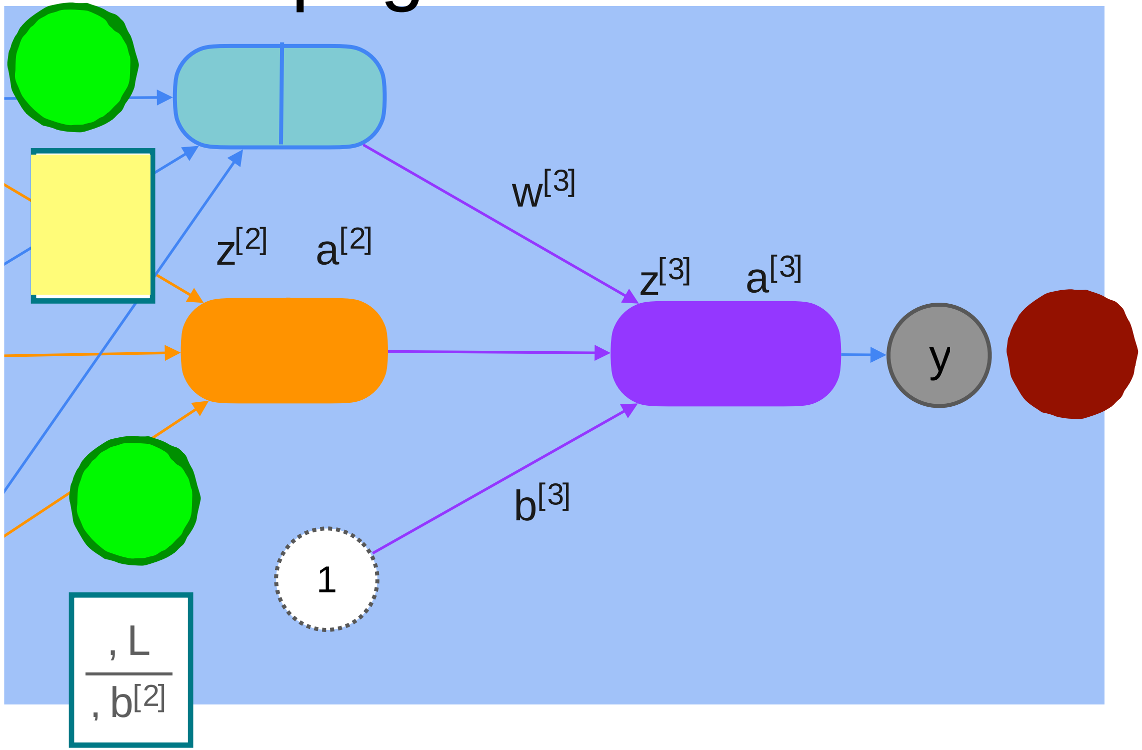
$$\frac{,\ L}{,\ b^{[3]}} = \frac{,\ z^{[3]}}{,\ b^{[3]}} \cdot \frac{,\ a^{[3]}}{,\ z^{[3]}} \cdot \frac{,\ L}{,\ a^{[3]}}$$

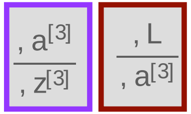
$\frac{,\ a^{[3]}}{,\ z^{[3]}}$ $\frac{,\ L}{,\ a^{[3]}}$

Back Propagation Introduction

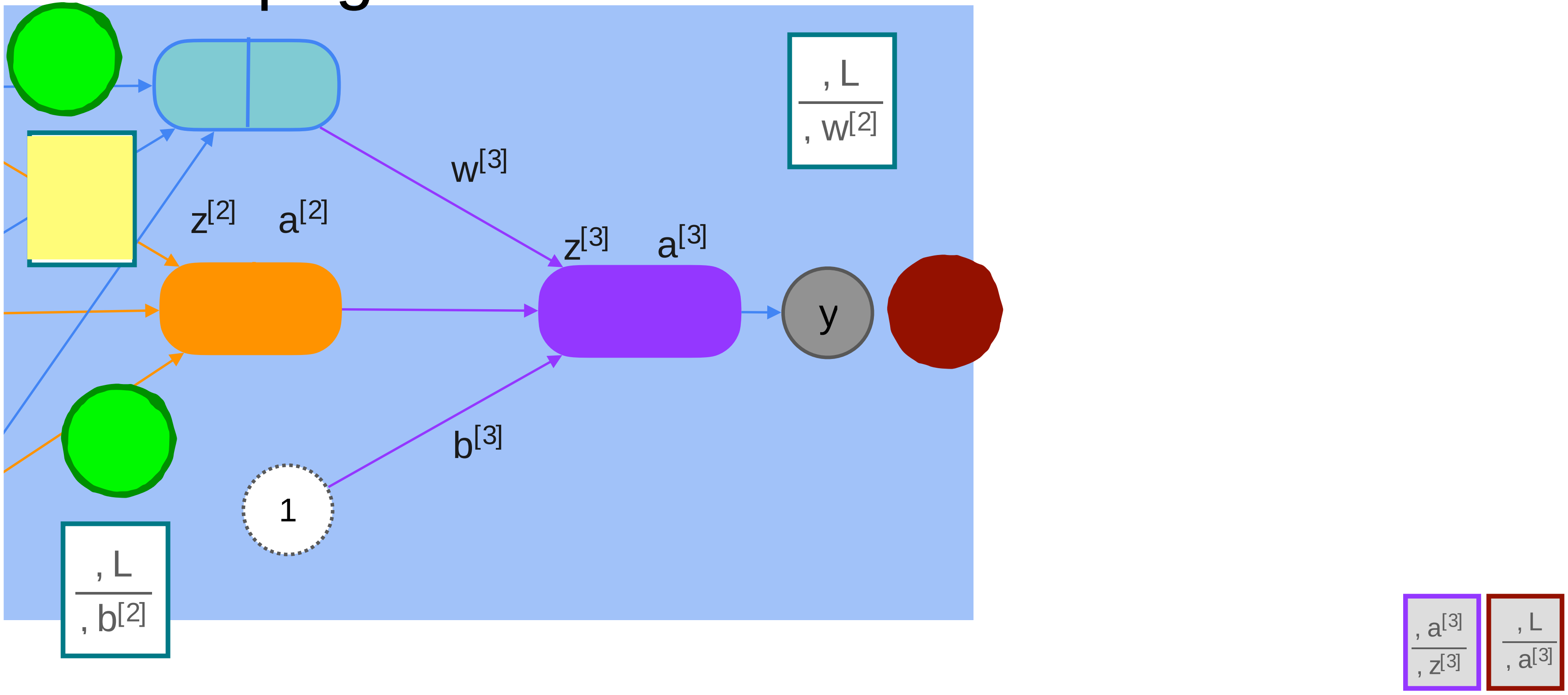


Back Propagation Introduction

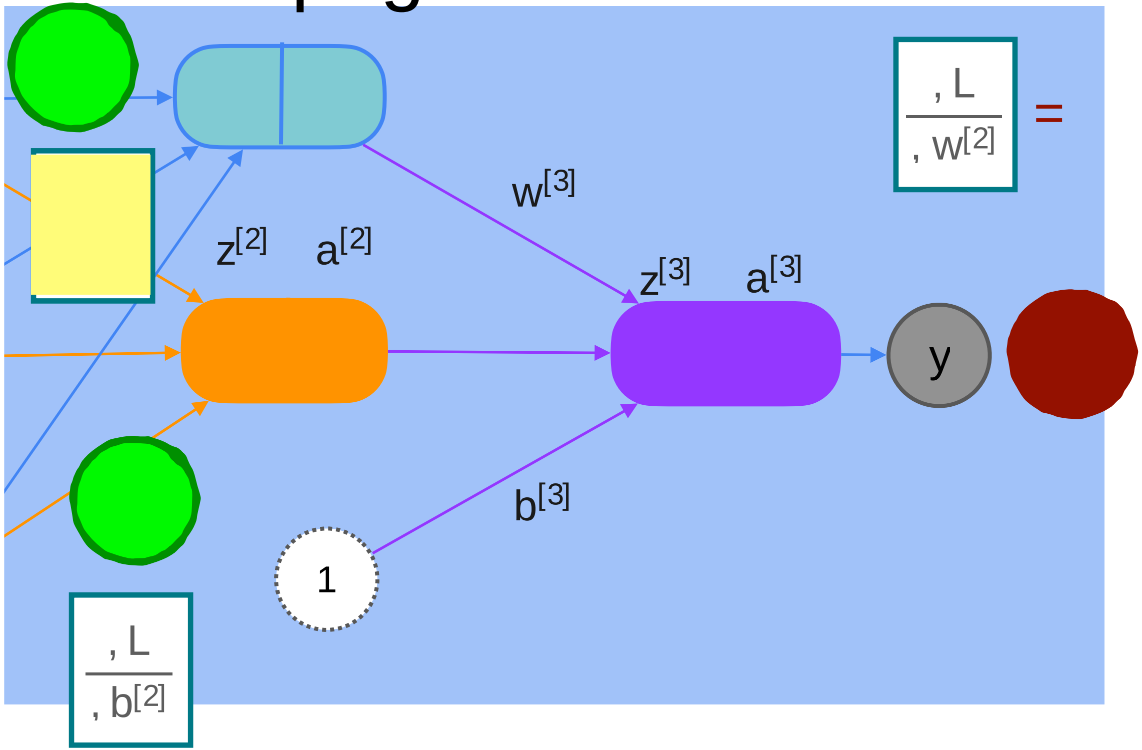


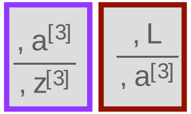


Back Propagation Introduction

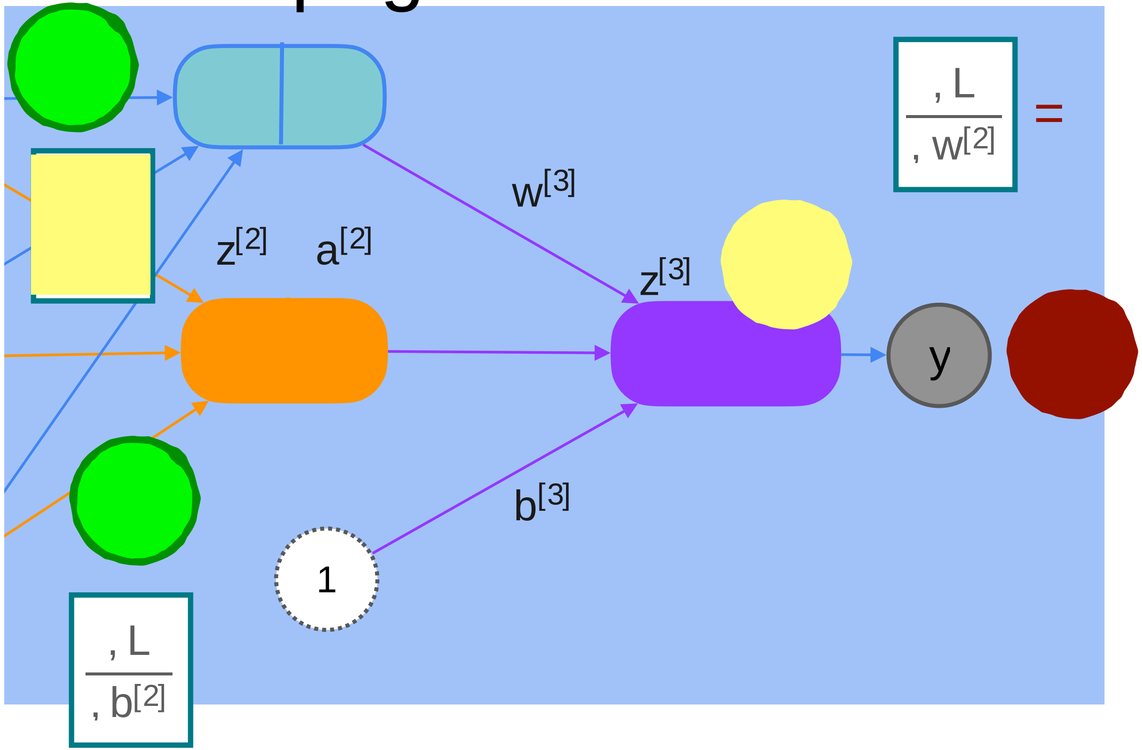


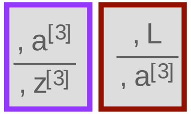
Back Propagation Introduction



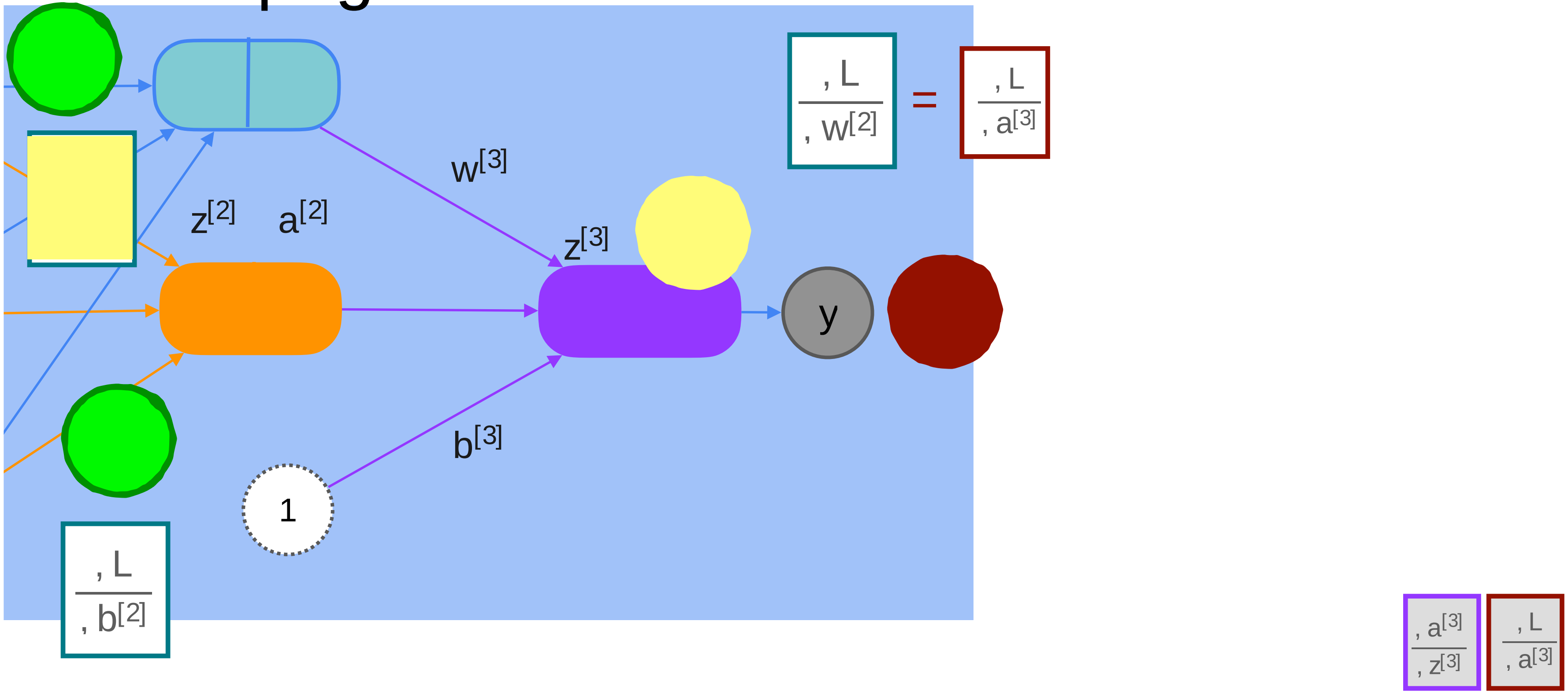


Back Propagation Introduction

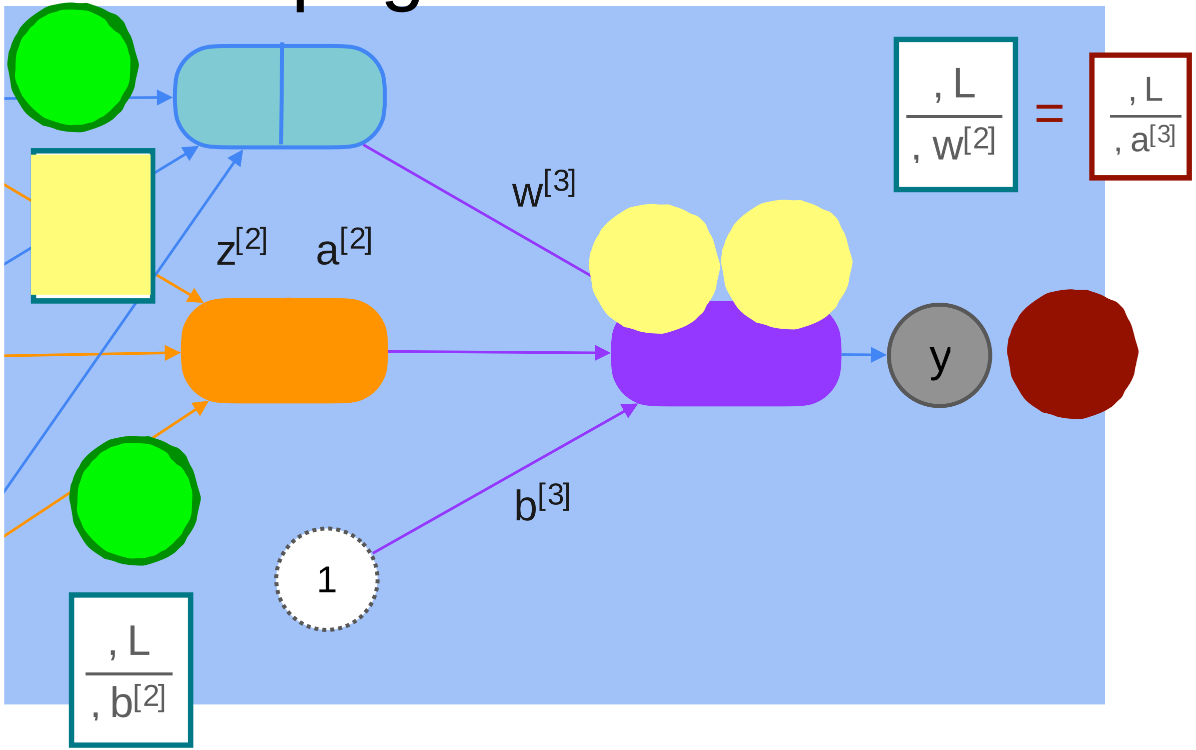


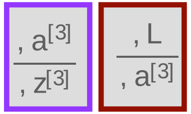


Back Propagation Introduction

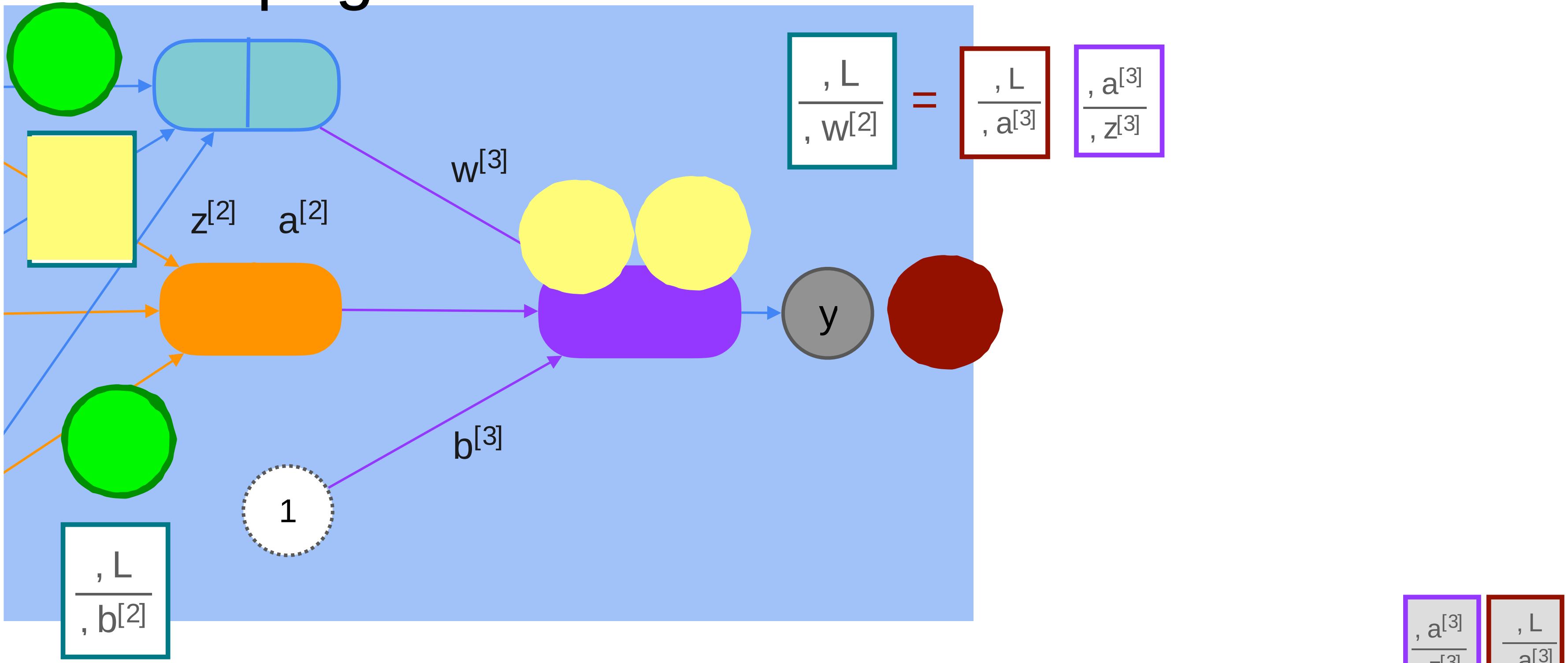


Back Propagation Introduction

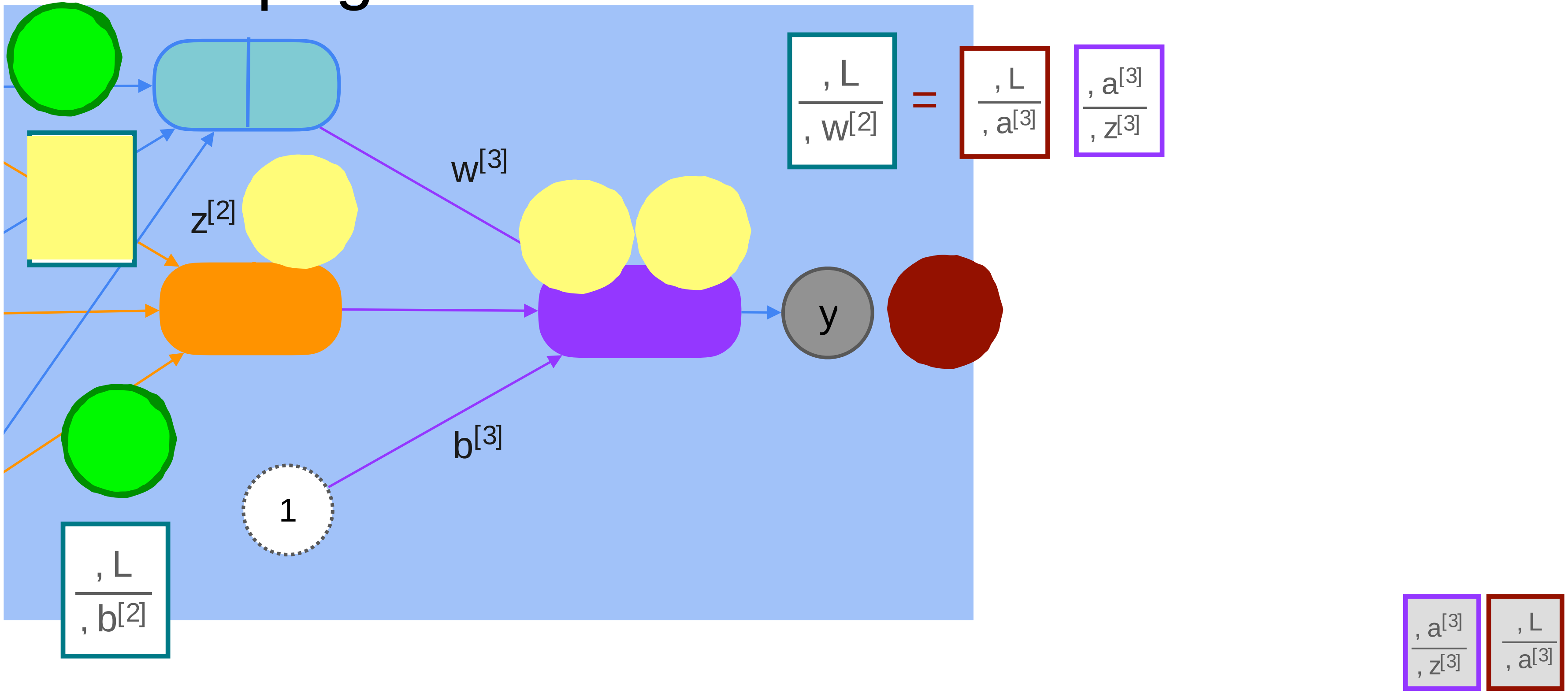




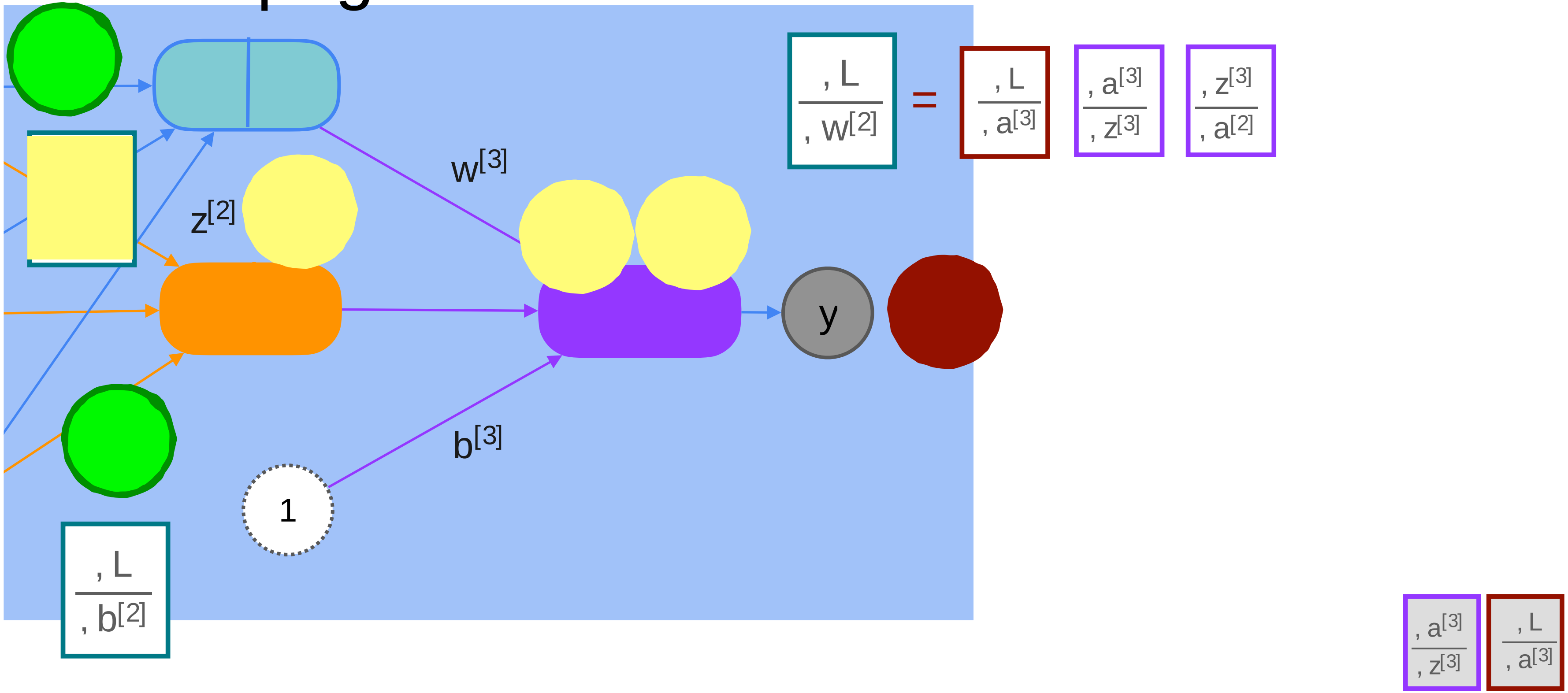
Back Propagation Introduction



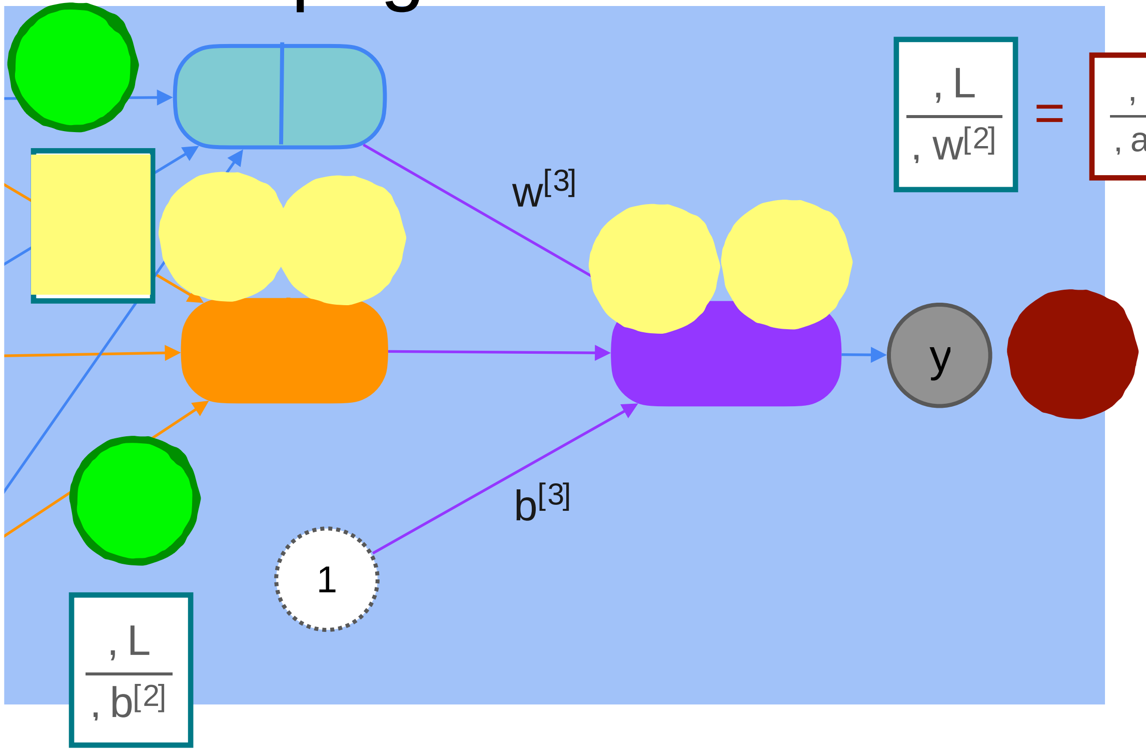
Back Propagation Introduction



Back Propagation Introduction



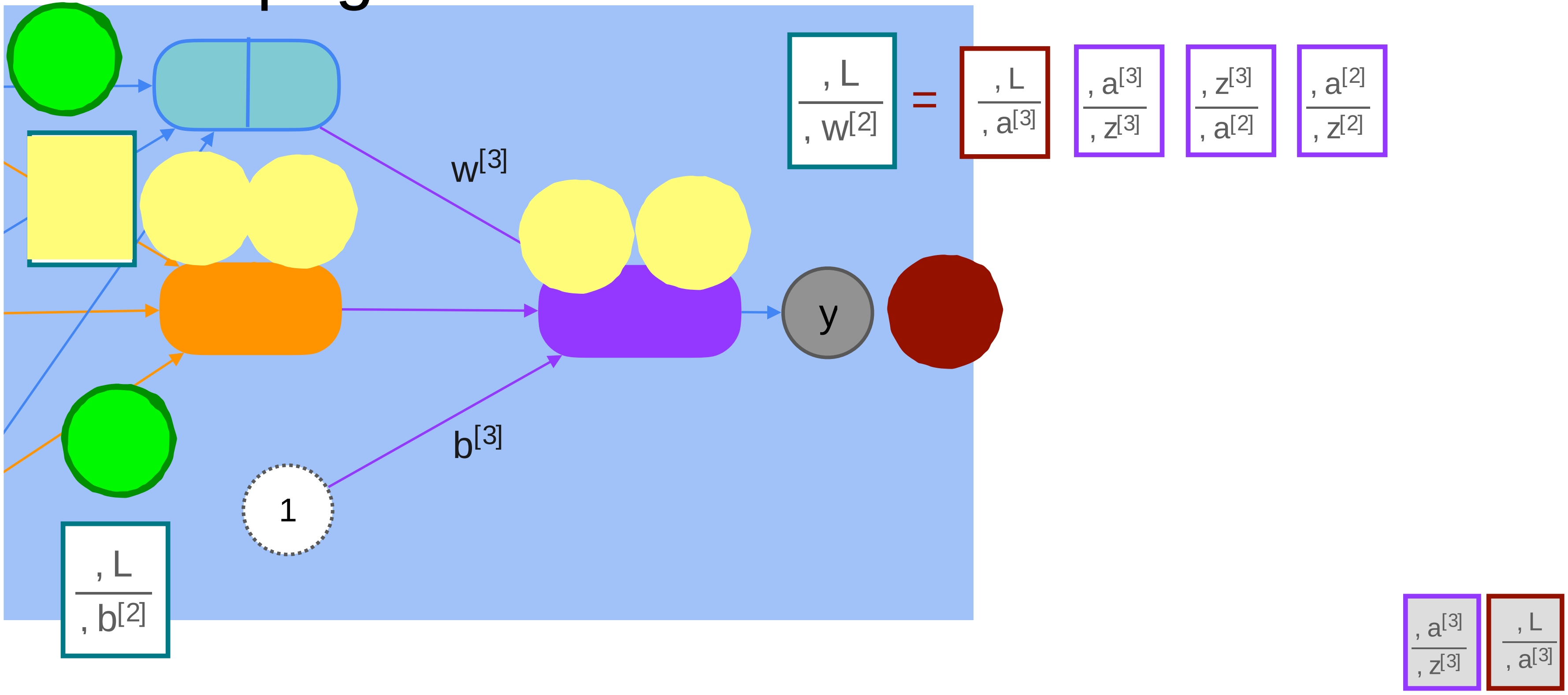
Back Propagation Introduction



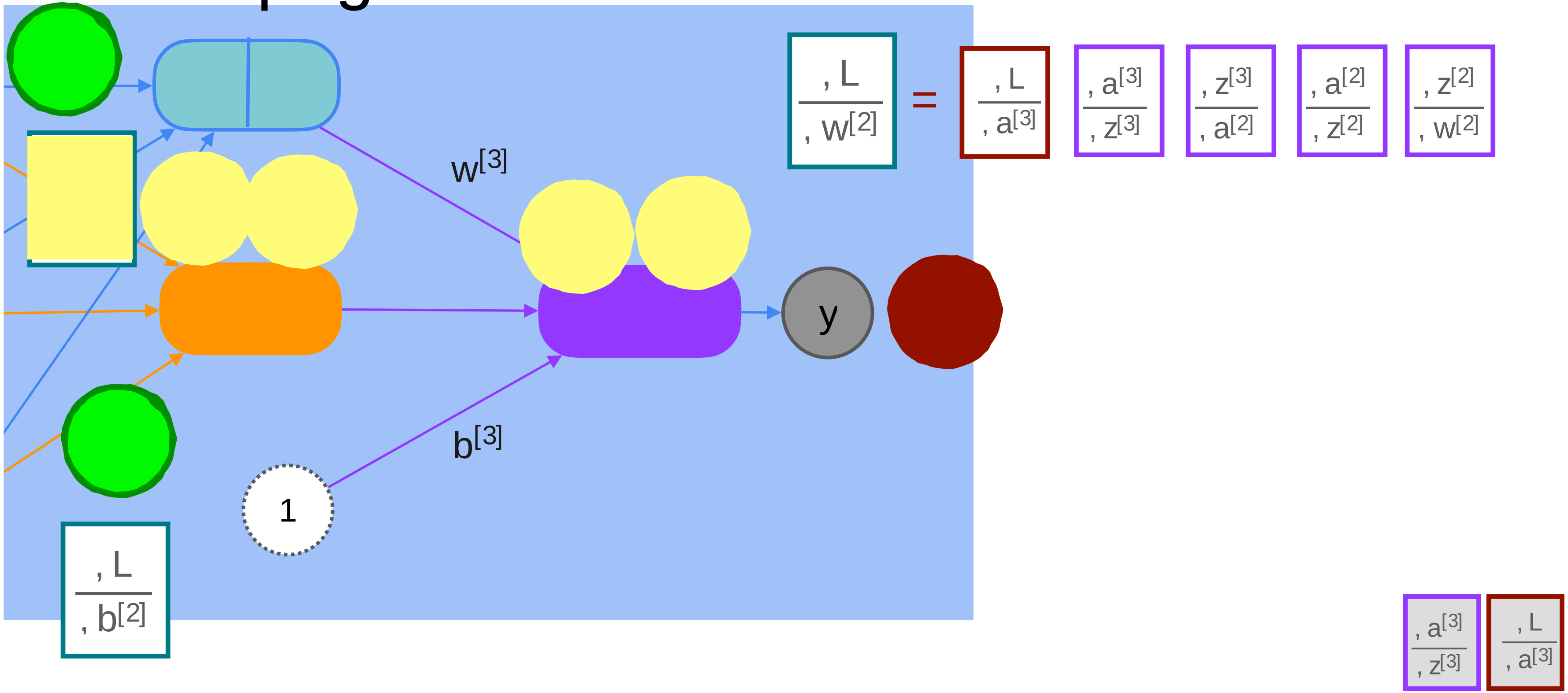
$$\frac{, L}{, w^{[2]}} = \frac{, L}{, a^{[3]}} \quad \frac{, a^{[3]}}{, z^{[3]}} \quad \frac{, z^{[3]}}{, a^{[2]}}$$

$$\frac{, a^{[3]}}{, z^{[3]}} \quad \frac{, L}{, a^{[3]}}$$

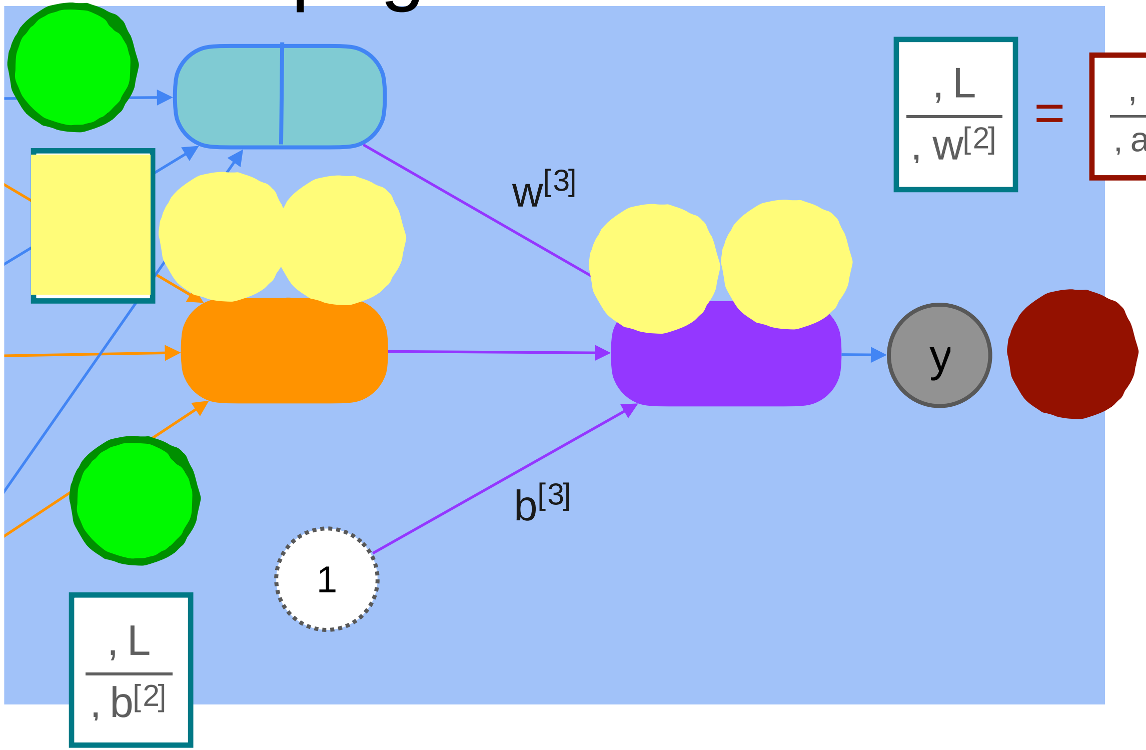
Back Propagation Introduction



Back Propagation Introduction



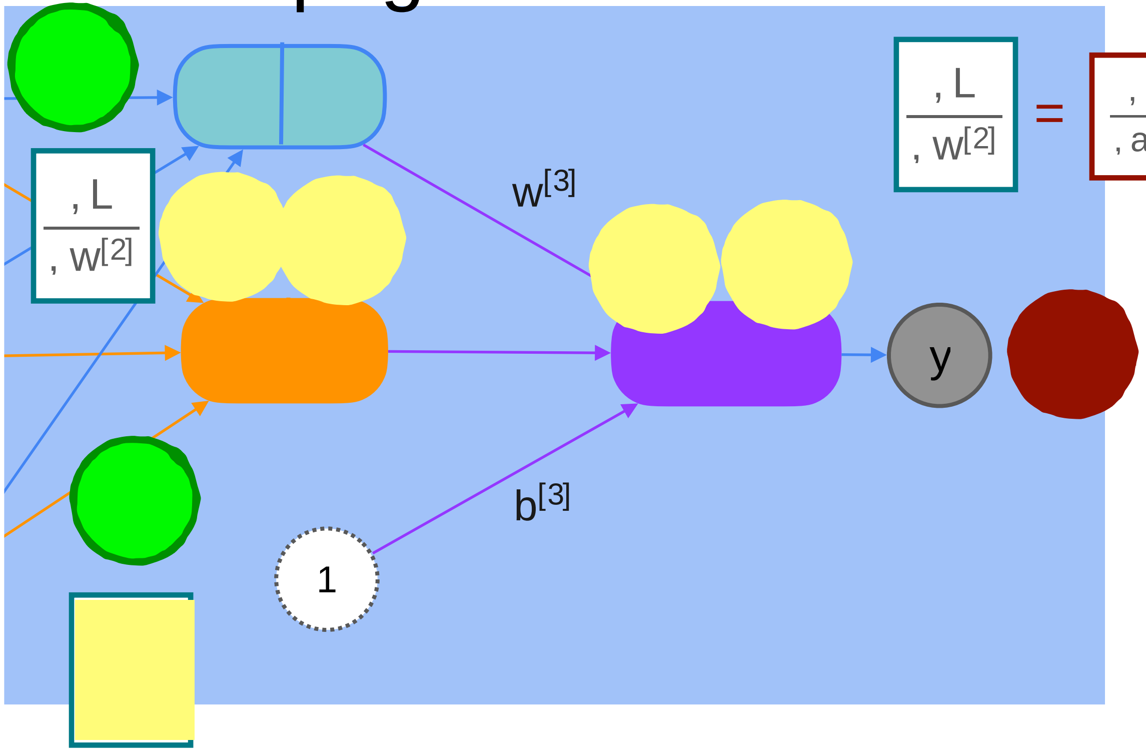
Back Propagation Introduction



$$\frac{, L}{, w^{[2]}} = \frac{, L}{, a^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, z^{[3]}}{, a^{[2]}} \cdot \frac{, a^{[2]}}{, z^{[2]}} \cdot \frac{, z^{[2]}}{, w^{[2]}}$$

$$\frac{, a^{[3]}}{, z^{[3]}} \quad \frac{, L}{, a^{[3]}}$$

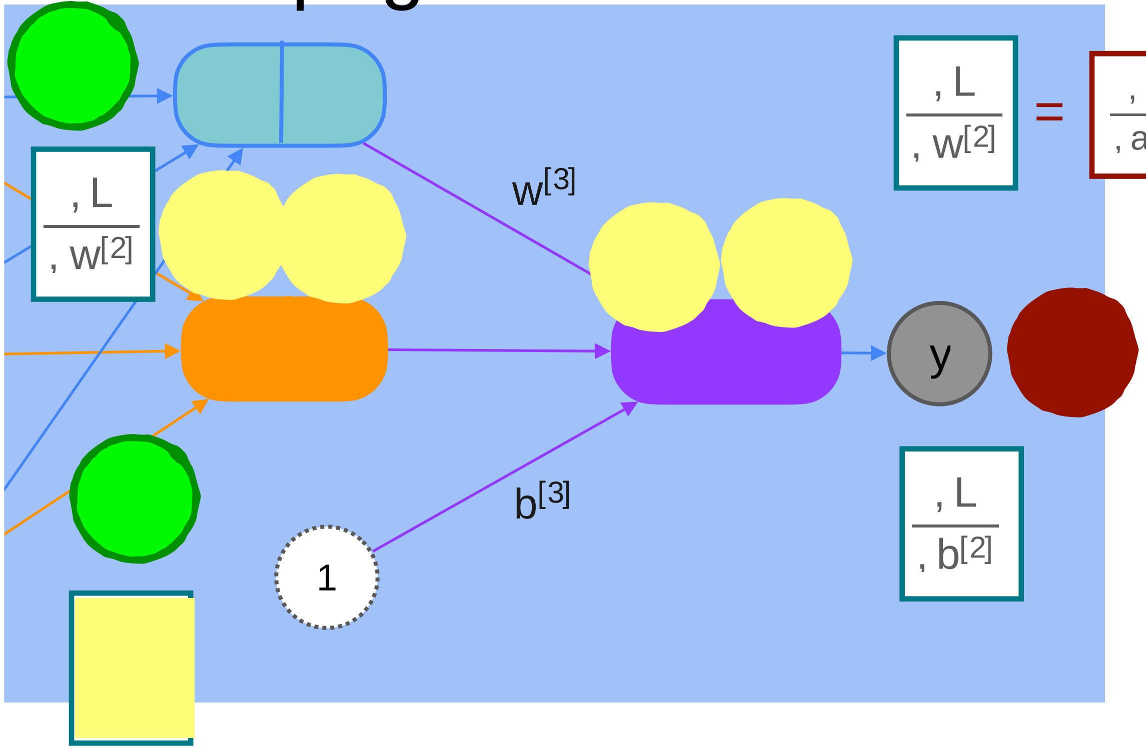
Back Propagation Introduction



$$\frac{, L}{, w^{[2]}} = \frac{, L}{, a^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, z^{[3]}}{, a^{[2]}} \cdot \frac{, a^{[2]}}{, z^{[2]}} \cdot \frac{, z^{[2]}}{, w^{[2]}}$$

$$\frac{, a^{[3]}}{, z^{[3]}} \quad \frac{, L}{, a^{[3]}}$$

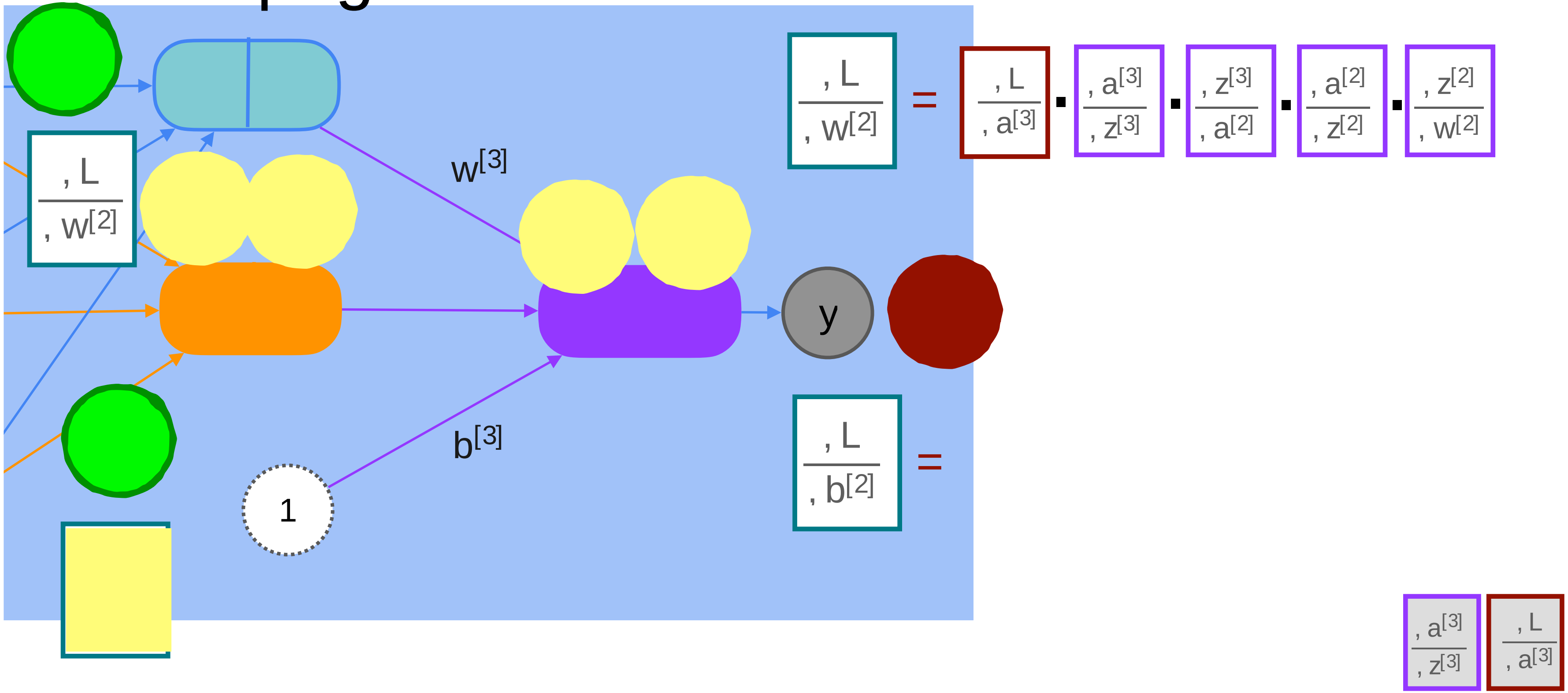
Back Propagation Introduction



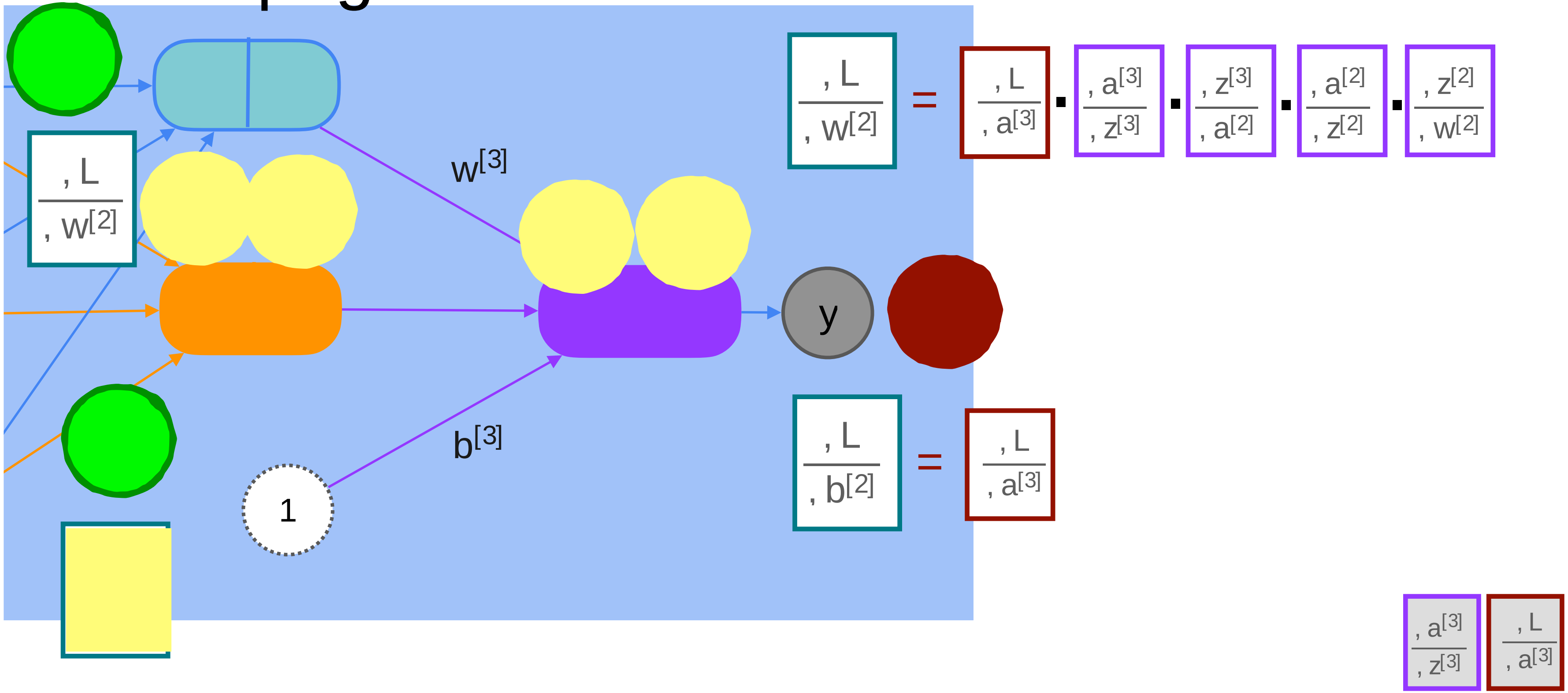
$$\frac{, L}{, w^{[2]}} = \frac{, L}{, a^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, z^{[3]}}{, a^{[2]}} \cdot \frac{, a^{[2]}}{, z^{[2]}} \cdot \frac{, z^{[2]}}{, w^{[2]}}$$

$$\frac{, a^{[3]}}{, z^{[3]}} \quad \frac{, L}{, a^{[3]}}$$

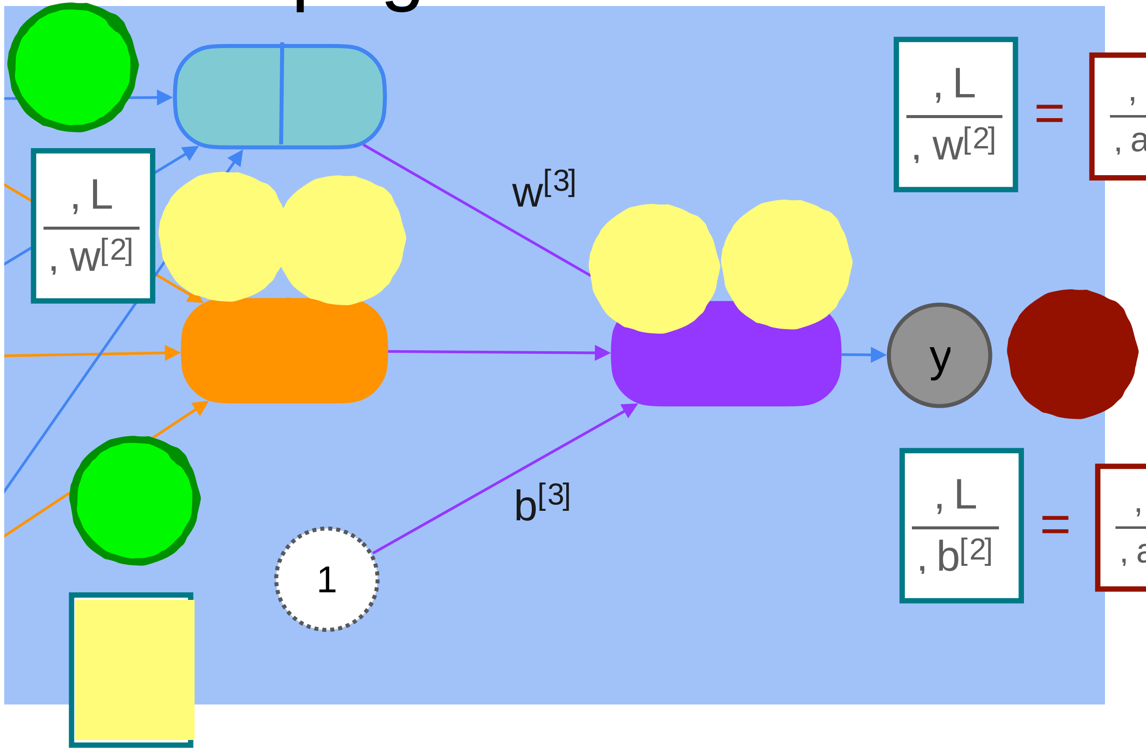
Back Propagation Introduction



Back Propagation Introduction



Back Propagation Introduction

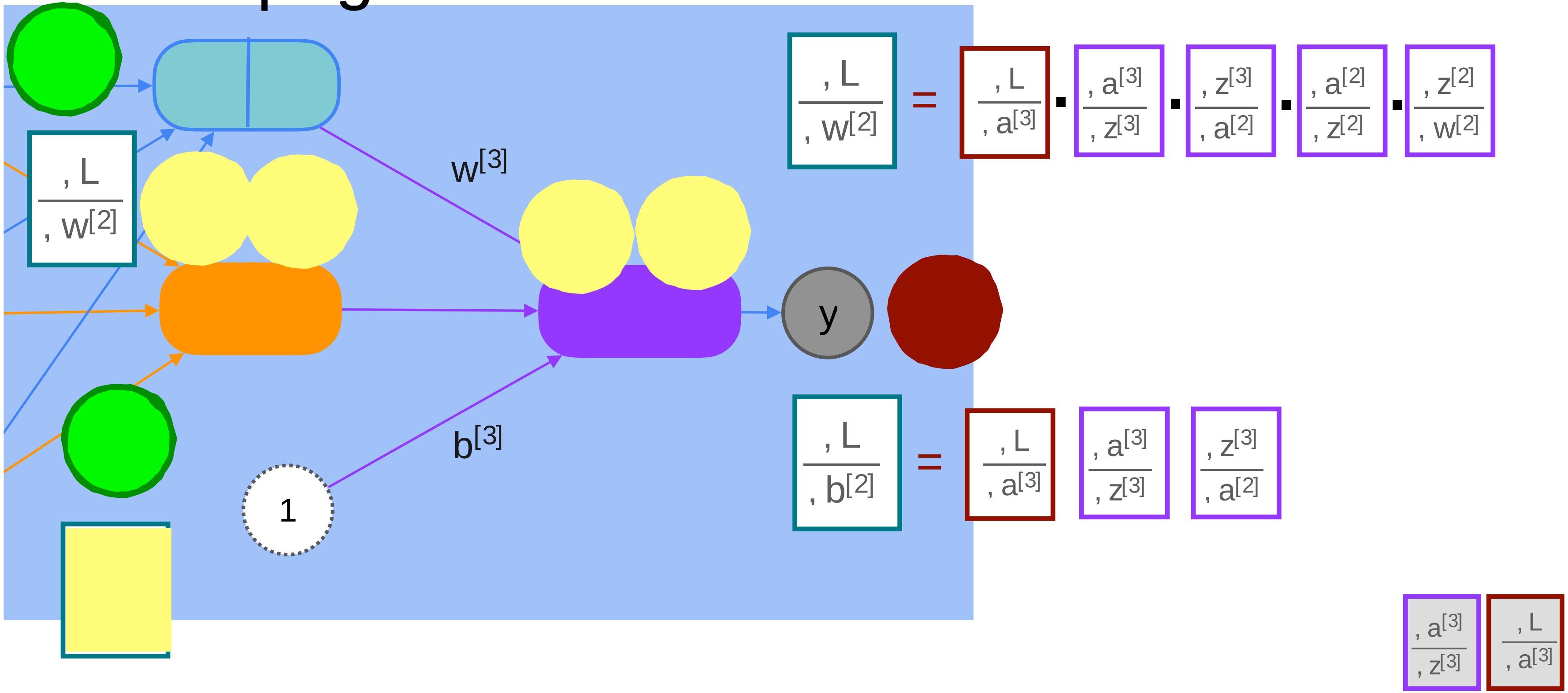


$$\frac{, L}{, w^{[2]}} = \frac{, L}{, a^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, z^{[3]}}{, a^{[2]}} \cdot \frac{, a^{[2]}}{, z^{[2]}} \cdot \frac{, z^{[2]}}{, w^{[2]}}$$

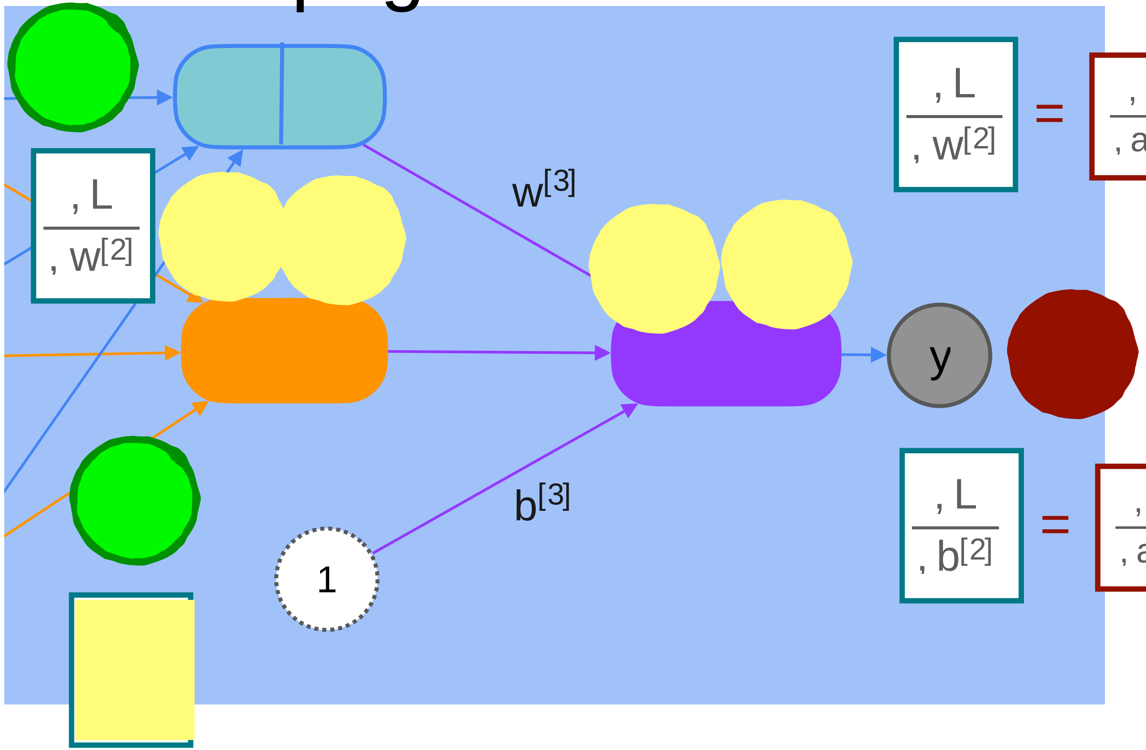
$$\frac{, L}{, b^{[2]}} = \frac{, L}{, a^{[3]}} \quad \frac{, a^{[3]}}{, z^{[3]}}$$

$$\frac{, a^{[3]}}{, z^{[3]}} \quad \frac{, L}{, a^{[3]}}$$

Back Propagation Introduction



Back Propagation Introduction

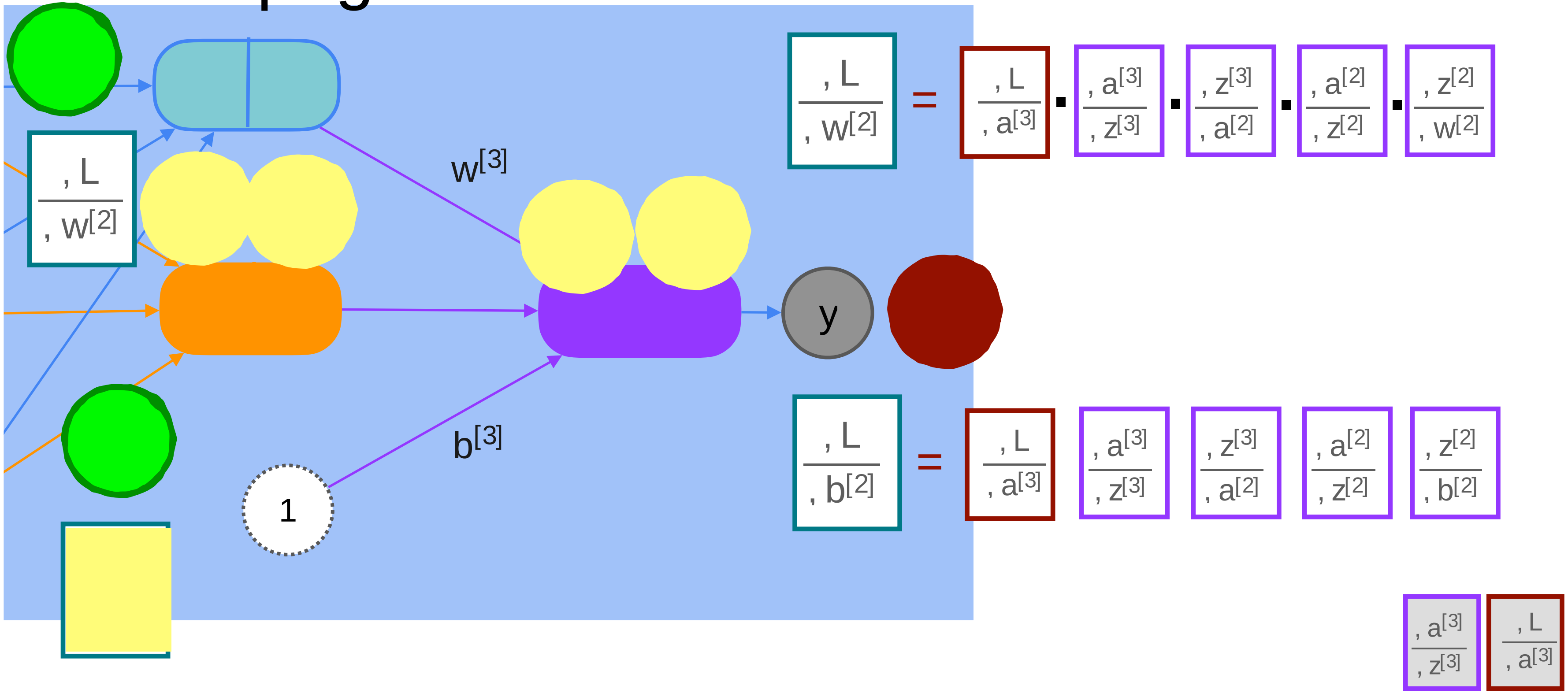


$$\frac{, L}{, w^{[2]}} = \frac{, L}{, a^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, z^{[3]}}{, a^{[2]}} \cdot \frac{, a^{[2]}}{, z^{[2]}} \cdot \frac{, z^{[2]}}{, w^{[2]}}$$

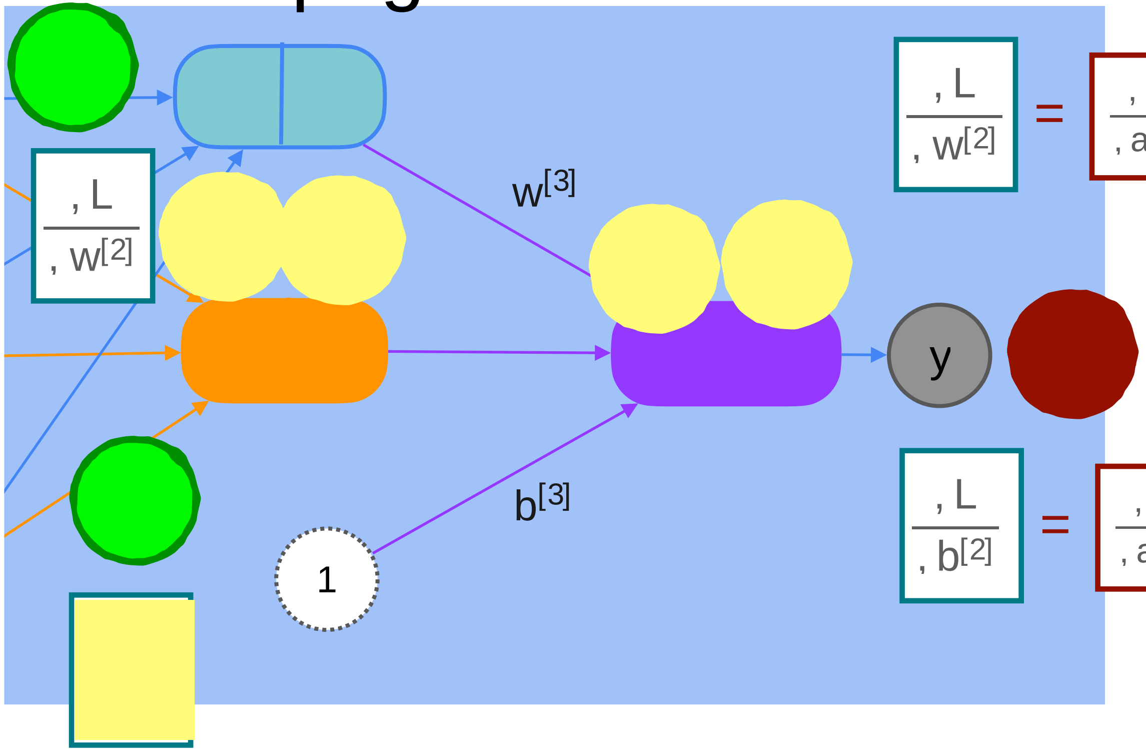
$$\frac{, L}{, b^{[2]}} = \frac{, L}{, a^{[3]}} \frac{, a^{[3]}}{, z^{[3]}} \frac{, z^{[3]}}{, a^{[2]}} \frac{, a^{[2]}}{, z^{[2]}}$$

$$\frac{, a^{[3]}}{, z^{[3]}} \quad \frac{, L}{, a^{[3]}}$$

Back Propagation Introduction



Back Propagation Introduction

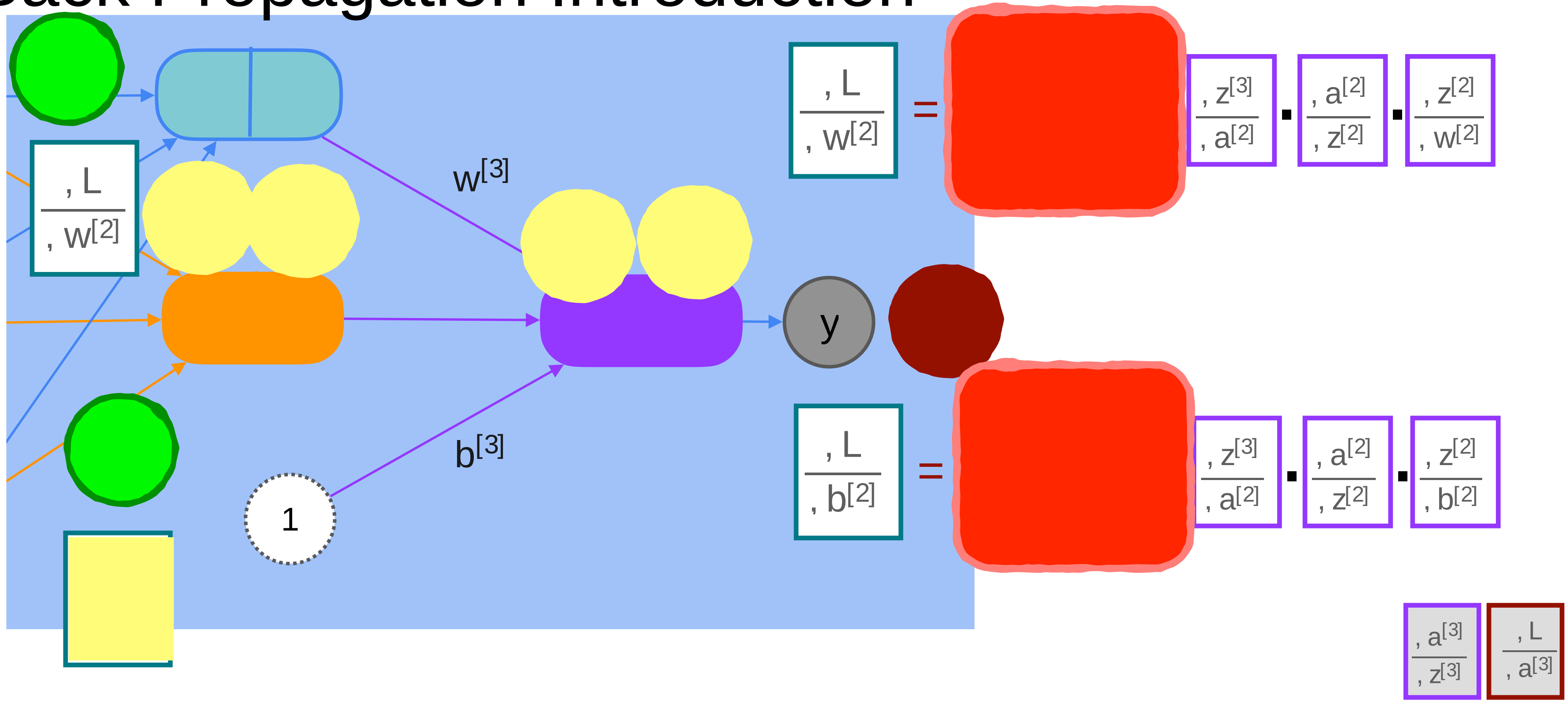


$$\frac{, L}{, w^{[2]}} = \frac{, L}{, a^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, z^{[3]}}{, a^{[2]}} \cdot \frac{, a^{[2]}}{, z^{[2]}} \cdot \frac{, z^{[2]}}{, w^{[2]}}$$

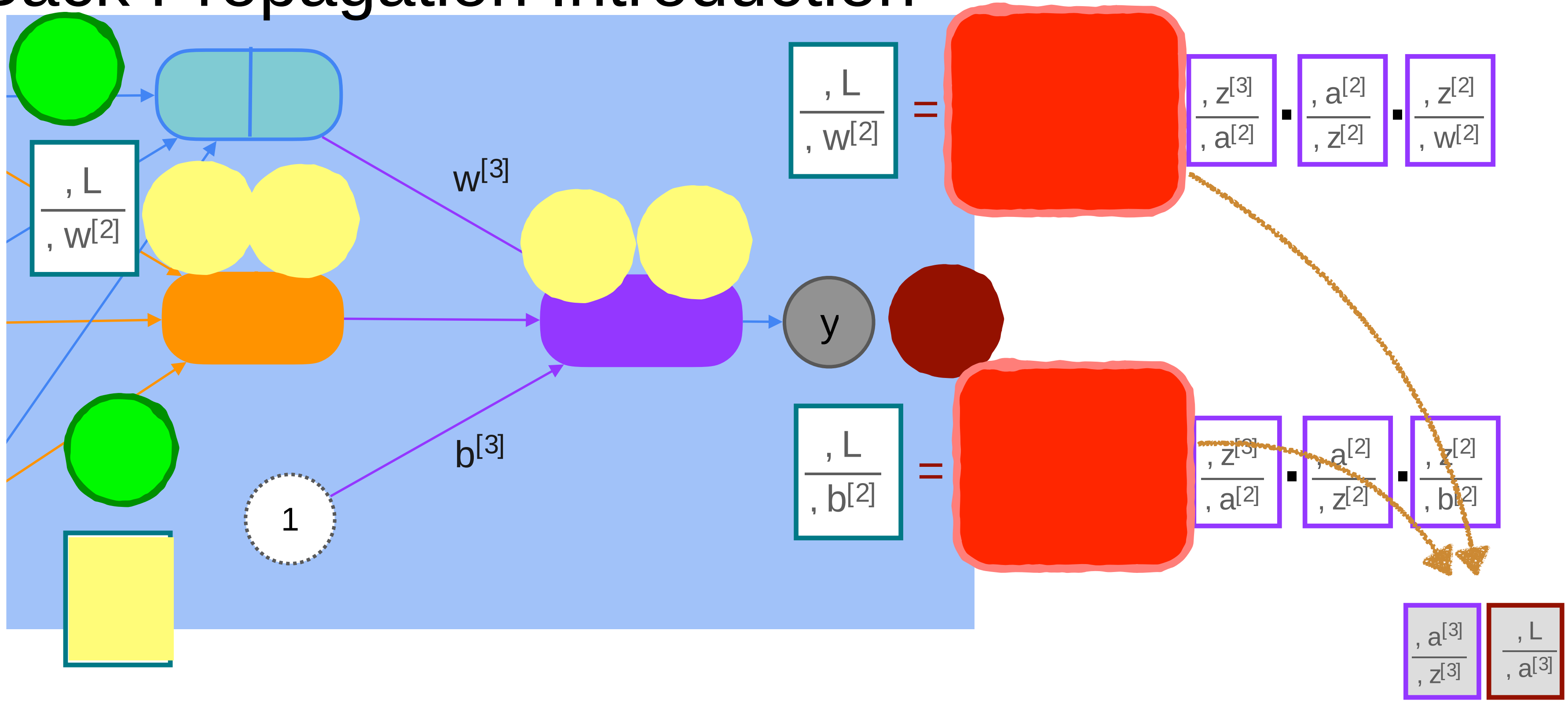
$$\frac{, L}{, b^{[2]}} = \frac{, L}{, a^{[3]}} \cdot \frac{, a^{[3]}}{, z^{[3]}} \cdot \frac{, z^{[3]}}{, a^{[2]}} \cdot \frac{, a^{[2]}}{, z^{[2]}} \cdot \frac{, z^{[2]}}{, b^{[2]}}$$

$\frac{, a^{[3]}}{, z^{[3]}}$ $\frac{, L}{, a^{[3]}}$

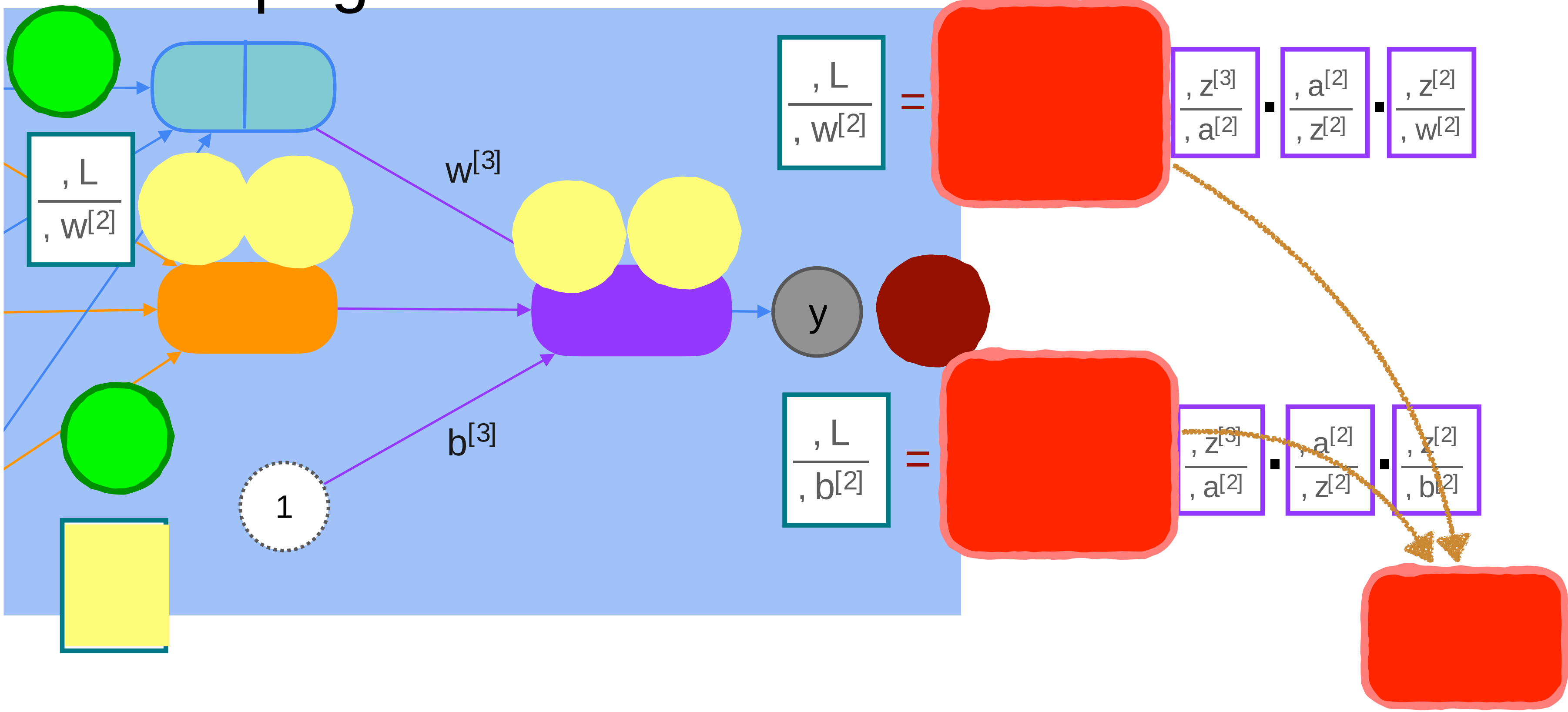
Back Propagation Introduction



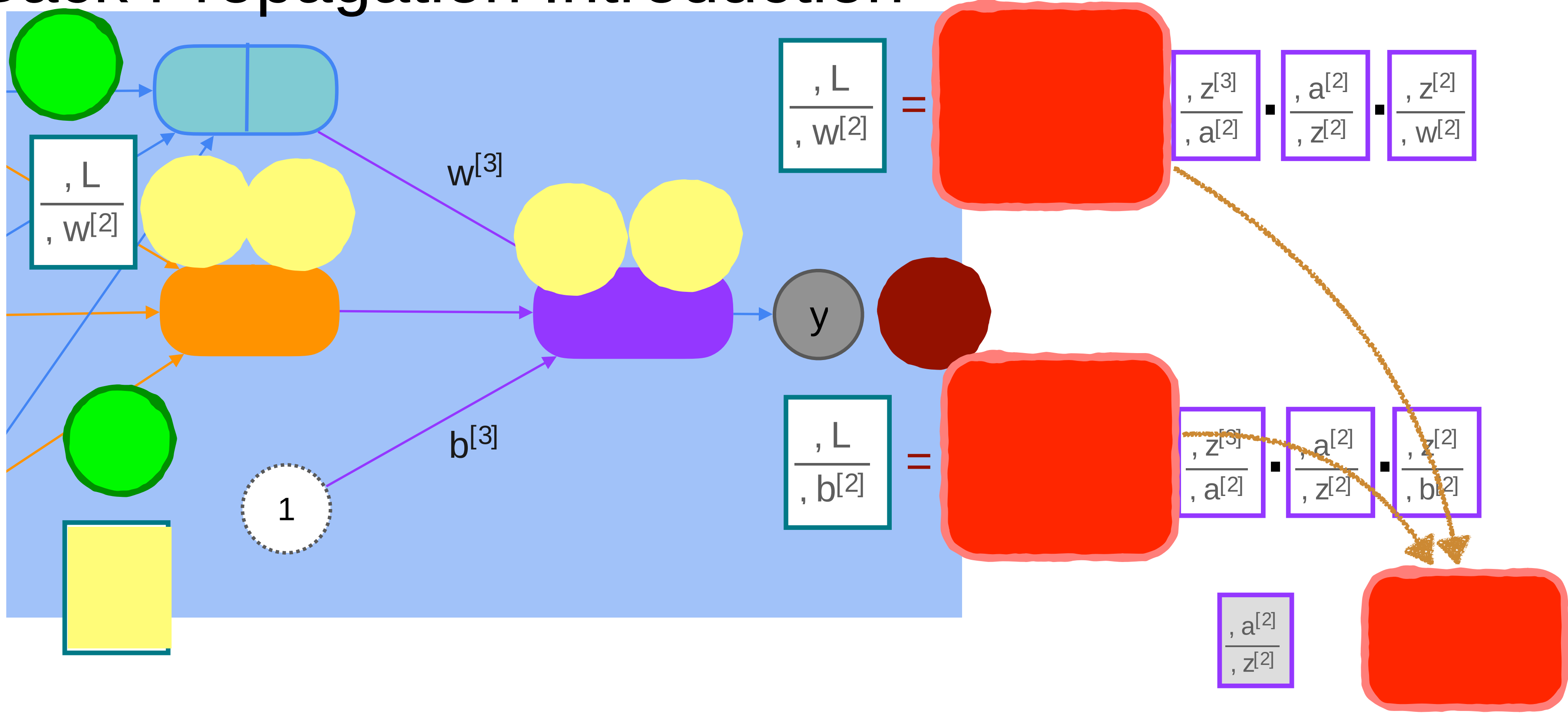
Back Propagation Introduction



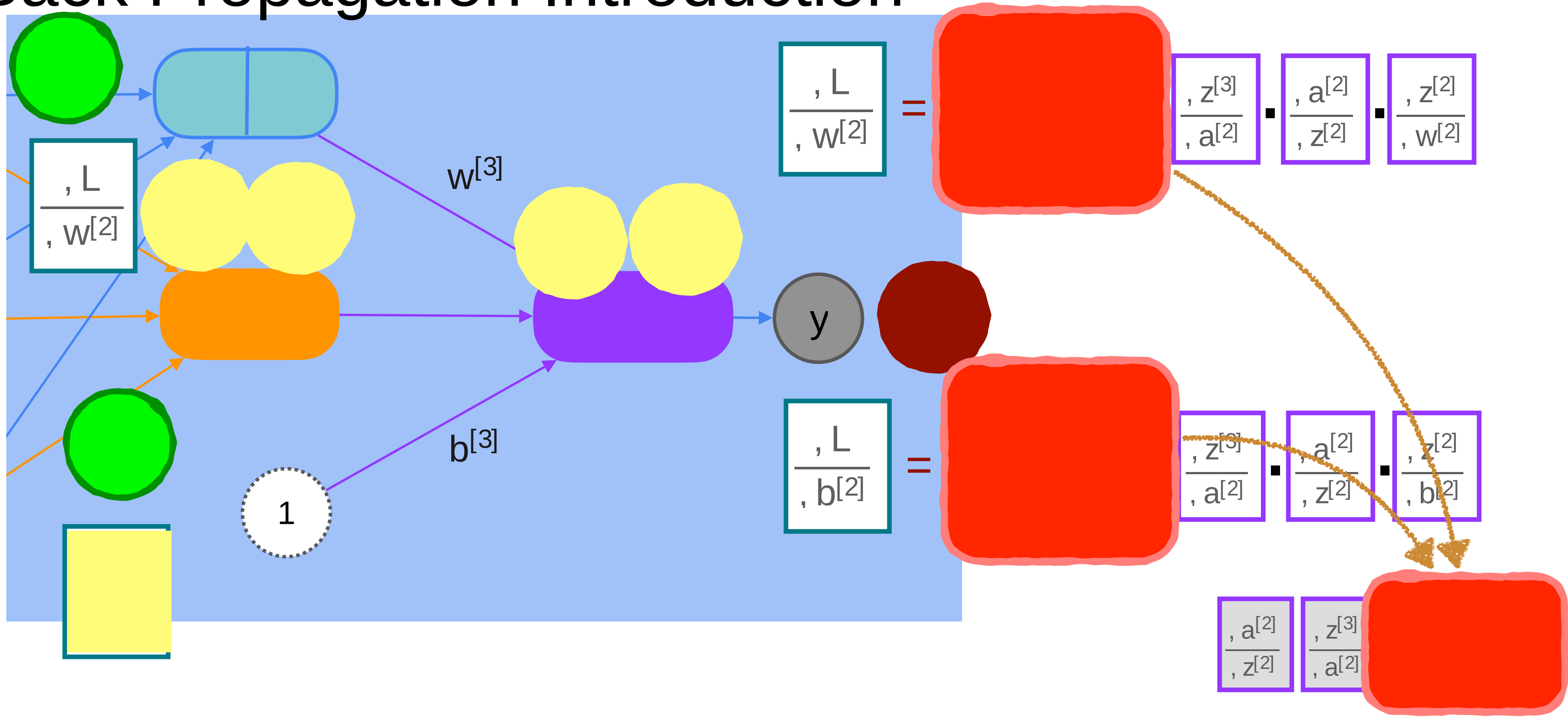
Back Propagation Introduction



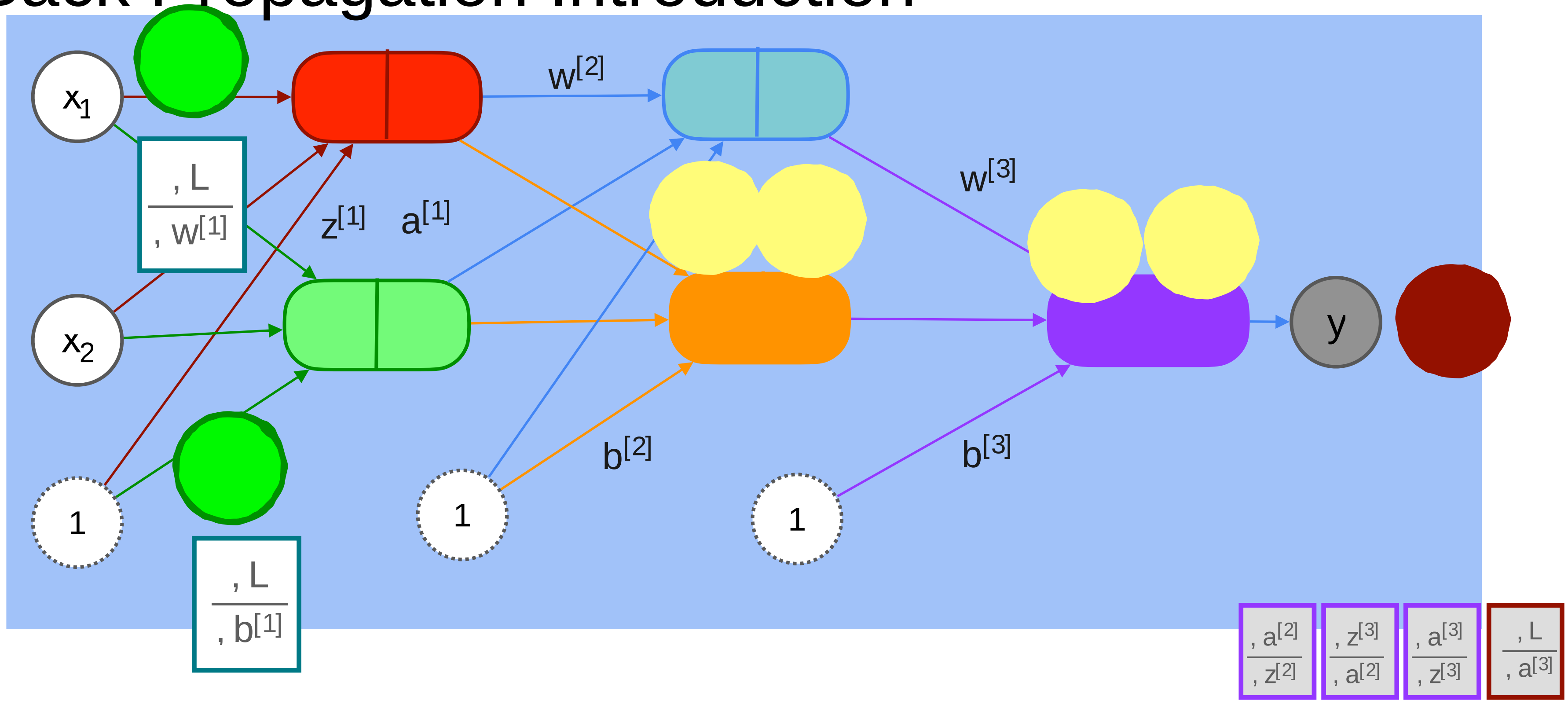
Back Propagation Introduction



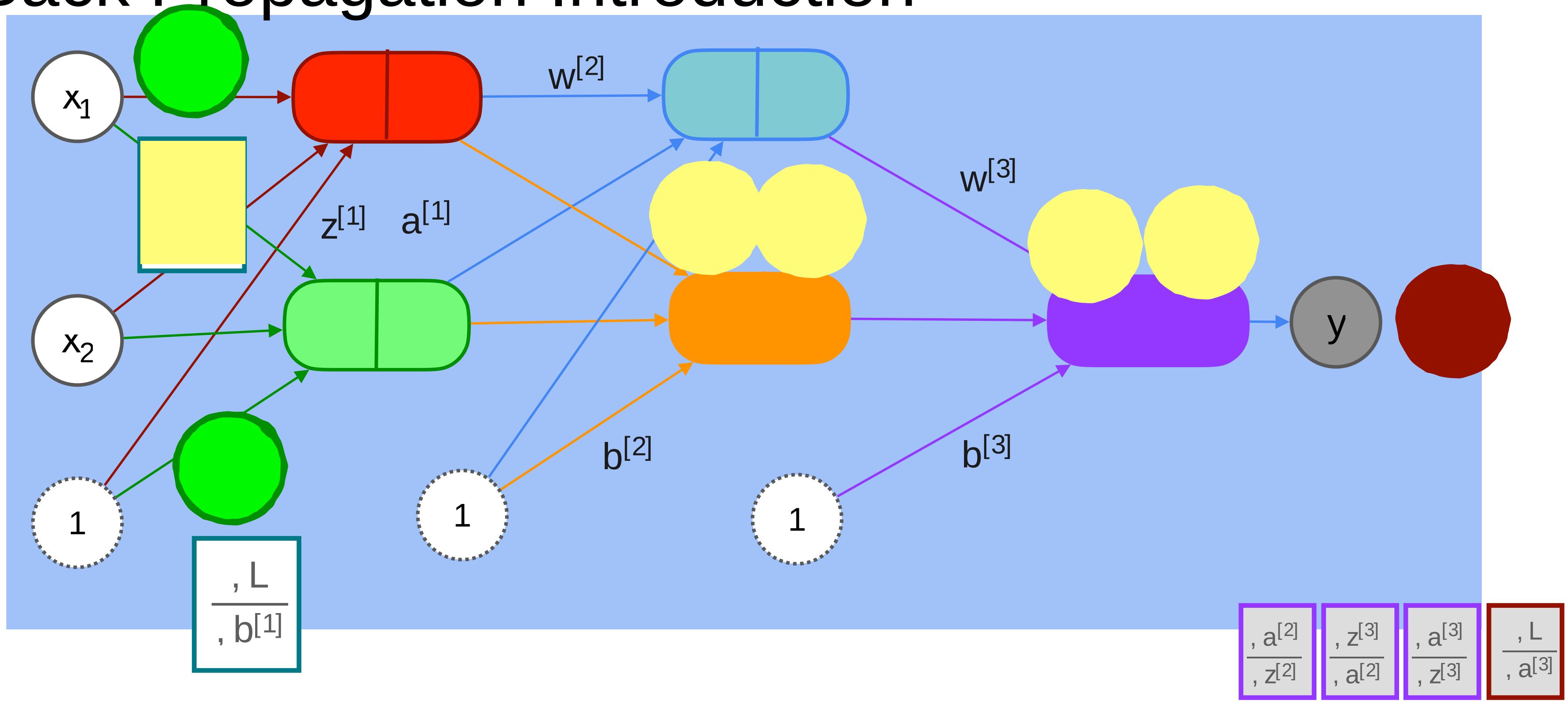
Back Propagation Introduction



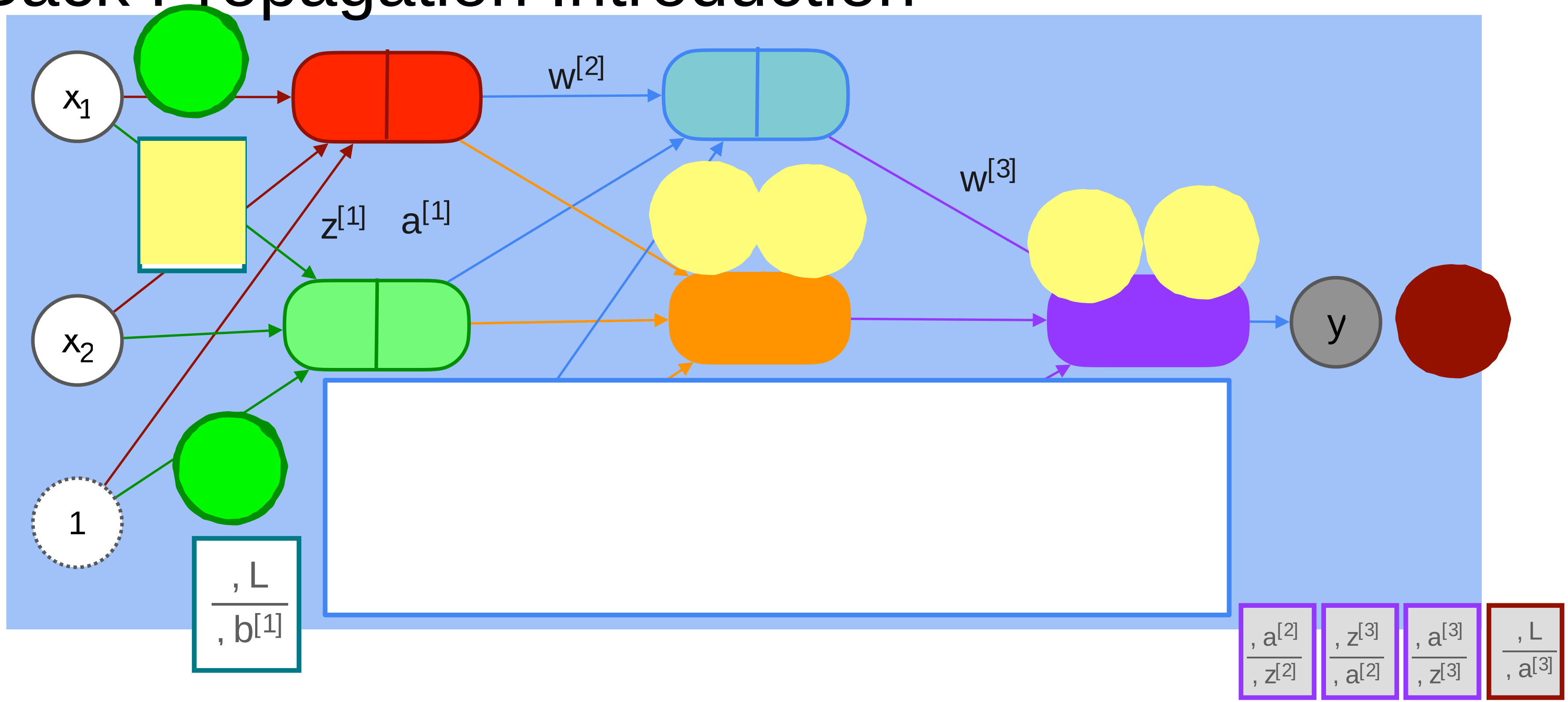
Back Propagation Introduction



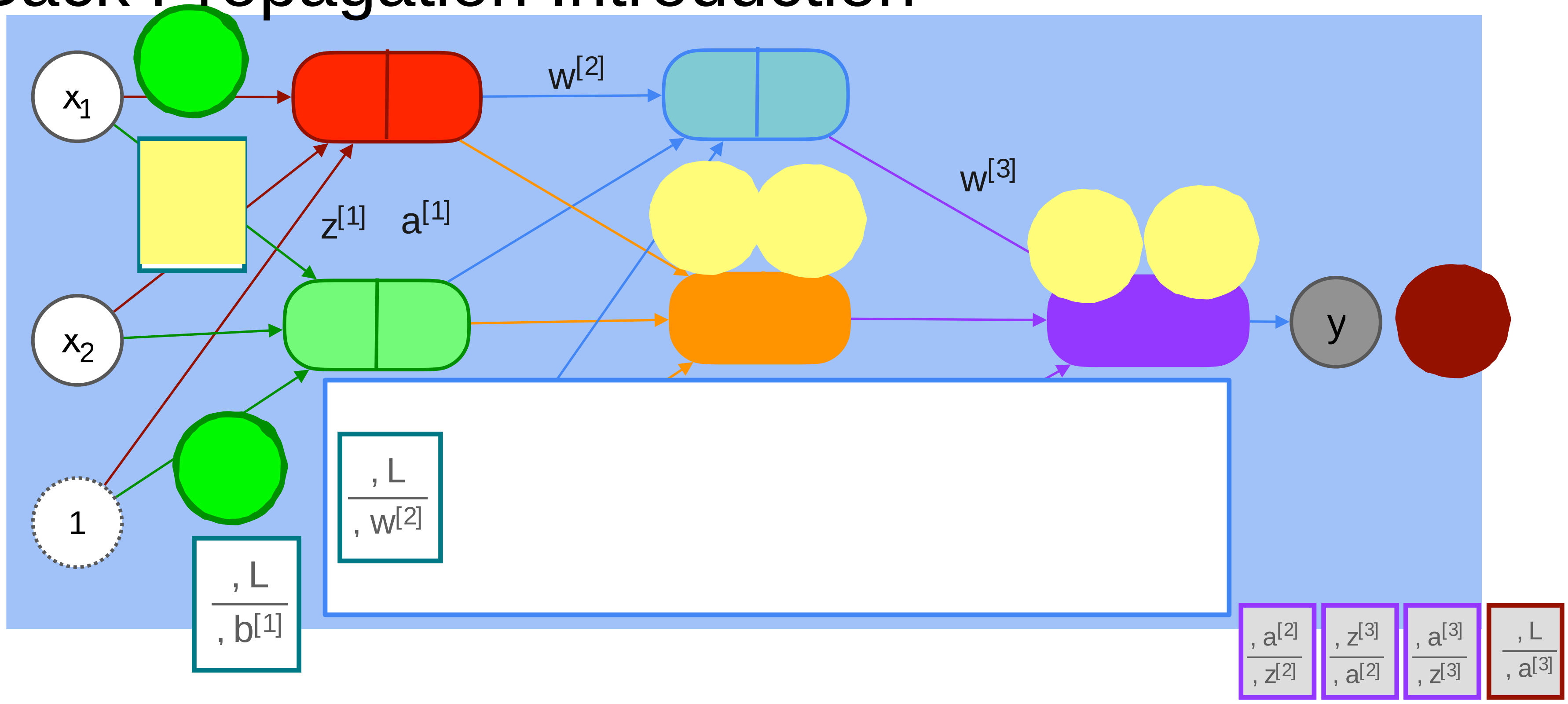
Back Propagation Introduction



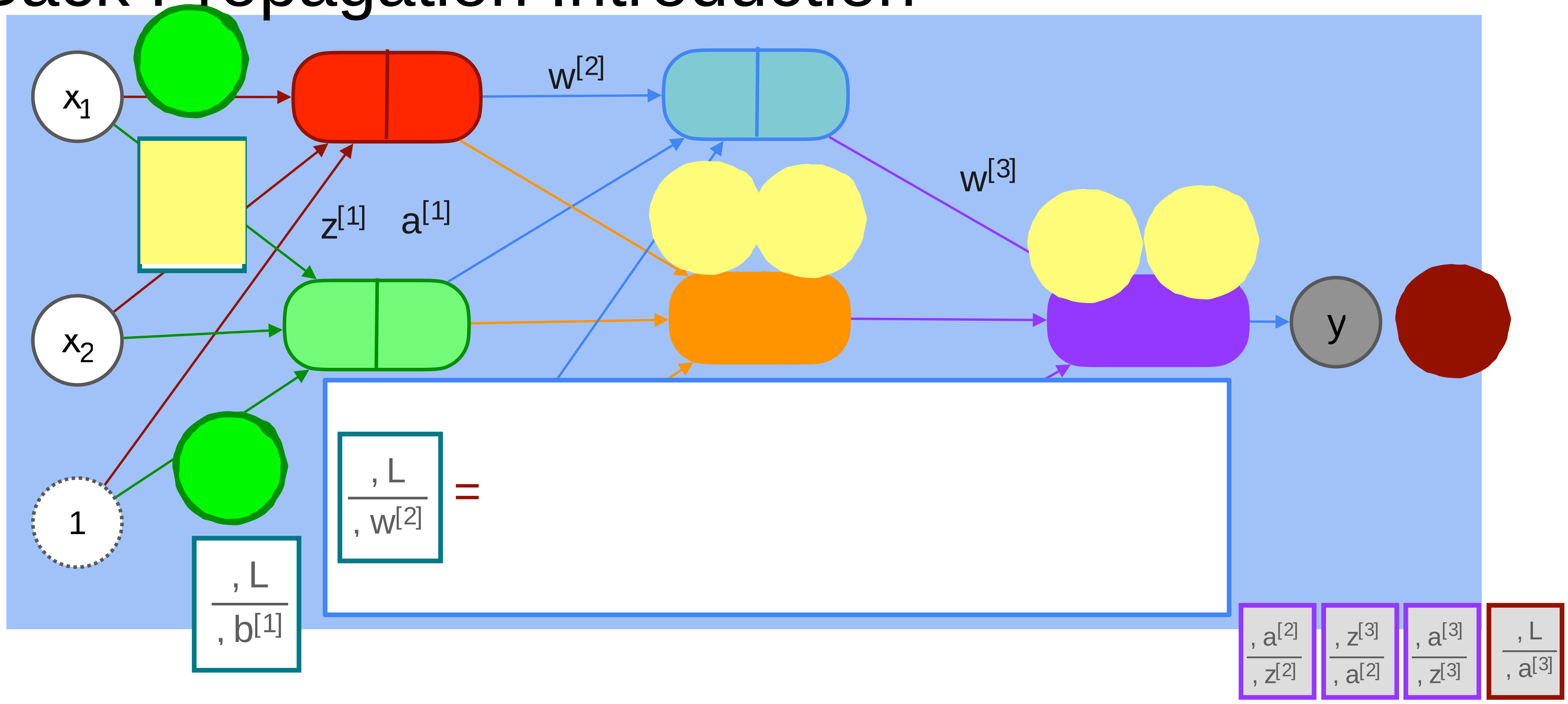
Back Propagation Introduction



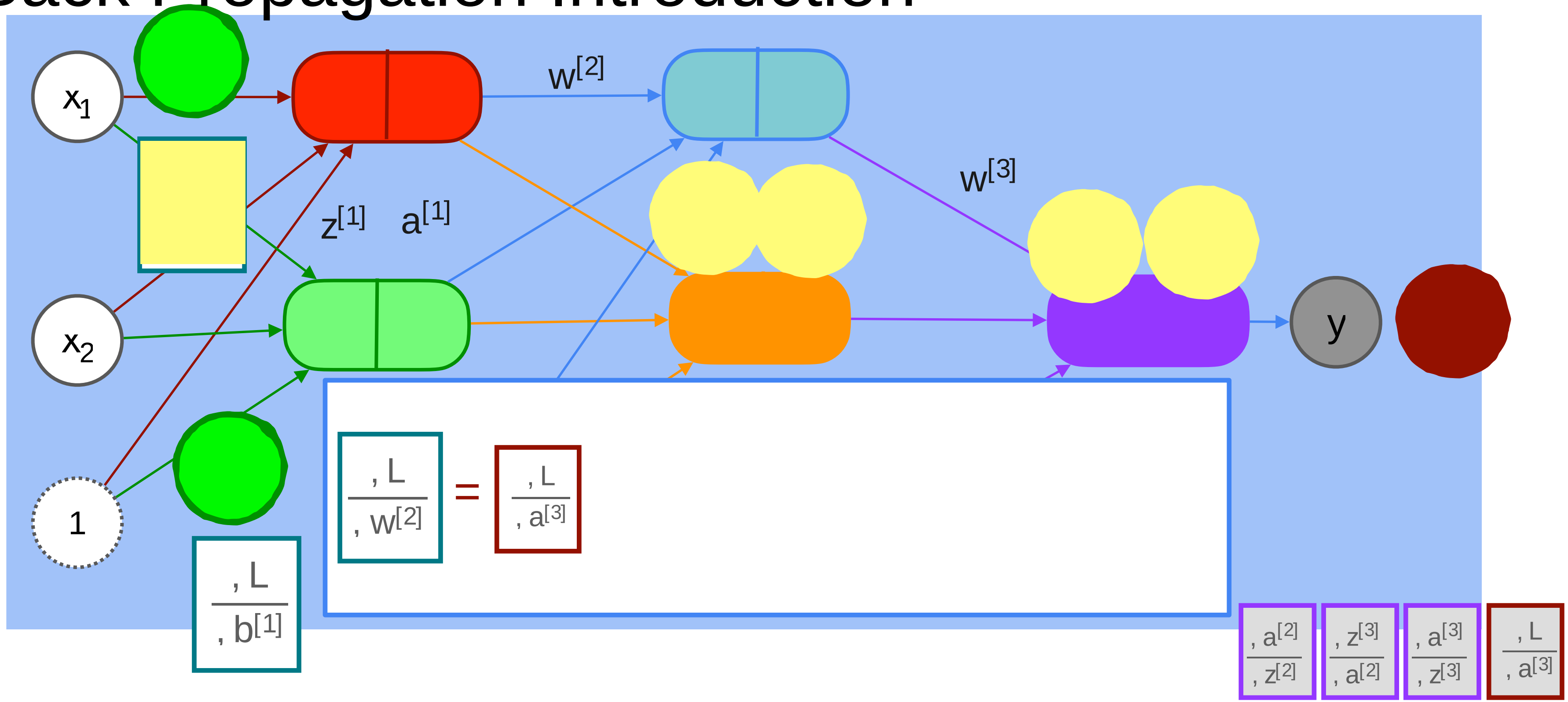
Back Propagation Introduction



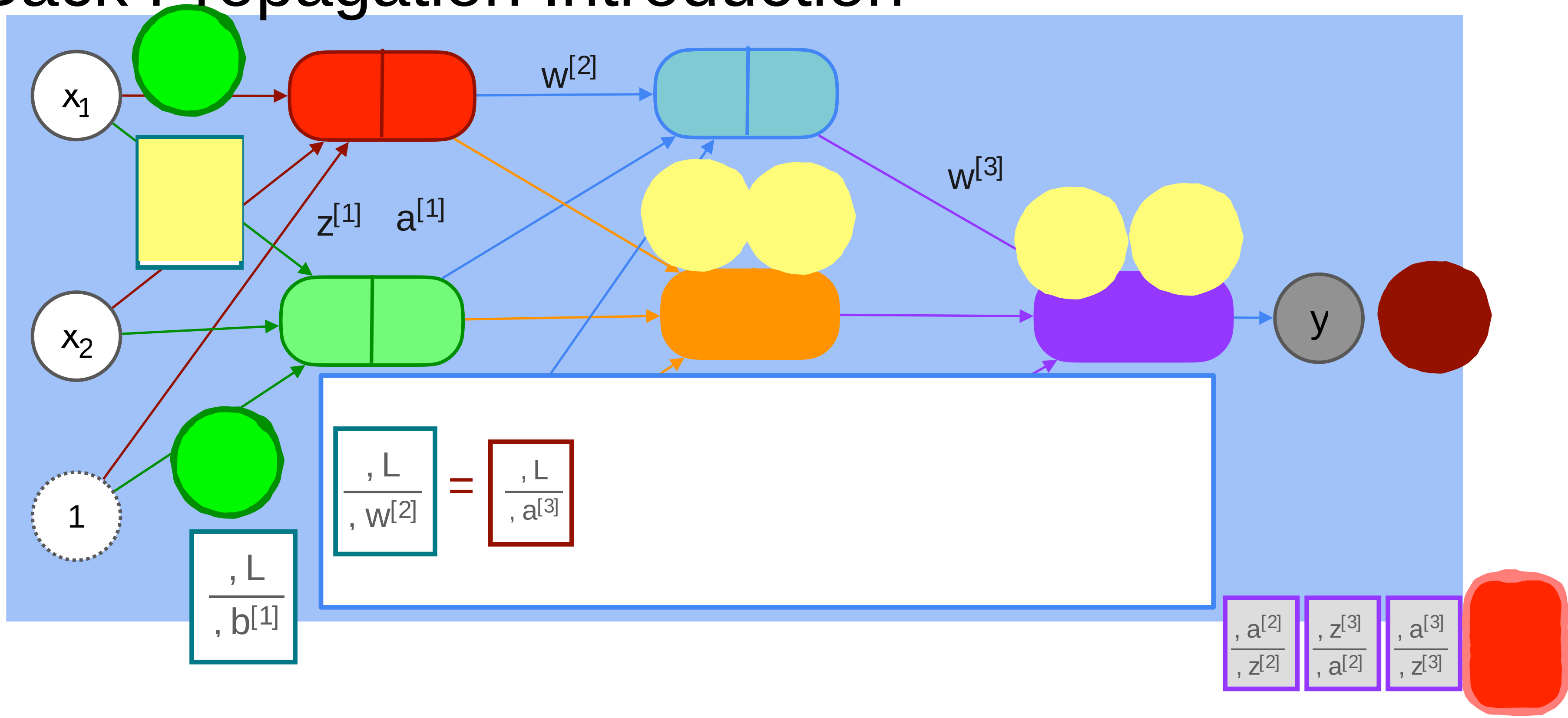
Back Propagation Introduction



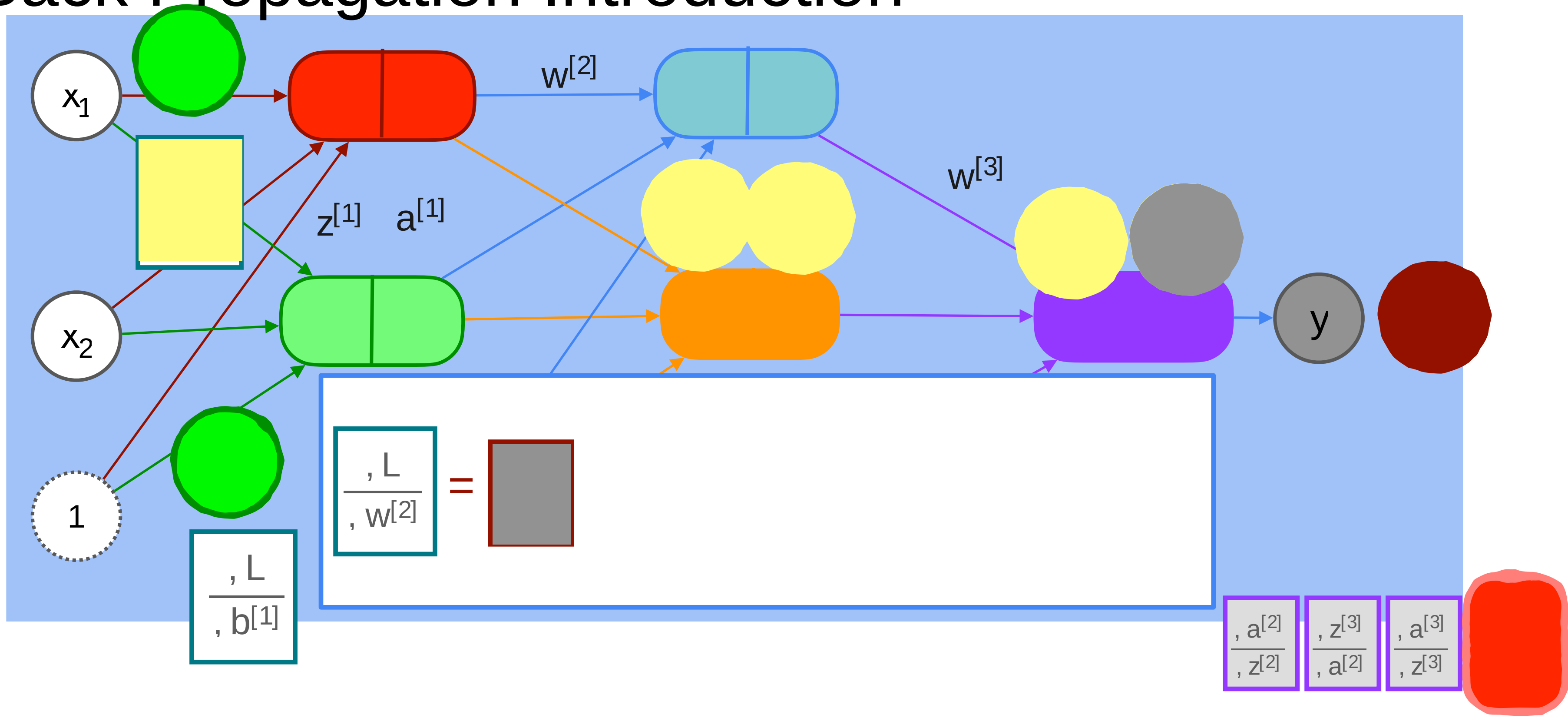
Back Propagation Introduction



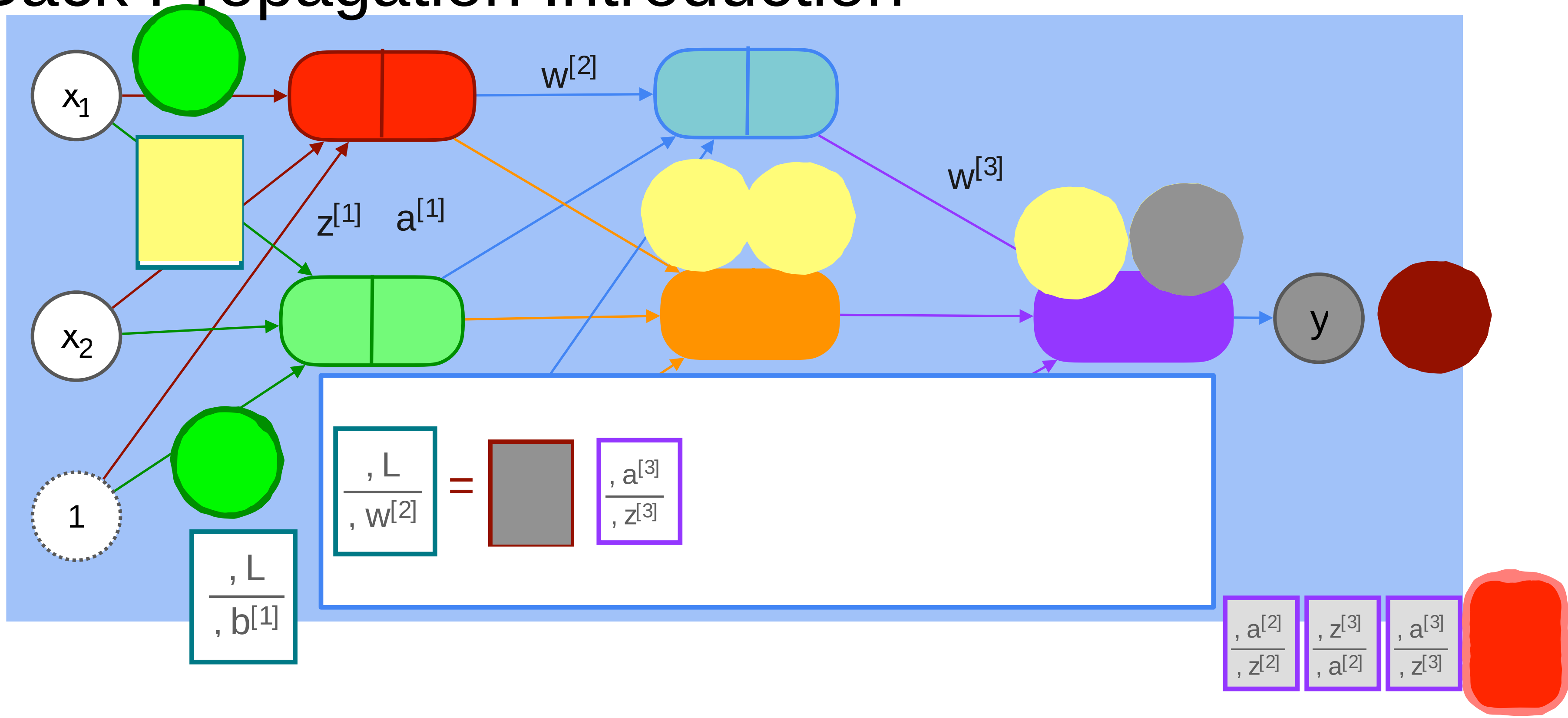
Back Propagation Introduction



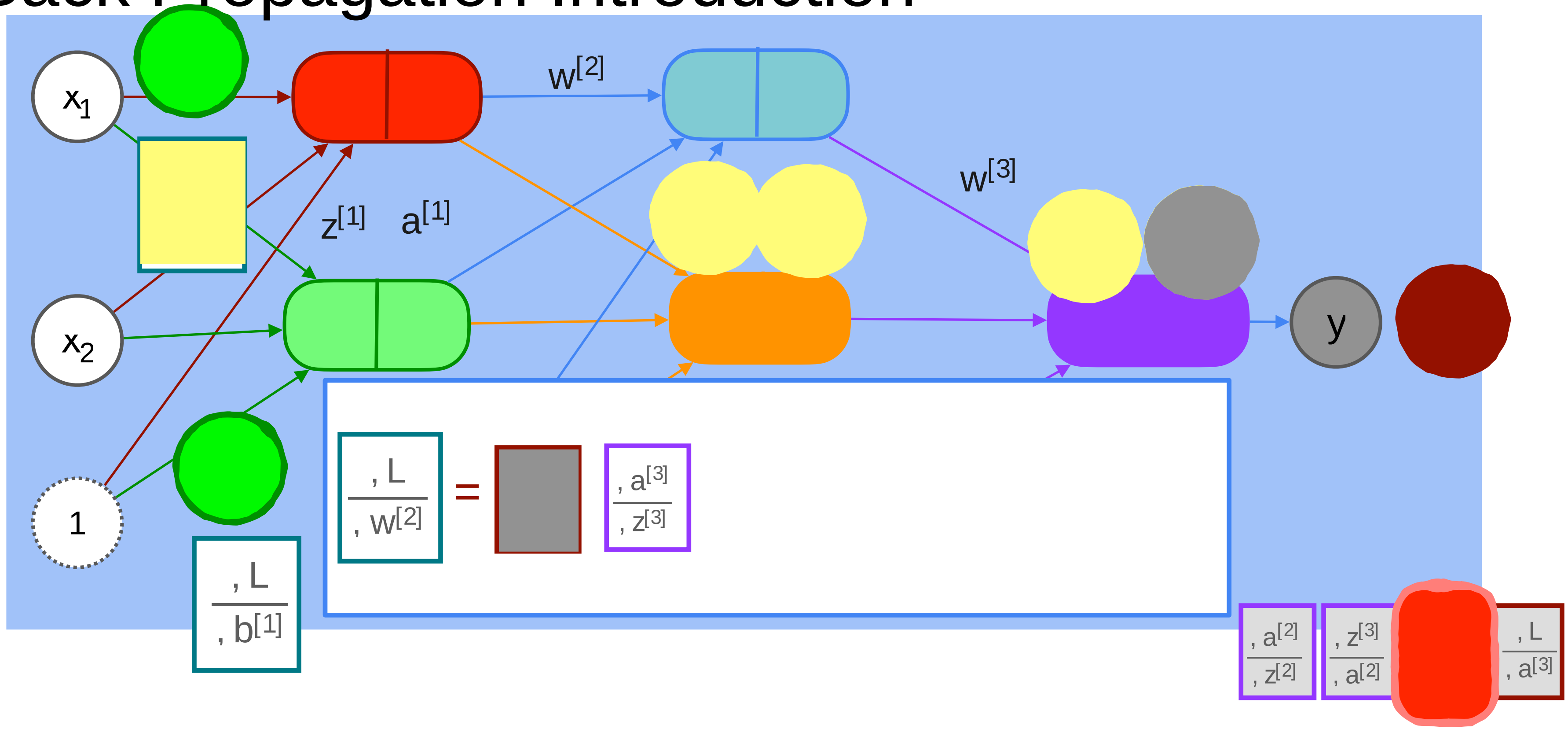
Back Propagation Introduction



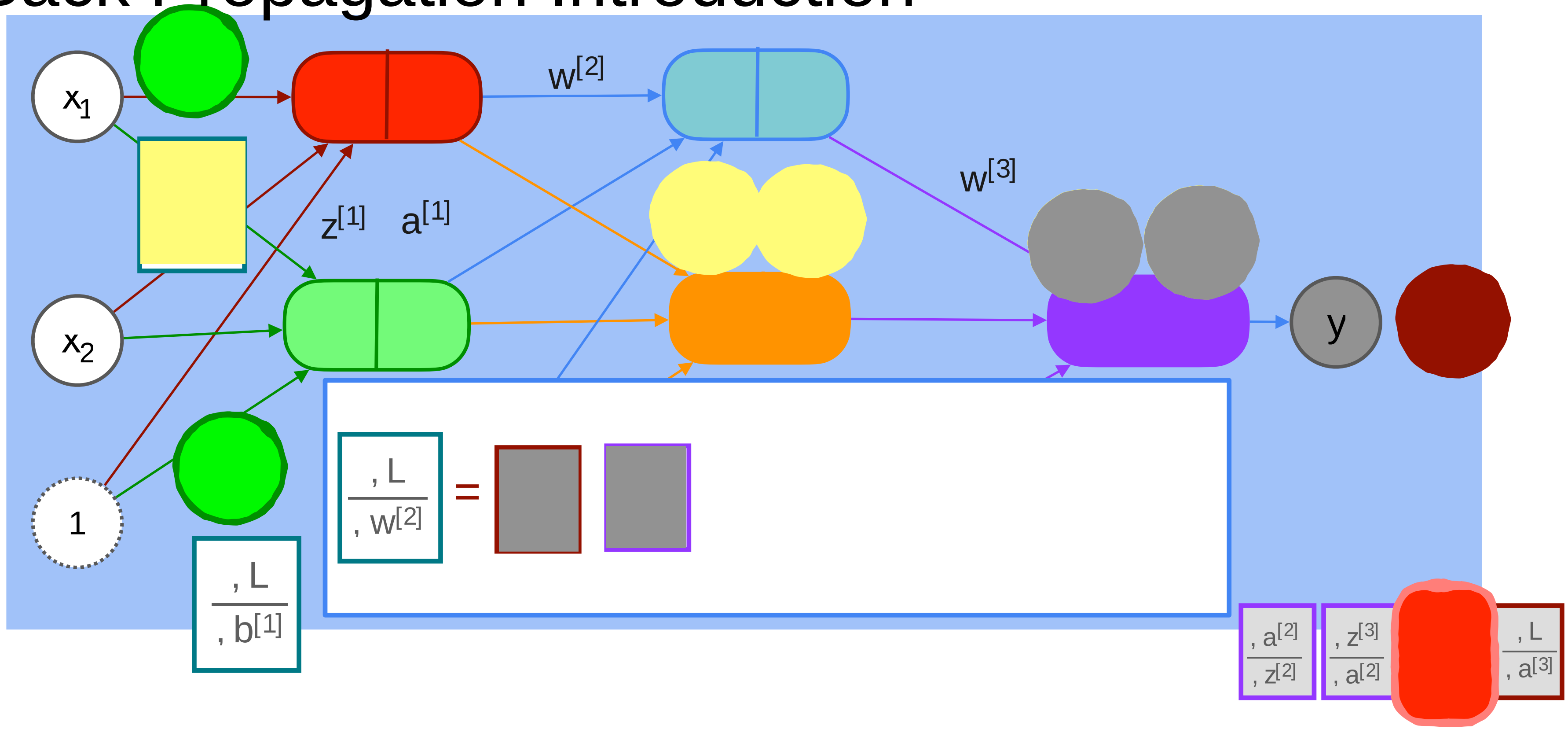
Back Propagation Introduction



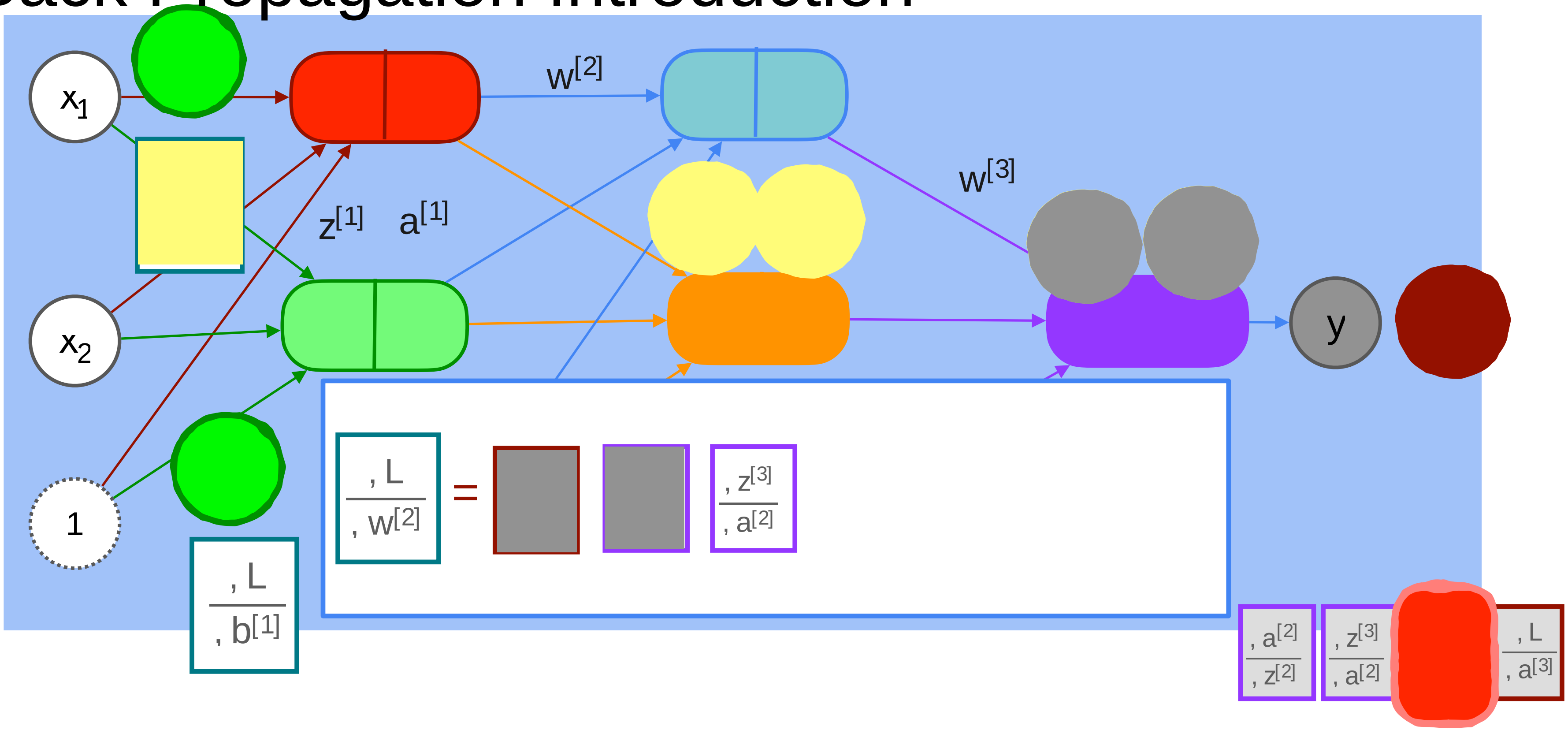
Back Propagation Introduction



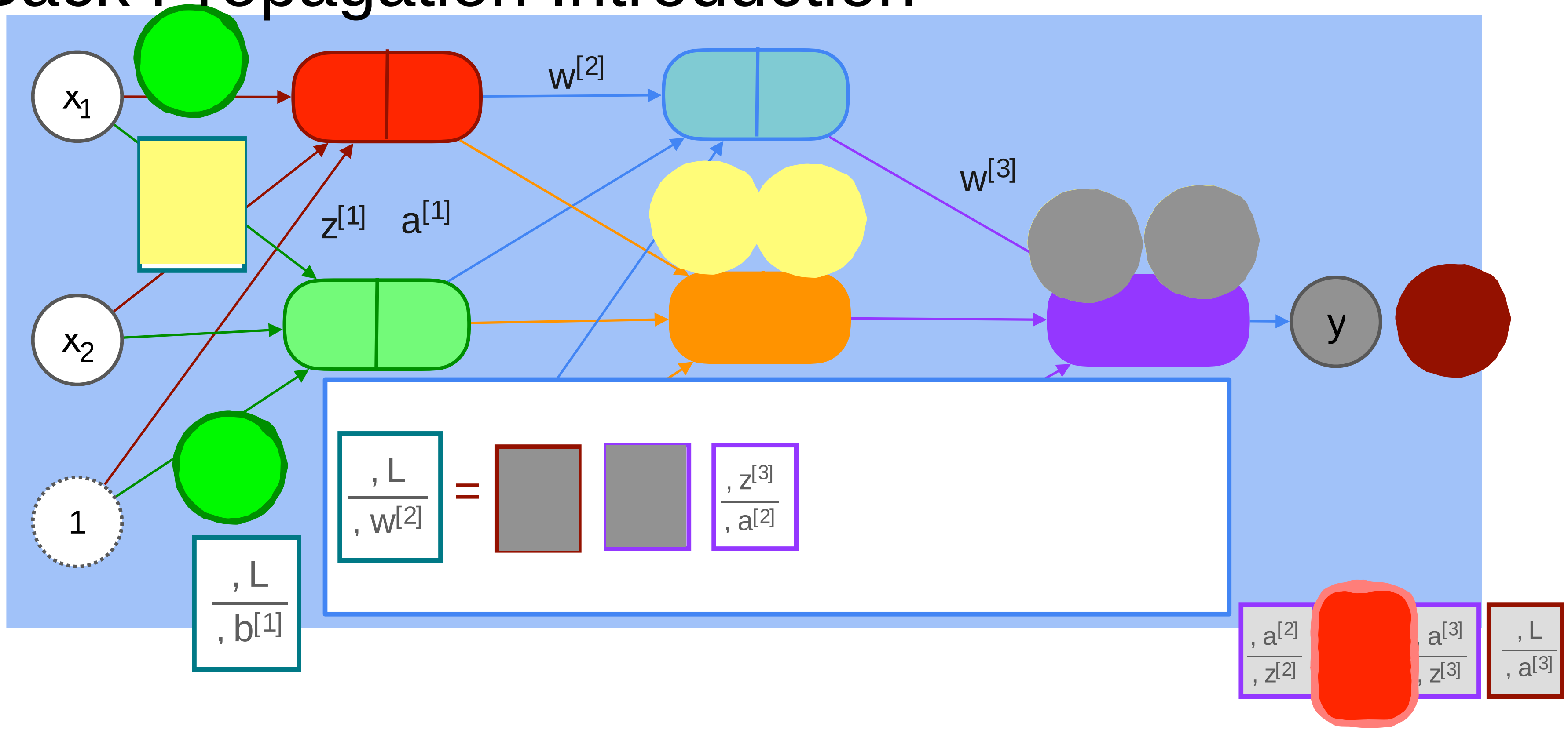
Back Propagation Introduction



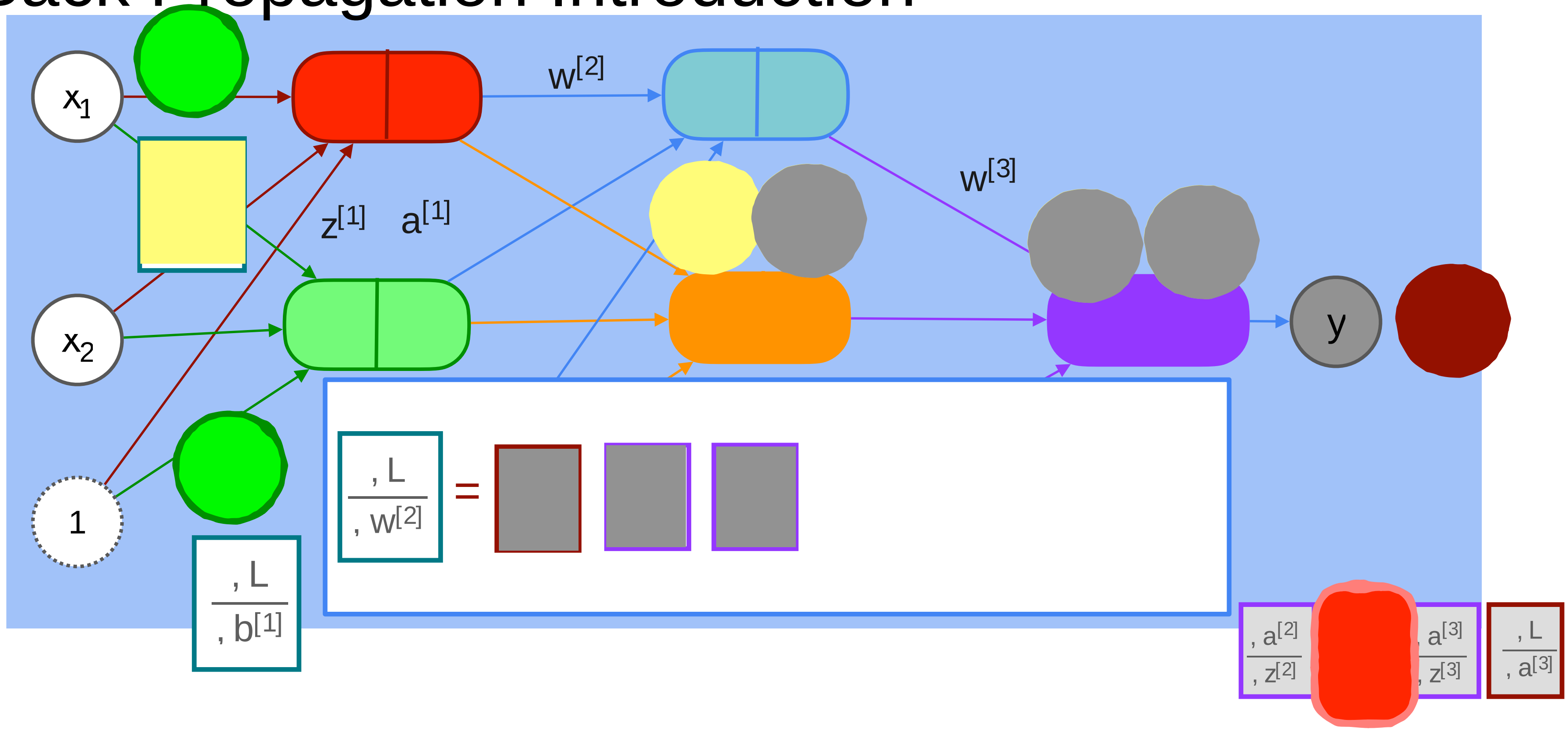
Back Propagation Introduction



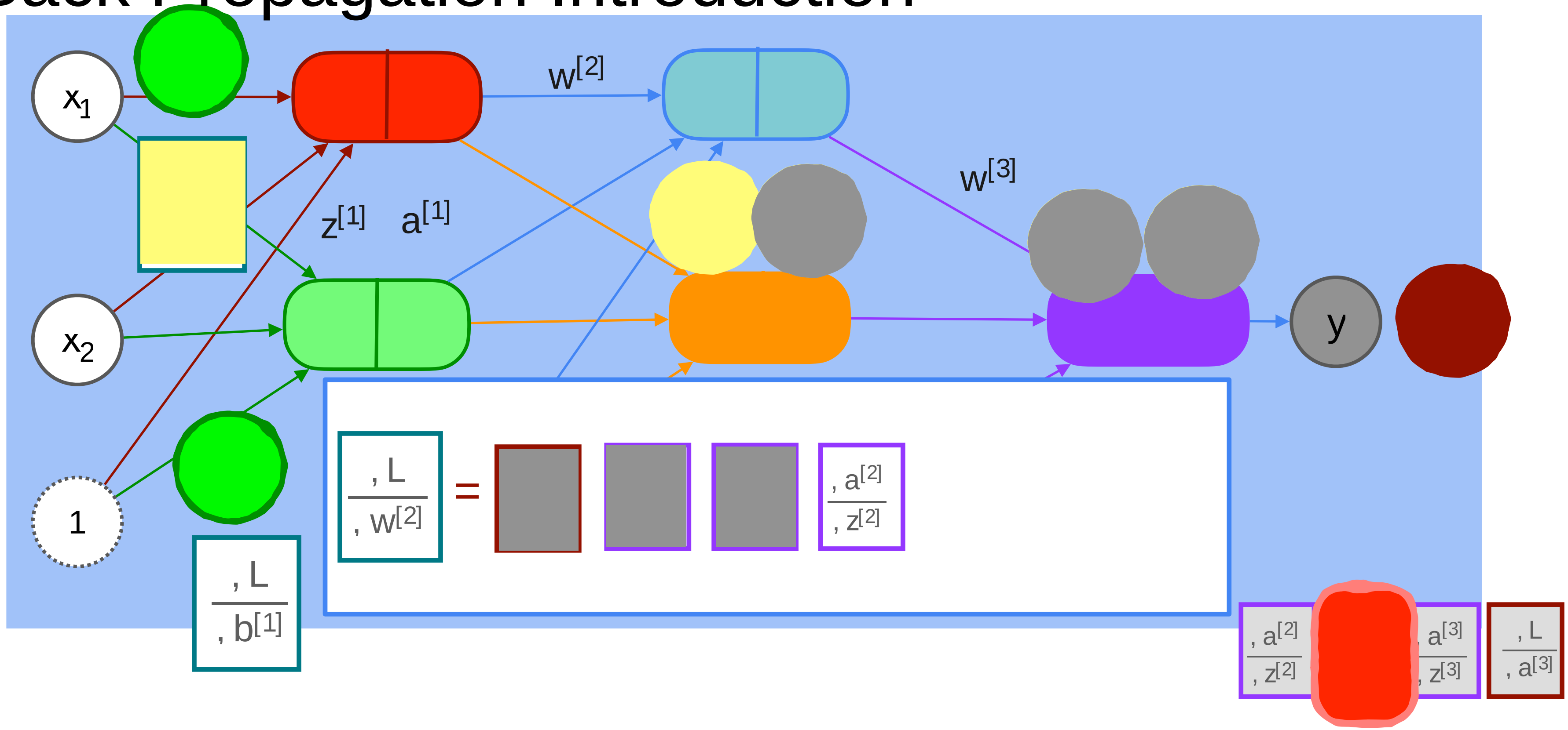
Back Propagation Introduction



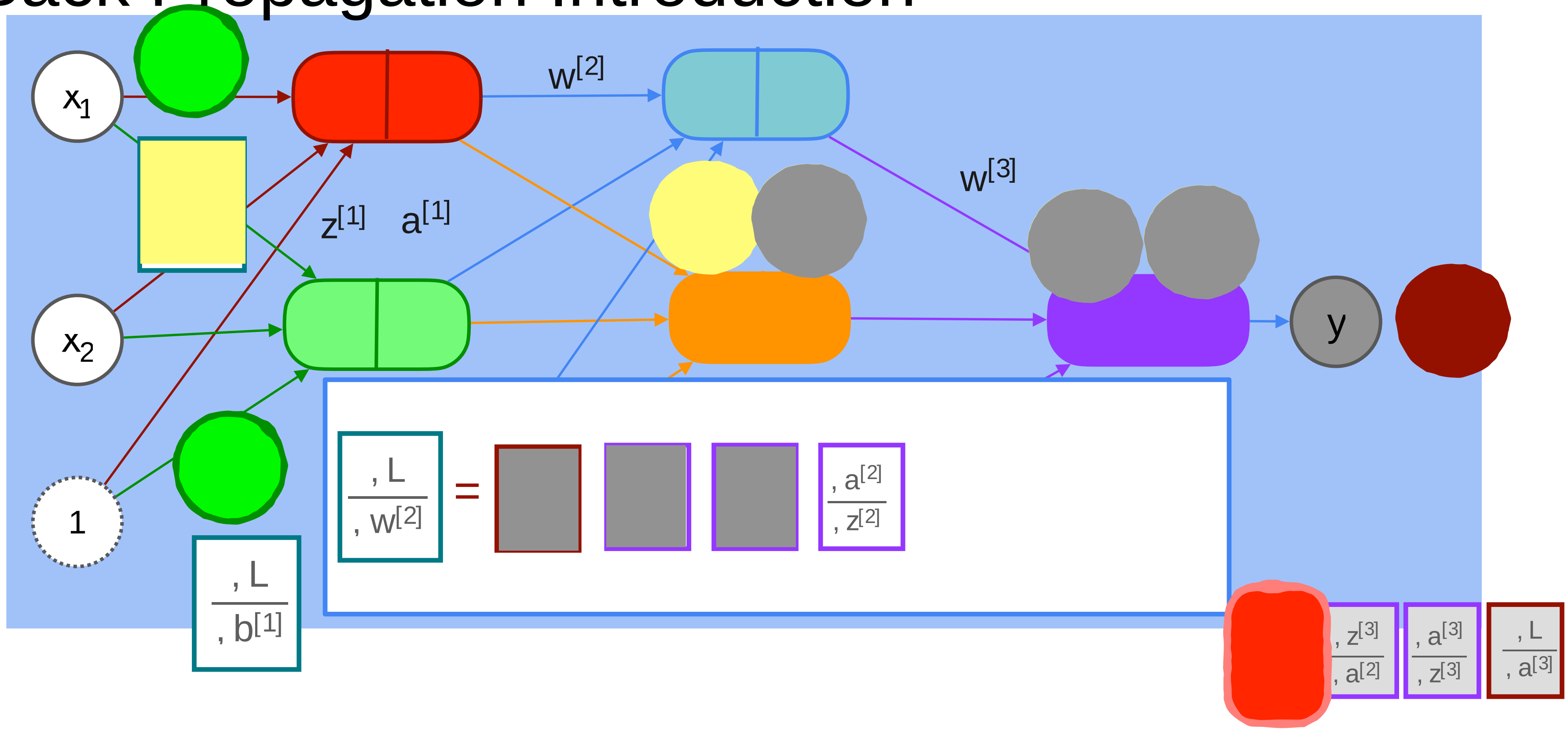
Back Propagation Introduction



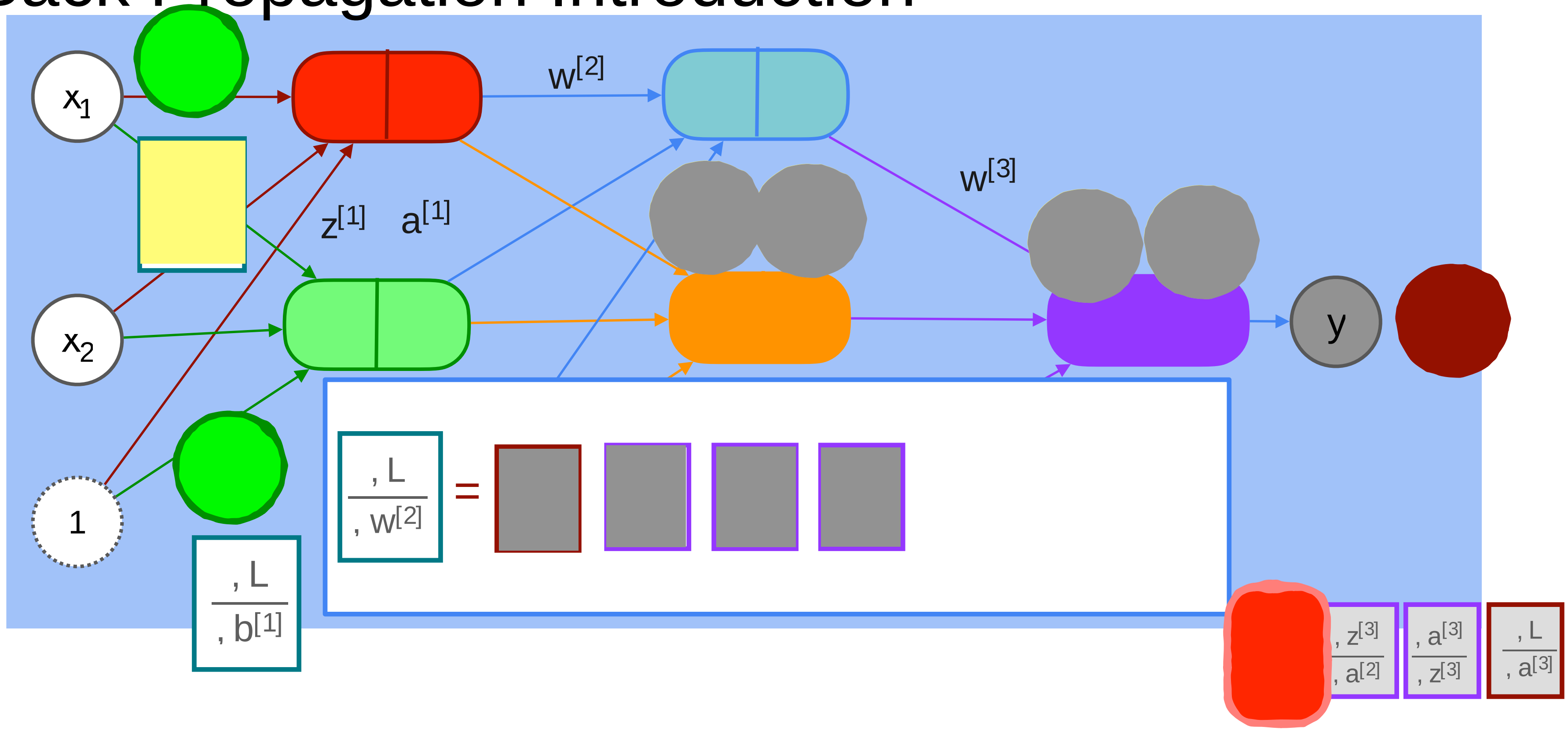
Back Propagation Introduction



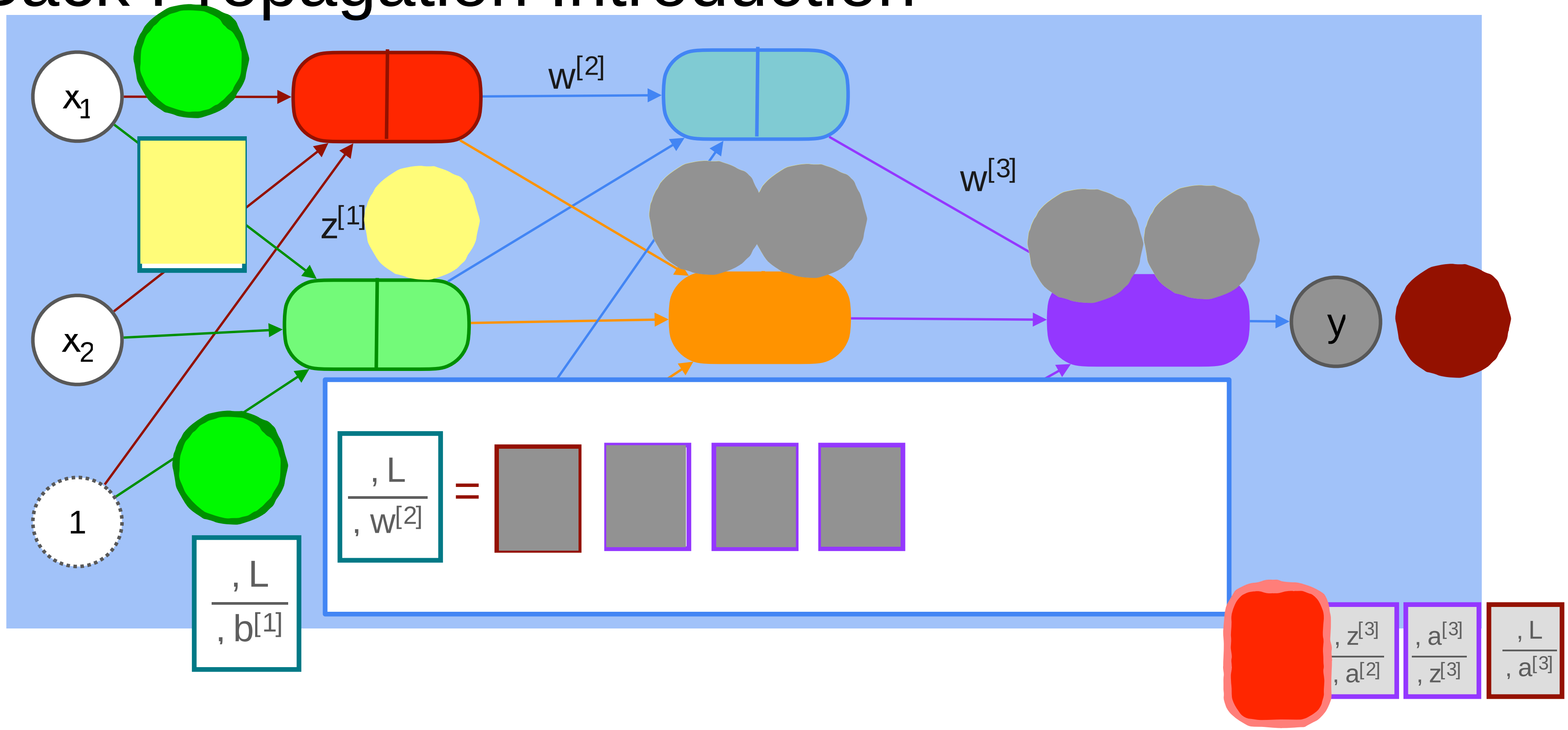
Back Propagation Introduction



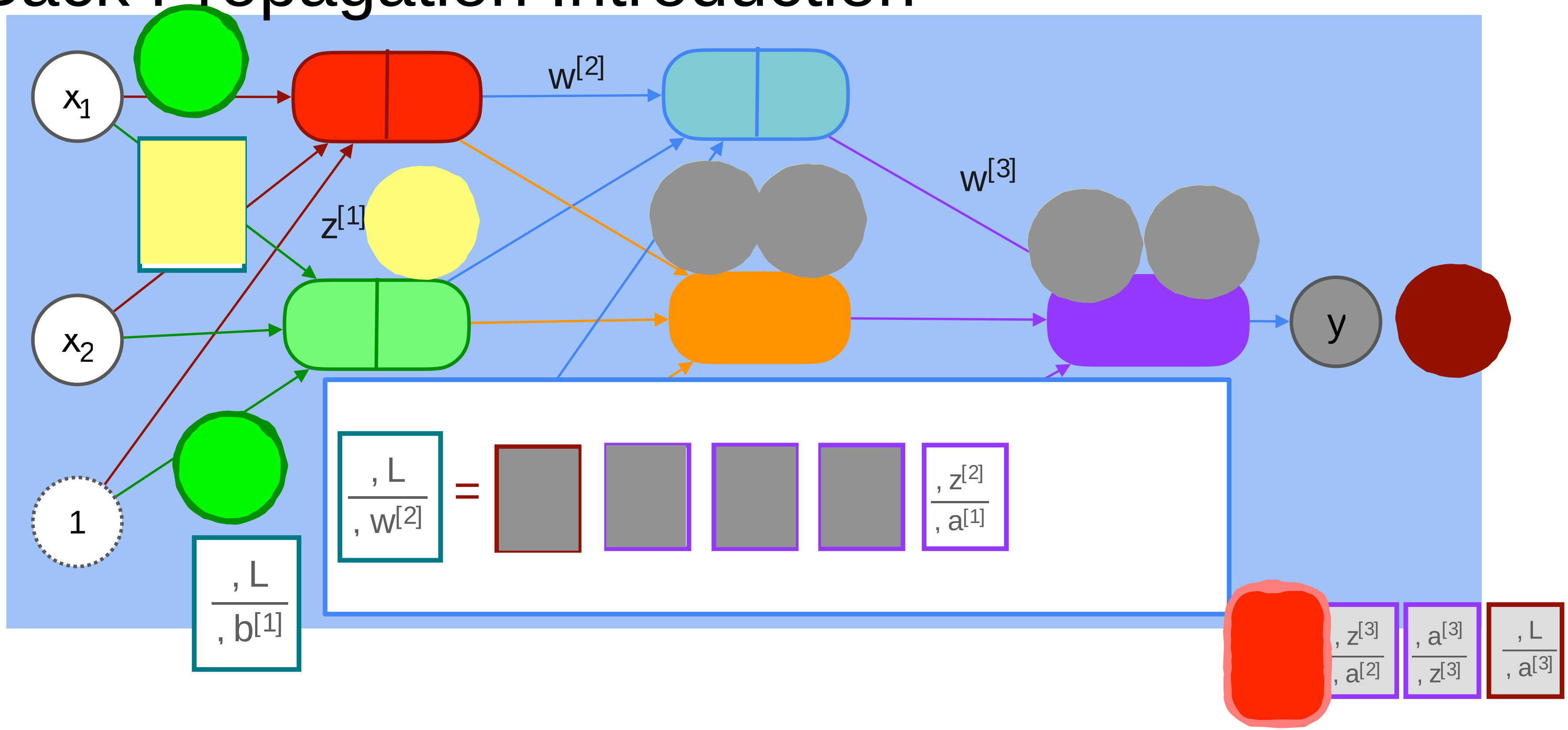
Back Propagation Introduction



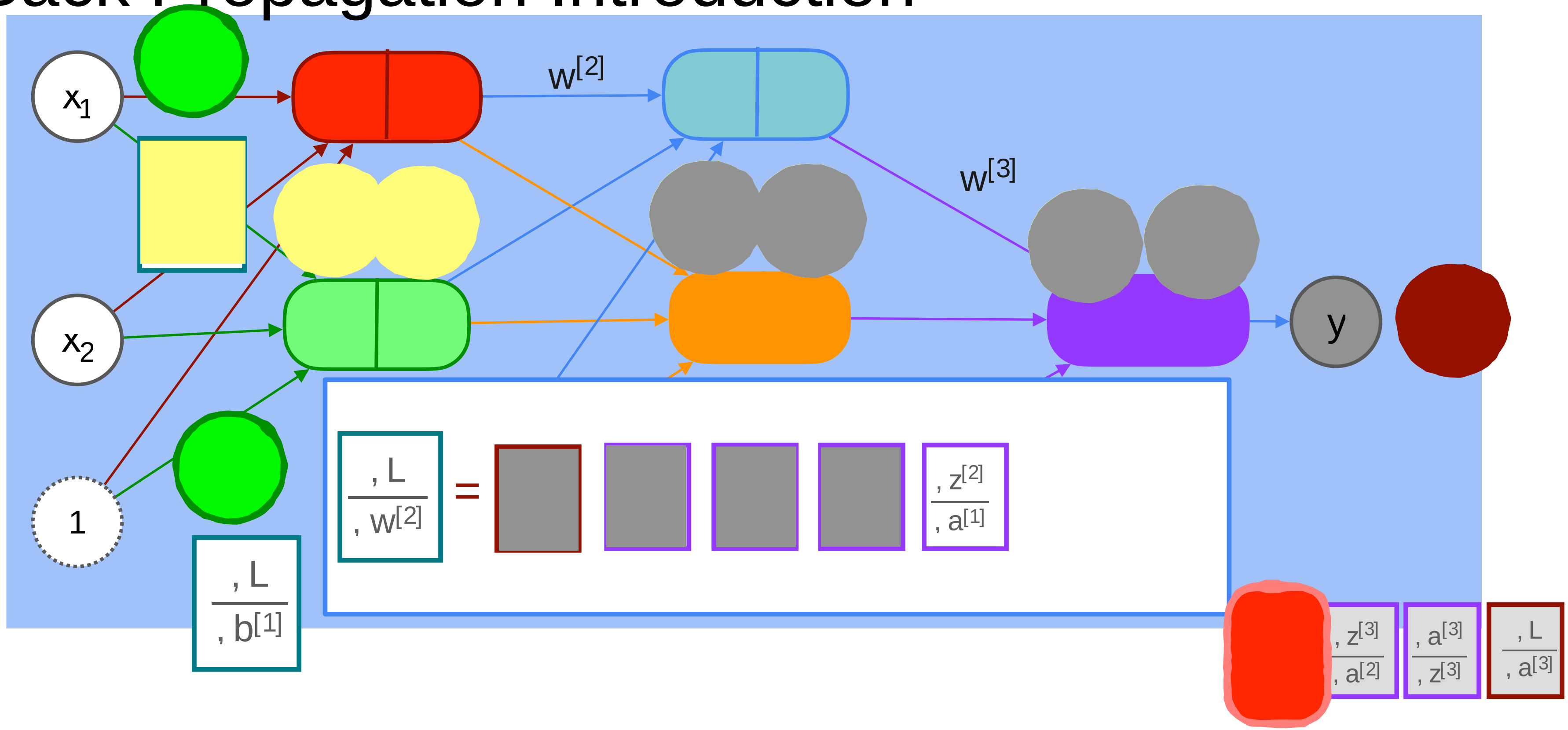
Back Propagation Introduction



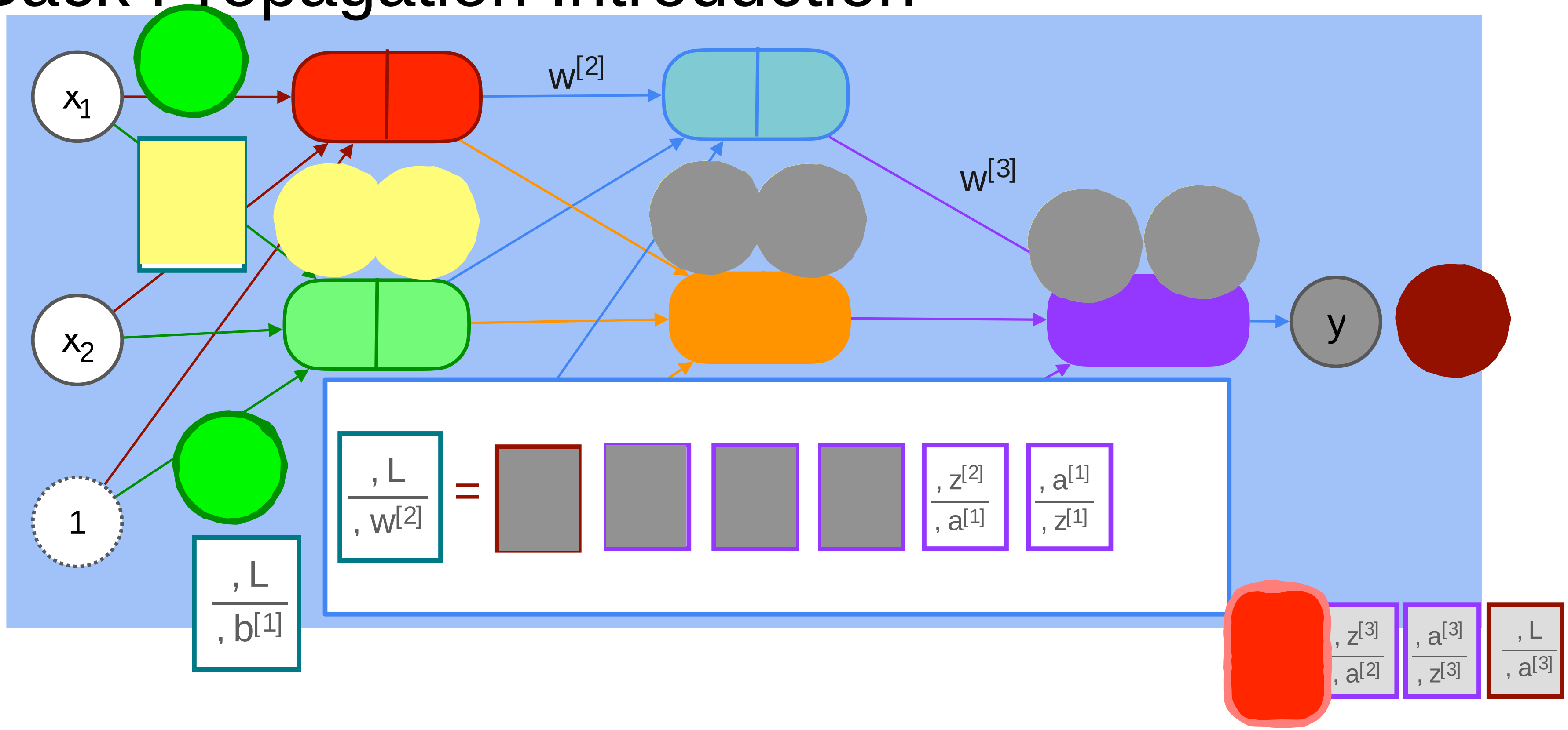
Back Propagation Introduction



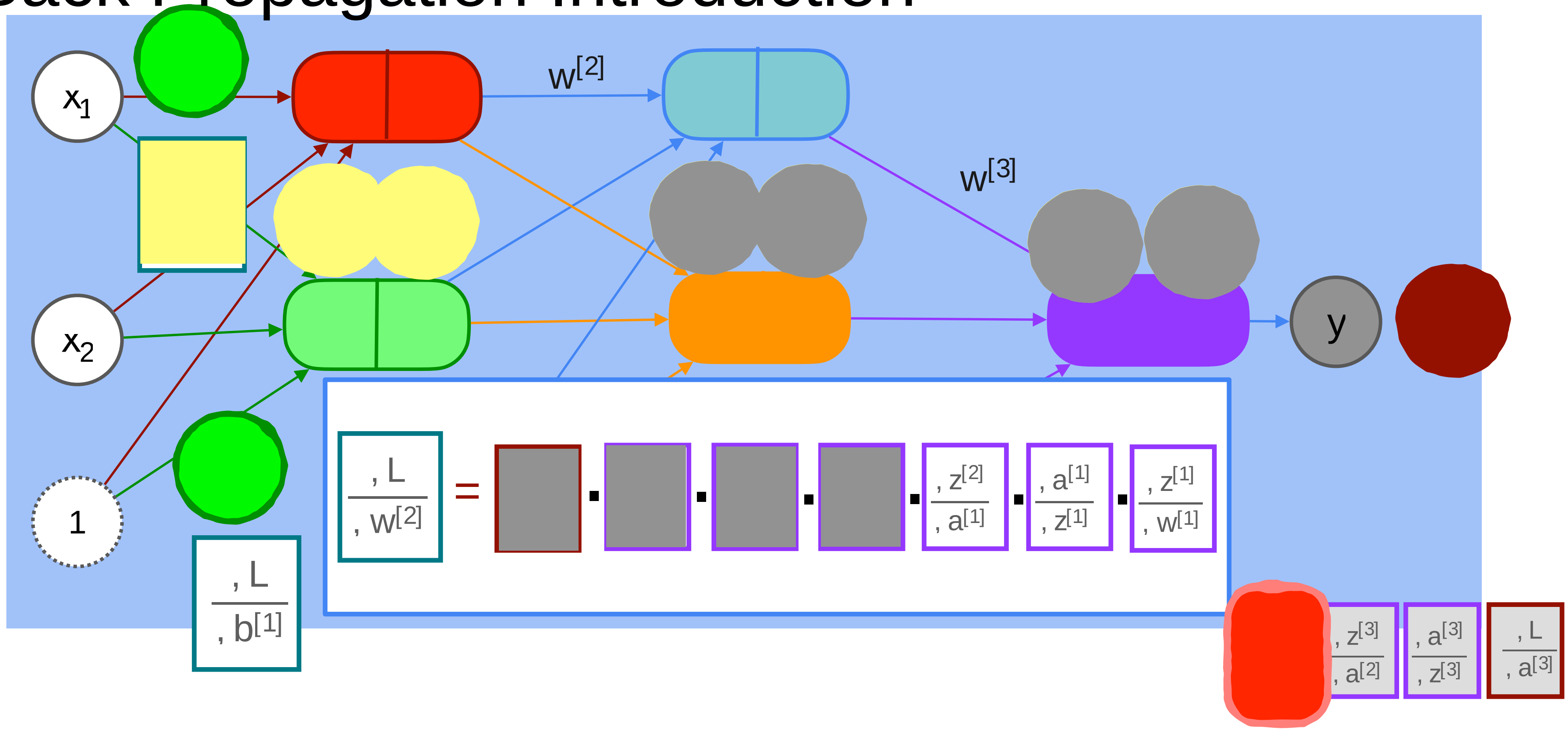
Back Propagation Introduction



Back Propagation Introduction



Back Propagation Introduction



Optimization in Neural Networks and NewtonÕs Method

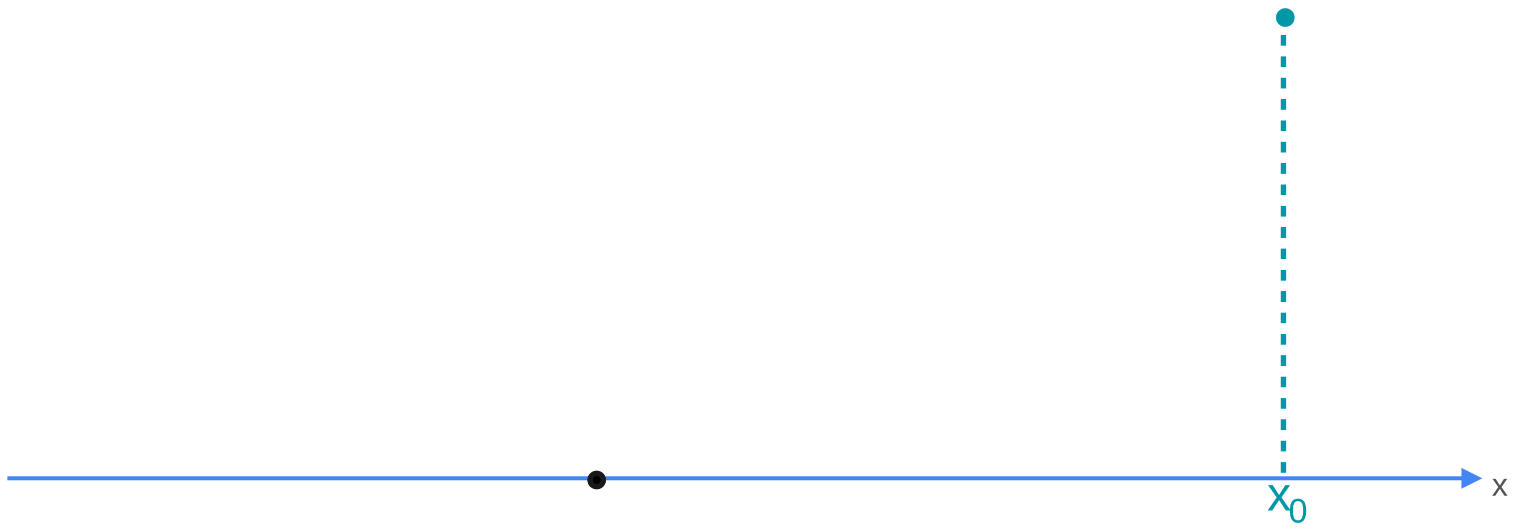
NewtonÕs method

NewtonÕs Method

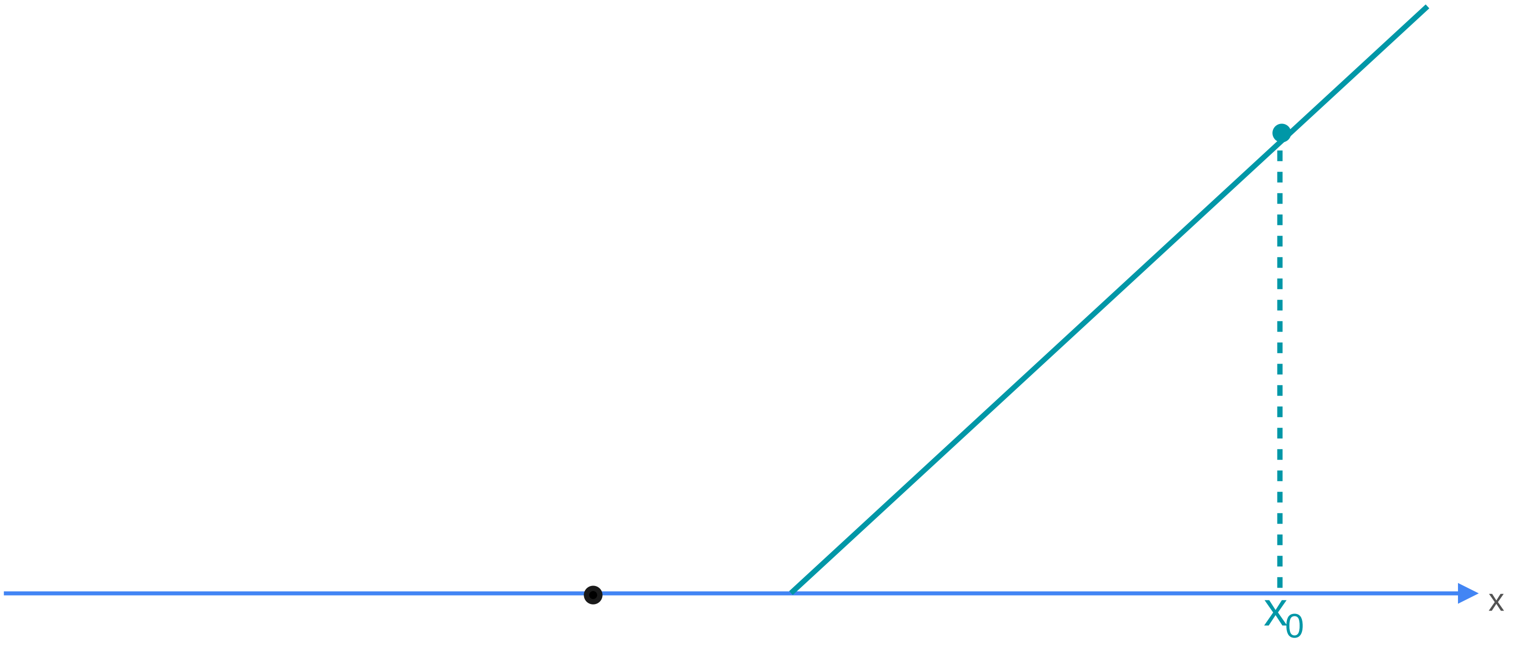
NewtonÕs Method

NewtonÕs Method

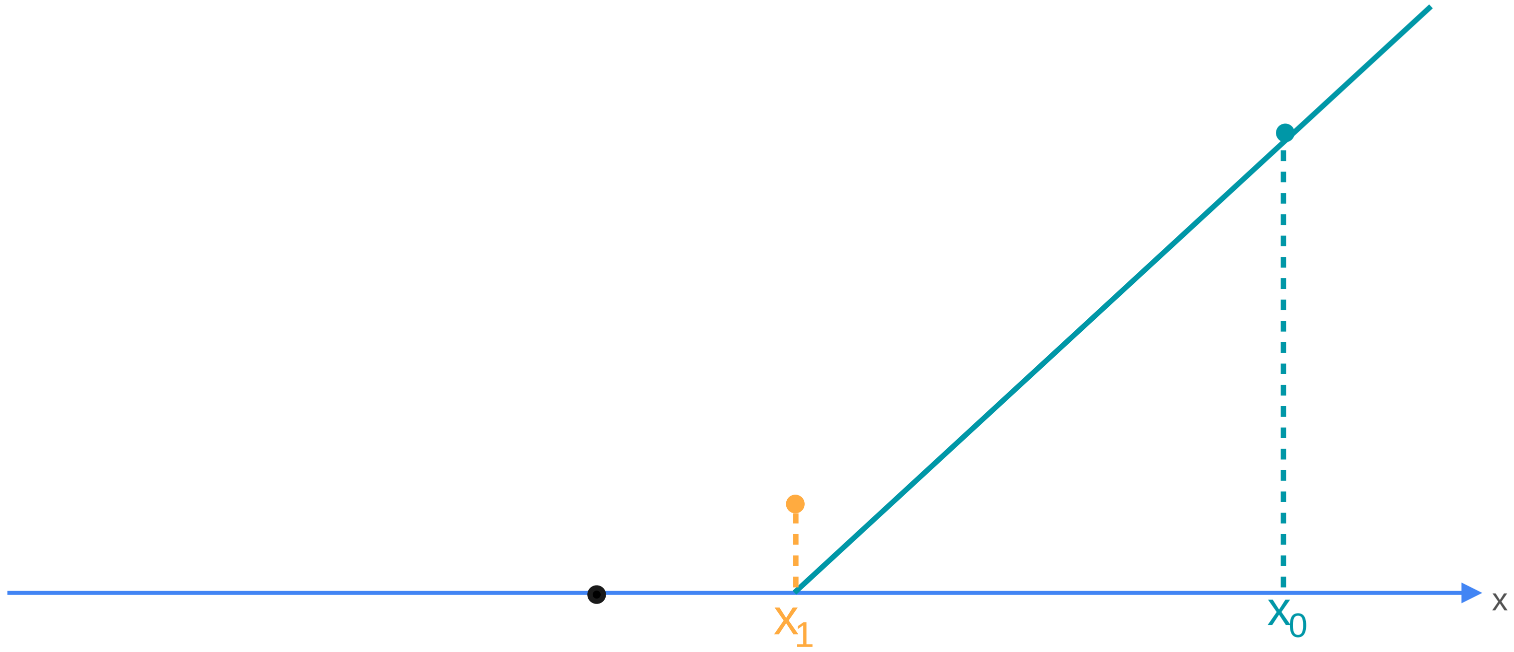
NewtonÕs Method



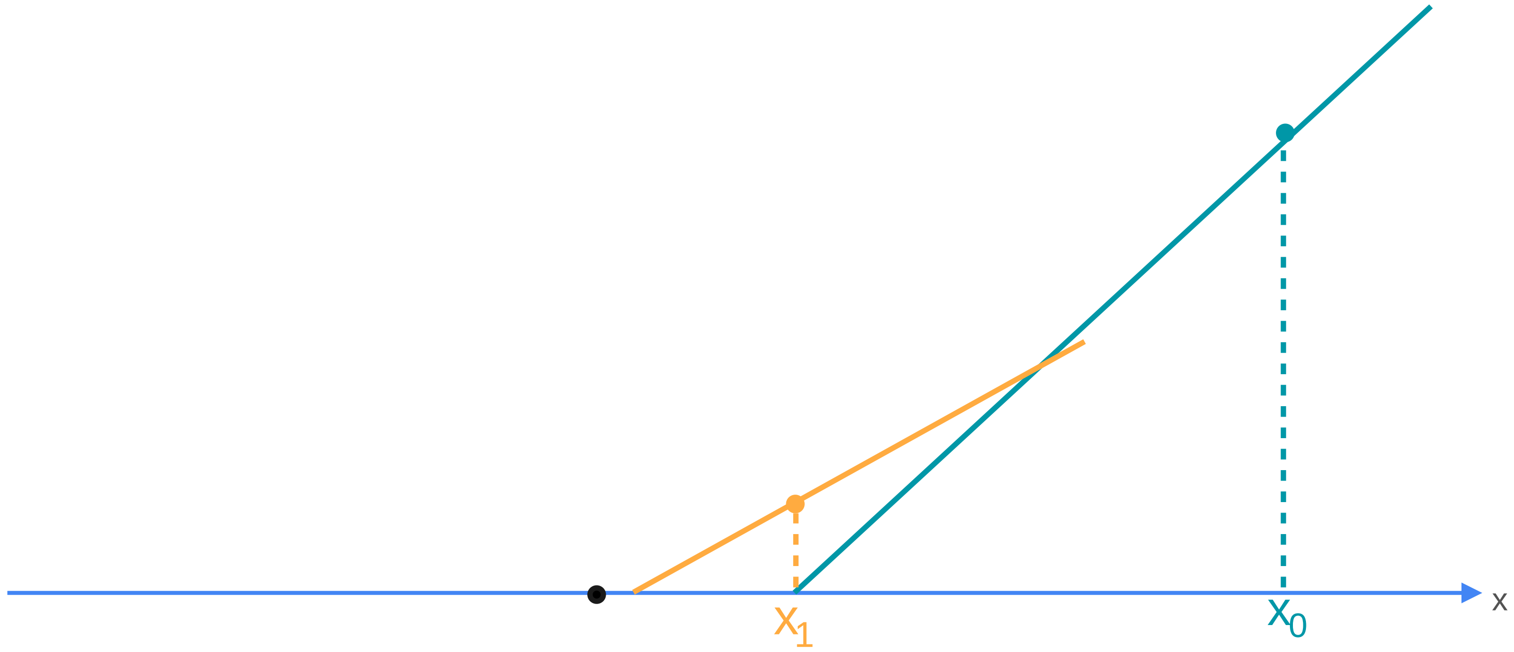
NewtonÕs Method



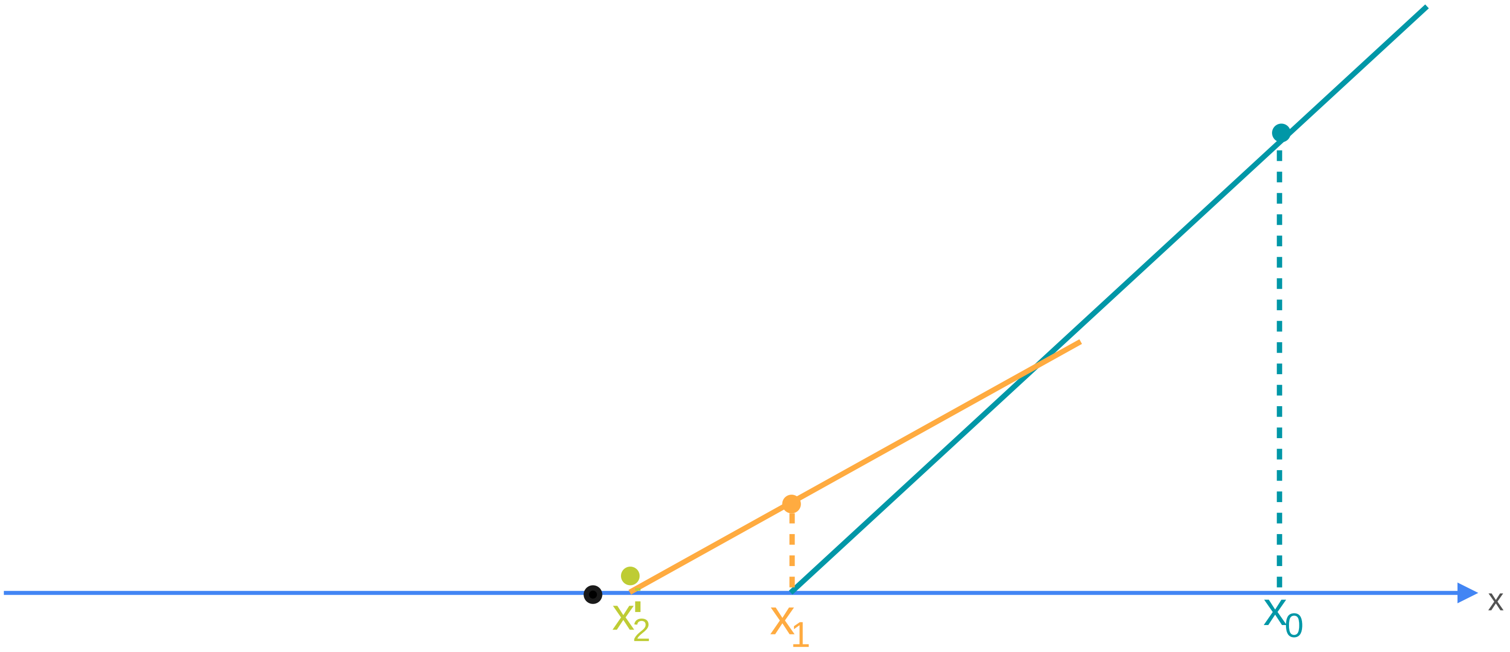
NewtonÕs Method



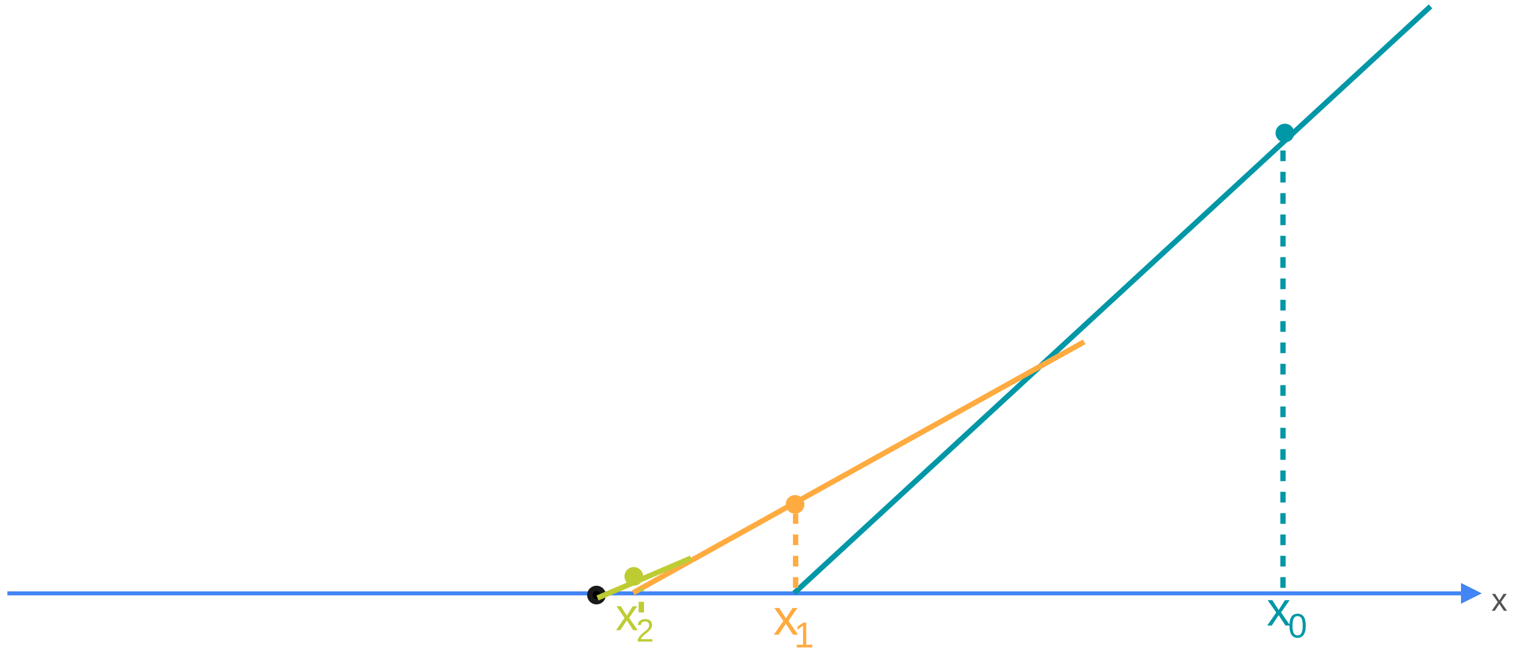
NewtonÕs Method



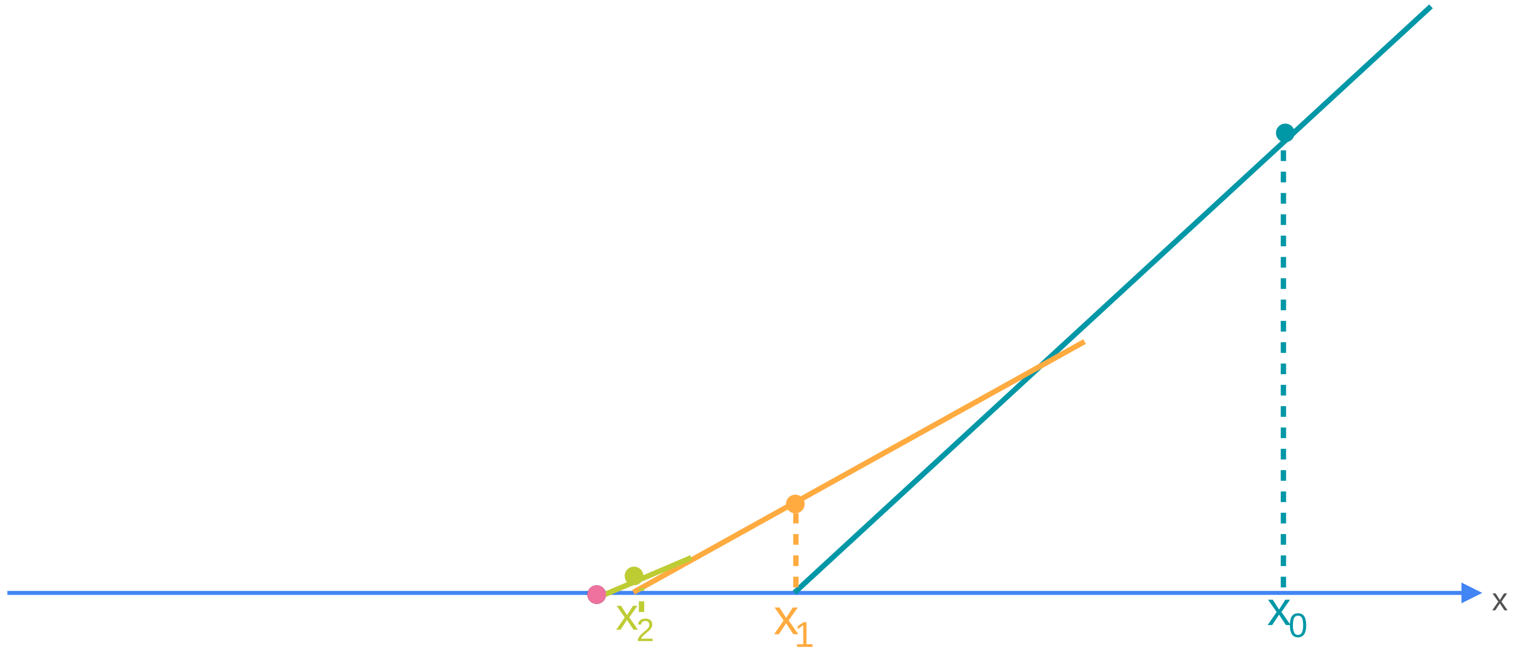
NewtonÕs Method



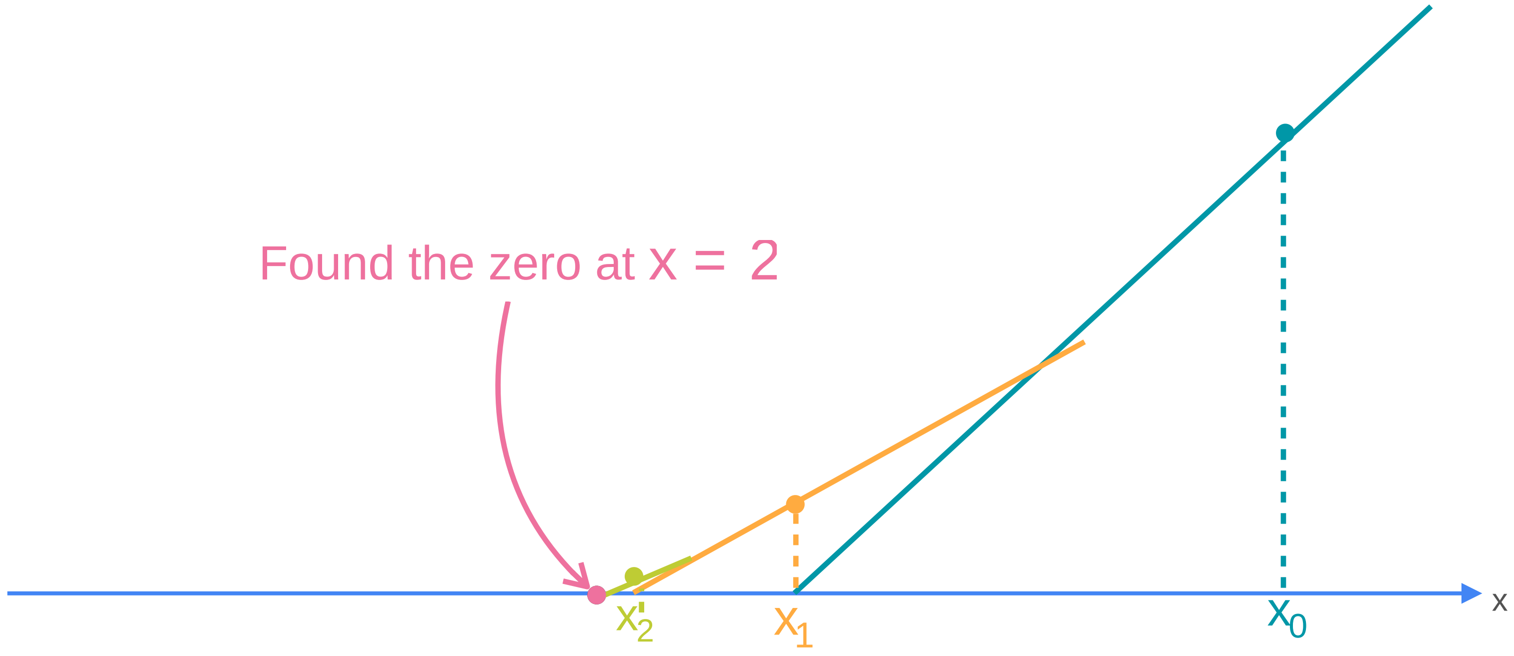
NewtonÕs Method



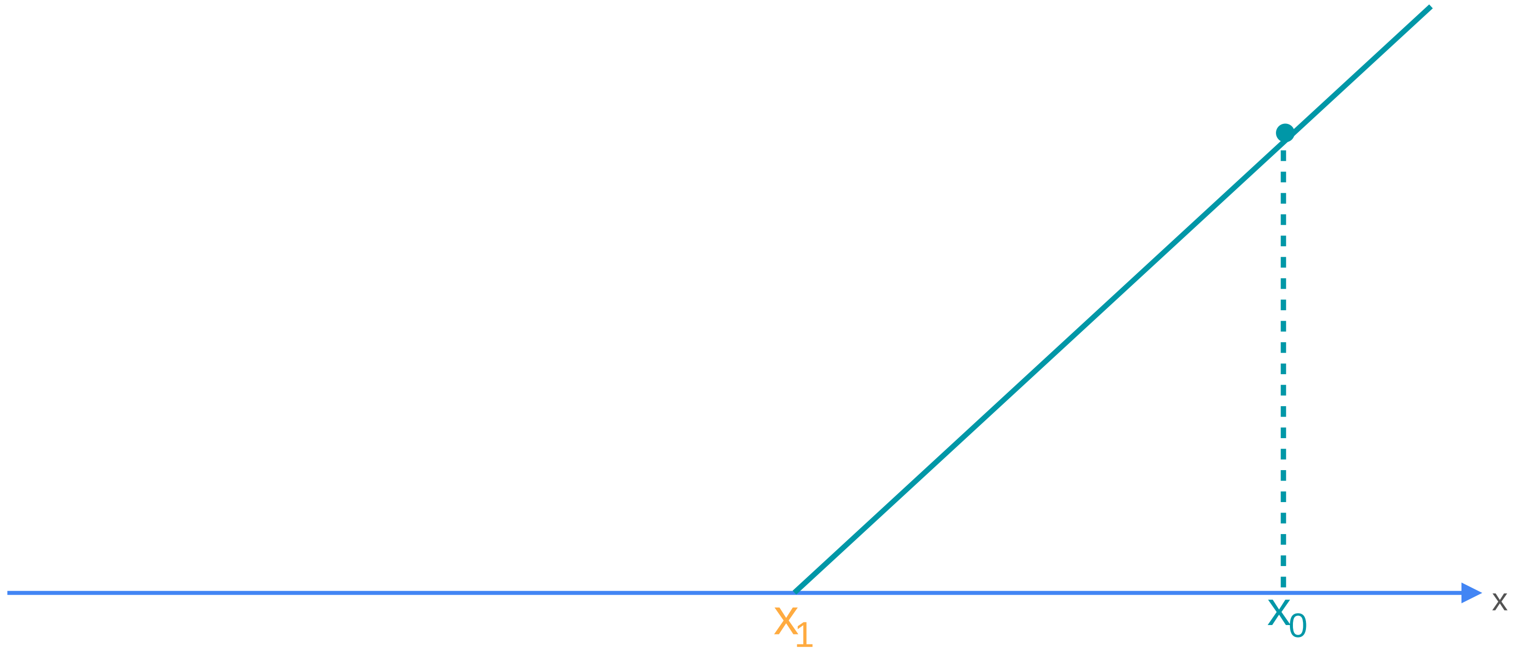
NewtonÕs Method



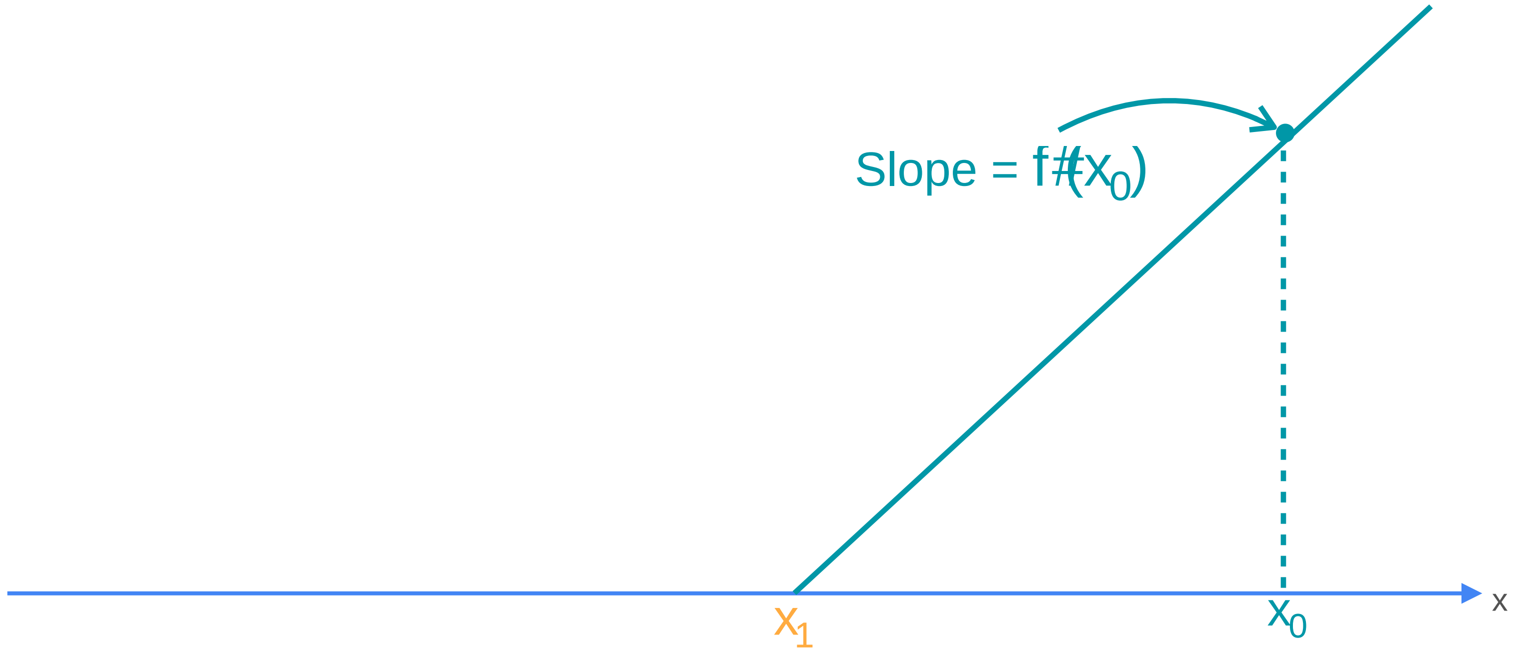
NewtonÕs Method



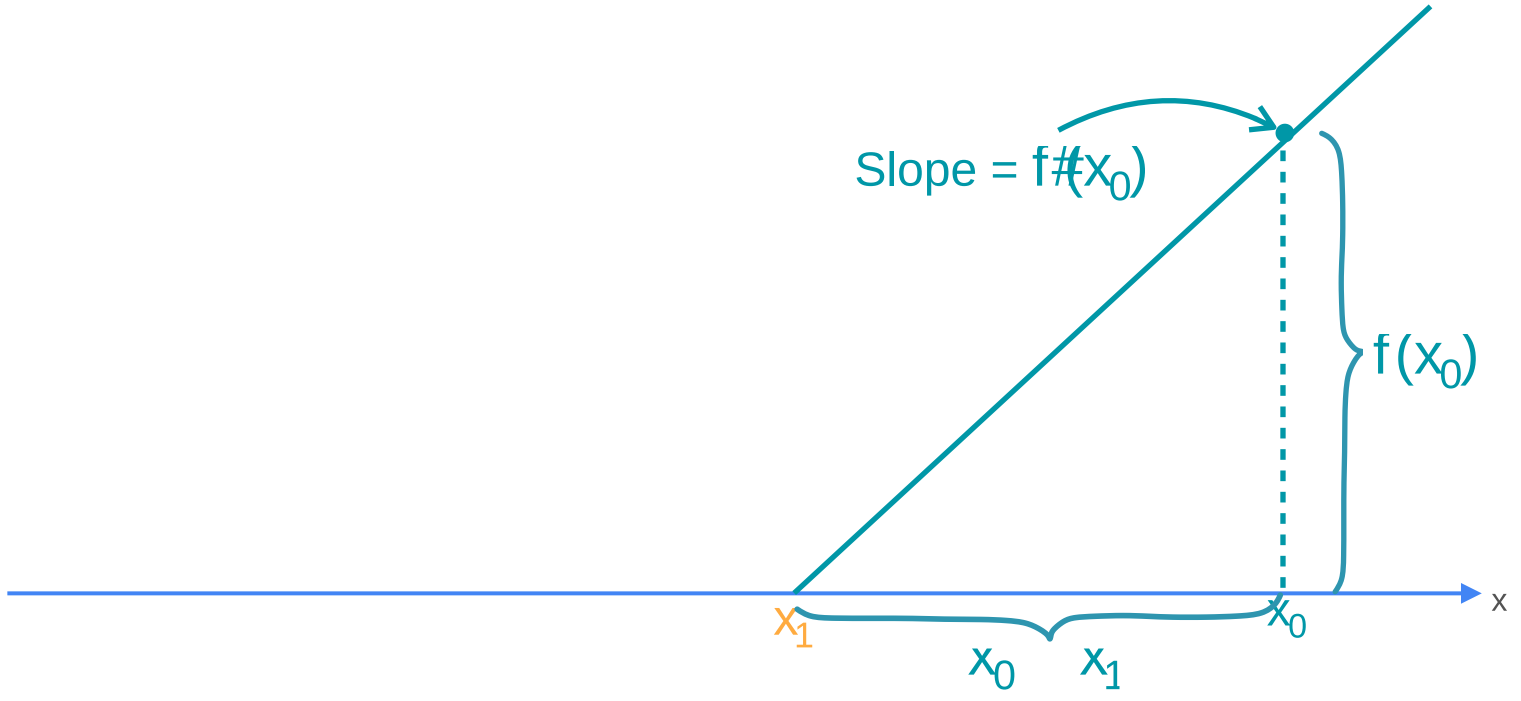
Update Approximation



Update Approximation

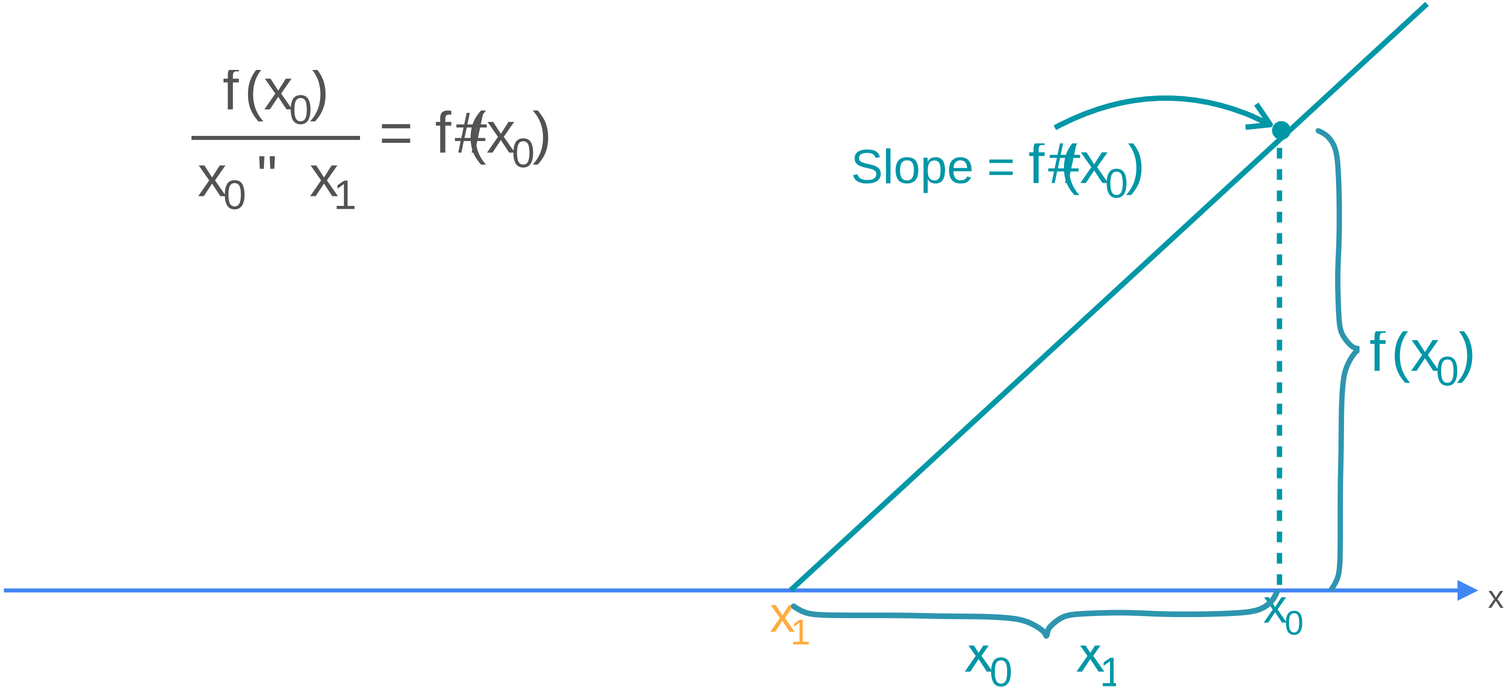


Update Approximation



Update Approximation

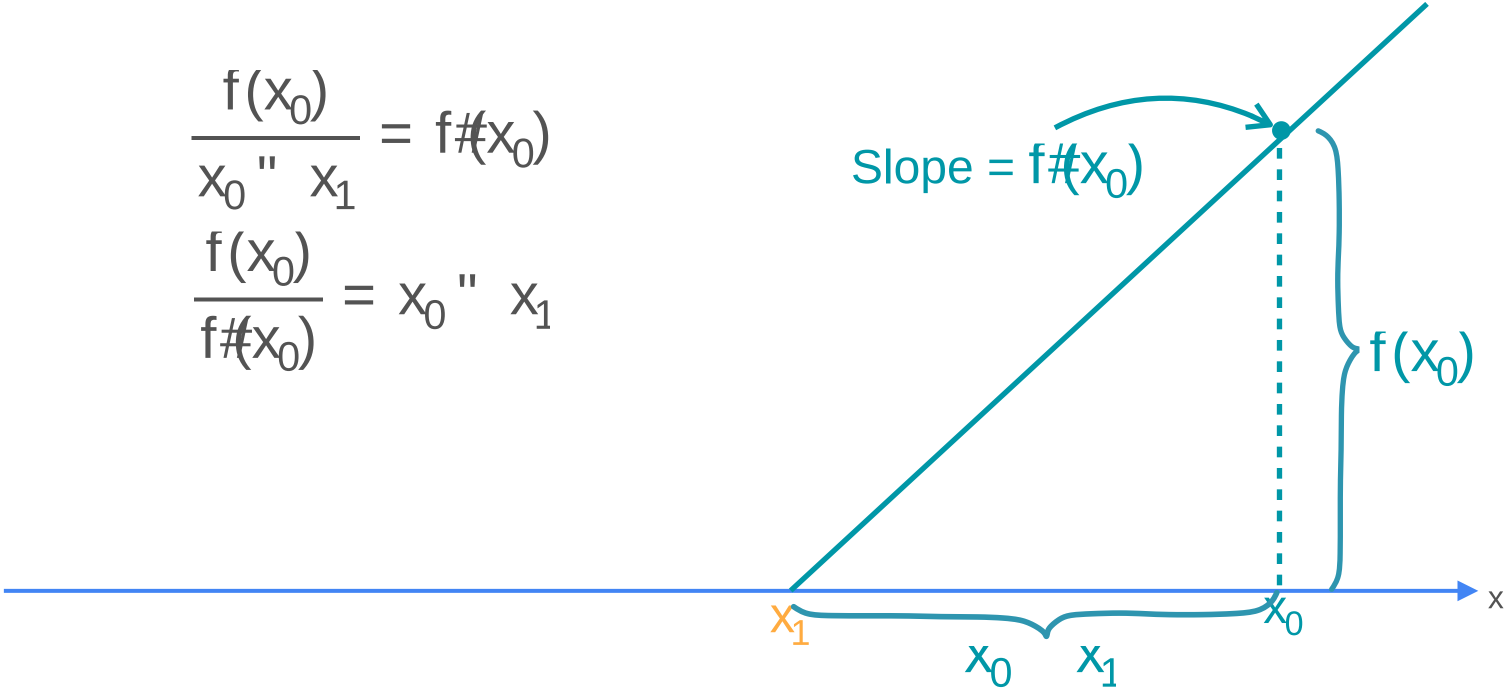
$$\frac{f(x_0)}{x_0 - x_1} = f'(x_0)$$



Update Approximation

$$\frac{f(x_0)}{x_0 - x_1} = f'(x_0)$$

$$\frac{f(x_0)}{f'(x_0)} = x_0 - x_1$$

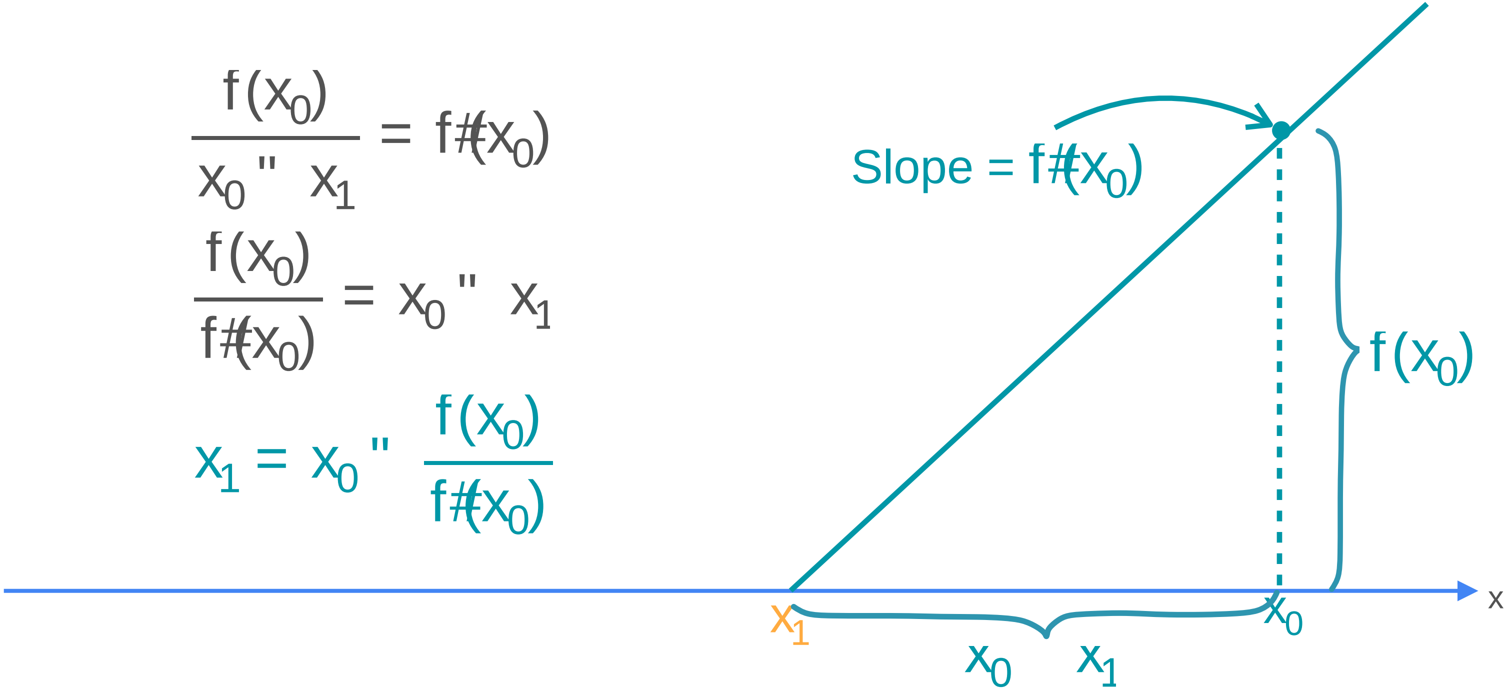


Update Approximation

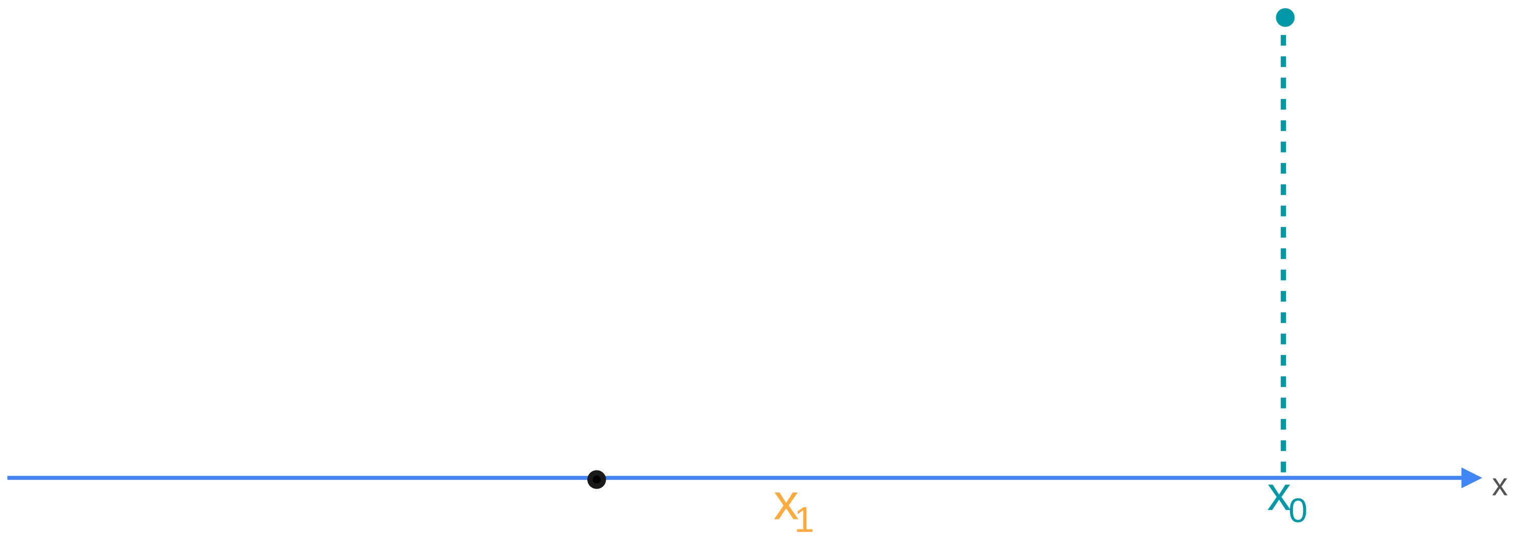
$$\frac{f(x_0)}{x_0 - x_1} = f'(x_0)$$

$$\frac{f(x_0)}{f'(x_0)} = x_0 - x_1$$

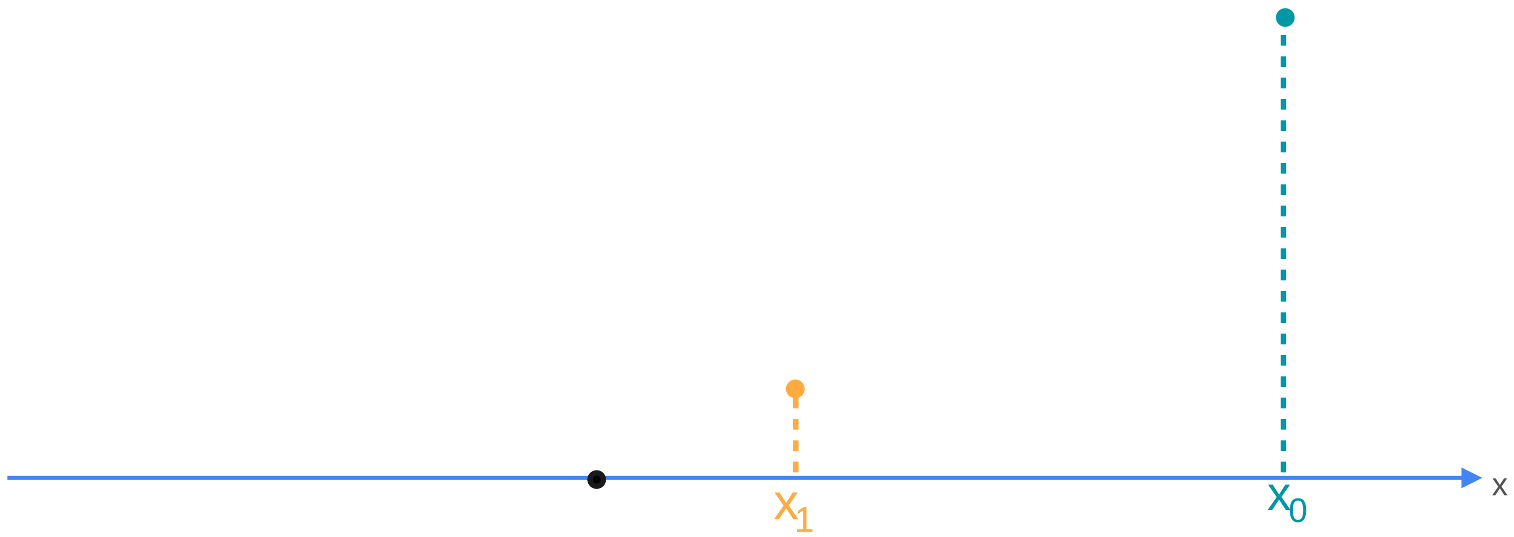
$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$



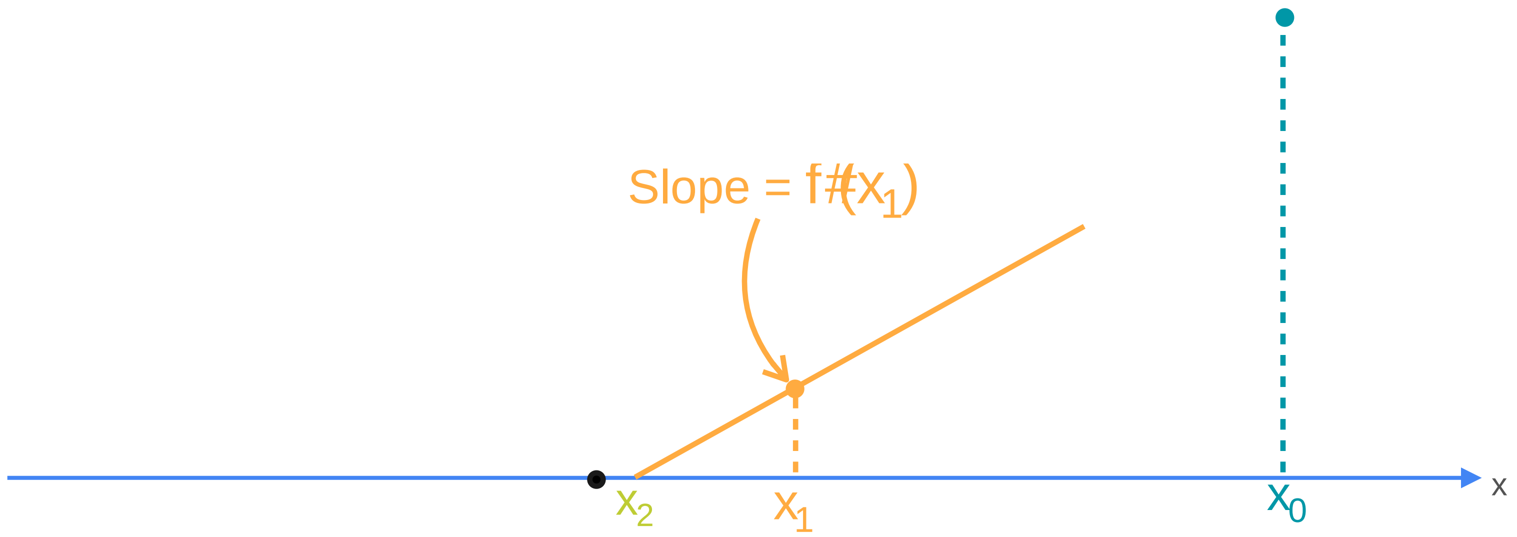
Update Approximation



Update Approximation

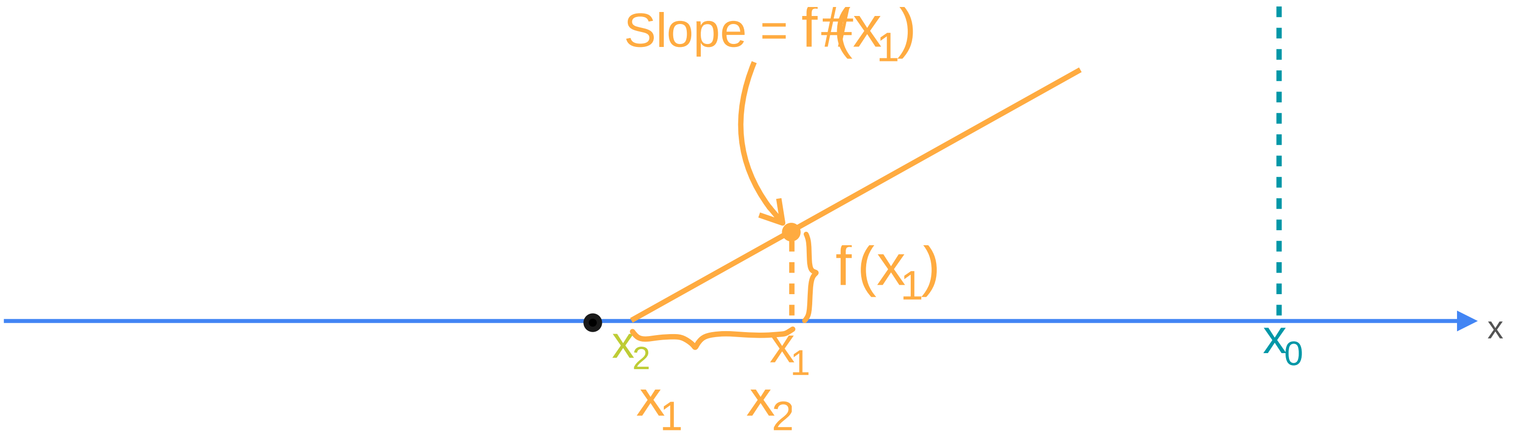


Update Approximation



Update Approximation

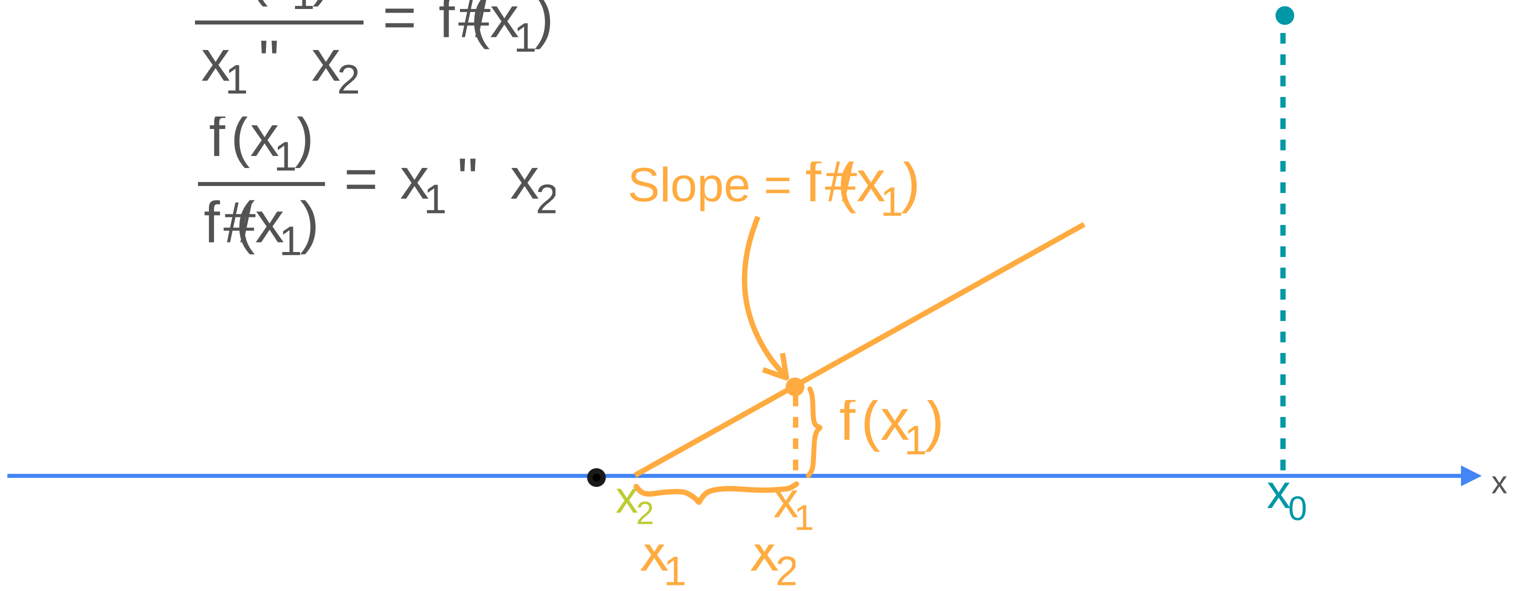
$$\frac{f(x_1)}{x_1 - x_2} = f'(x_1)$$



Update Approximation

$$\frac{f(x_1)}{x_1 - x_2} = f'(x_1)$$

$$\frac{f(x_1)}{f'(x_1)} = x_1 - x_2$$

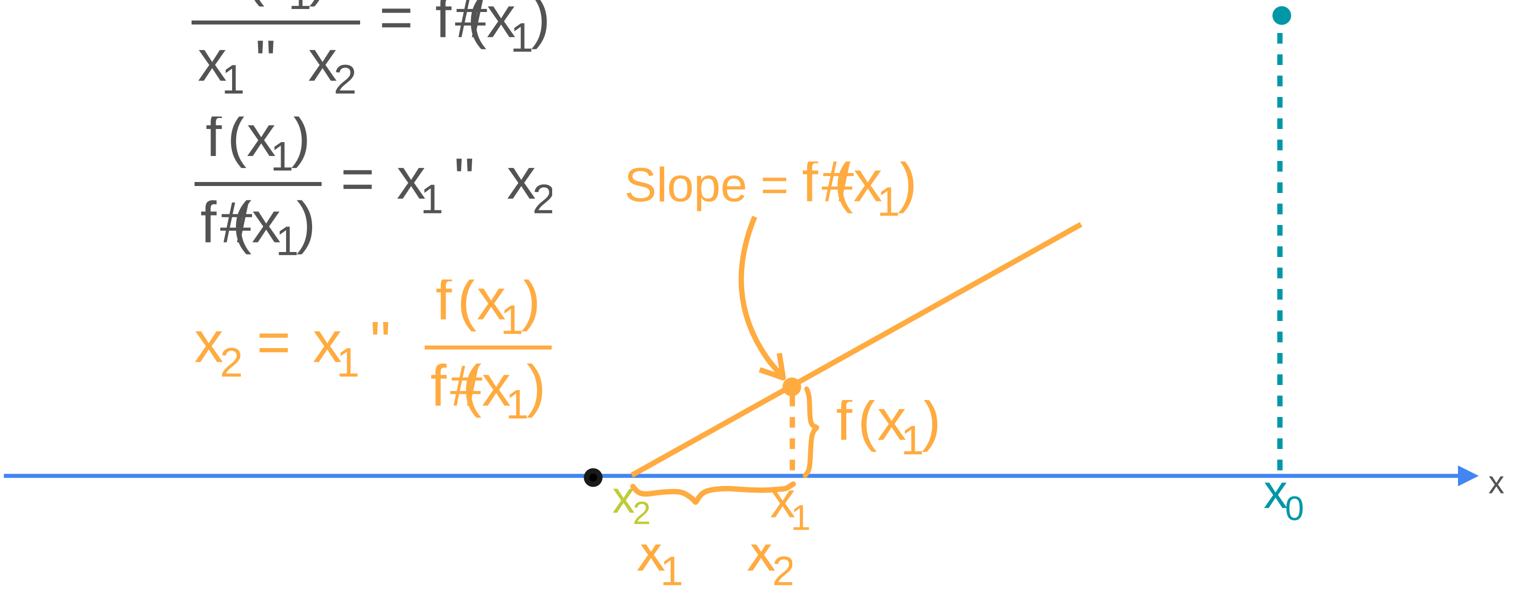


Update Approximation

$$\frac{f(x_1)}{x_1 - x_2} = f'(x_1)$$

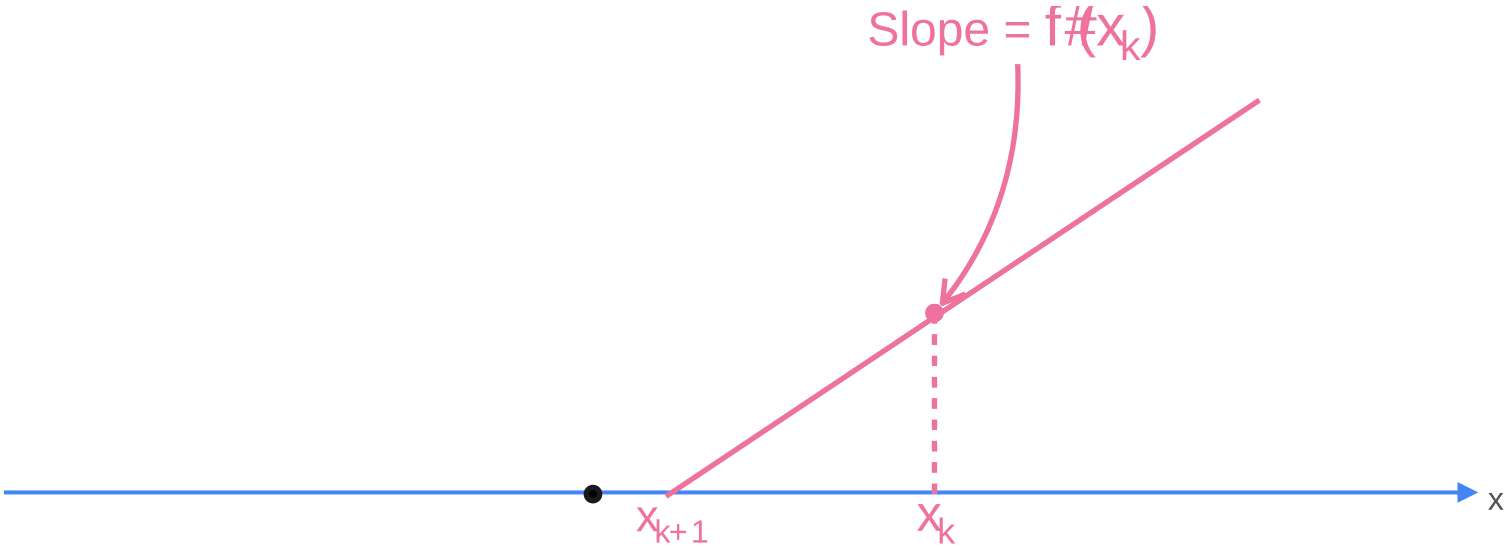
$$\frac{f(x_1)}{f'(x_1)} = x_1 - x_2$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

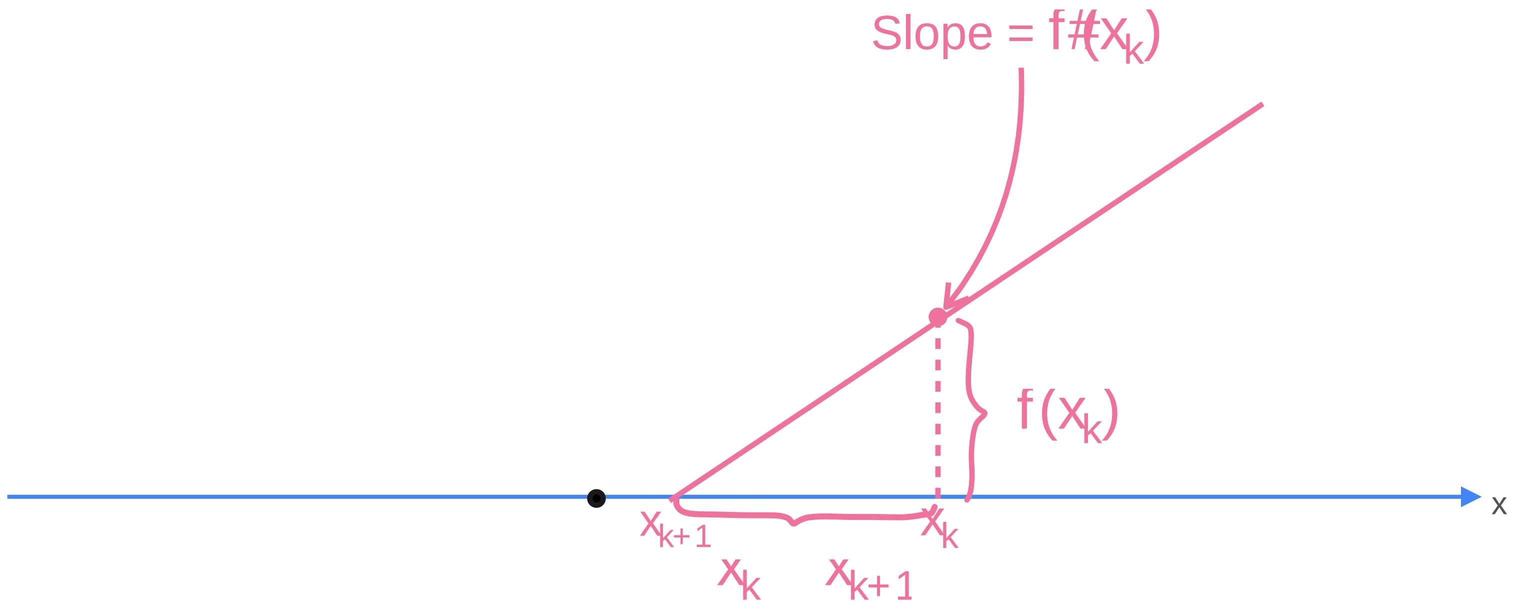


Update Approximation

Update Approximation

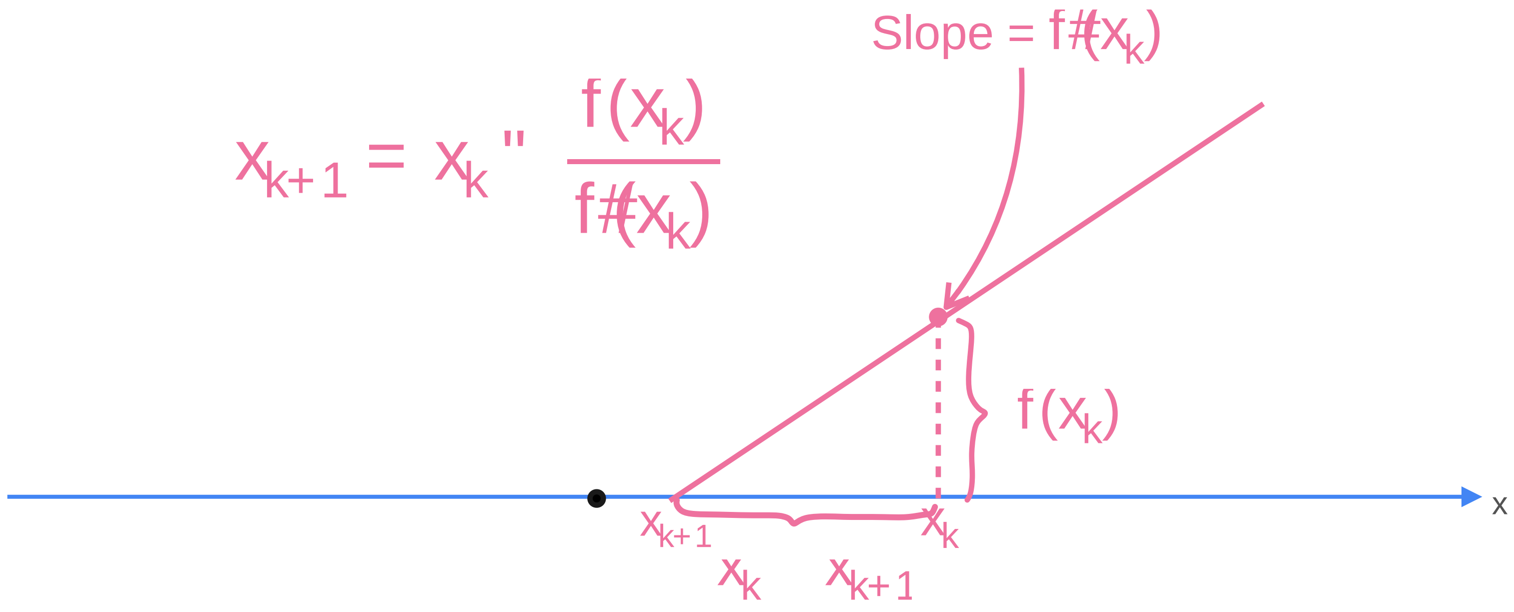


Update Approximation



Update Approximation

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$



NewtonÕs Method for Optimization

NewtonÕs Method for Optimization

NewtonÕs Method for Optimization

NewtonÕs method

Goal: find a zero of $f(x)$

NewtonÕs Method for Optimization

NewtonÕs method

Goal: find a zero of $f(x)$

NM for Optimization

Goal: minimize $g(x)$ $\Rightarrow$ find zeros of g#(x)

NewtonÕs Method for Optimization

NewtonÕs method

Goal: find a zero of f (x)

NM for Optimization

Goal: minimize g(x) find zeros of g#(x)

f (x) / g#(x)   f#(x) / (g#(x))#

NewtonÕs Method for Optimization

NewtonÕs method

Goal: find a zero of f (x)

1) Start with some x_0

NM for Optimization

Goal: minimize g(x) ⇒ find zeros of g#(x)

f (x) / g#(x) f#(x) / (g#(x))#

1) Start with some x_0

NewtonÕs Method for Optimization

NewtonÕs method

Goal: find a zero of $f(x)$

1) Start with some x_0

2) Update:

$$x_{k+1} = x_k \text{"} \frac{f(x_k)}{f\#(x_k)}$$

NM for Optimization

Goal: minimize $g(x)$ $\Rightarrow$ find zeros of $g\#(x)$

$f(x)$ / $g\#(x)$ $\quad$ $f\#(x)$ / $(g\#(x))\#$

1) Start with some x_0

2) Update:

NewtonÕs Method for Optimization

NewtonÕs method

Goal: find a zero of $f(x)$

1) Start with some x_0

2) Update:

$$x_{k+1} = x_k \text{"} \frac{f(x_k)}{f\#(x_k)}$$

NM for Optimization

Goal: minimize $g(x)$ $\Rightarrow$ find zeros of $g\#(x)$

$f(x)$ / $g\#(x)$ $\quad$ $f\#(x)$ / $(g\#(x))\#$

1) Start with some x_0

2) Update:

$$x_{k+1} = x_k \text{"} \frac{g\#(x_k)}{(g\#(x_k))\#}$$

NewtonÕs Method for Optimization

NewtonÕs method

Goal: find a zero of $f(x)$

1) Start with some x_0

2) Update:

$$x_{k+1} = x_k \text{ " } \frac{f(x_k)}{f\#(x_k)}$$

3) Repeat 2) until you find the root.

NM for Optimization

Goal: minimize $g(x) \Longrightarrow$ find zeros of $g\#(x)$

$f(x)\ /\ g\#(x) \qquad f\#(x)\ /\ (g\#(x))\#$

1) Start with some x_0

2) Update:

$$x_{k+1} = x_k \text{ " } \frac{g\#(x_k)}{(g\#(x_k))\#}$$

3) Repeat 2) until you find the root.

Optimization in Neural Networks and NewtonÕs Method

NewtonÕs method: An example

NewtonÕs Method for Optimization

NewtonÕs Method for Optimization

$$g(x) = e^{x} \text{ " } \log(x)$$

NewtonÕs Method for Optimization

$g(x) = e^{x}$ " $\log(x)$ $g\#(x) = e^{x}$ " $1/x$

NewtonÕs Method for Optimization

$g(x) = e^{x}$ " $\log(x)$ $\quad$ $g\#(x) = e^{x}$ " $1/x$

Minimum: $x^{*} = 0.567$

NewtonÕs Method for Optimization

$g(x) = e^x - \log(x) \qquad g'(x) = e^x - 1/x$

Minimum: $x^* = 0.567$

NewtonÕs Method for Optimization

$$g(x) = e^x - \log(x) \qquad g'(x) = e^x - 1/x$$

Minimum: $x^* = 0.567$

$$(g'(x))' = e^x + \frac{1}{x^2}$$

NewtonÕs Method for Optimization

$$g(x) = e^x \text{ " } \log(x) \qquad \overbrace{g\#(x) = e^x \text{ " } 1/x}^{f(x)}$$

Minimum: $x^* = 0.567$

$$\underbrace{(g\#(x))\# = e^x + \frac{1}{x^2}}_{f\#(x)}$$

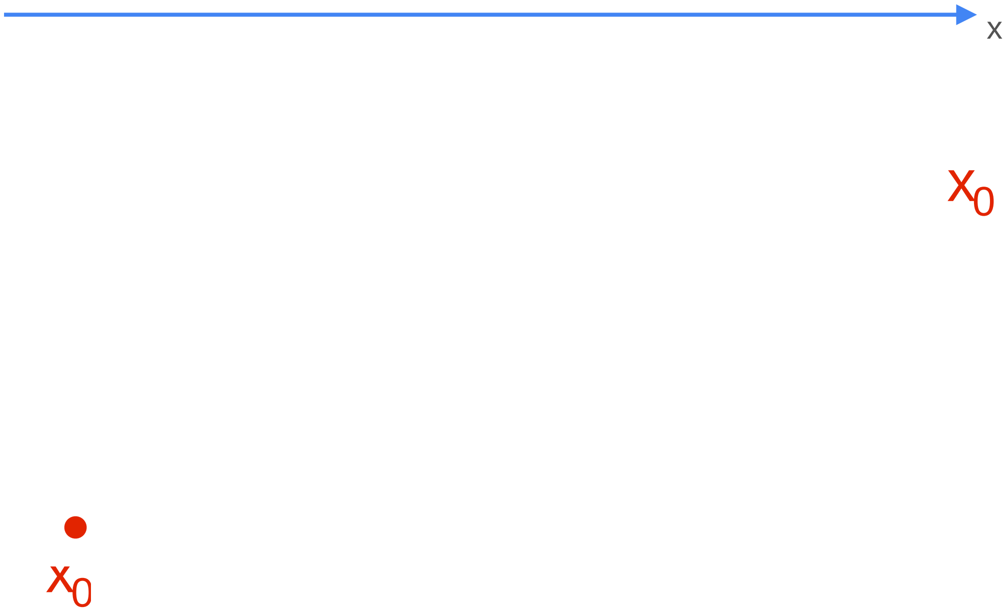
NewtonÕs Method for Optimization

$g(x) = e^x - \log(x)$ $\quad \overbrace{g'(x) = e^x - 1/x}^{f(x)}$

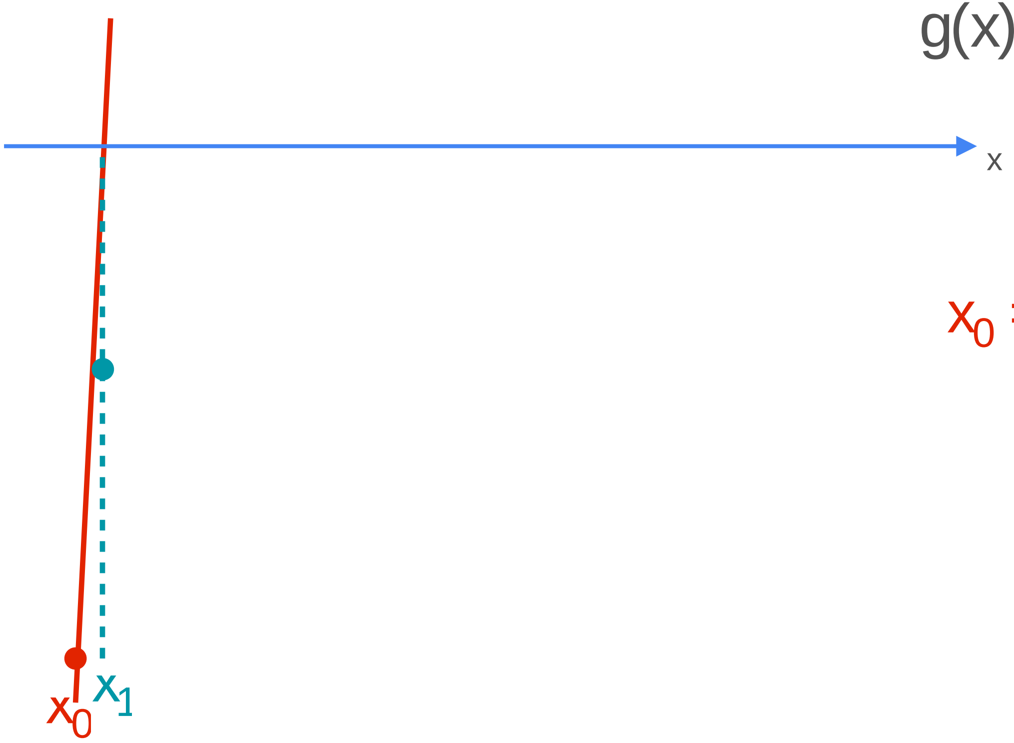
Minimum: $x^* = 0.567$

$\underbrace{(g'(x))' = e^x + \frac{1}{x^2}}_{f'(x)}$

$x_0 = 0.05$



NewtonÕs Method for Optimization



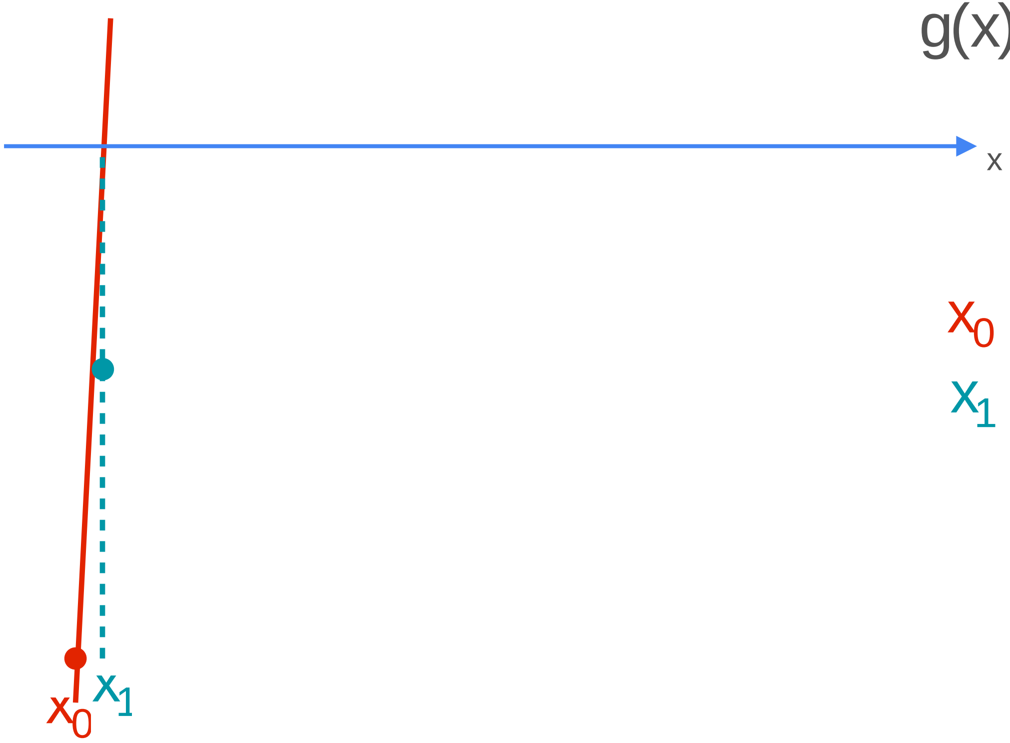
$$g(x) = e^x \text{"}\ \log(x) \qquad \overbrace{g\#(x) = e^x \text{"}\ 1/x}^{f(x)}$$

Minimum: $x^* = 0.567$

$$\underbrace{(g\#(x))\# = e^x + \frac{1}{x^2}}_{f\#(x)}$$

$x_0 = 0.05$

NewtonÕs Method for Optimization



$$g(x) = e^x \text{"} \ \log(x) \qquad \overbrace{g\#(x) = e^x \text{"} \ 1/x}^{f(x)}$$

Minimum: $x^* = 0.567$

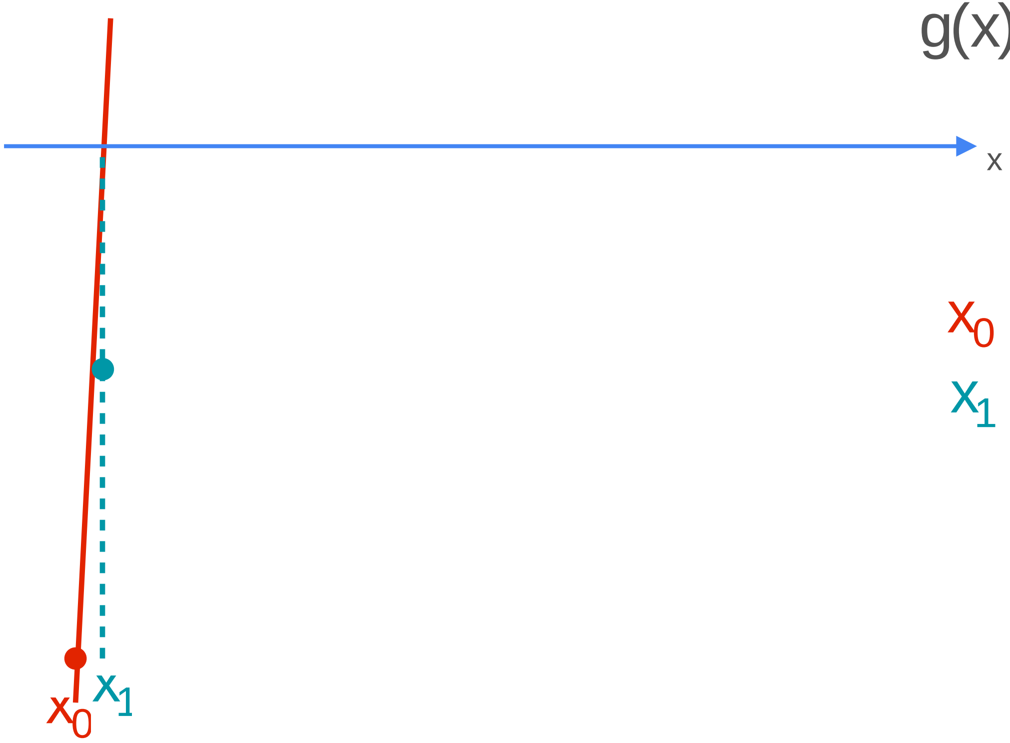
$$\underbrace{(g\#(x))\# = e^x + \frac{1}{x^2}}_{f\#(x)}$$

$x_0 = 0.05$

$$x_1 = x_0 \text{"} \ \frac{g\#(x_0)}{(g\#(x_0))\#}$$

$$= 0.05 \text{"} \ \frac{\left(e^{0.05} \text{"} \ \frac{1}{0.05}\right)}{\left(e^{0.05} + \frac{1}{0.05^2}\right)}$$

NewtonÕs Method for Optimization



$$g(x) = e^x \text{"} \log(x) \qquad \overbrace{g\#(x) = e^x \text{"} 1/x}^{f(x)}$$

Minimum: $x^* = 0.567$

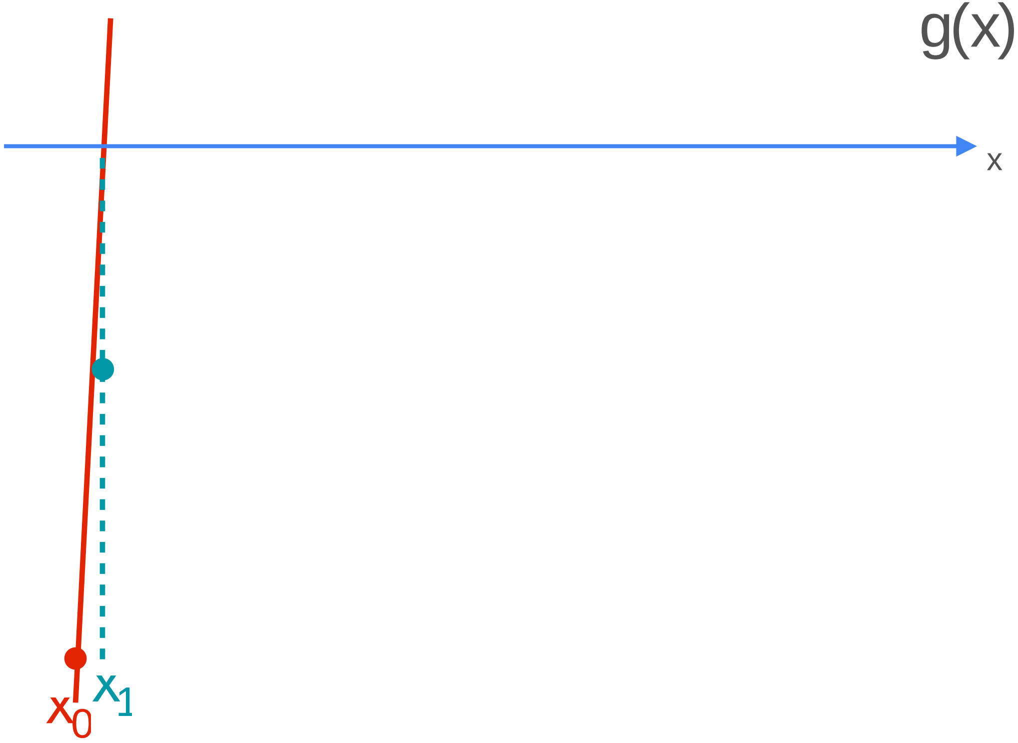
$$\underbrace{(g\#(x))\# = e^x + \frac{1}{x^2}}_{f\#(x)}$$

$x_0 = 0.05$

$$x_1 = x_0 \text{"} \frac{g\#(x_0)}{(g\#(x_0))\#}$$

$$= 0.05 \text{"} \frac{(e^{0.05} \text{"} \frac{1}{0.05})}{(e^{0.05} + \frac{1}{0.05^2})} = 0.097$$

NewtonÕs Method for Optimization

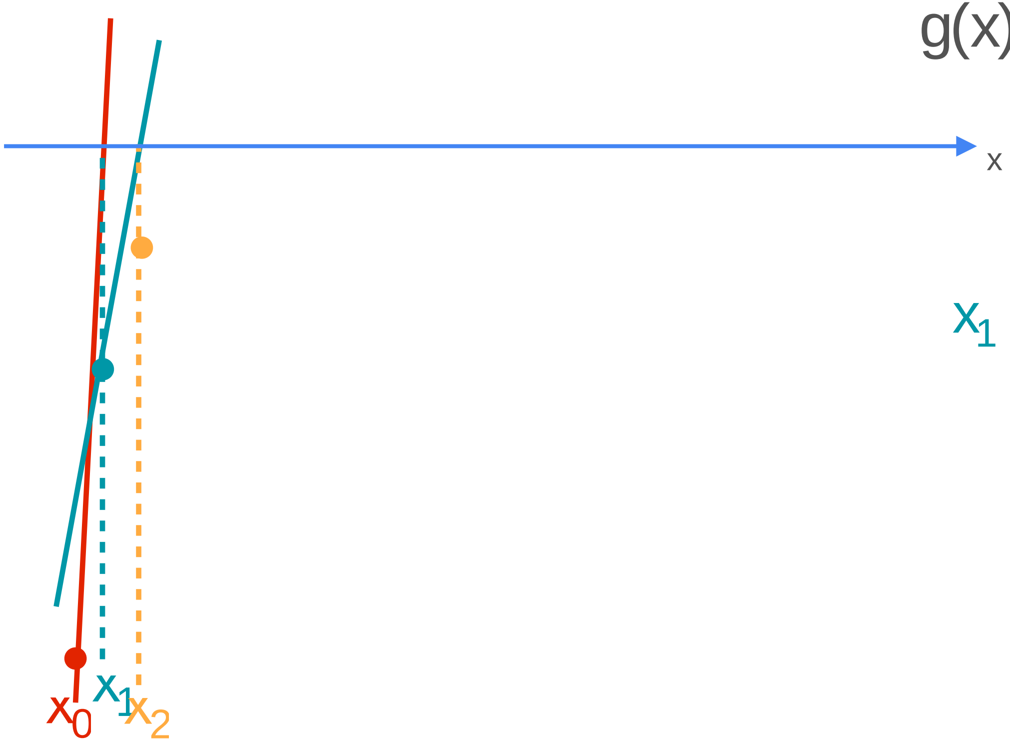


$g(x) = e^x \text{ " } \log(x)$ $g\#(x) = e^x \text{ " } 1/x$

Minimum: $x^* = 0.567$

$$(g\#(x))\# = e^x + \frac{1}{x^2}$$

NewtonÕs Method for Optimization



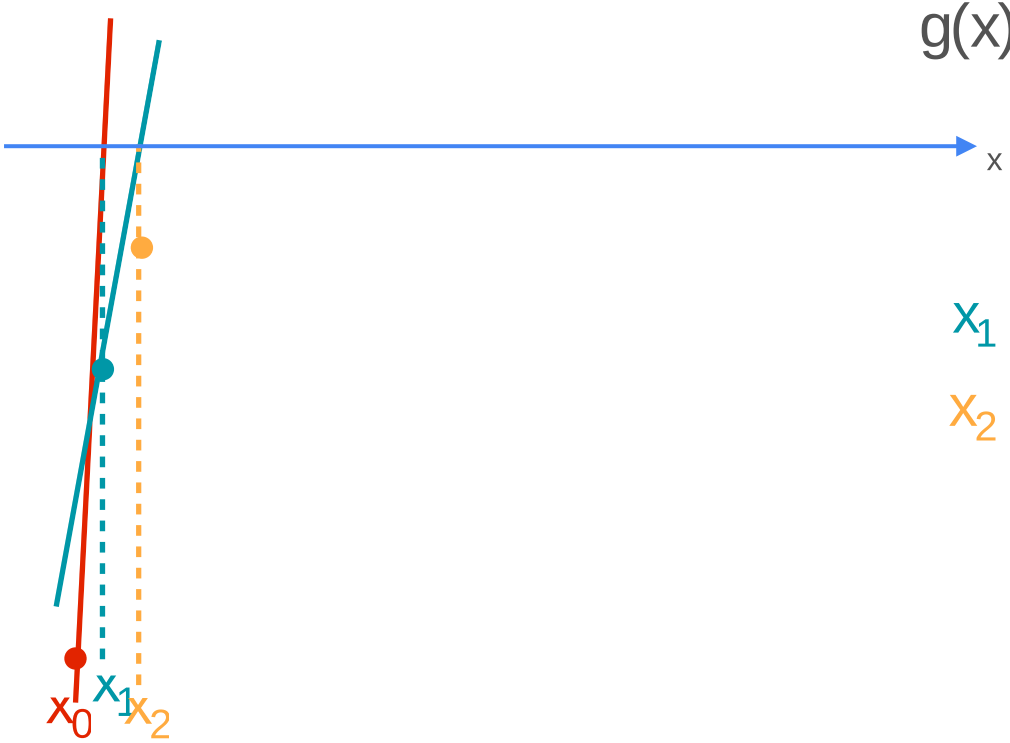
$g(x) = e^x - \log(x)$ $\quad g'(x) = e^x - 1/x$

Minimum: $x^* = 0.567$

$(g'(x))' = e^x + \dfrac{1}{x^2}$

$x_1 = 0.097$

NewtonÕs Method for Optimization



$$g(x) = e^x - \log(x) \qquad g'(x) = e^x - 1/x$$

Minimum: $x^* = 0.567$

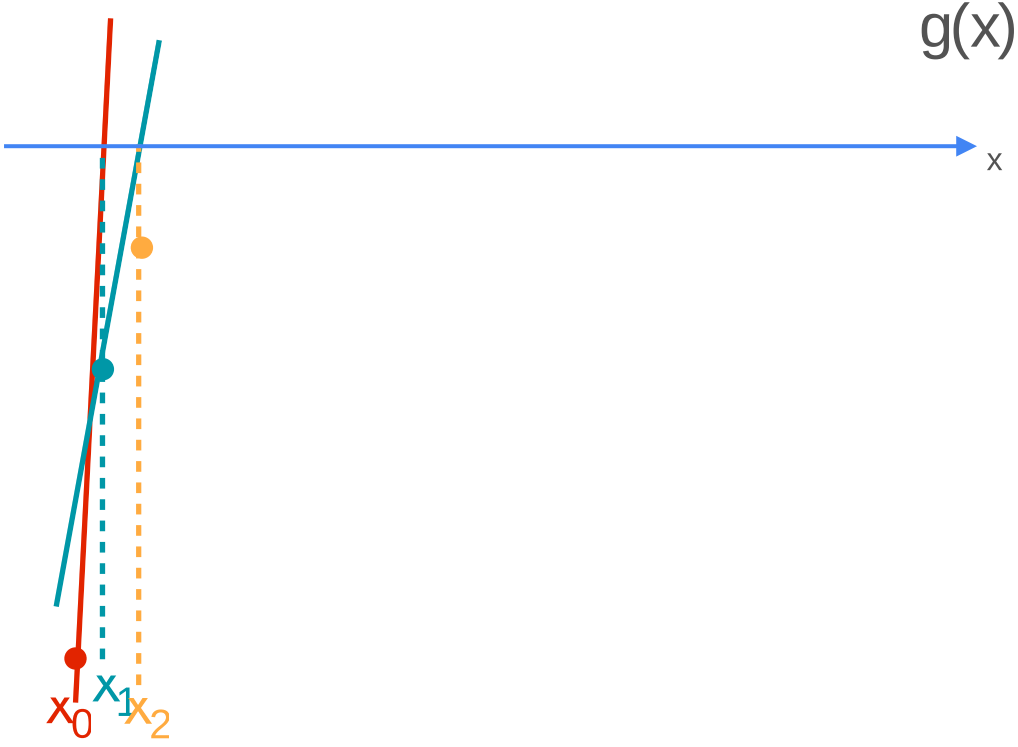
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_1 = 0.097$$

$$x_2 = x_1 - \frac{g'(x_1)}{(g'(x_1))'}$$

$$= 0.097 - \frac{\left(e^{0.097} - \frac{1}{0.097}\right)}{\left(e^{0.097} + \frac{1}{0.097^2}\right)} = 0.183$$

NewtonÕs Method for Optimization

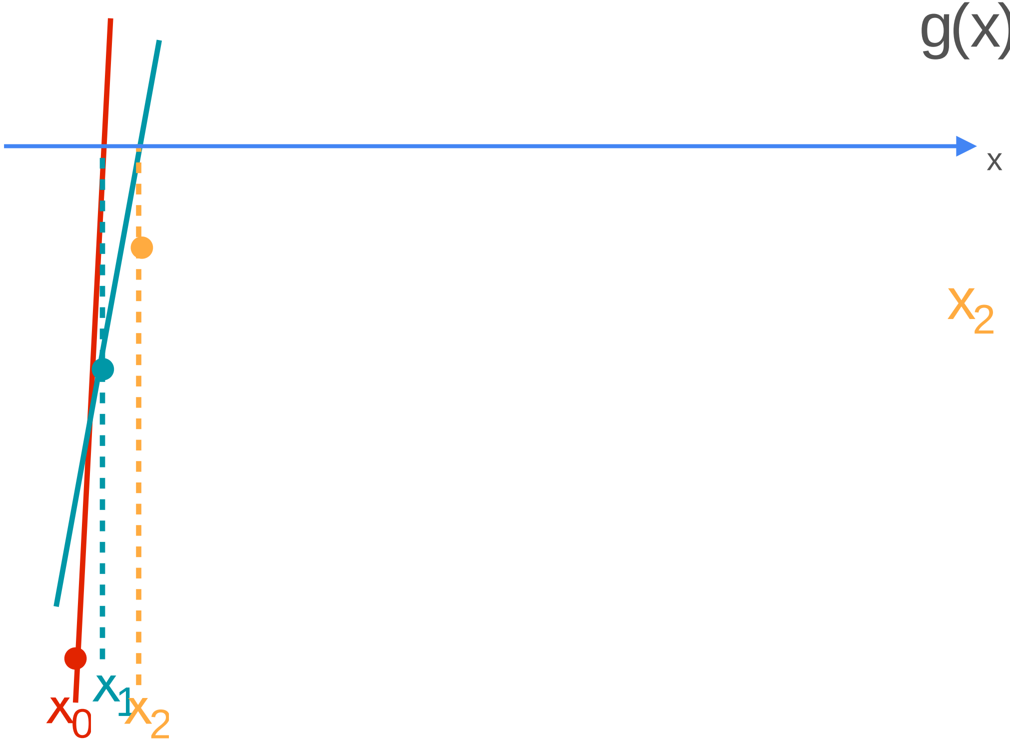


$g(x) = e^x \text{ " } \log(x)$ $\quad g\#(x) = e^x \text{ " } 1/x$

Minimum: $x^* = 0.567$:

$$(g\#(x))\# = e^x + \frac{1}{x^2}$$

NewtonÕs Method for Optimization



$g(x) = e^x$ " $\log(x)$ $g\#(x) = e^x$ " $1/x$

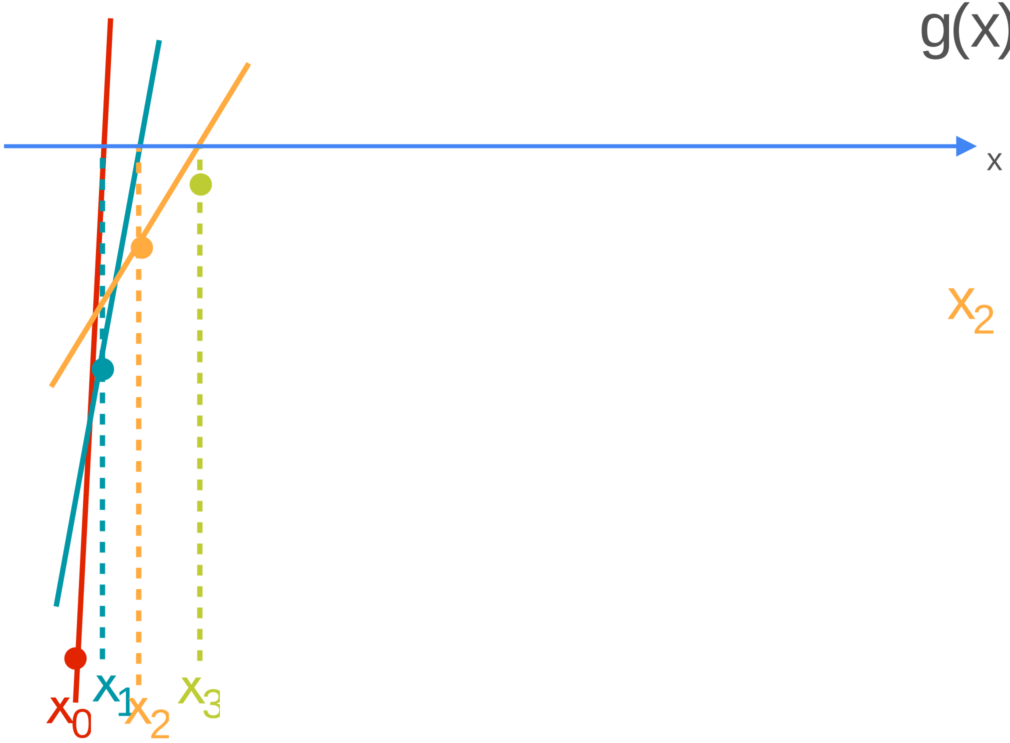
Minimum: $x^* = 0.567$

$(g\#(x))\# = e^x + \dfrac{1}{x^2}$

$x_2 = 0.183$

NewtonÕs Method for Optimization

$g(x) = e^x$ " $\log(x)$ $g\#(x) = e^x$ " $1/x$

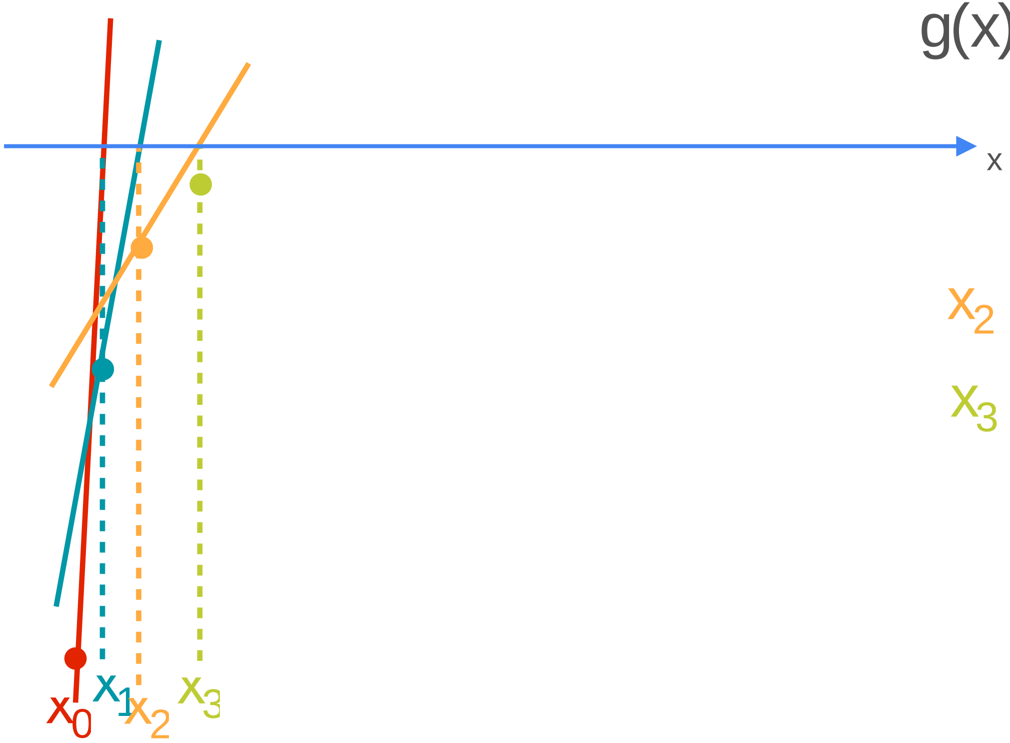


Minimum: $x^* = 0.567$

$(g\#(x))\# = e^x + \frac{1}{x^2}$

$x_2 = 0.183$

NewtonÕs Method for Optimization



$$g(x) = e^x - \log(x) \qquad g'(x) = e^x - 1/x$$

Minimum: $x^* = 0.567$

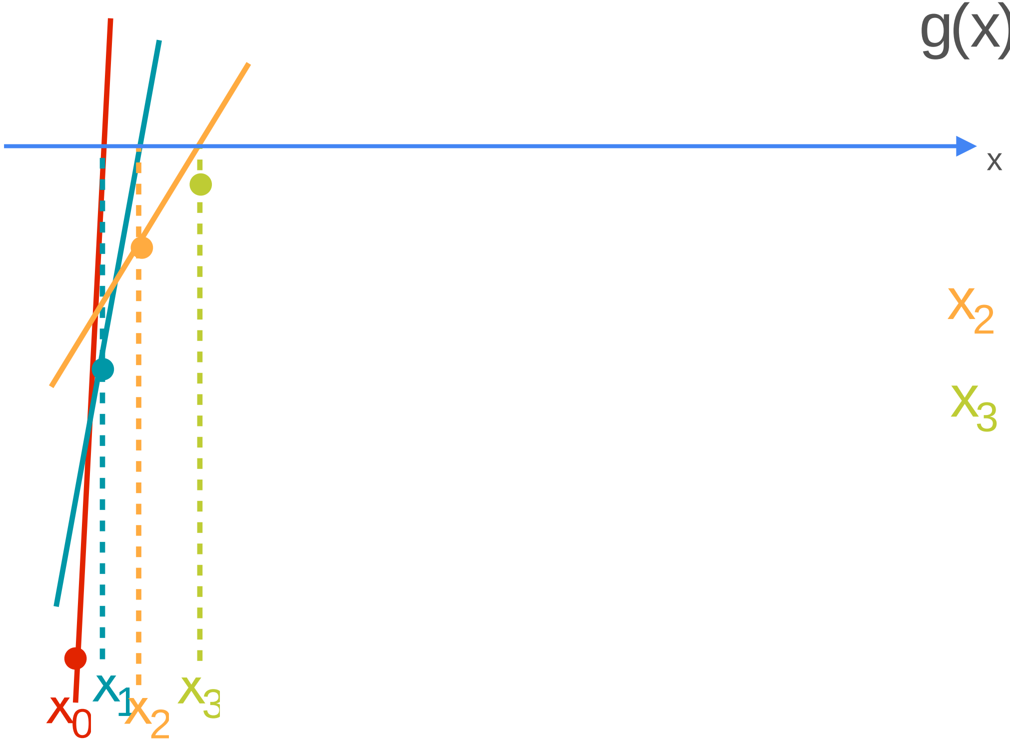
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$x_2 = 0.183$

$$x_3 = x_2 - \frac{g'(x_2)}{(g'(x_2))'}$$

$$= 0.183 - \frac{\left(e^{0.183} - \frac{1}{0.183}\right)}{\left(e^{0.183} + \frac{1}{0.183^2}\right)}$$

NewtonÕs Method for Optimization



$g(x) = e^x \text{ " } \log(x)$ $\quad g\#(x) = e^x \text{ " } 1/x$

Minimum: $x^* = 0.567$

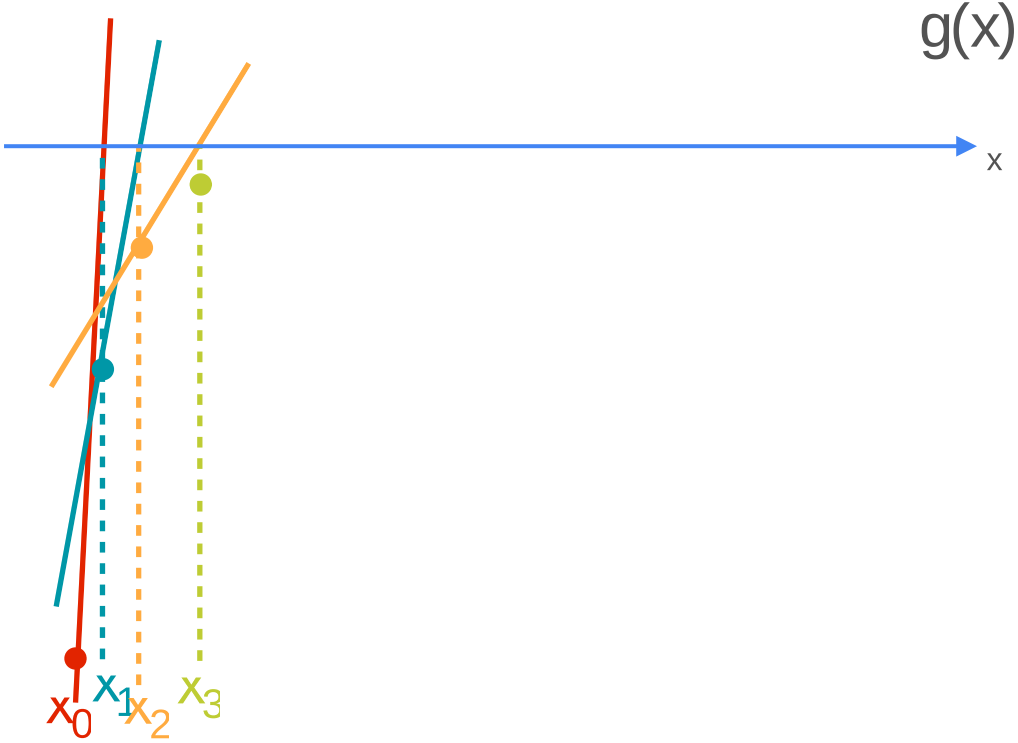
$$(g\#(x))\# = e^x + \frac{1}{x^2}$$

$x_2 = 0.183$

$$x_3 = x_2 \text{ " } \frac{g\#(x_2)}{(g\#(x_2))\#}$$

$$= 0.183 \text{"} \frac{\left(e^{0.183 \text{"}} \frac{1}{0.183}\right)}{\left(e^{0.183} + \frac{1}{0.183^2}\right)} = 0.320$$

NewtonÕs Method for Optimization



$g(x) = e^x \text{ " } \log(x)$ $\quad$ $g\#(x) = e^x \text{ " } 1/x$

Minimum: $x^* = 0.567$

$$(g\#(x))\# = e^x + \frac{1}{x^2}$$

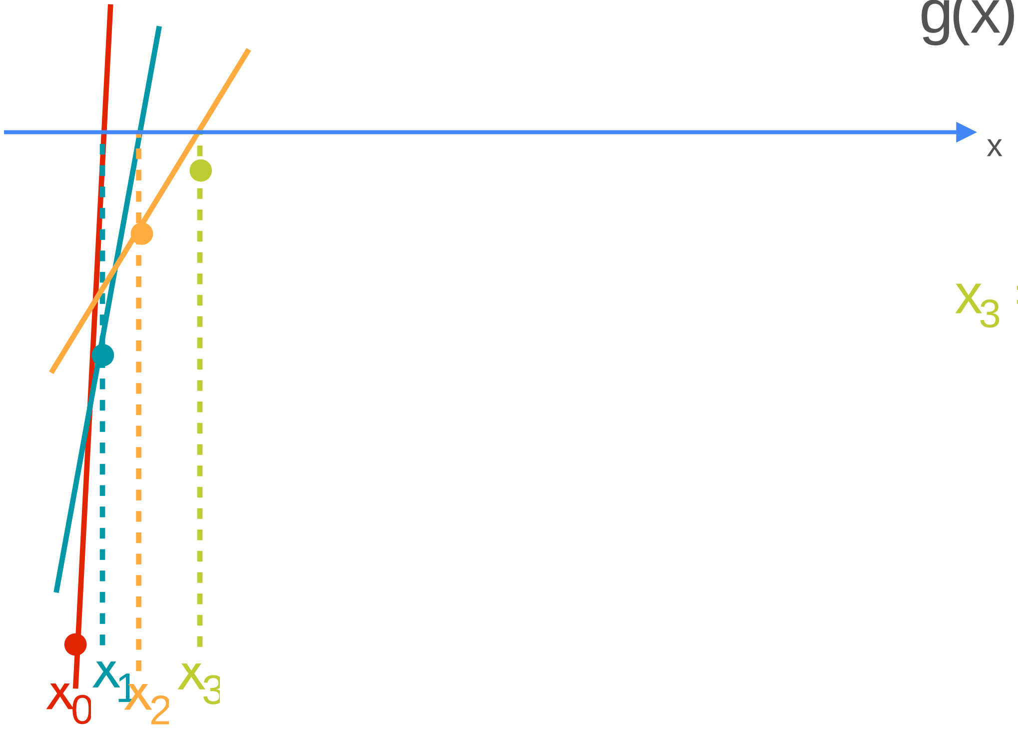
NewtonÕs Method for Optimization

$g(x) = e^x \text{ " } \log(x)$ $\quad$ $g\#(x) = e^x \text{ " } 1/x$

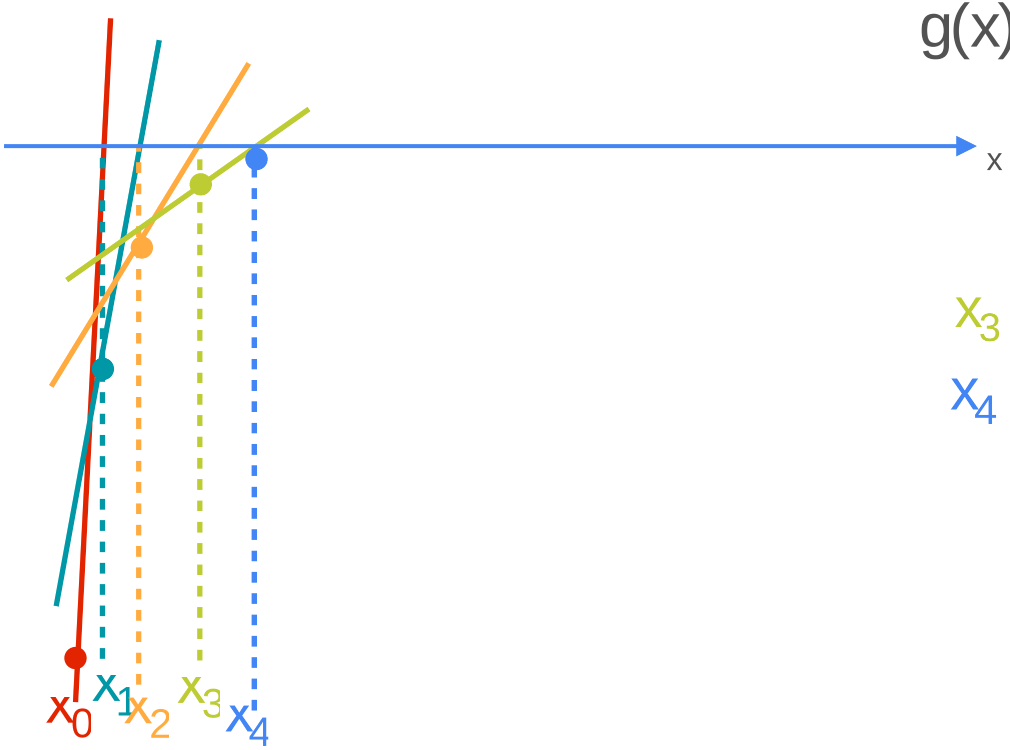
Minimum: $x^* = 0.567$

$(g\#(x))\# = e^x + \dfrac{1}{x^2}$

$x_3 = 0.320$



NewtonÕs Method for Optimization



$g(x) = e^x \text{ " } \log(x)$ $g\#(x) = e^x \text{ " } 1/x$

Minimum: $x^* = 0.567$

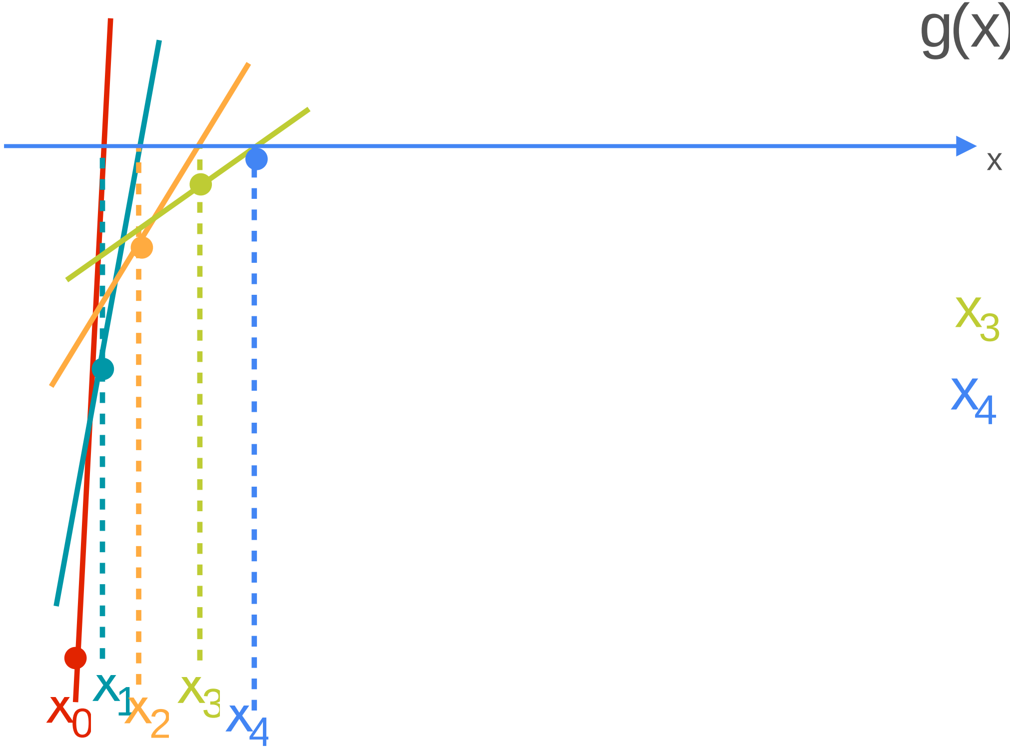
$$(g\#(x))\# = e^x + \frac{1}{x^2}$$

$x_3 = 0.320$

$$x_4 = x_3 \text{ " } \frac{g\#(x_3)}{(g\#(x_3))\#}$$

$$= 0.320 \text{ " } \frac{\left(e^{0.320} \text{ " } \frac{1}{0.320}\right)}{\left(e^{0.320} + \frac{1}{0.320^2}\right)}$$

Newton's Method for Optimization



$$g(x) = e^x - \log(x) \qquad g'(x) = e^x - 1/x$$

Minimum: $x^* = 0.567$

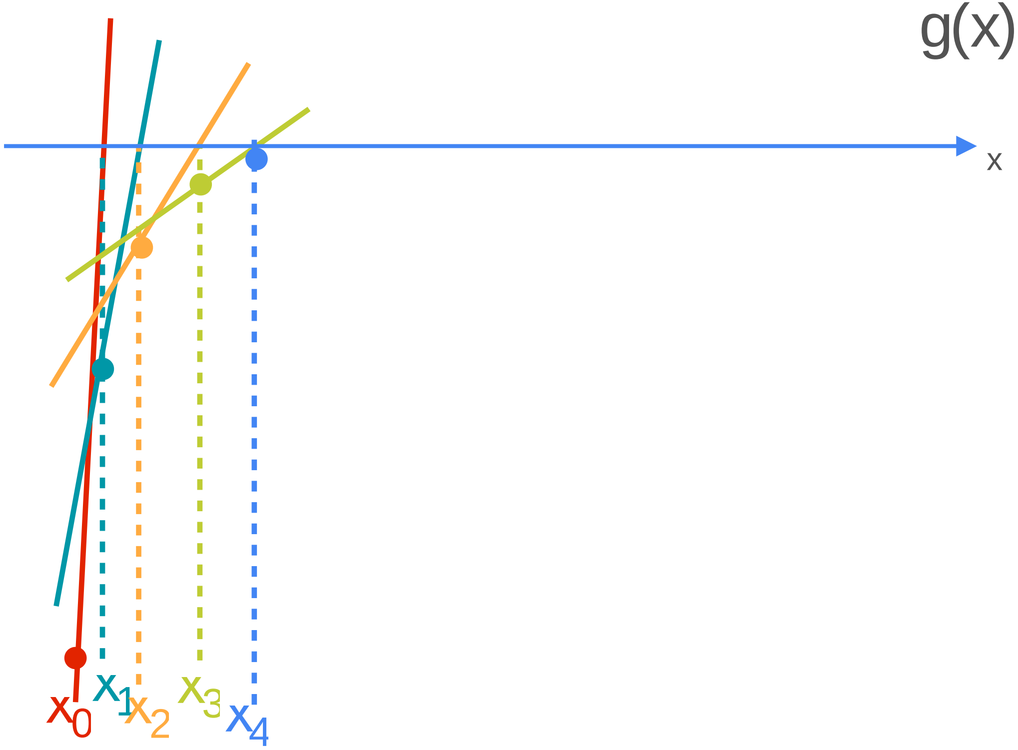
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_3 = 0.320$$

$$x_4 = x_3 - \frac{g'(x_3)}{(g'(x_3))'}$$

$$= 0.320 - \frac{\left(e^{0.320} - \frac{1}{0.320}\right)}{\left(e^{0.320} + \frac{1}{0.320^2}\right)} = 0.477$$

NewtonÕs Method for Optimization

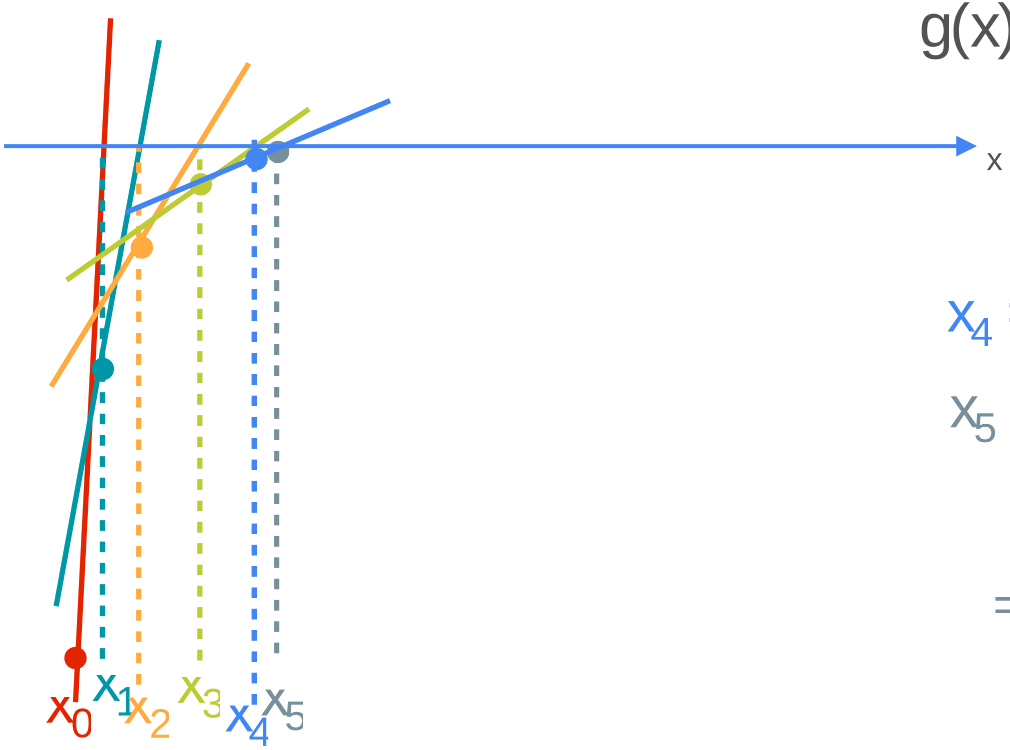


$g(x) = e^x \text{ " } \log(x)$ $\quad$ $g\#(x) = e^x \text{ " } 1/x$

Minimum: $x^* = 0.567$:

$$(g\#(x))\# = e^x + \frac{1}{x^2}$$

NewtonÕs Method for Optimization



$$g(x) = e^x - \log(x) \qquad g'(x) = e^x - 1/x$$

Minimum: $x^* = 0.567$

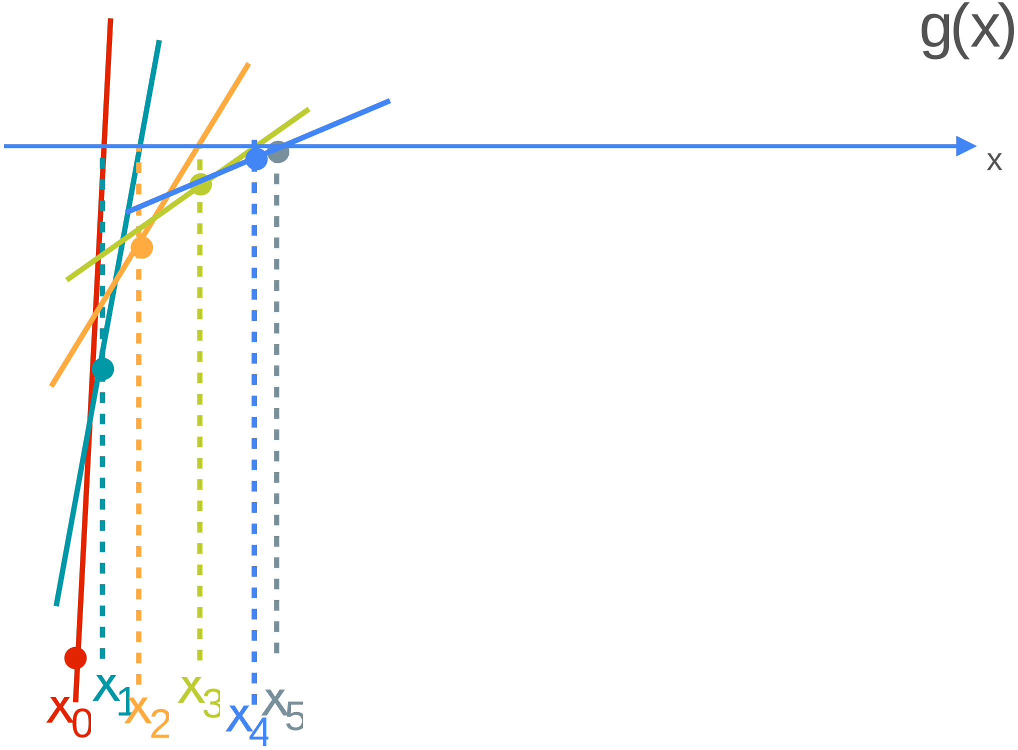
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$x_4 = 0.477$

$$x_5 = x_4 - \frac{g'(x_4)}{(g'(x_4))'}$$

$$= 0.447 - \frac{\left(e^{0.447} - \frac{1}{0.447}\right)}{\left(e^{0.447} + \frac{1}{0.447^2}\right)} = 0.558$$

NewtonÕs Method for Optimization

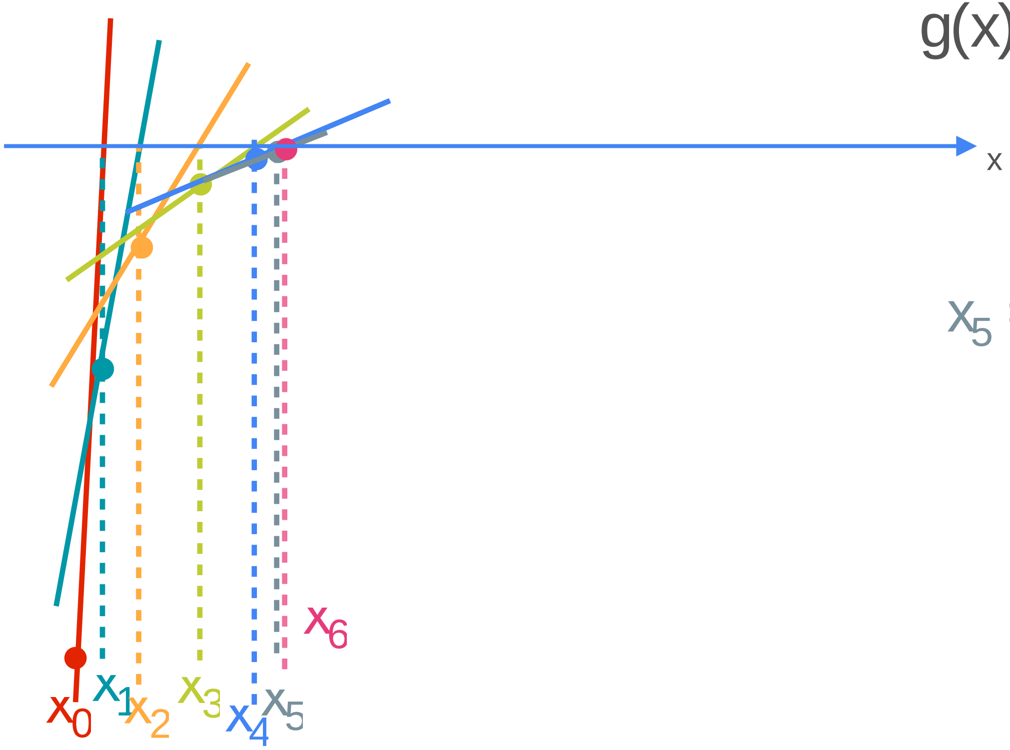


$g(x) = e^x$ " $\log(x)$ $\quad$ $g\#(x) = e^x$ " $1/x$

Minimum: $x^* = 0.567$

$(g\#(x))\# = e^x + \dfrac{1}{x^2}$

NewtonÕs Method for Optimization



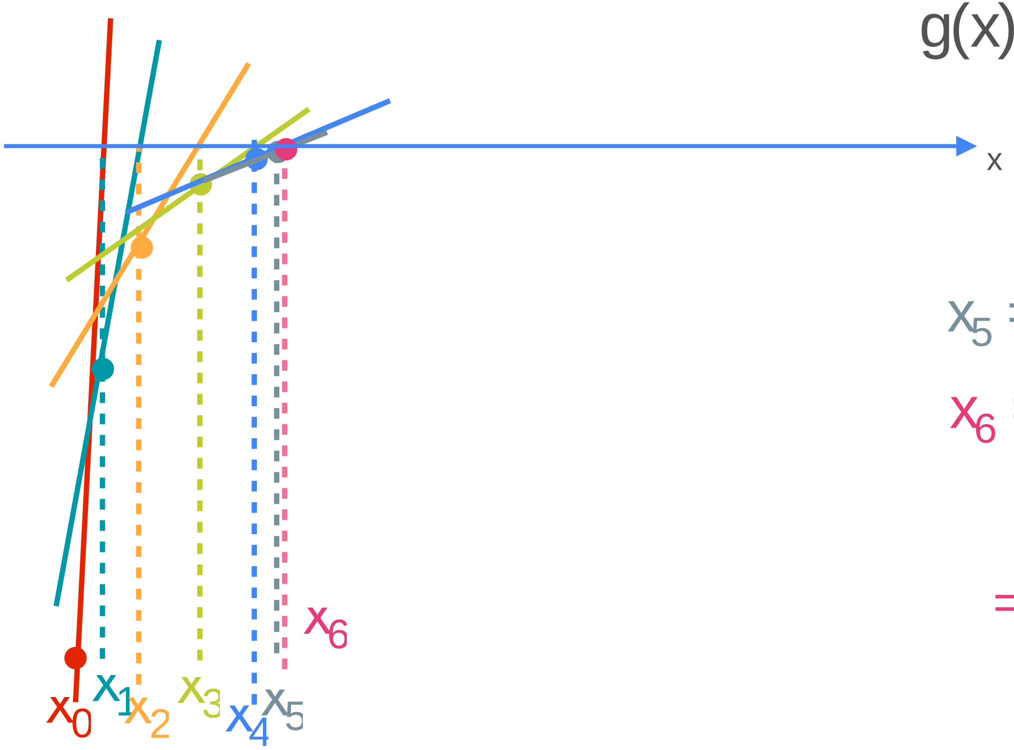
$g(x) = e^{x}$ " $\log(x)$ $g\#(x) = e^{x}$ " $1/x$

Minimum: $x^* = 0.567$

$(g\#(x))\# = e^{x} + \dfrac{1}{x^{2}}$

$x_5 = 0.558$

NewtonÕs Method for Optimization



$g(x) = e^x \text{ " } \log(x)$ $\quad$ $g\#(x) = e^x \text{ " } 1/x$

Minimum: $x^* = 0.567$

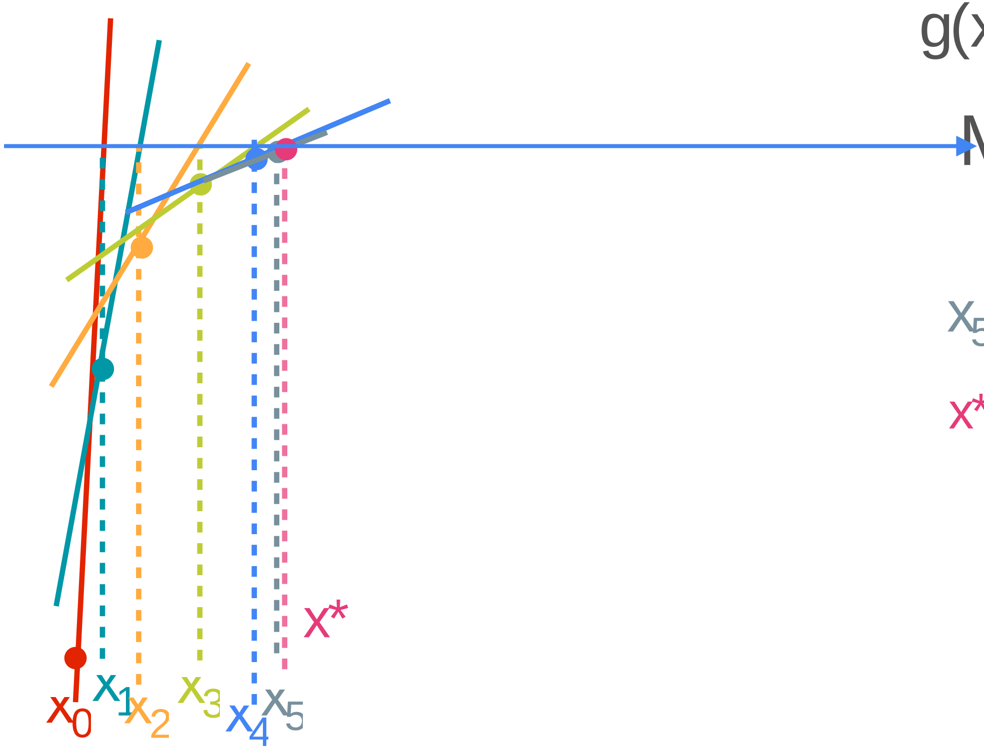
$$(g\#(x))\# = e^x + \frac{1}{x^2}$$

$x_5 = 0.558$

$$x_6 = x_5 \text{ " } \frac{g\#(x_5)}{(g\#(x_5))\#}$$

$$= 0.558 \text{ " } \frac{\left(e^{0.558} \text{ " } \frac{1}{0.558}\right)}{\left(e^{0.558} + \frac{1}{0.558^2}\right)}$$

NewtonÕs Method for Optimization



$$g(x) = e^x \text{ " } \log(x) \qquad g\#(x) = e^x \text{ " } 1/x$$

Minimum: $x^* = 0.567$

$$(g\#(x))\# = e^x + \frac{1}{x^2}$$

$$x_5 = 0.558$$

$$x^* = x_5 \text{ " } \frac{g\#(x_5)}{(g\#(x_5))\#}$$

$$= 0.558 \text{ " } \frac{\left(e^{0.558} \text{ " } \frac{1}{0.558}\right)}{\left(e^{0.558} + \frac{1}{0.558^2}\right)} = 0.567$$

Optimization in Neural Networks and NewtonÕs Method

The second derivative

Second Derivative

Second Derivative

NewtonÕs method:

Second Derivative

NewtonÕs method: $x_{k+1} = x_k \text{ " } \frac{g\#(x_k)}{(g\#(x_k))\#}$

Second Derivative

NewtonÕs method: $x_{k+1} = x_k \text{ " } \frac{g\#(x_k)}{\boxed{(g\#(x_k))\#}}$??

Second Derivative

NewtonÕs method: $x_{k+1} = x_k - \dfrac{g'(x_k)}{\boxed{(g'(x_k))'}}$??

Second derivative

Second Derivative

NewtonÕs method: $x_{k+1} = x_k \,"\, \dfrac{g\#(x_k)}{\boxed{(g\#(x_k))\#}}$??

Second derivative

Leibniz notation: $\dfrac{d^2 f(x)}{dx^2} = \dfrac{d}{dx}\left(\dfrac{df(x)}{dx}\right)$

Second Derivative

NewtonÕs method: $x_{k+1} = x_k - \dfrac{g'(x_k)}{\boxed{(g'(x_k))'}}$??

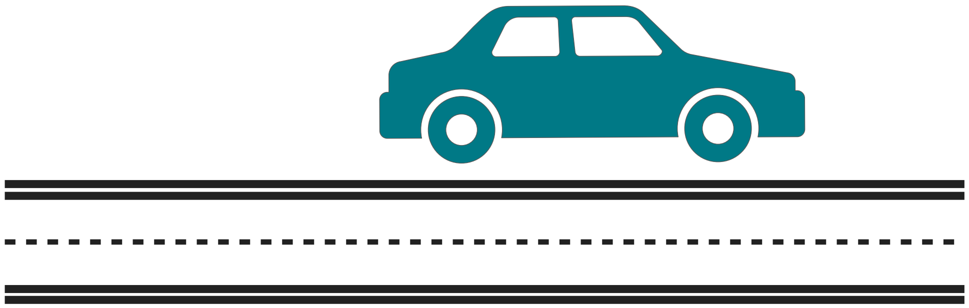
Second derivative

Leibniz notation: $\dfrac{d^2 f(x)}{dx^2} = \dfrac{d}{dx}\left(\dfrac{df(x)}{dx}\right)$

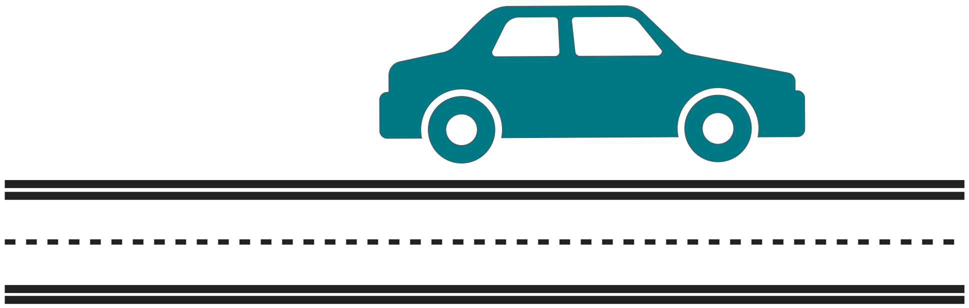
Lagrange notation: $f''(x)$

Understanding Second Derivative

Understanding Second Derivative

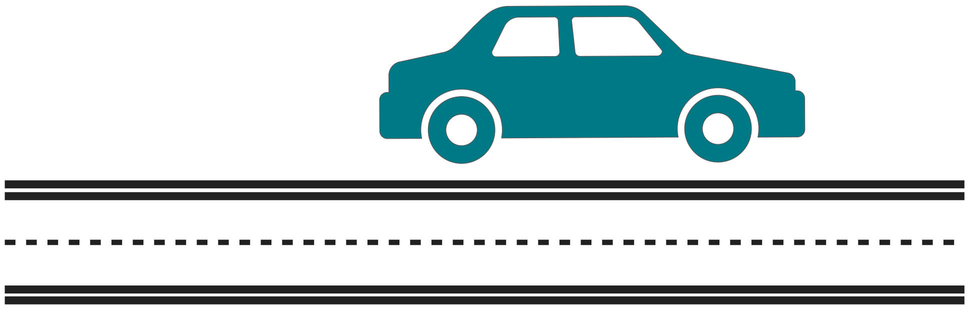


Understanding Second Derivative



X Distance

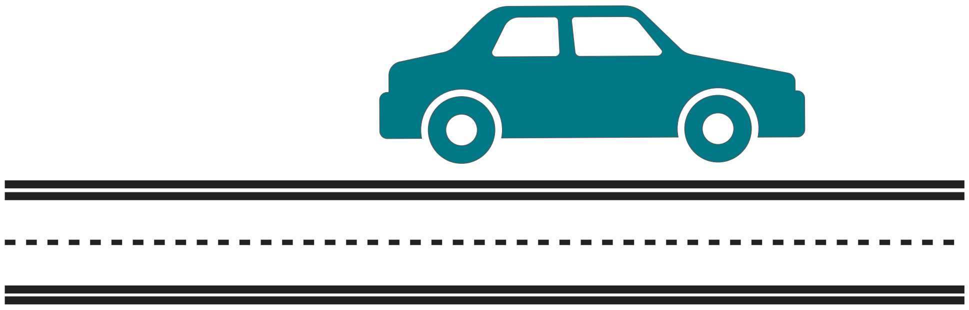
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

Understanding Second Derivative



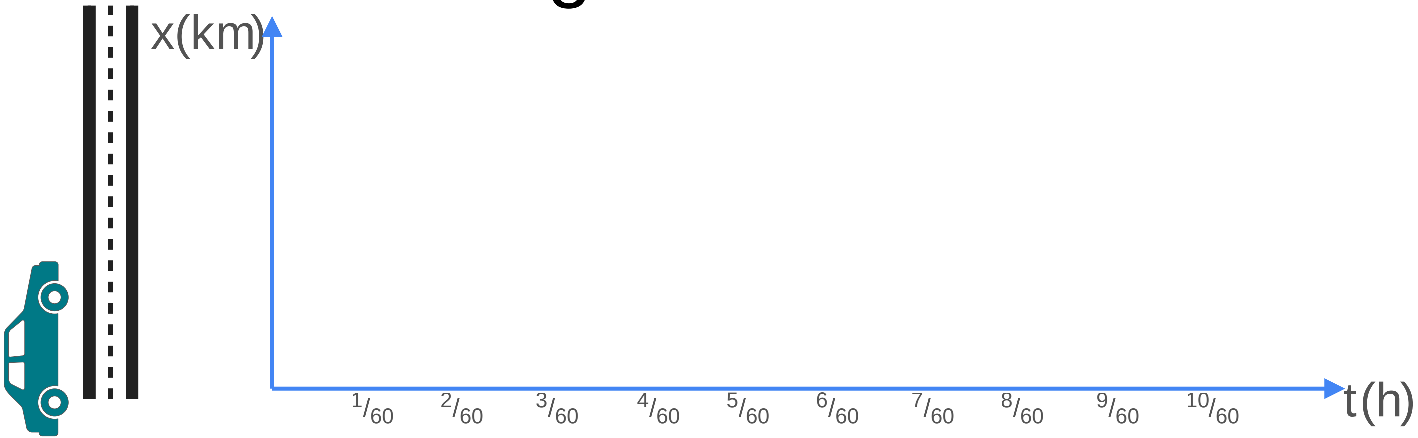
x Distance

v Velocity $\dfrac{dx}{dt}$

a Acceleration $\dfrac{dv}{dt} = \dfrac{d^2x}{dt^2}$

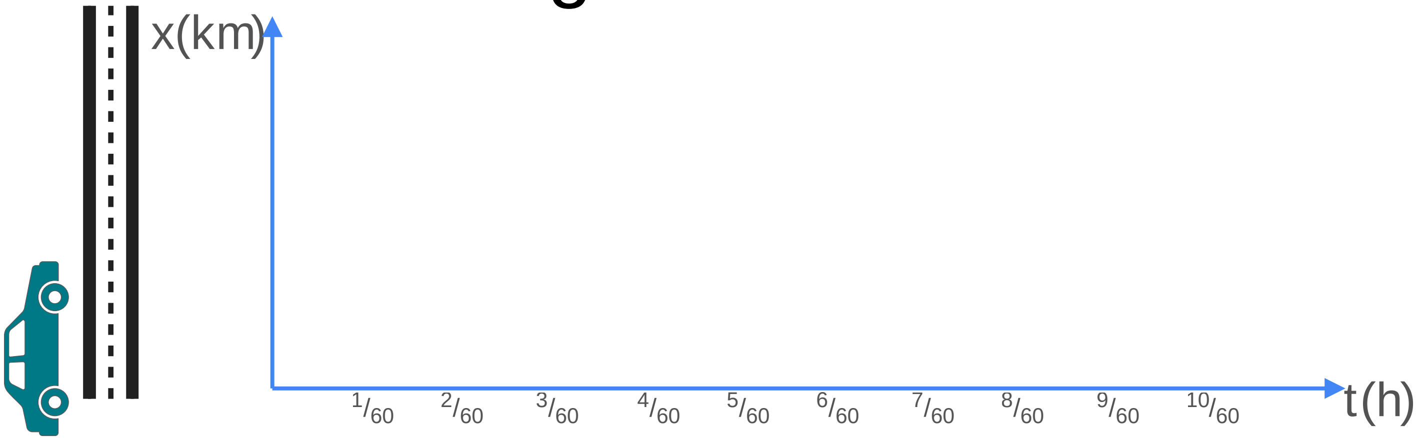
Understanding Second Derivative

Understanding Second Derivative



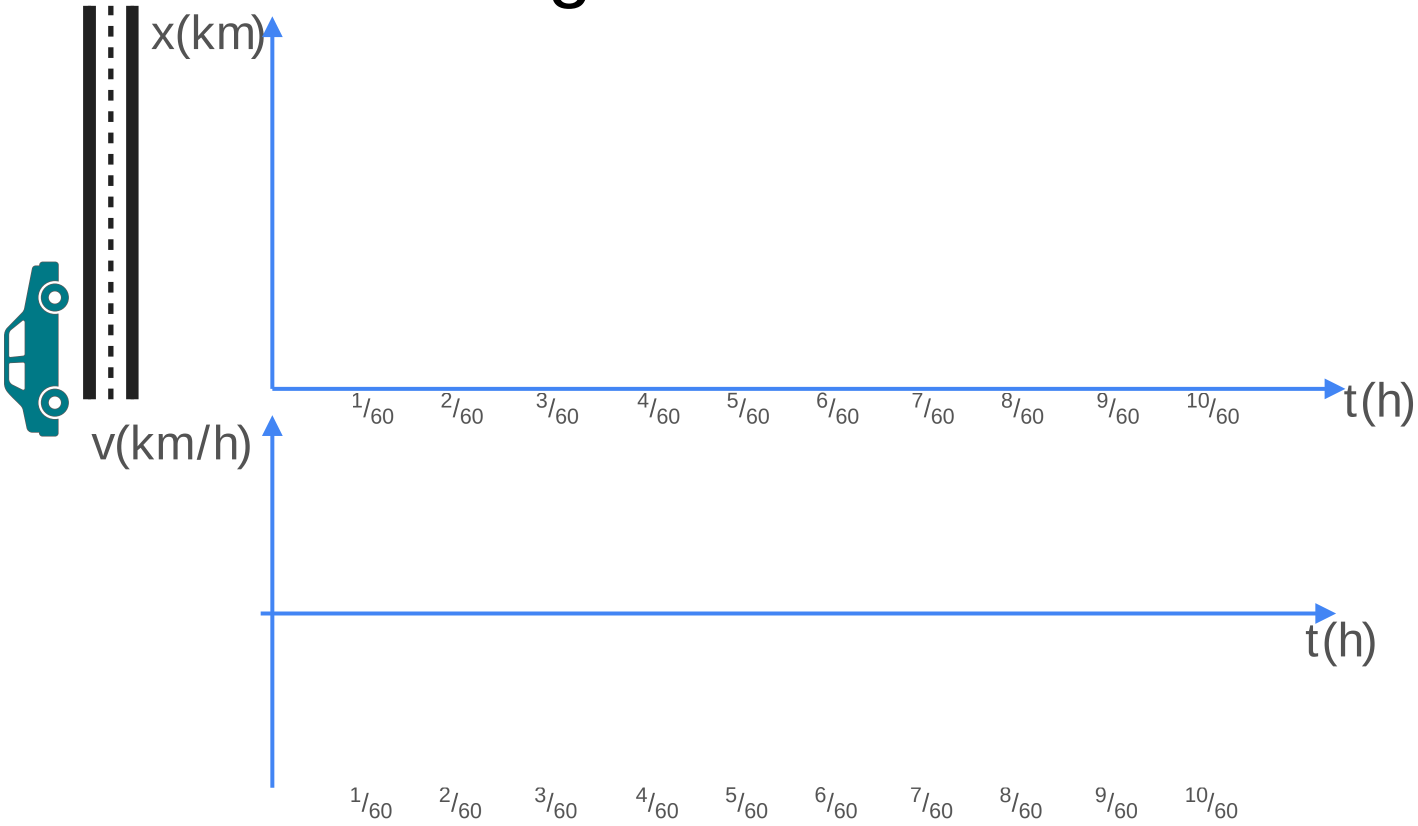
X Distance

Understanding Second Derivative



X Distance

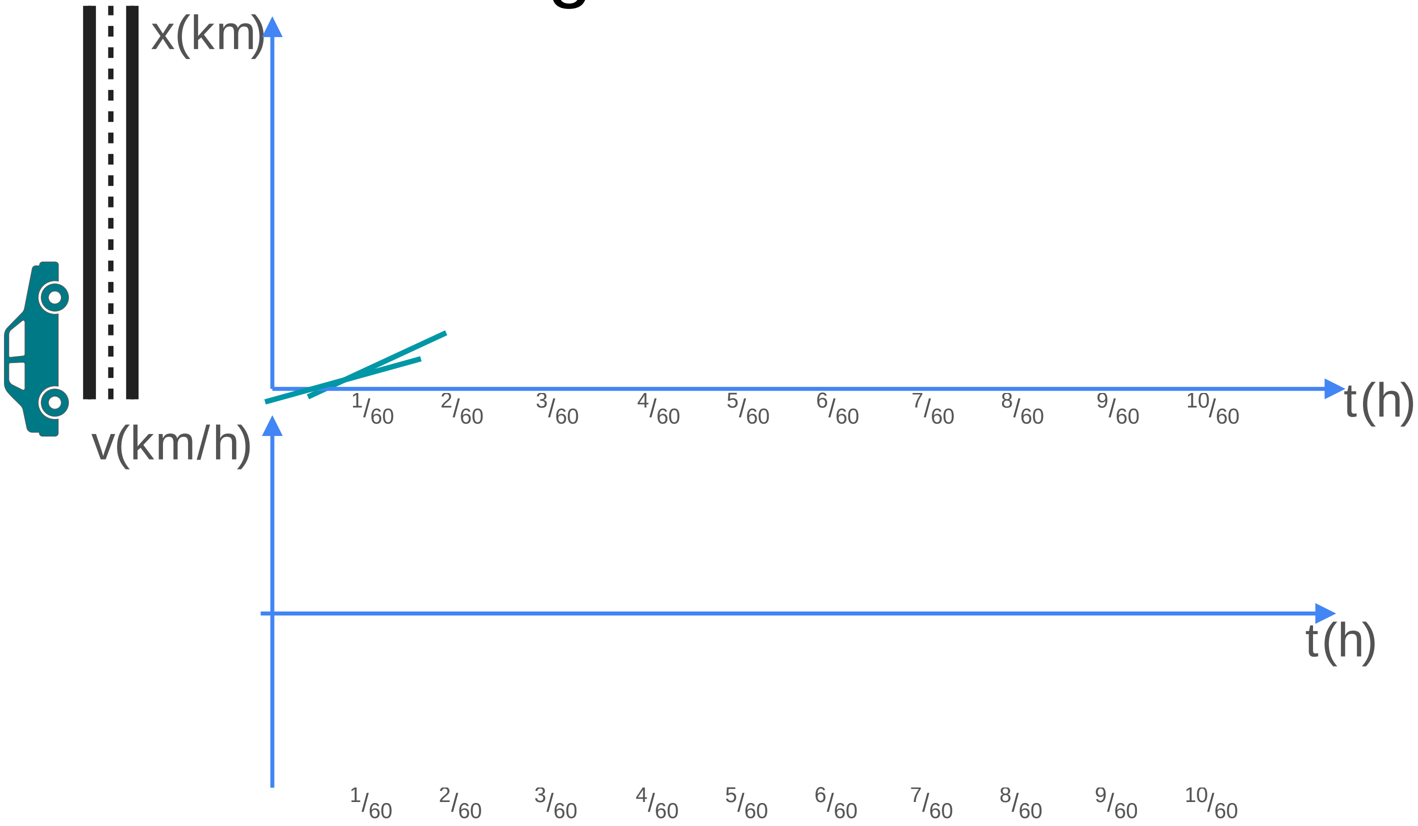
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

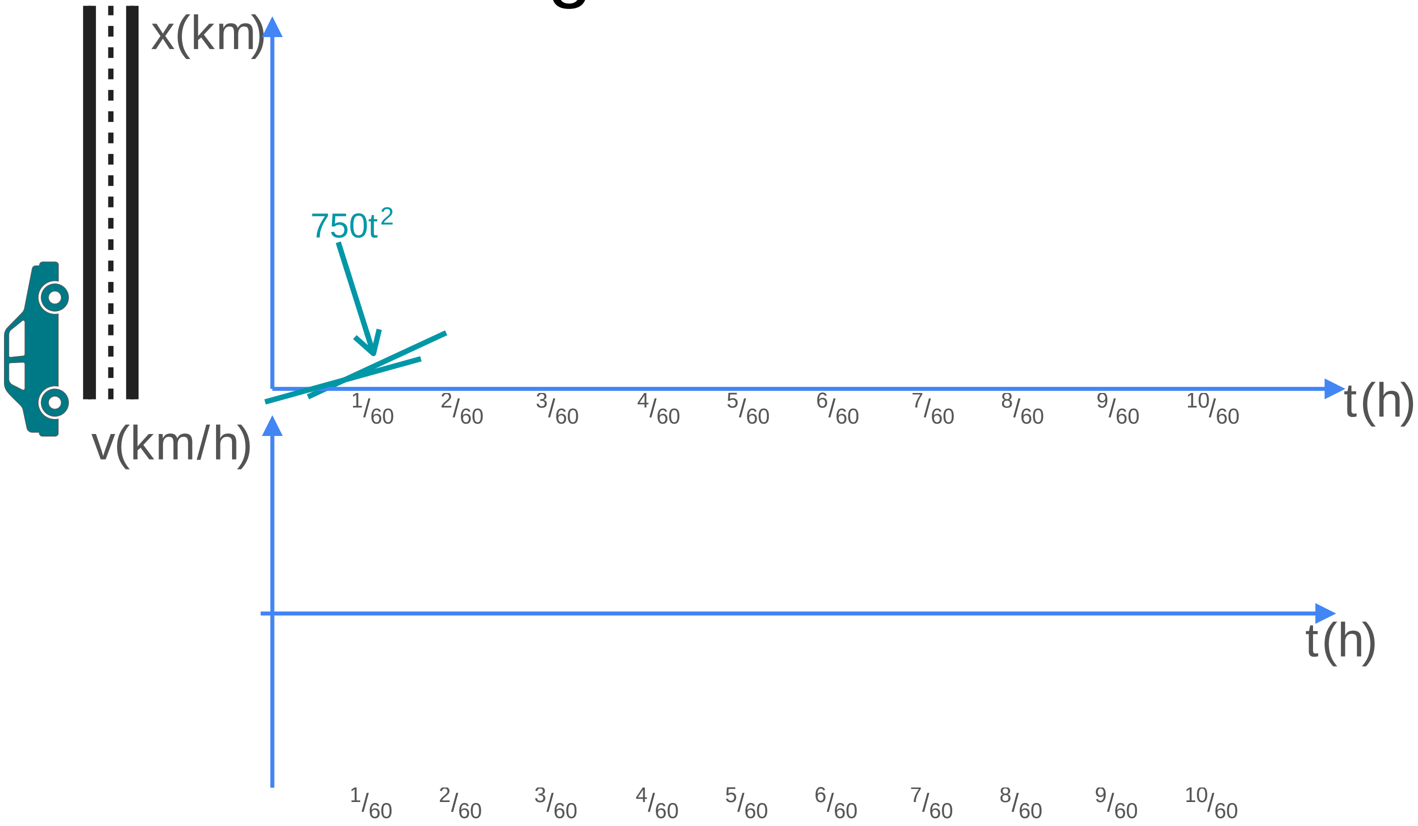
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

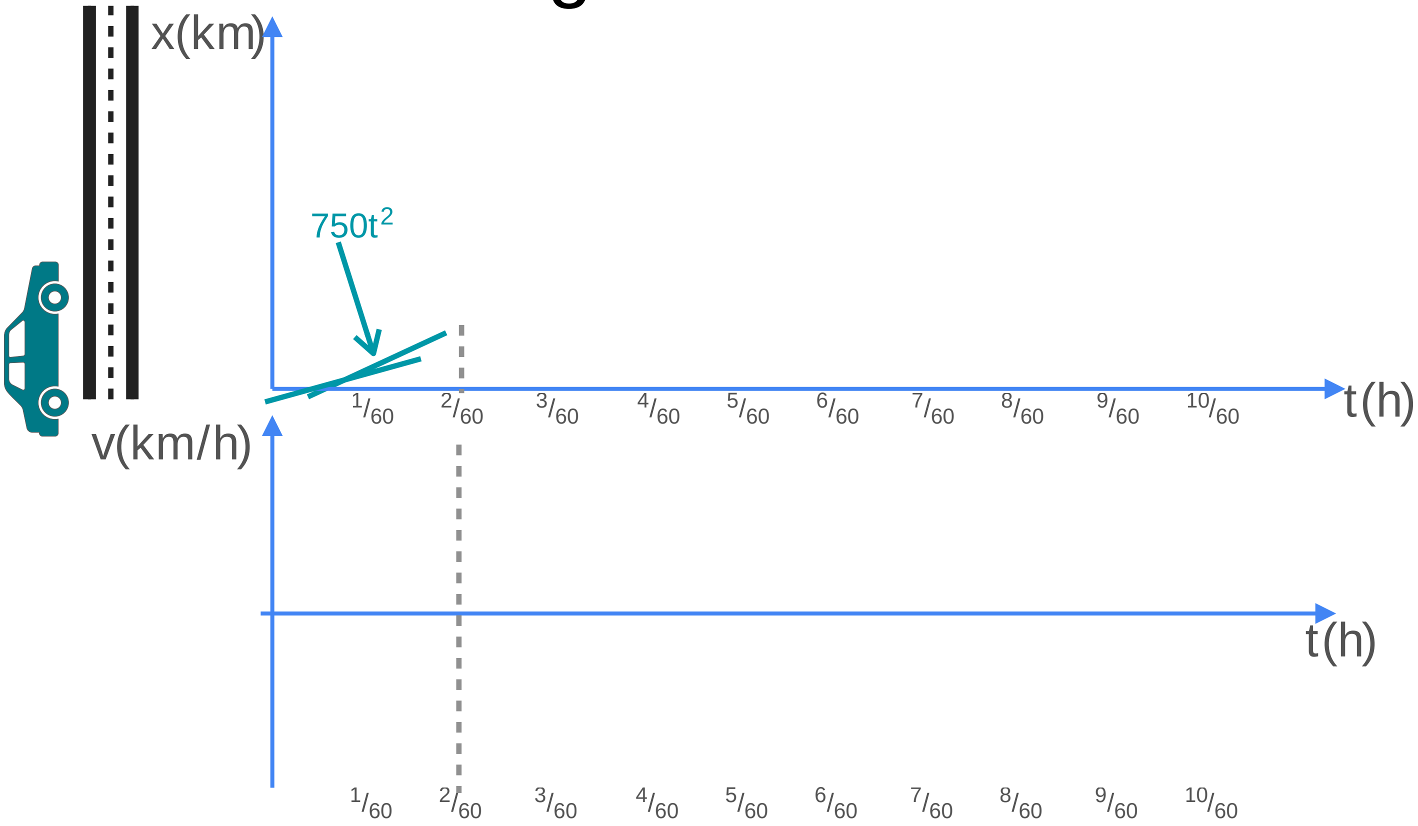
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

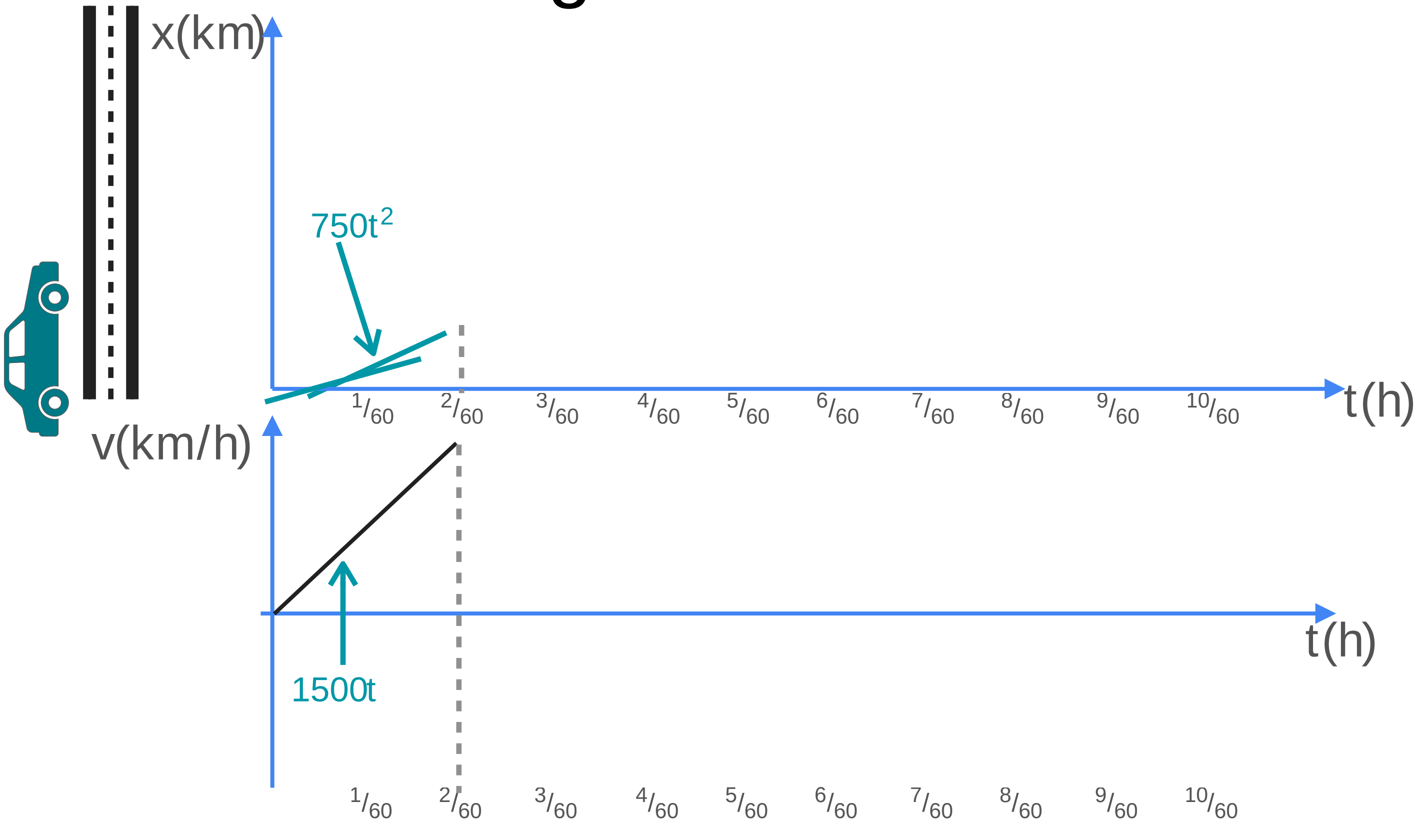
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

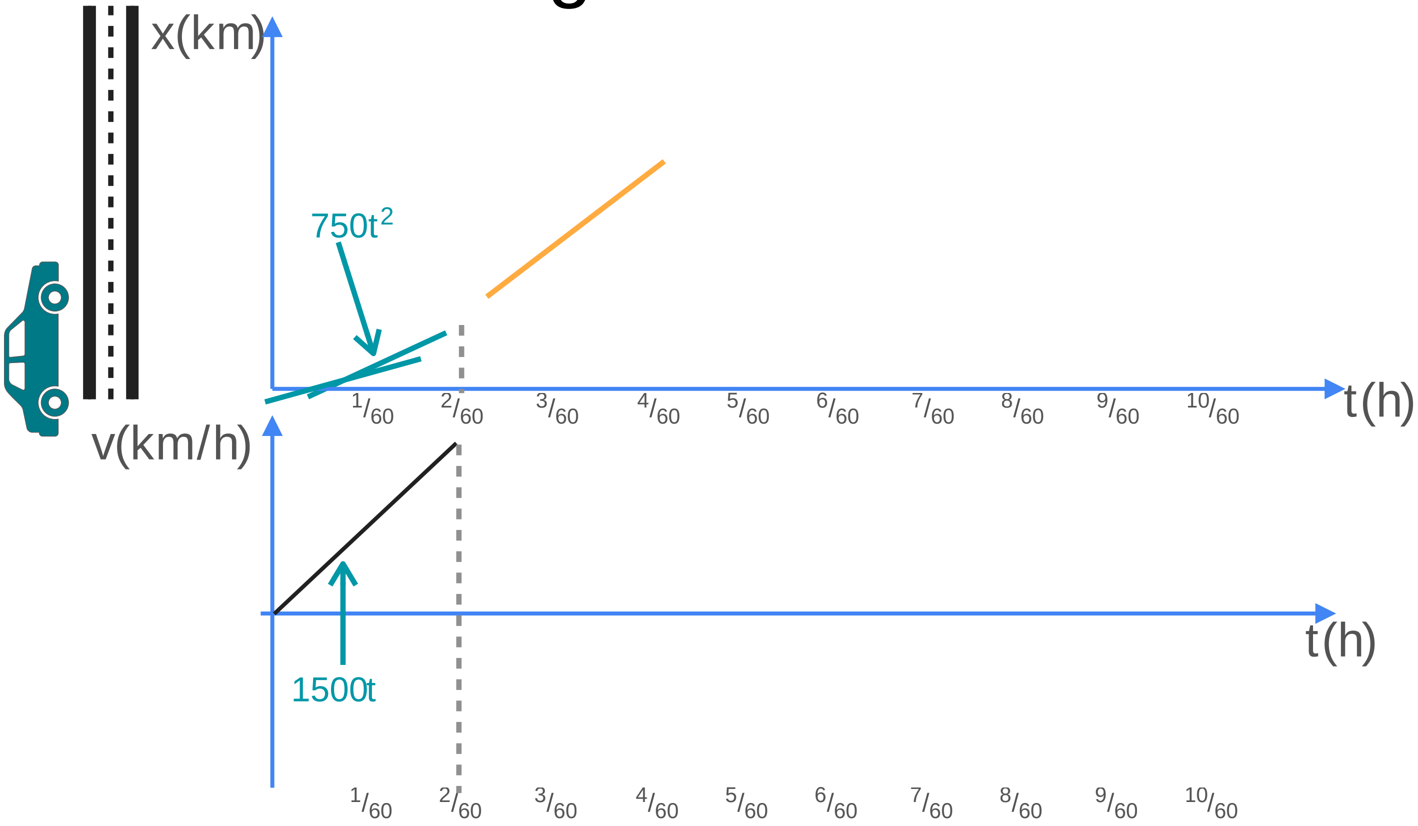
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

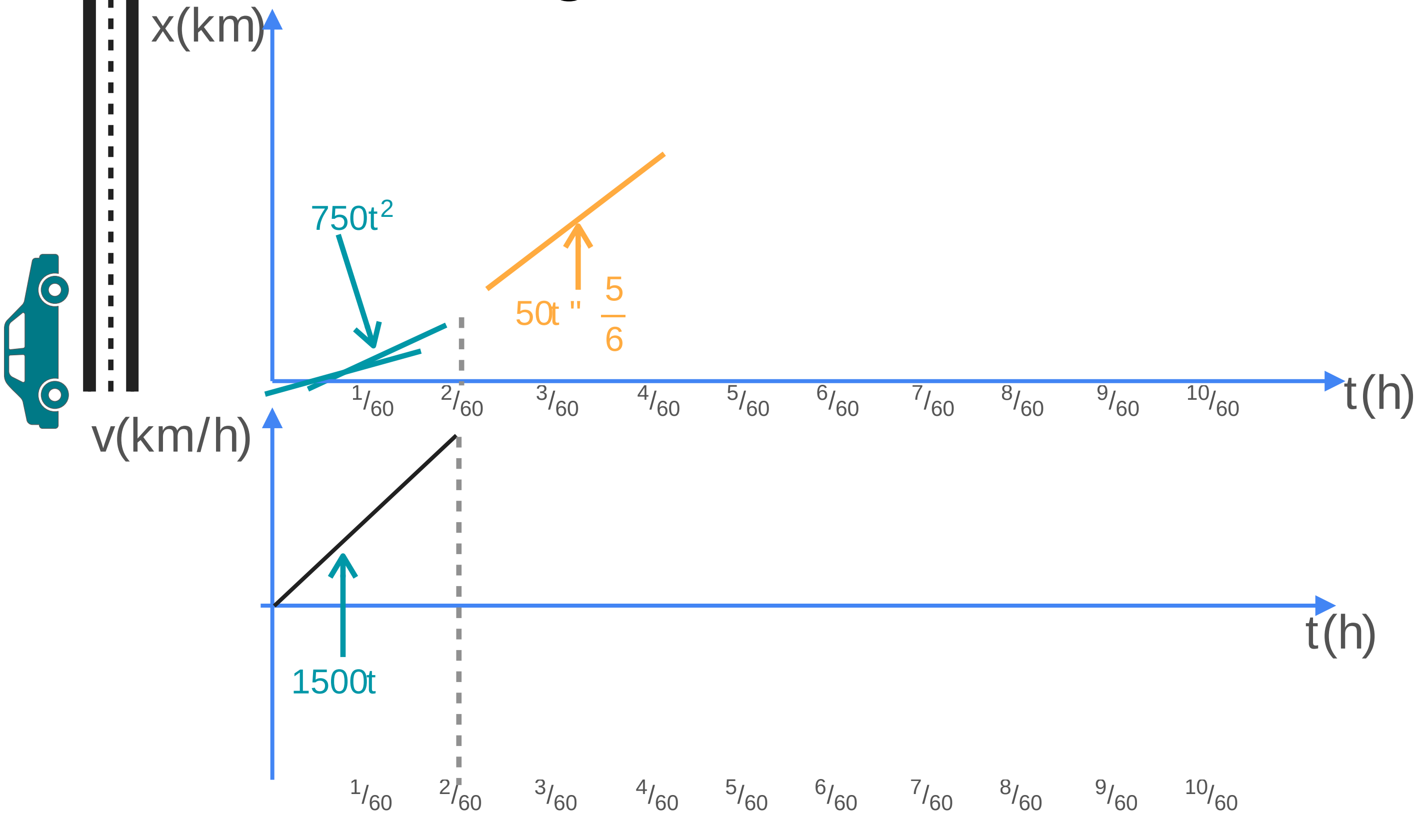
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

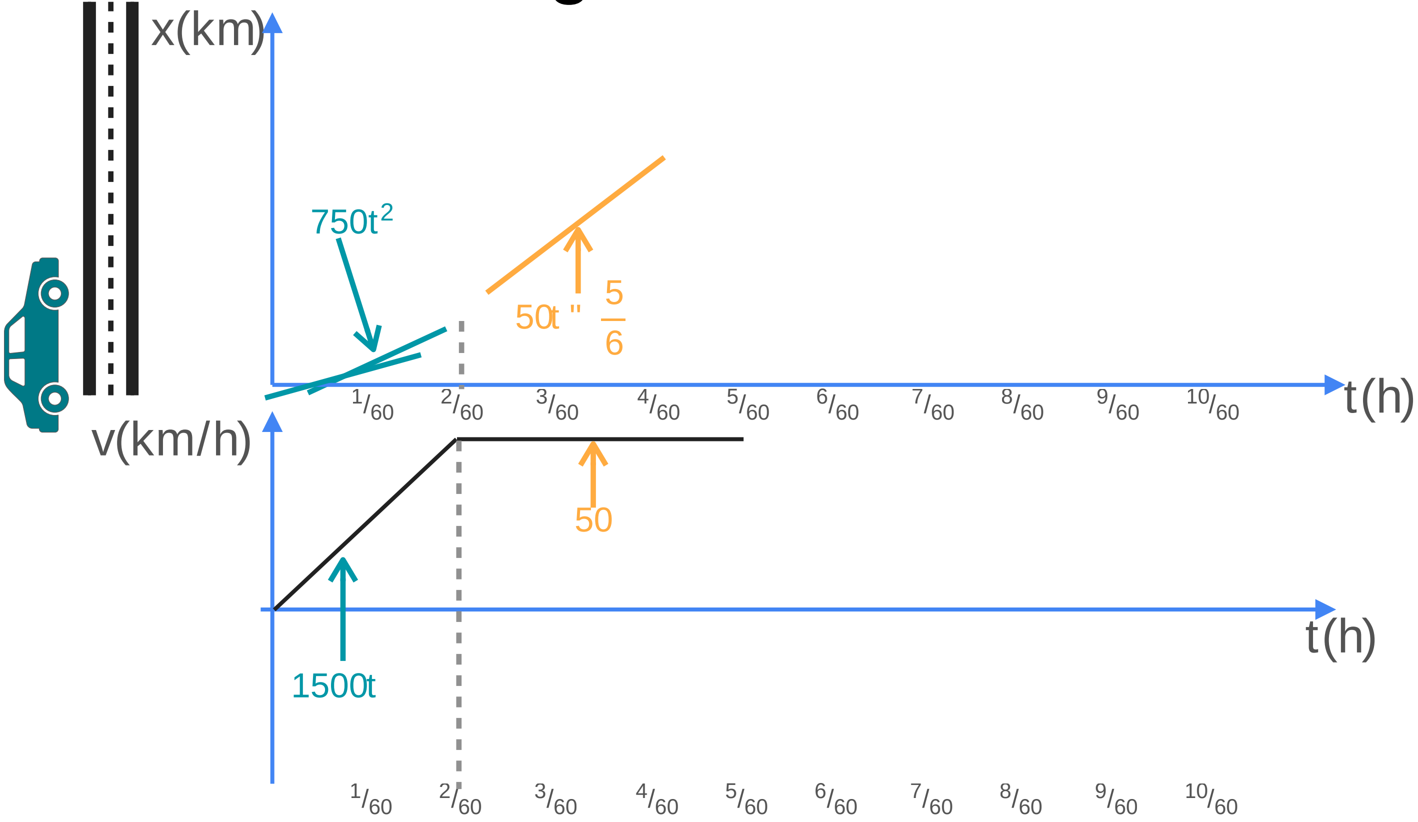
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

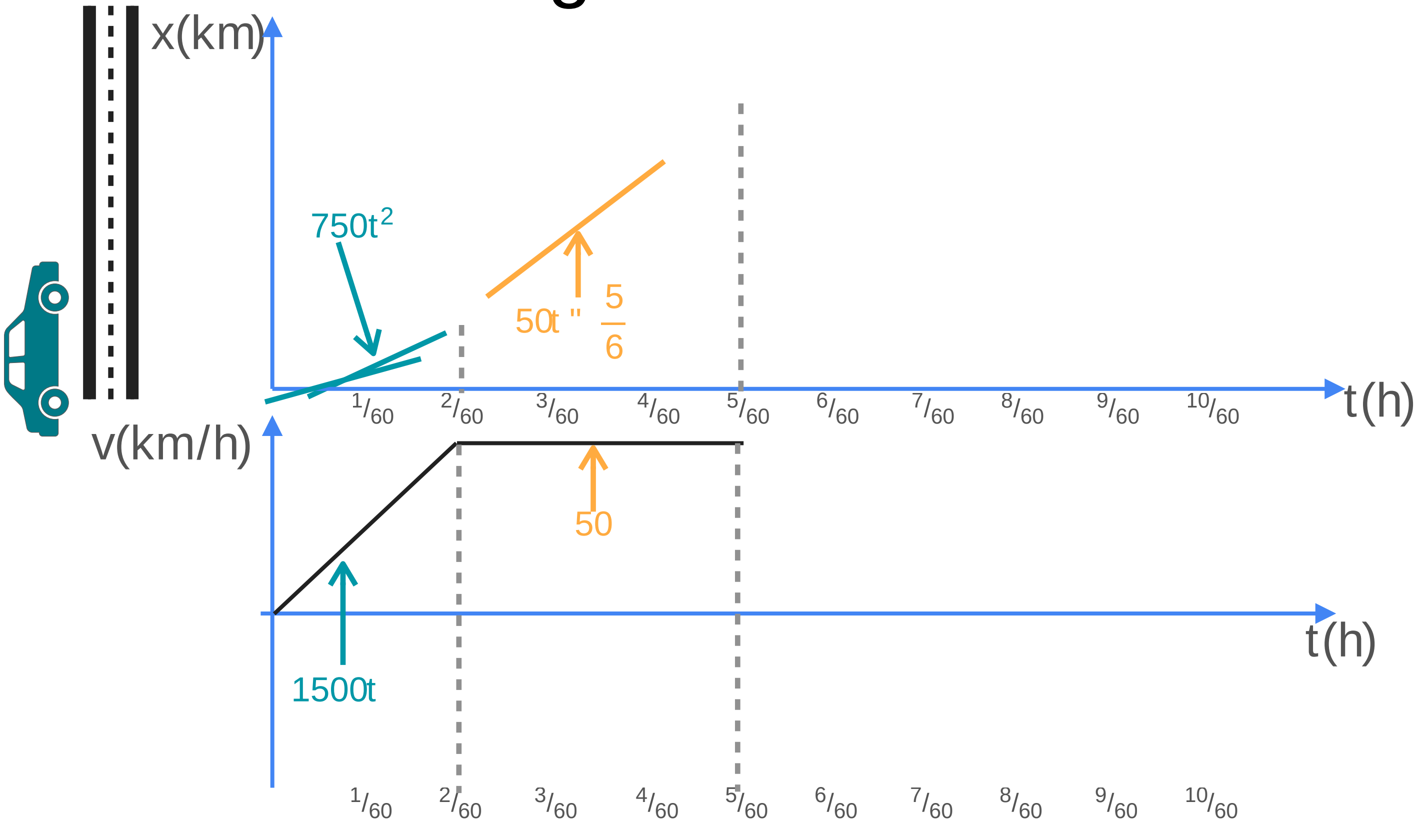
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

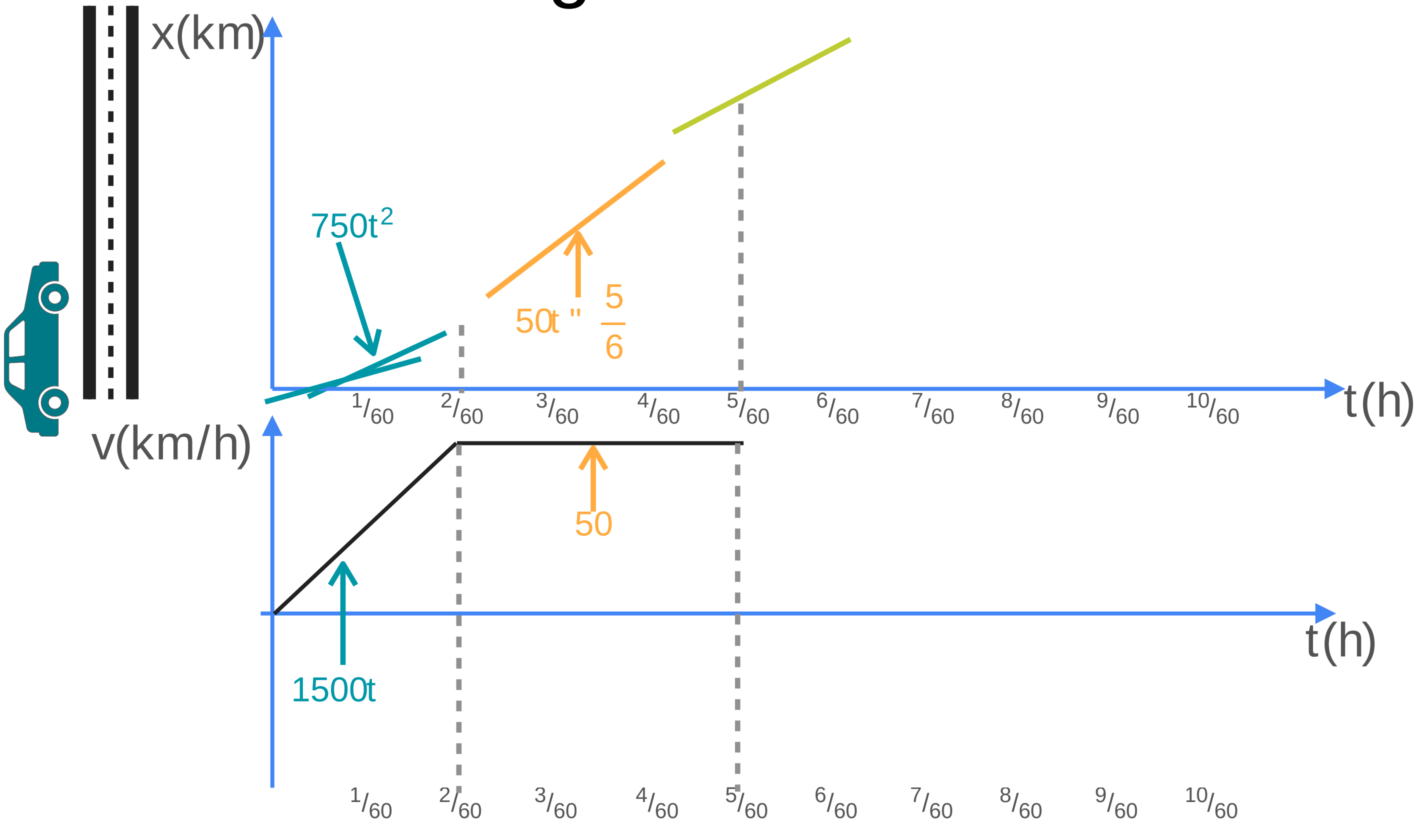
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

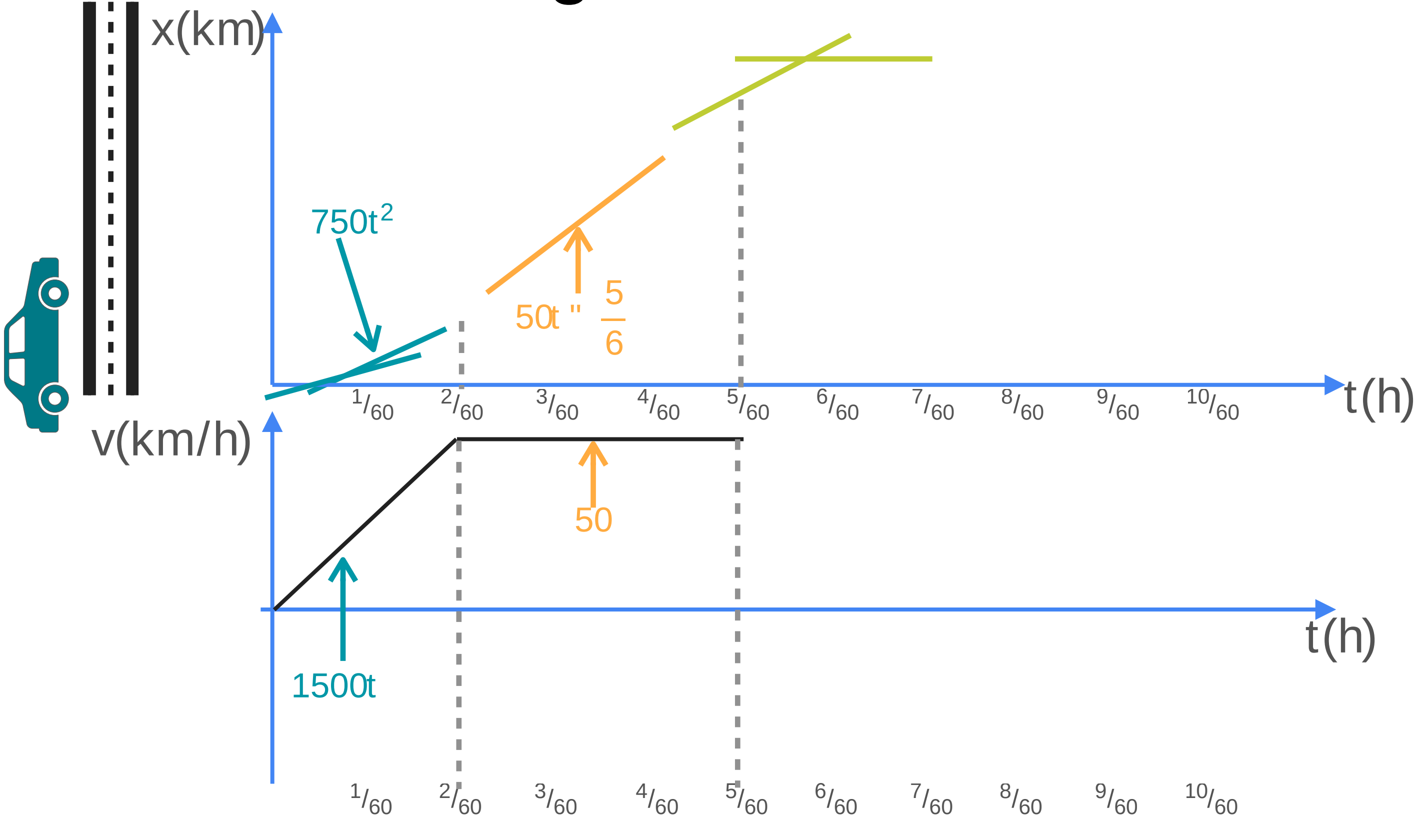
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

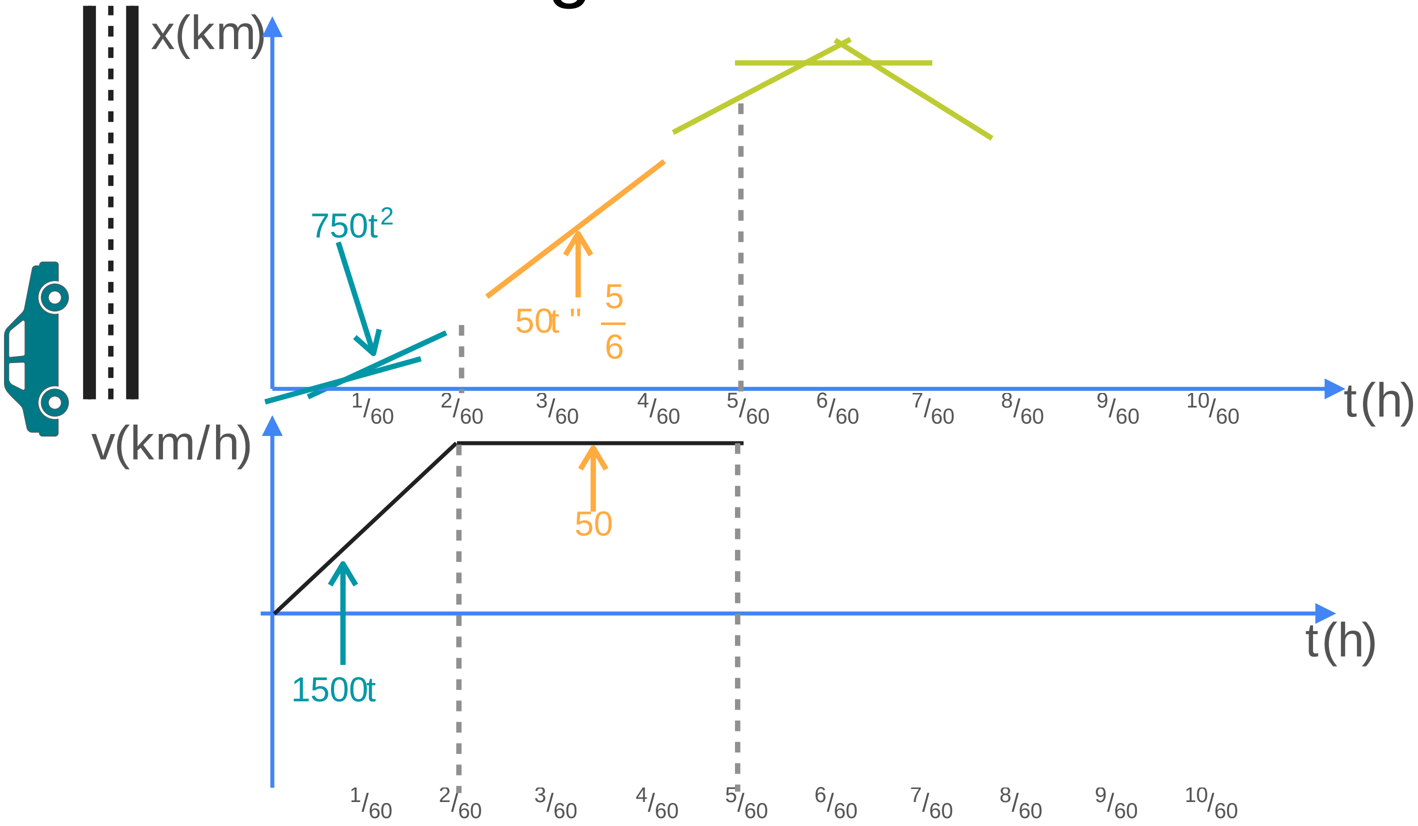
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

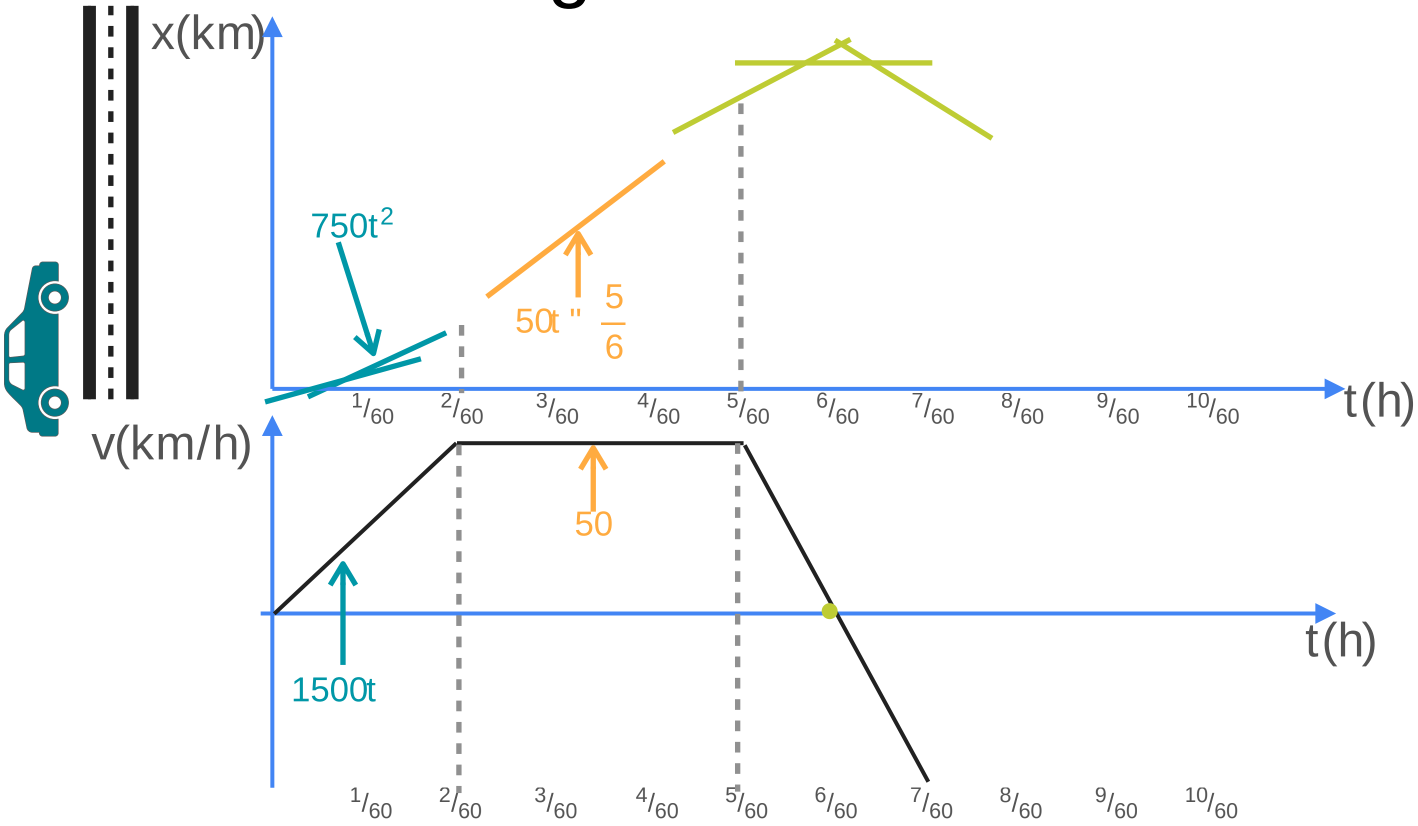
Understanding Second Derivative



X Distance

V Velocity $\frac{dx}{dt}$

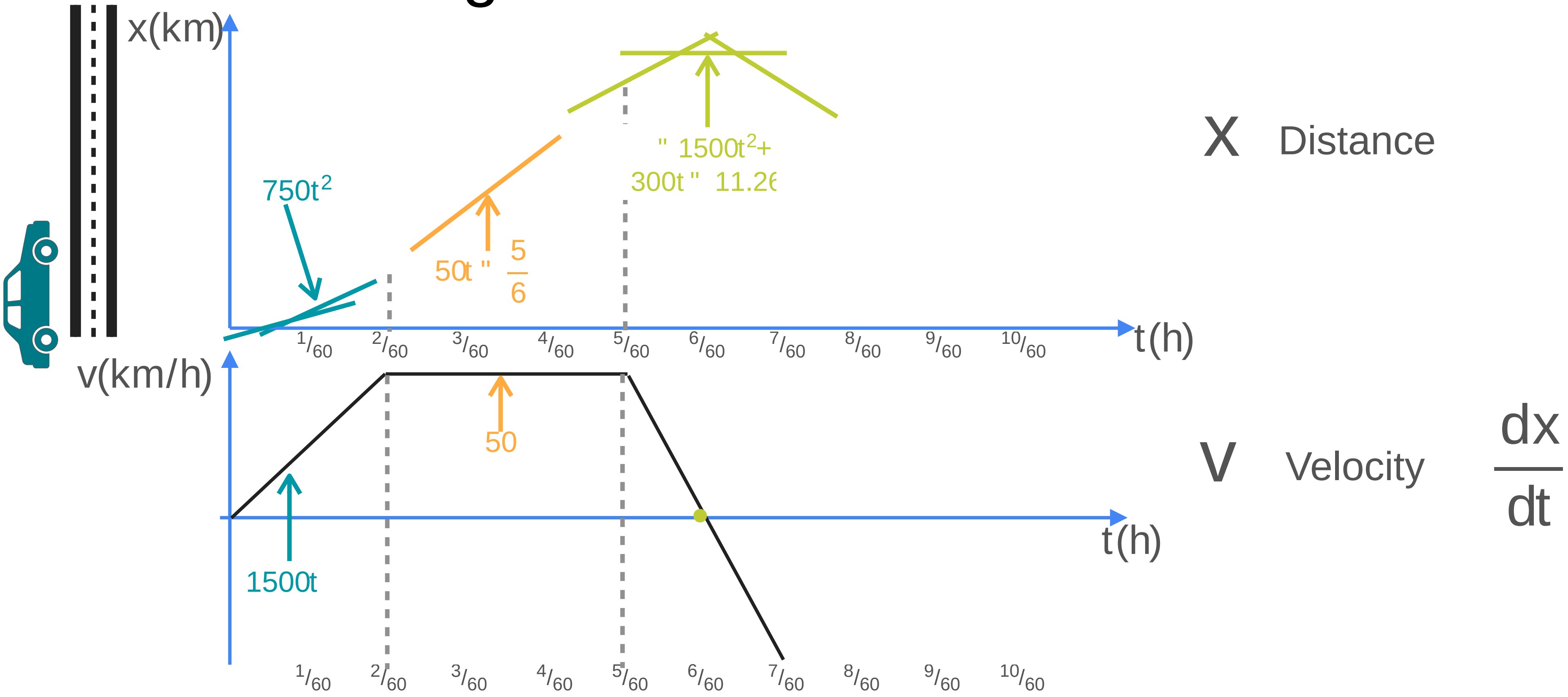
Understanding Second Derivative



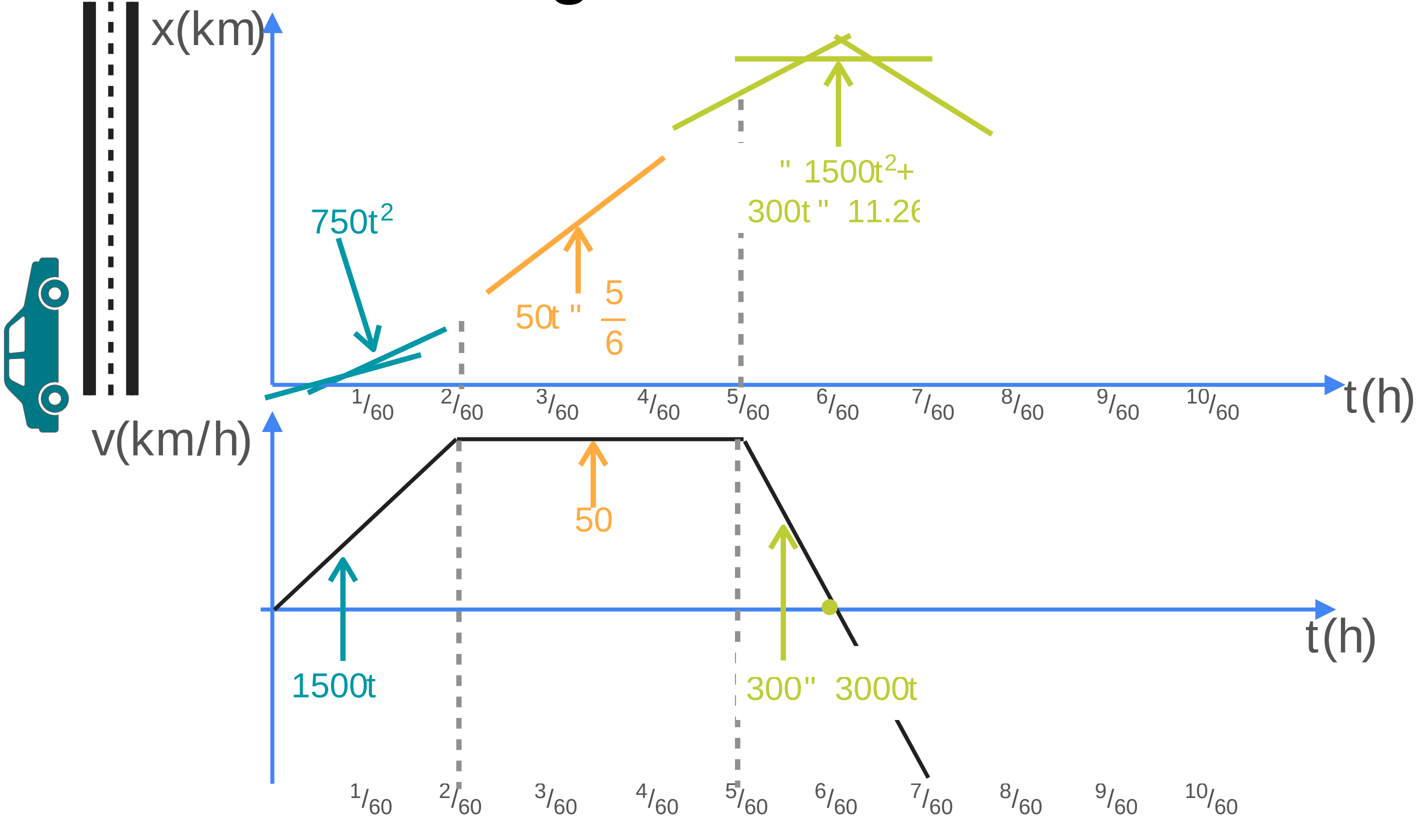
X Distance

V Velocity $\frac{dx}{dt}$

Understanding Second Derivative



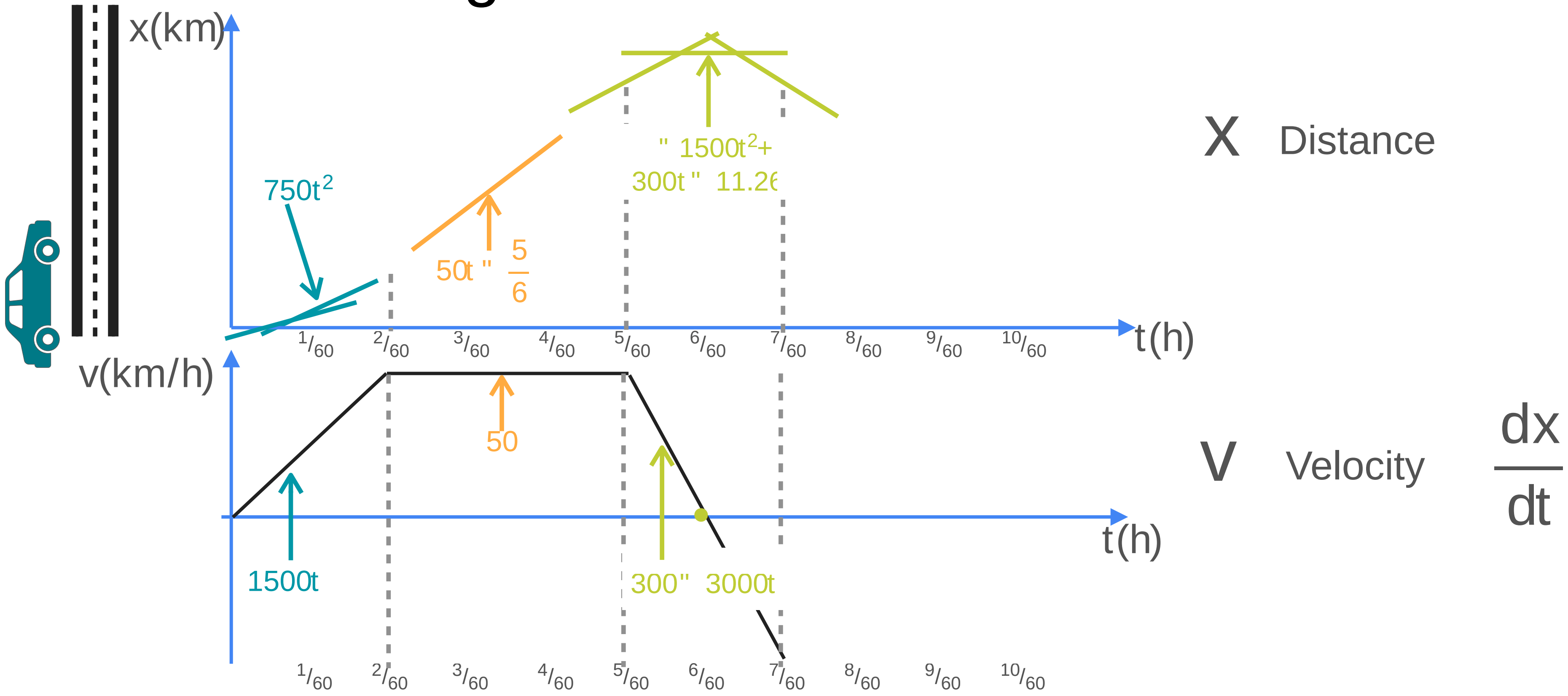
Understanding Second Derivative



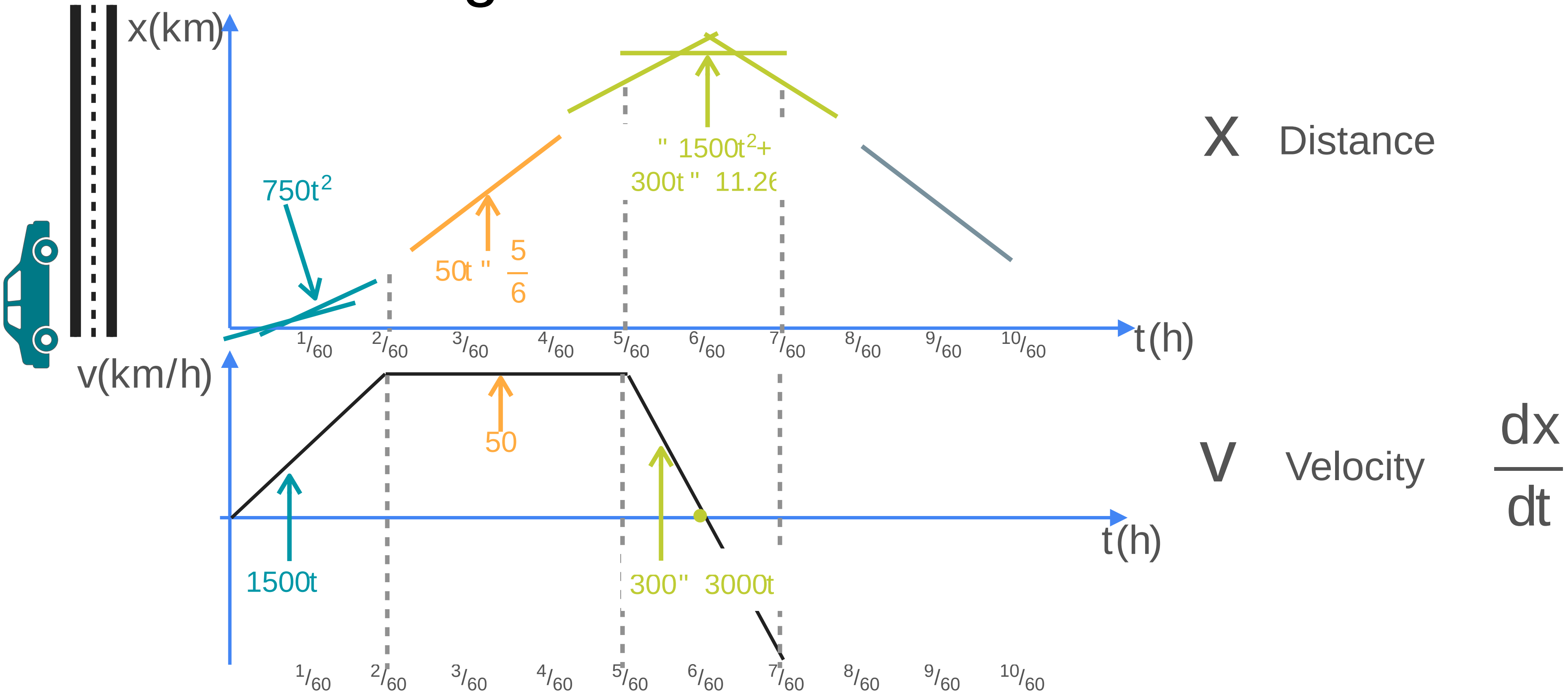
X Distance

V Velocity $\frac{dx}{dt}$

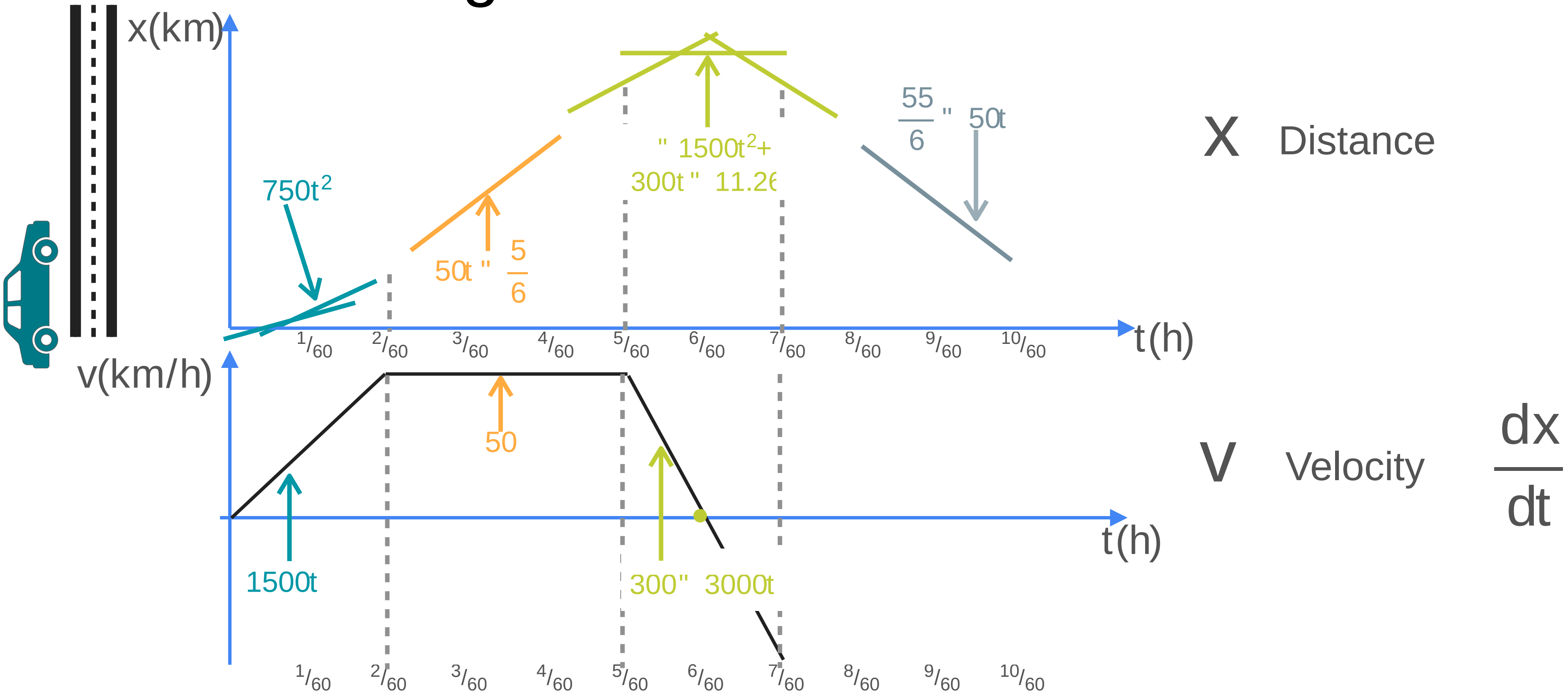
Understanding Second Derivative



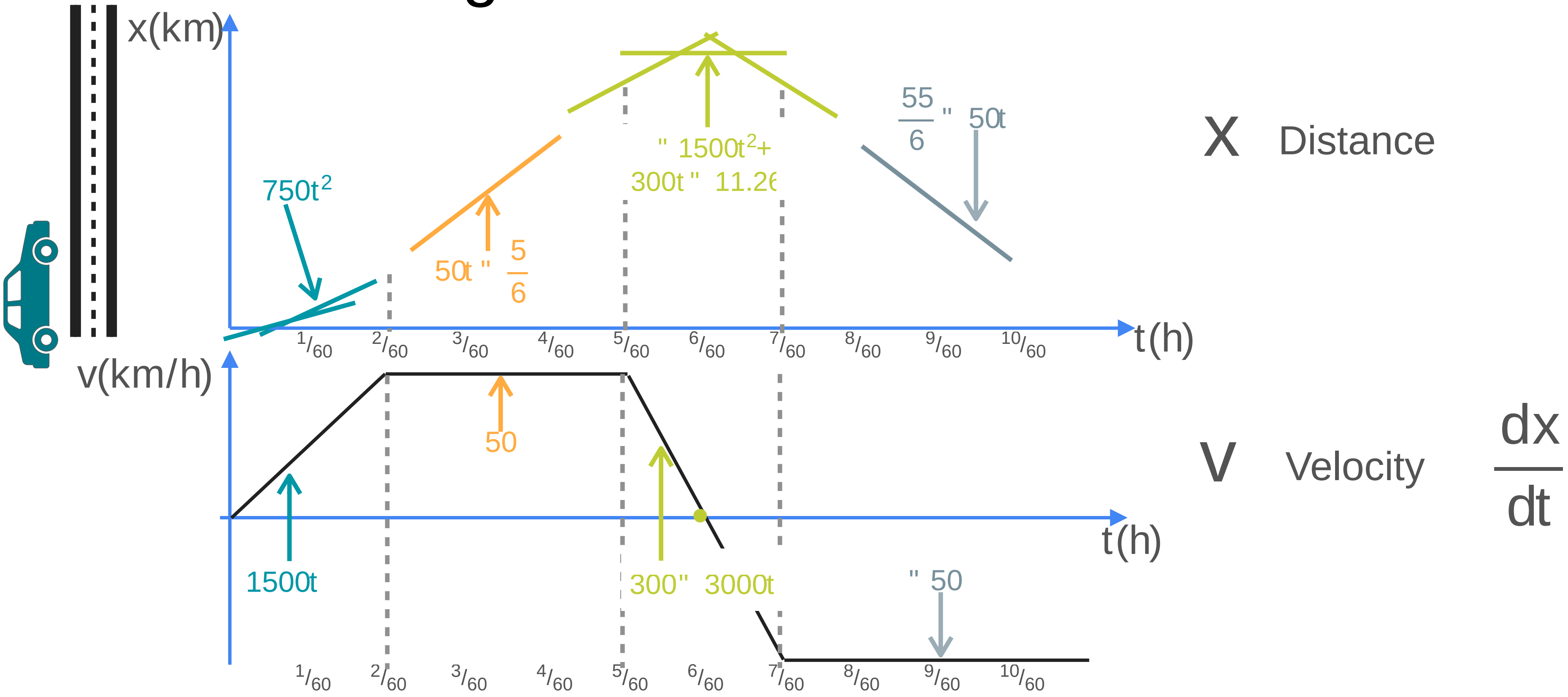
Understanding Second Derivative



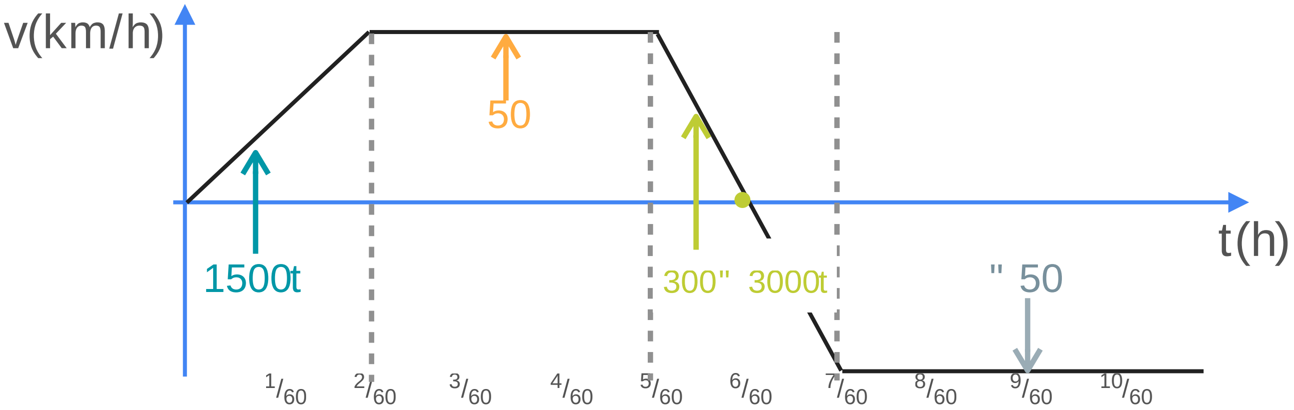
Understanding Second Derivative



Understanding Second Derivative

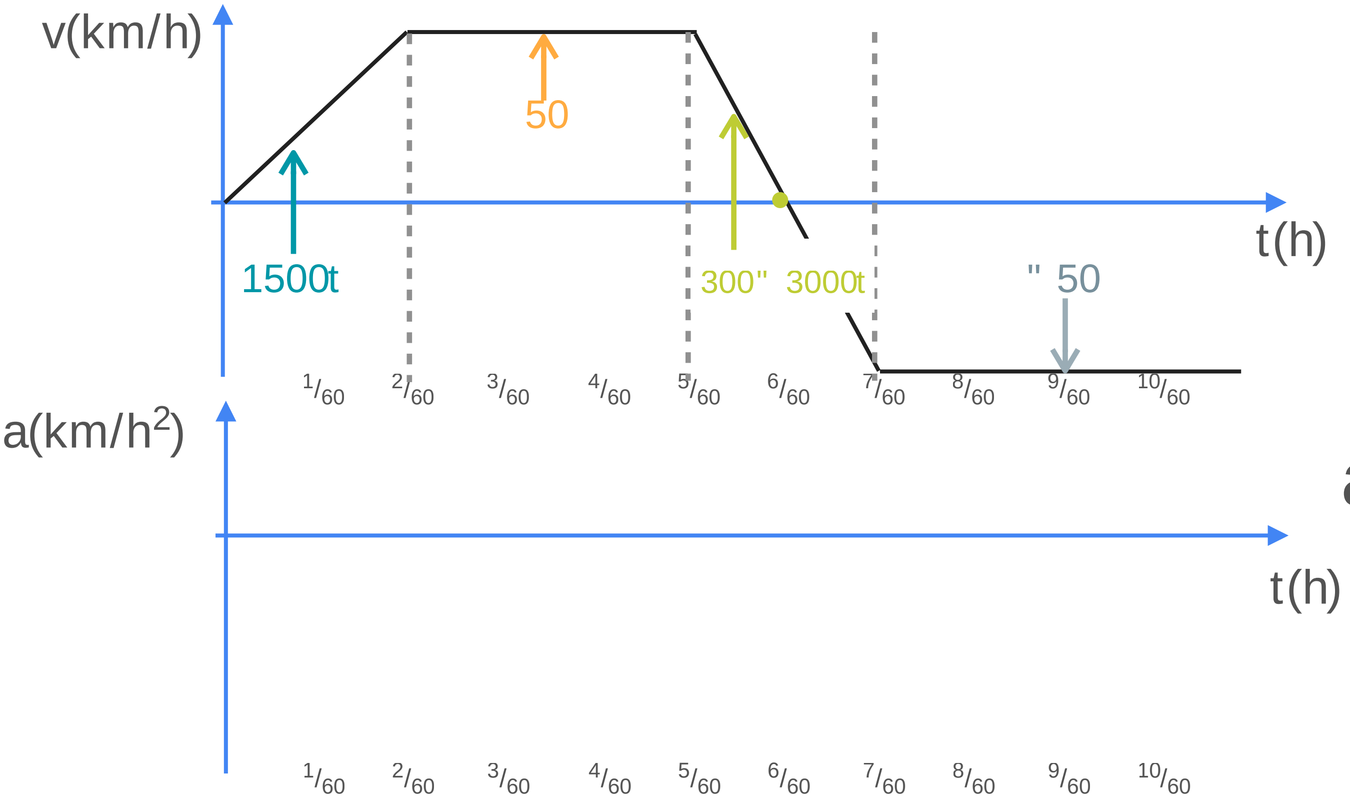


Understanding Second Derivative



V Velocity $\frac{dx}{dt}$

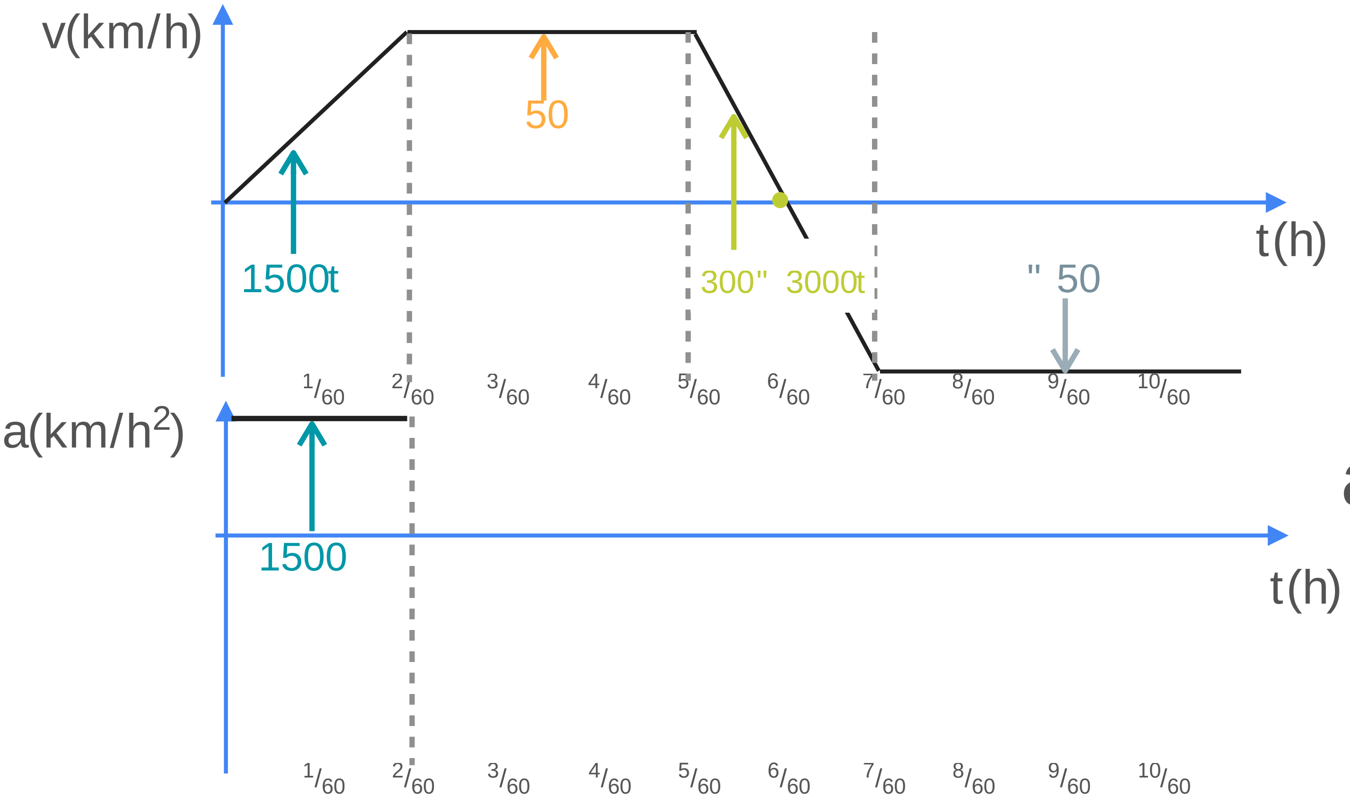
Understanding Second Derivative



v Velocity $\dfrac{dx}{dt}$

a Acceleration $\dfrac{dv}{dt}$

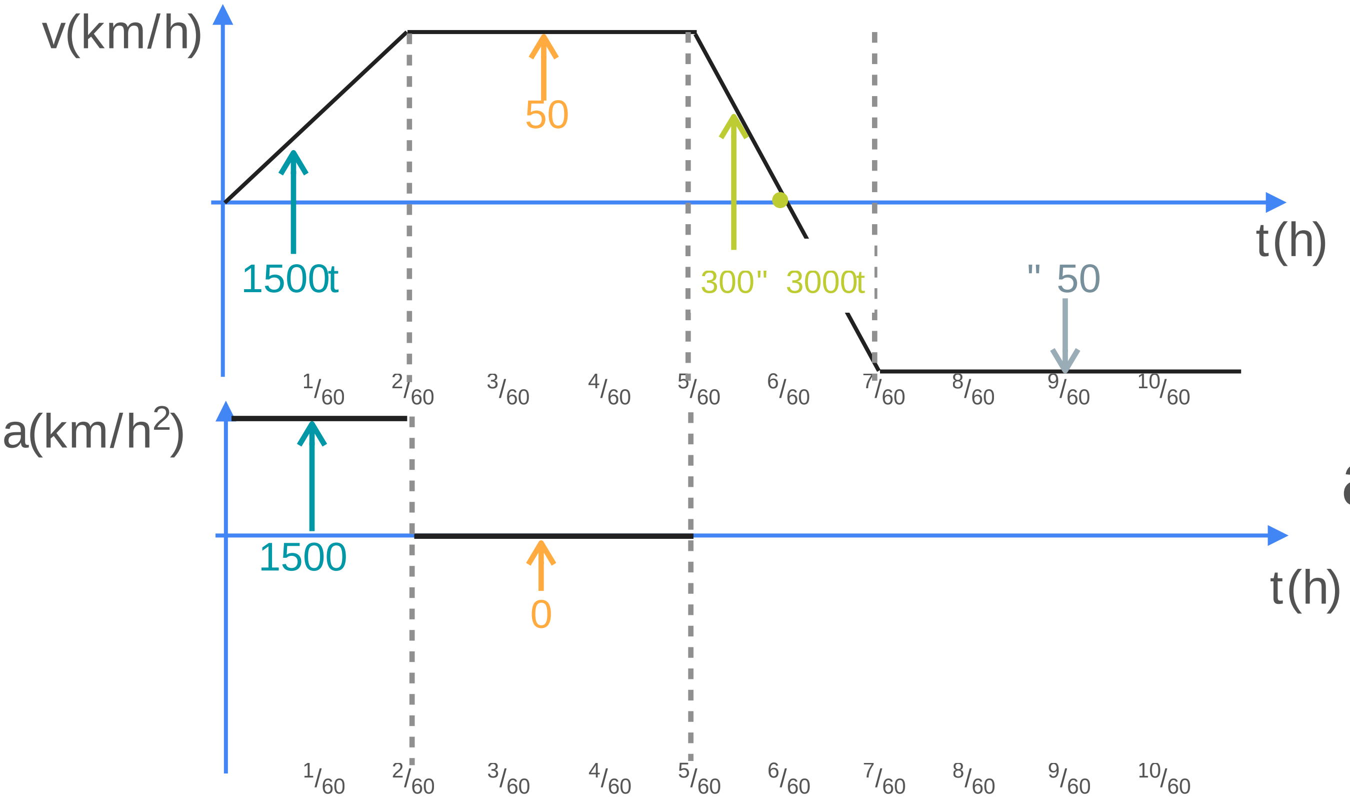
Understanding Second Derivative



v Velocity $\frac{dx}{dt}$

a Acceleration $\frac{dv}{dt}$

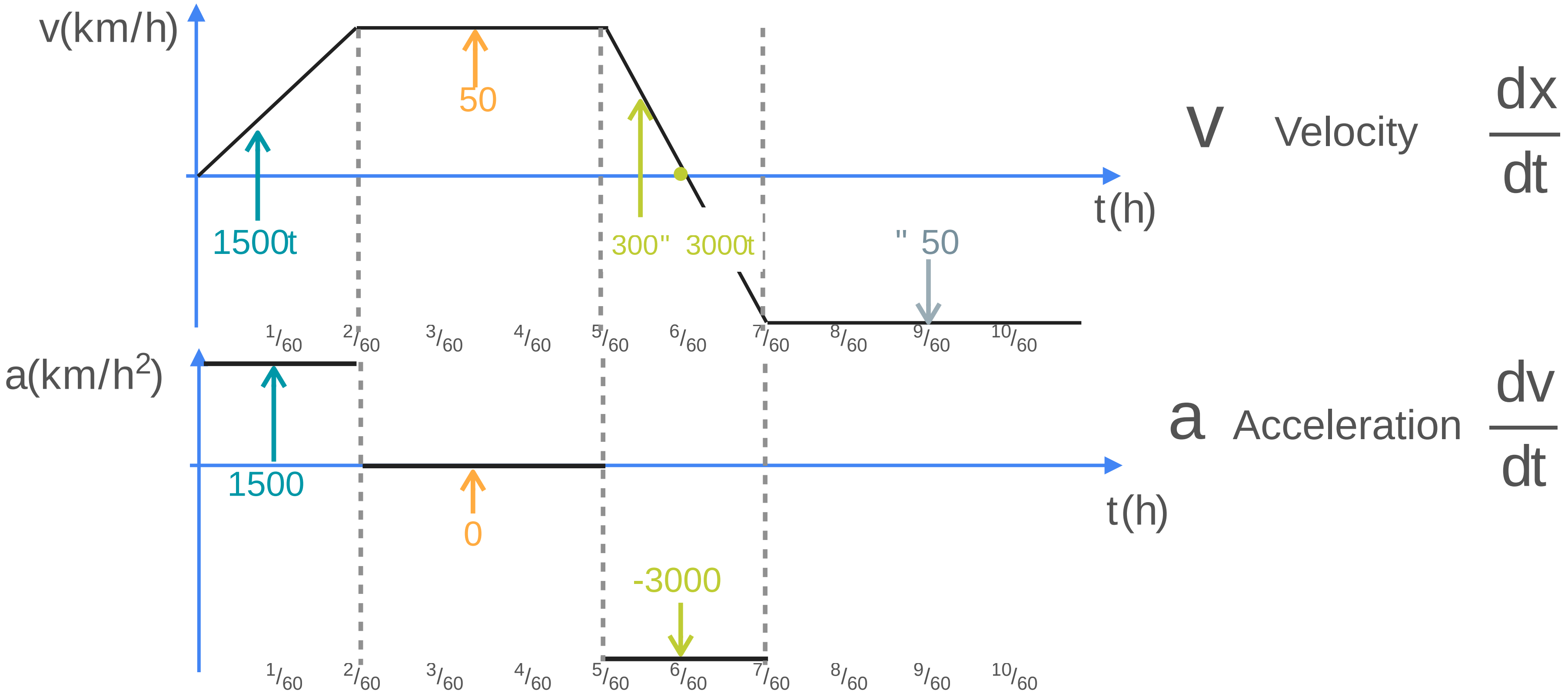
Understanding Second Derivative



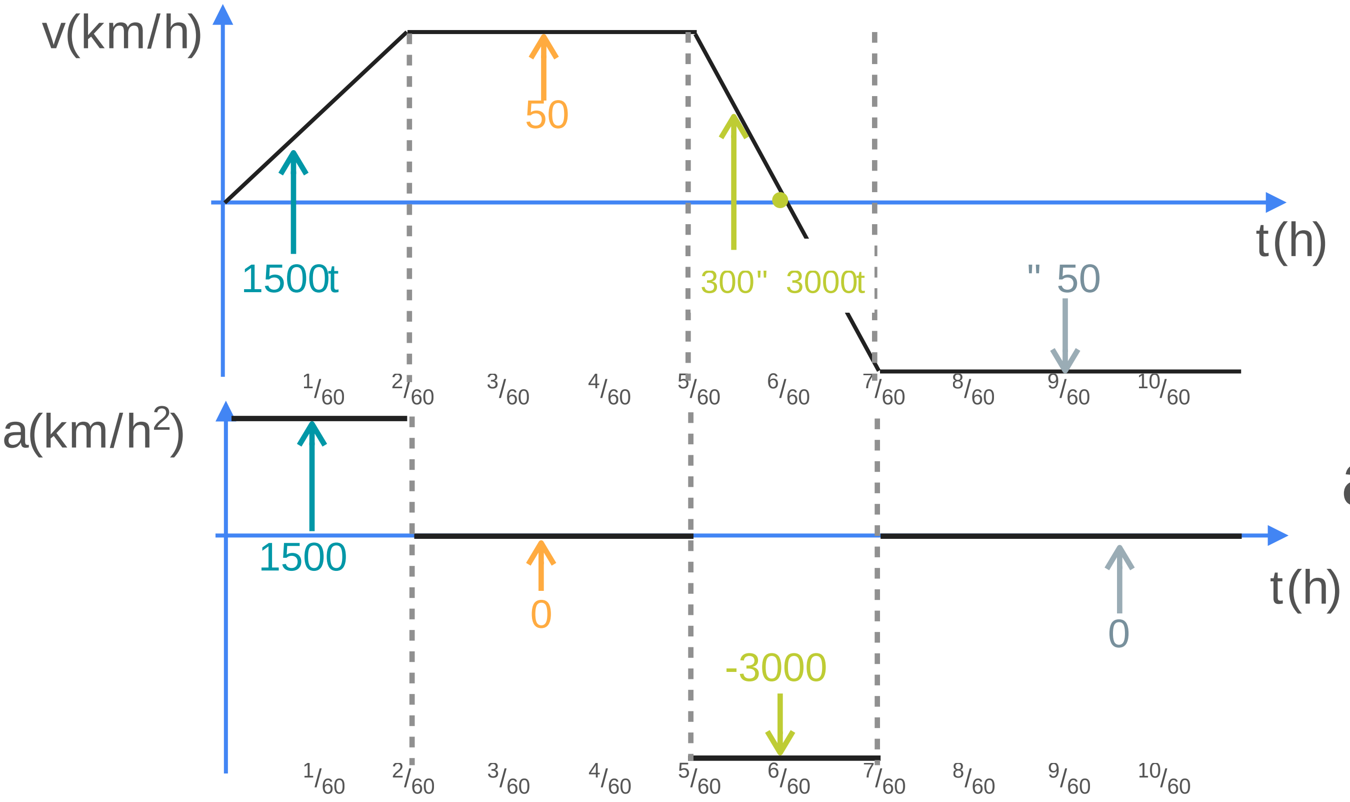
v Velocity $\frac{dx}{dt}$

a Acceleration $\frac{dv}{dt}$

Understanding Second Derivative



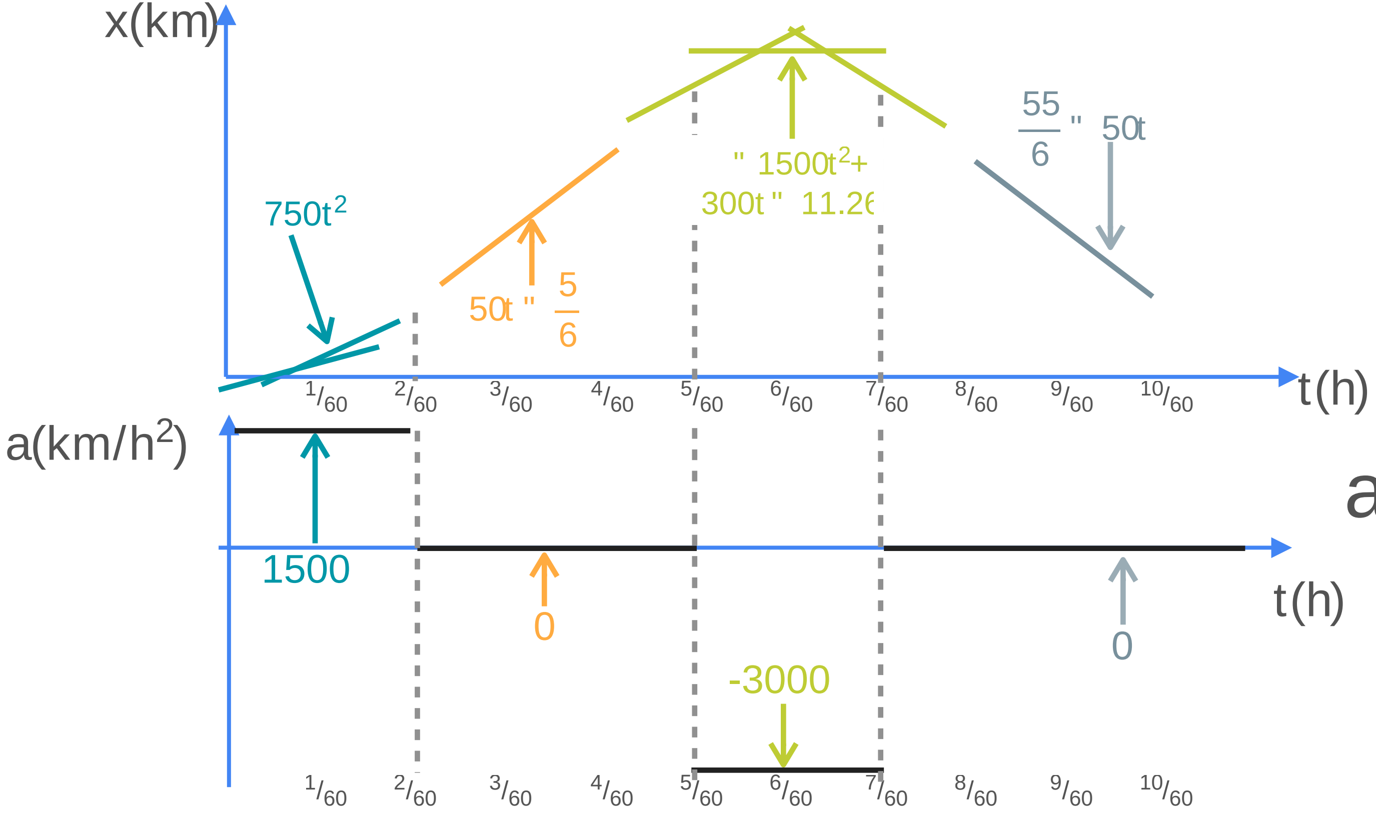
Understanding Second Derivative



v Velocity $\frac{dx}{dt}$

a Acceleration $\frac{dv}{dt}$

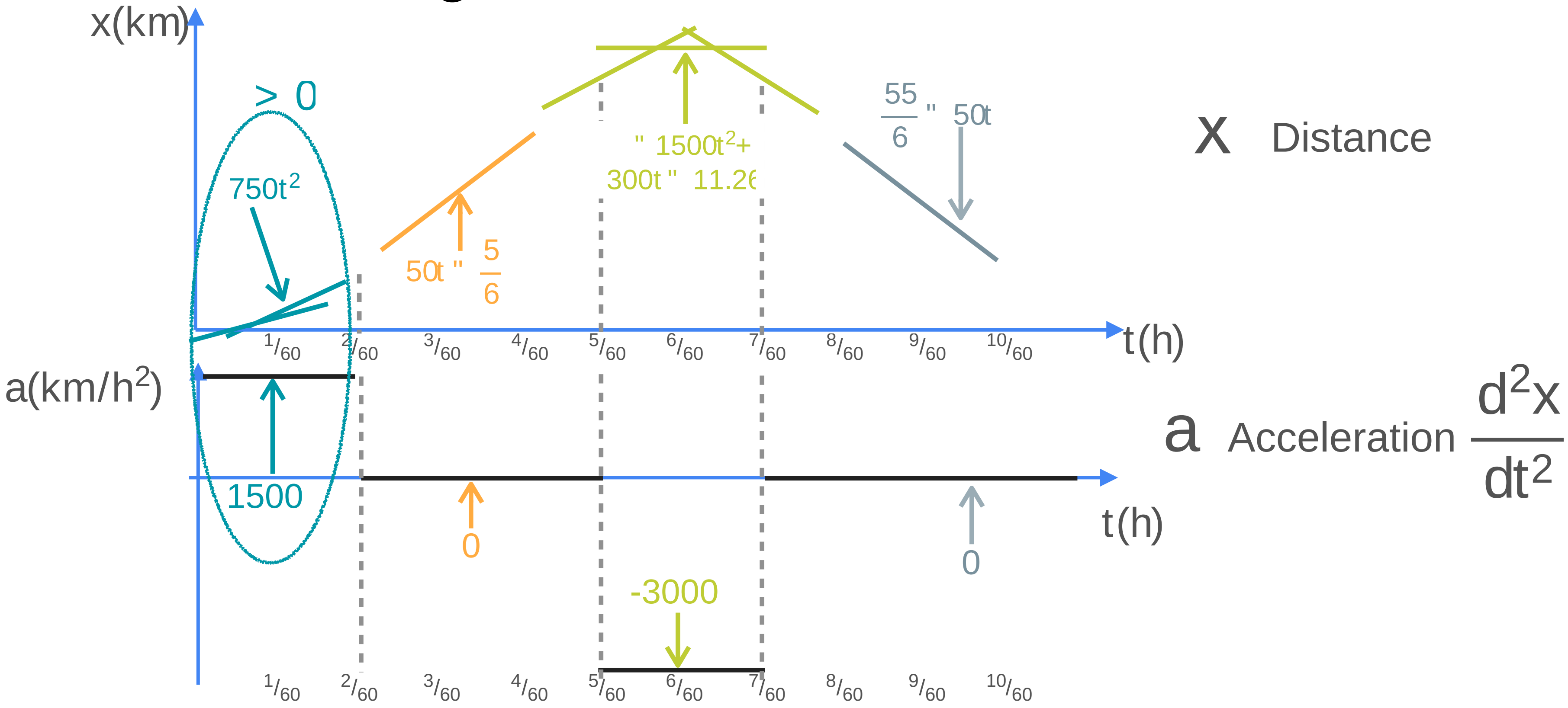
Understanding Second Derivative



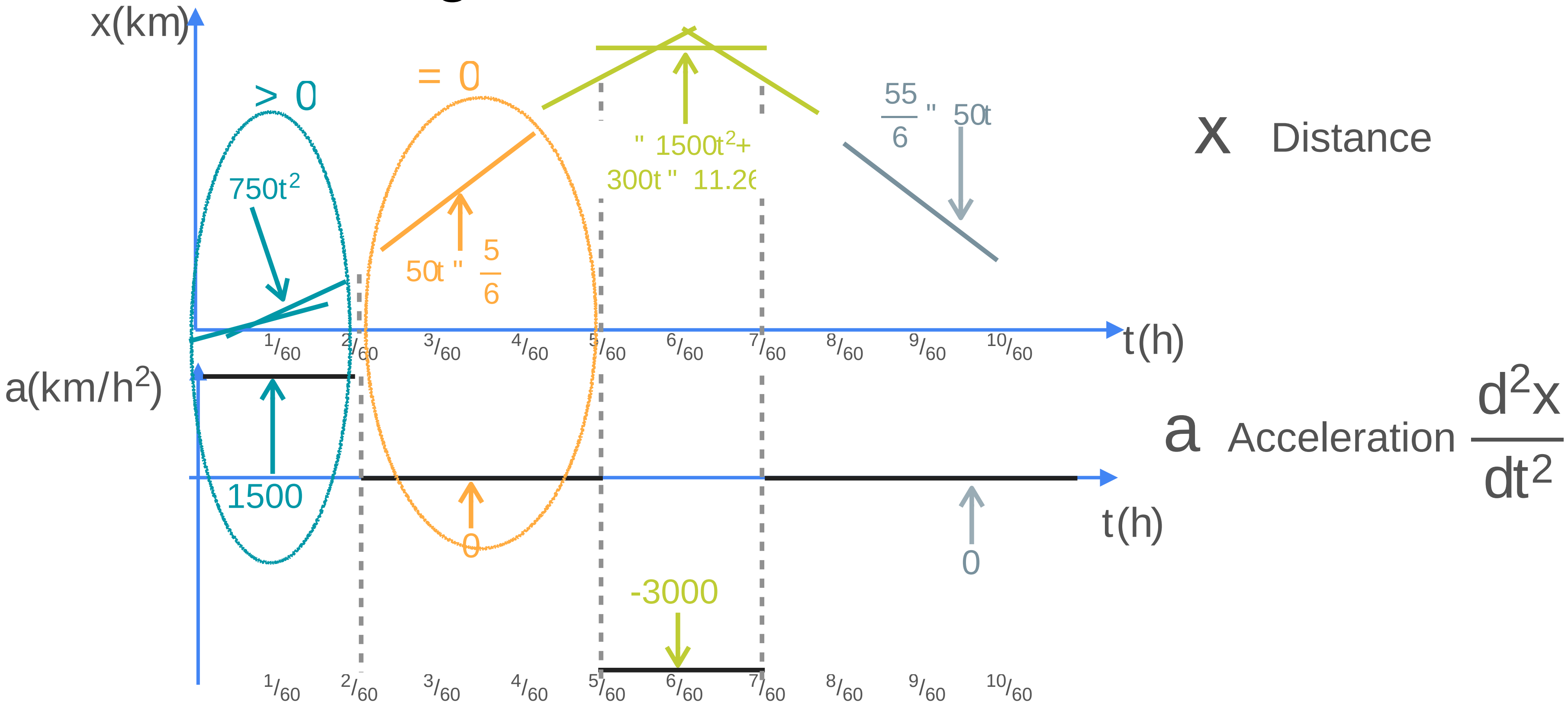
x Distance

a Acceleration $\frac{d^2x}{dt^2}$

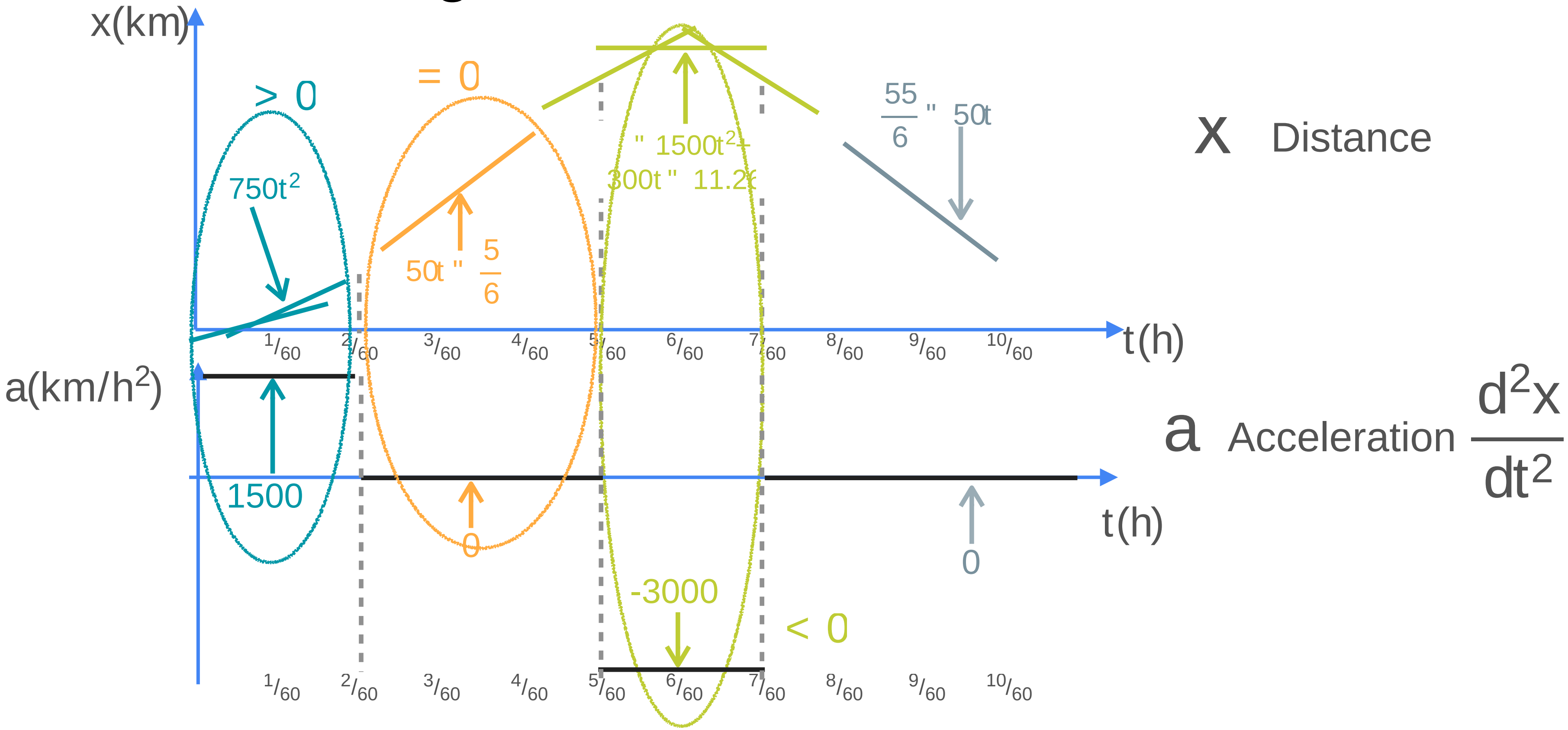
Understanding Second Derivative



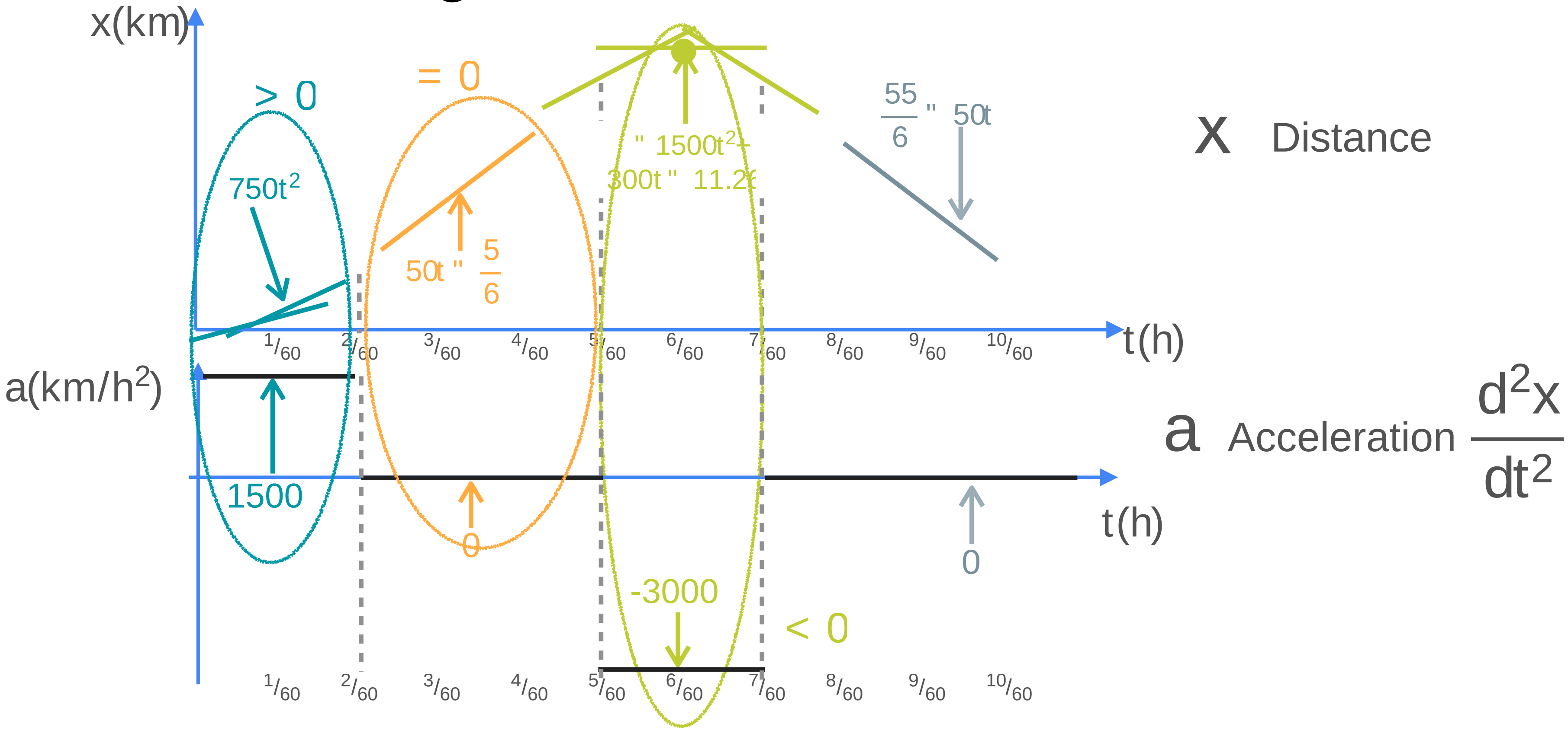
Understanding Second Derivative



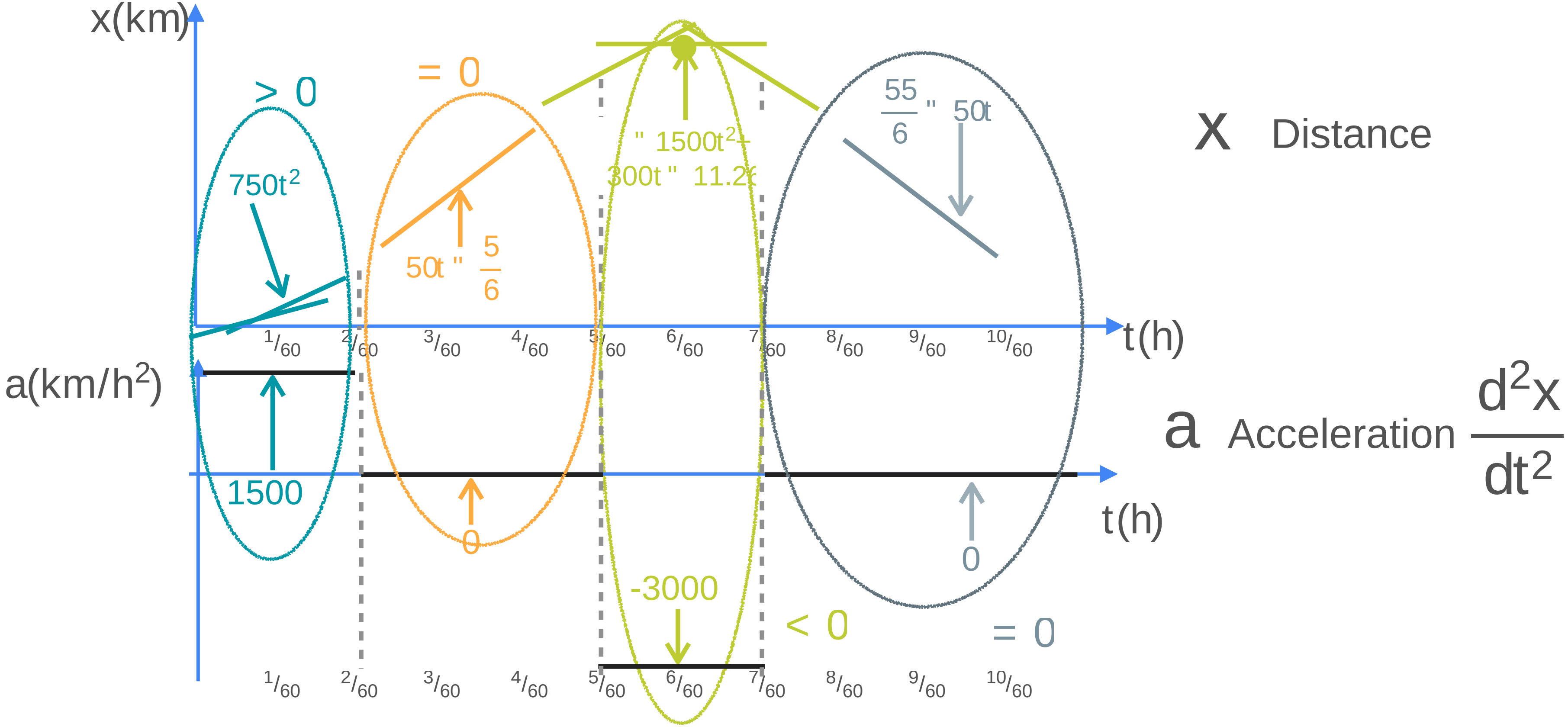
Understanding Second Derivative



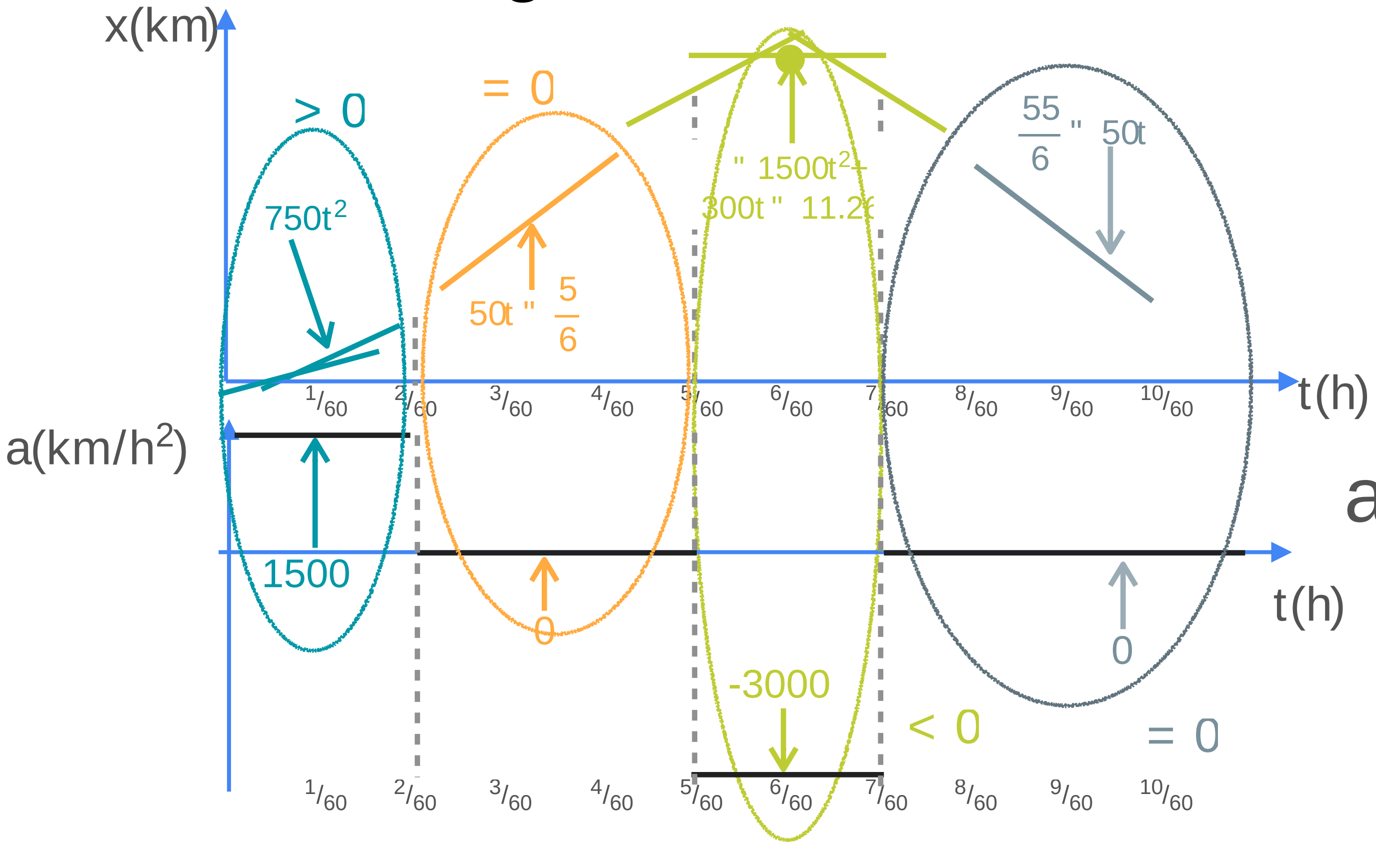
Understanding Second Derivative



Understanding Second Derivative



Understanding Second Derivative



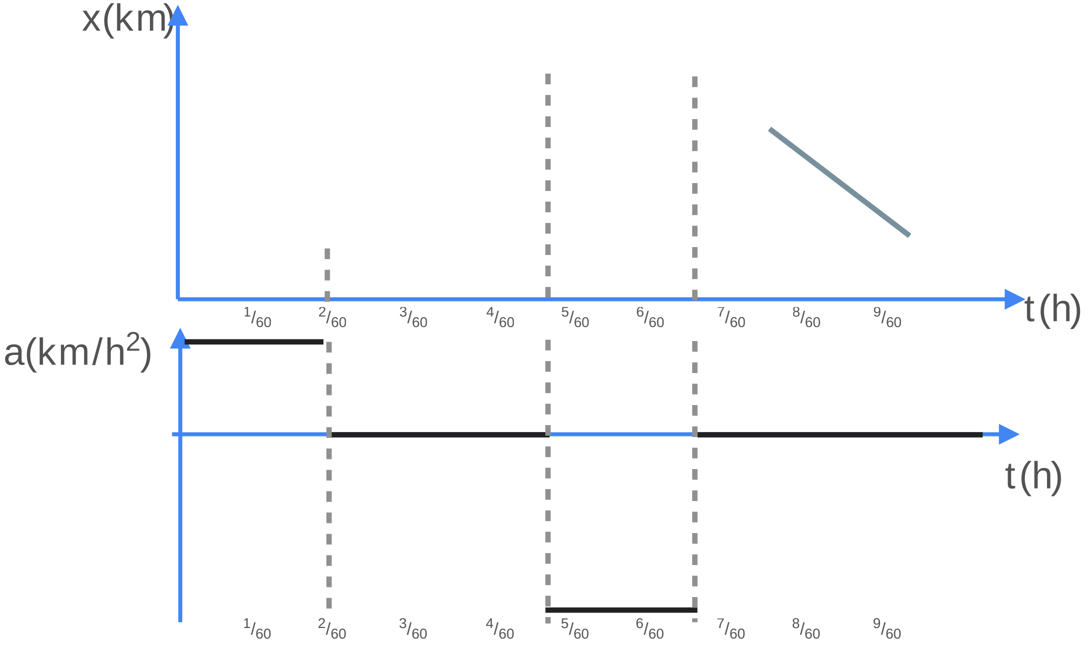
x Distance

Second derivative tells us about the curvature

a Acceleration $\frac{d^2x}{dt^2}$

Curvature

Curvature

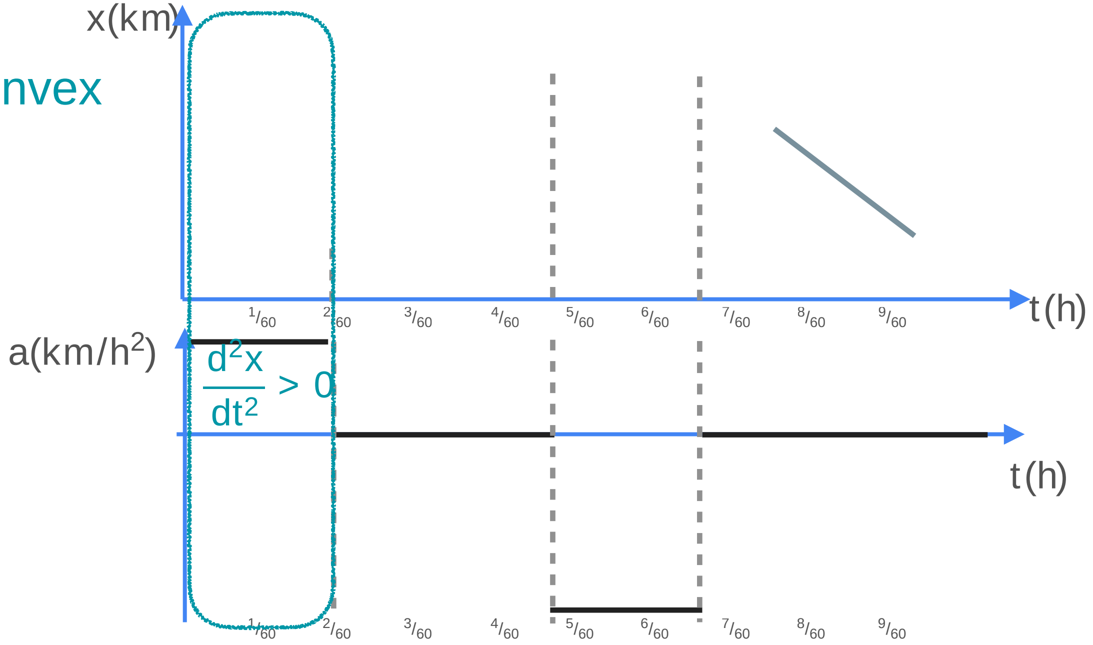


x(km)
t(h)
1/60
2/60
3/60
4/60
5/60
6/60
7/60
8/60
9/60
a(km/h2)
t(h)
1/60
2/60
3/60
4/60
5/60
6/60
7/60
8/60
9/60

Curvature

$$\frac{d^2x}{dt^2} > 0$$

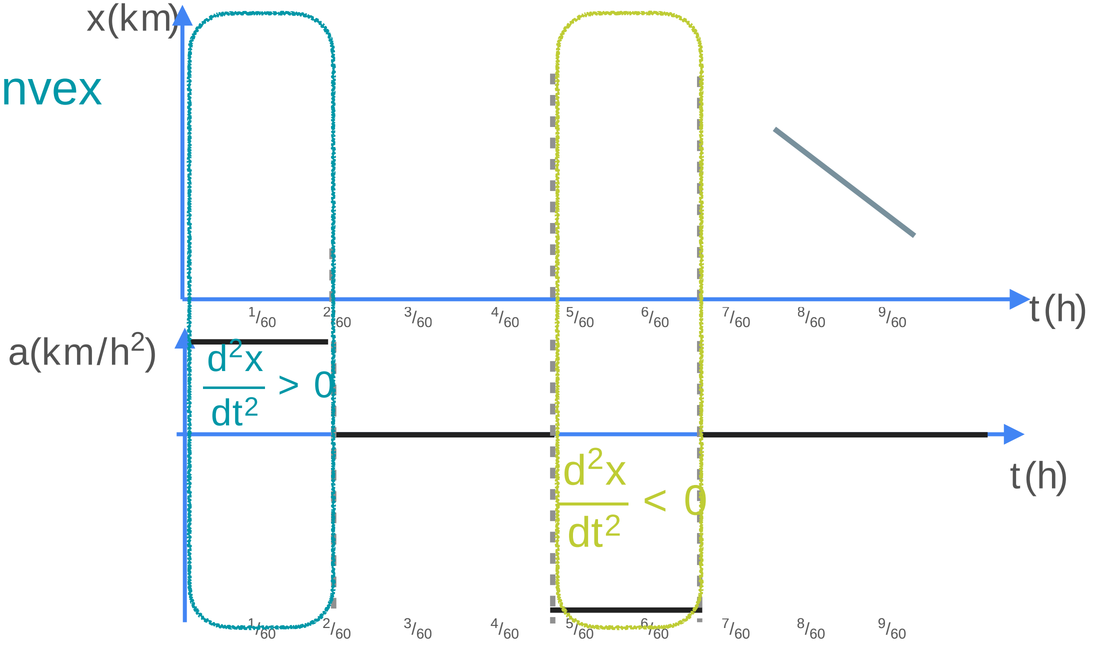
Concave up or convex



Curvature

$$\frac{d^2x}{dt^2} > 0$$ Concave up or convex

$$\frac{d^2x}{dt^2} < 0$$ Concave down

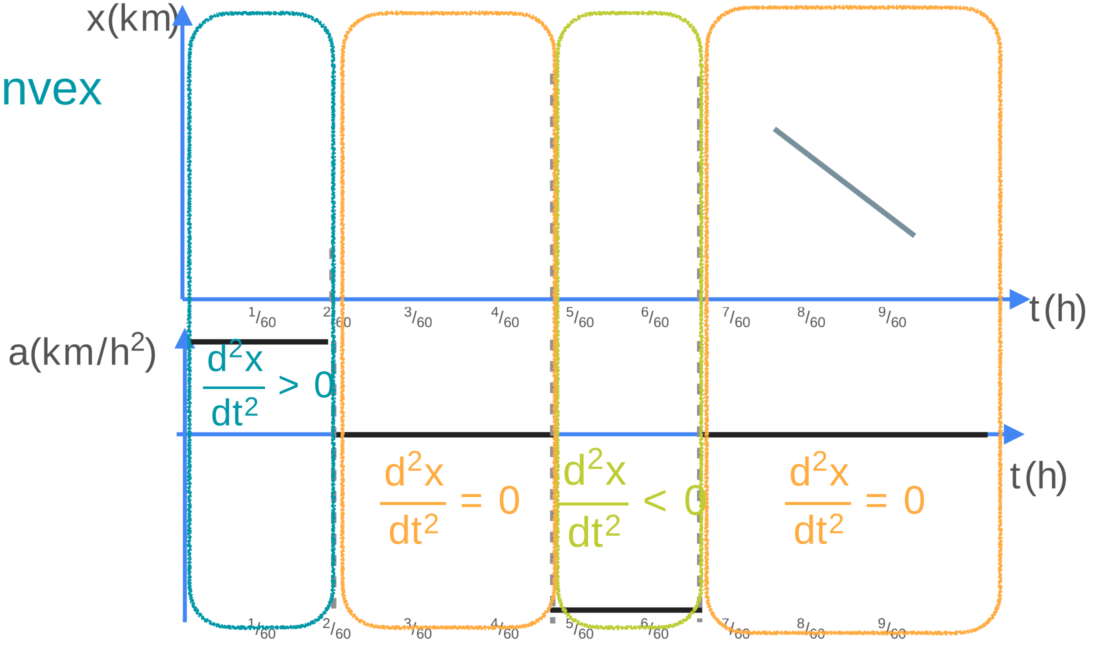


Curvature

$$\frac{d^2x}{dt^2} > 0$$ Concave up or convex

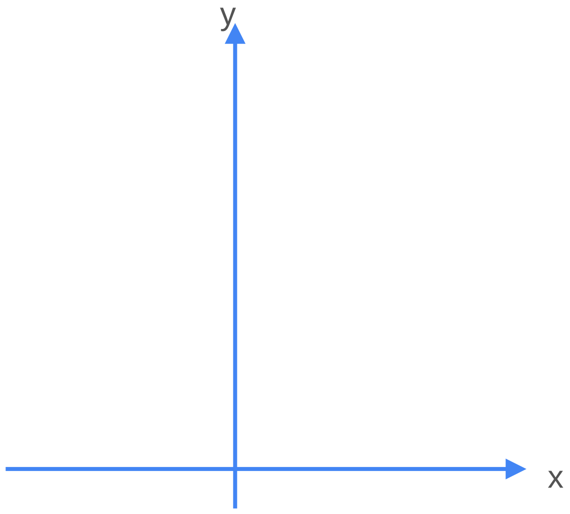
$$\frac{d^2x}{dt^2} < 0$$ Concave down

$$\frac{d^2x}{dt^2} = 0$$ Need more information

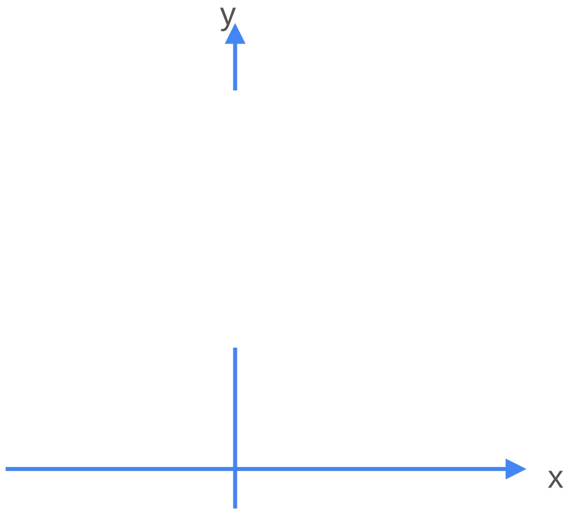


Curvature

Curvature



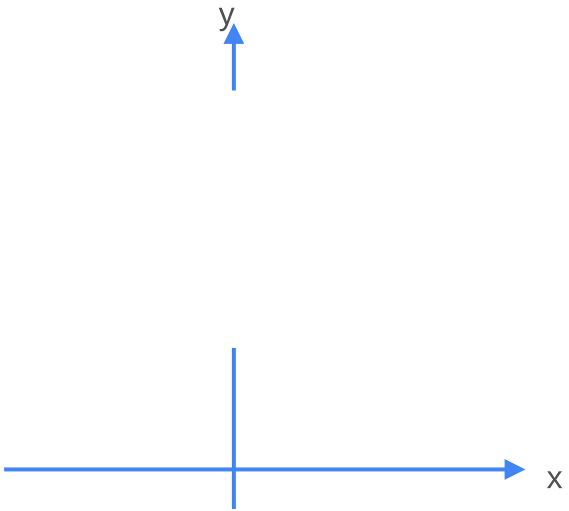
Curvature



Concave up or convex

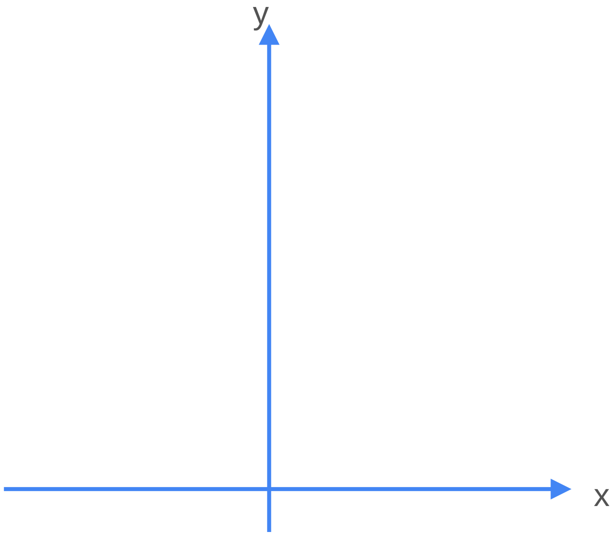
f#(0) > 0

Curvature

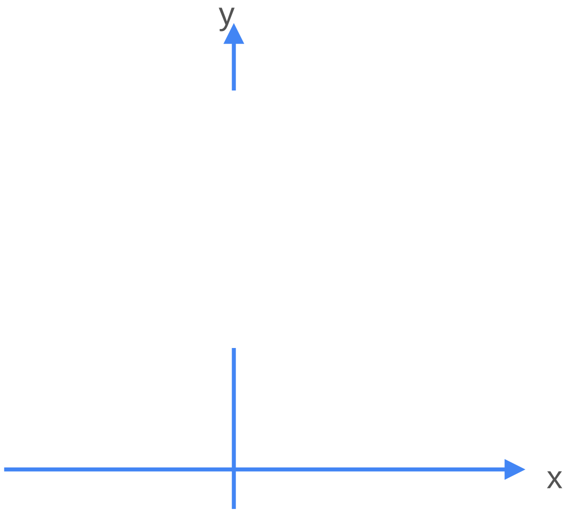


Concave up or convex

f#(0) > 0

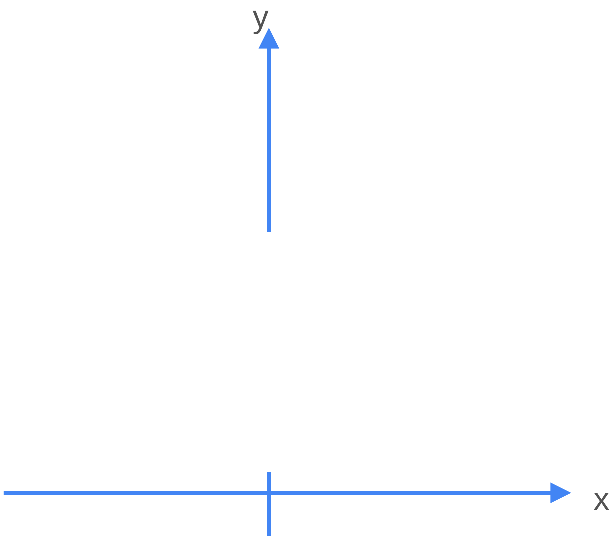


Curvature



Concave up or convex

f#(0) > 0



Concave down

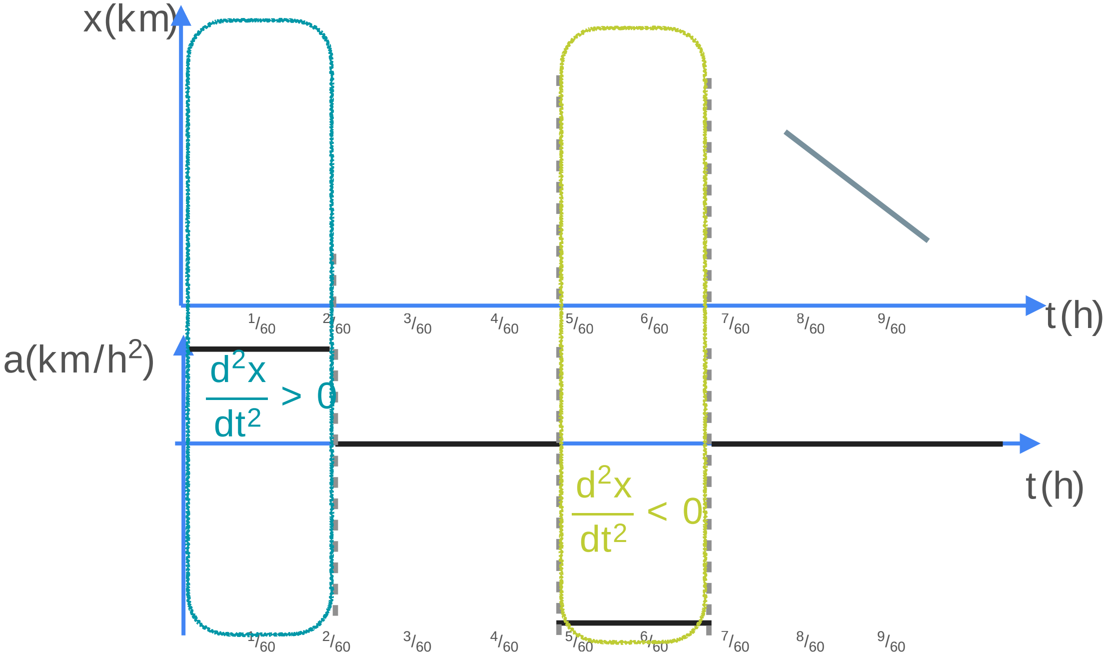
f#(0) < 0

Second Derivative and Optimization

$$\frac{d^2x}{dt^2} > 0$$

$$\frac{d^2x}{dt^2} < 0$$

$$\frac{d^2x}{dt^2} = 0$$

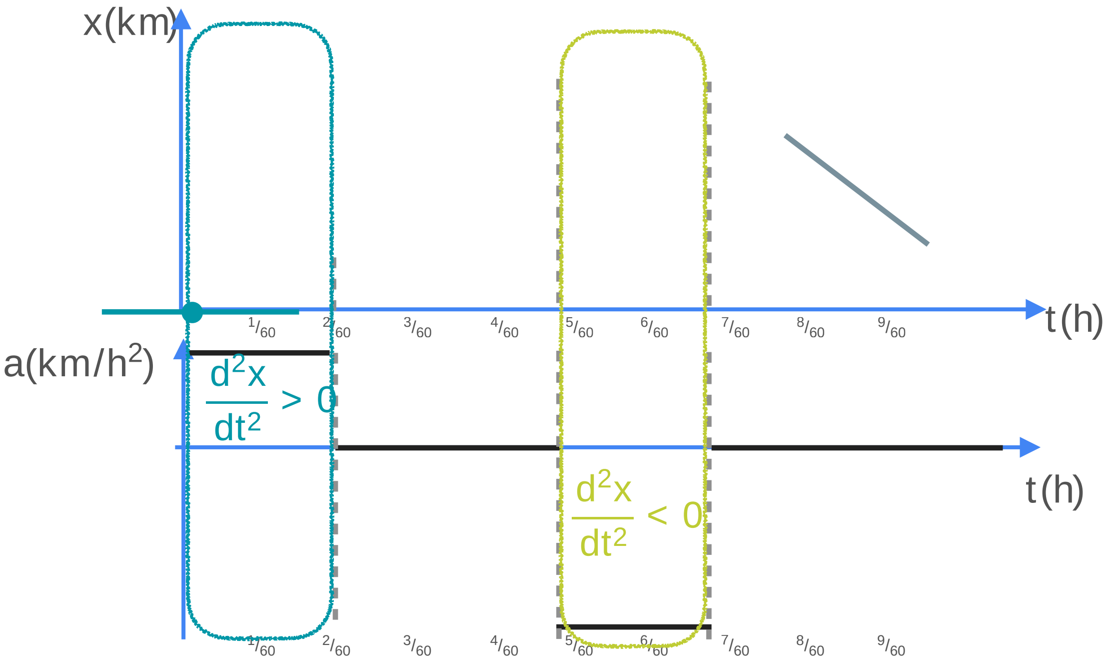


Second Derivative and Optimization

$$\frac{d^2x}{dt^2} > 0$$ (Local) Minimum

$$\frac{d^2x}{dt^2} < 0$$

$$\frac{d^2x}{dt^2} = 0$$

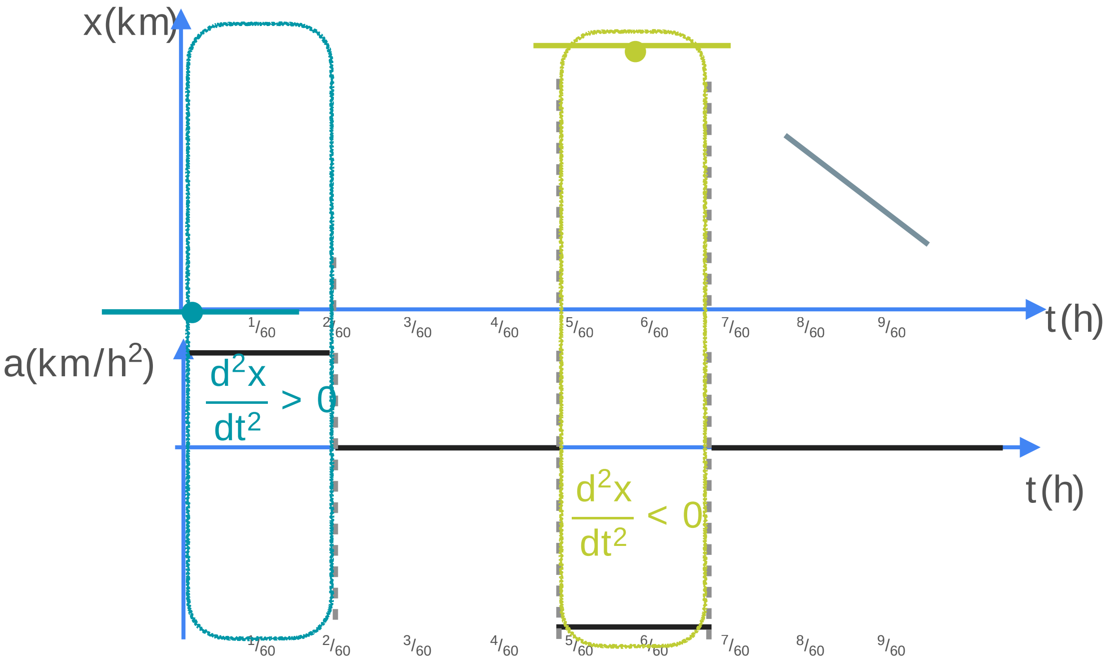


Second Derivative and Optimization

$$\frac{d^2x}{dt^2} > 0$$ (Local) Minimum

$$\frac{d^2x}{dt^2} < 0$$ (Local) maximum

$$\frac{d^2x}{dt^2} = 0$$

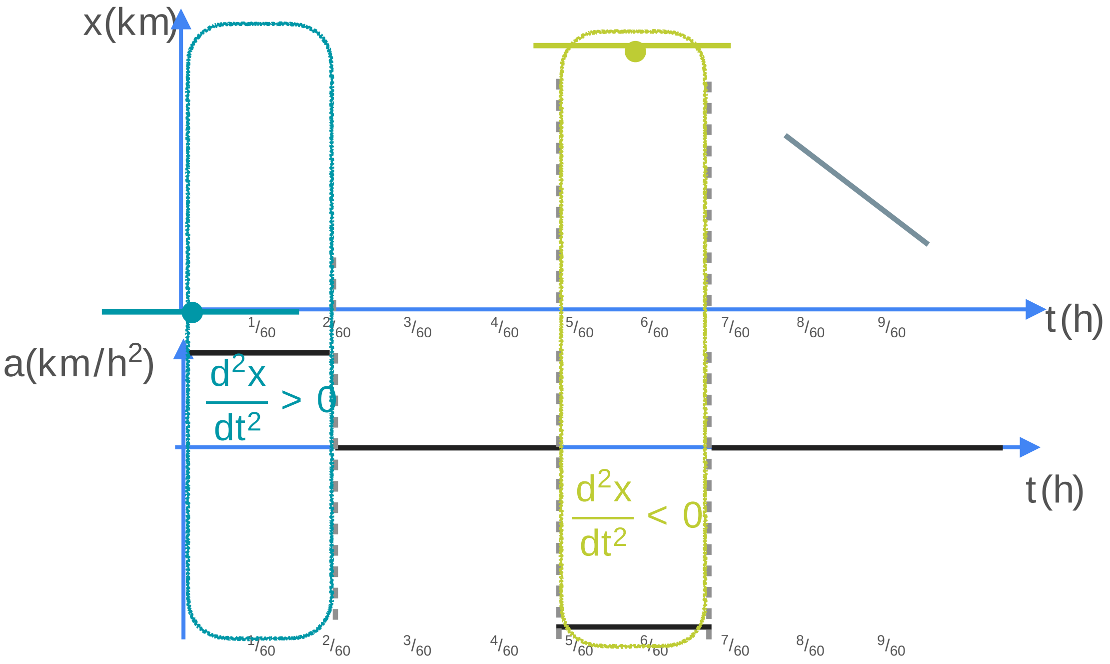


Second Derivative and Optimization

$$\frac{d^2x}{dt^2} > 0$$ (Local) Minimum

$$\frac{d^2x}{dt^2} < 0$$ (Local) maximum

$$\frac{d^2x}{dt^2} = 0$$ Inconclusive



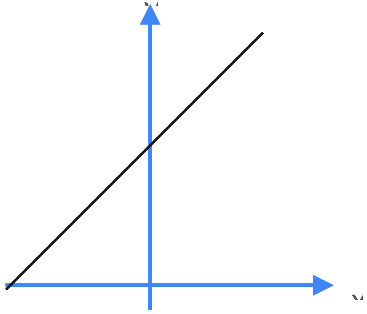
Curvature

First derivative

Second derivative

Curvature

First derivative



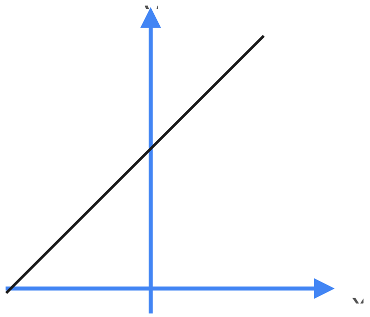
Increasing

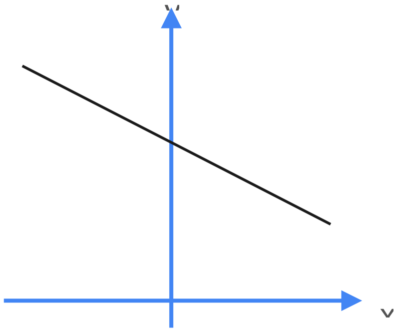
f#(0) > 0

Second derivative

Curvature

First derivative





Increasing

Decreasing

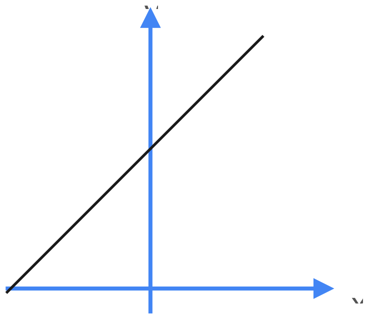
$f'(0) > 0$

$f'(0) < 0$

Second derivative

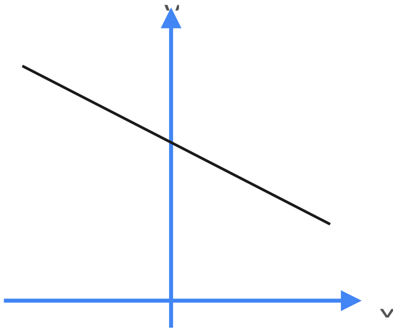
Curvature

First derivative



Increasing

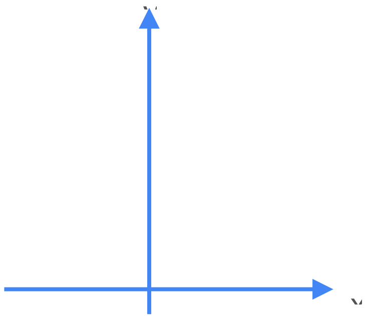
$f'(0) > 0$



Decreasing

$f'(0) < 0$

Second derivative

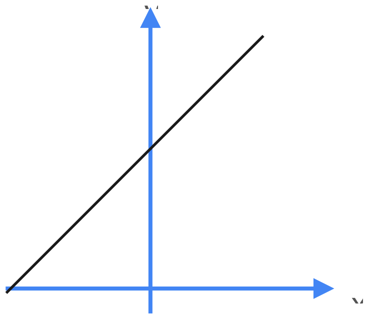


Concave up

$f''(0) > 0$

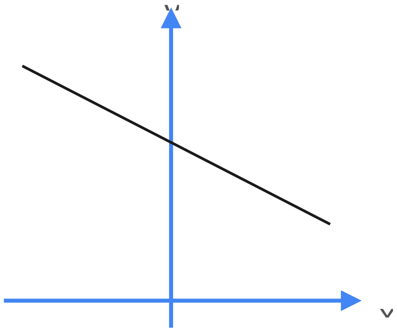
Curvature

First derivative



Increasing

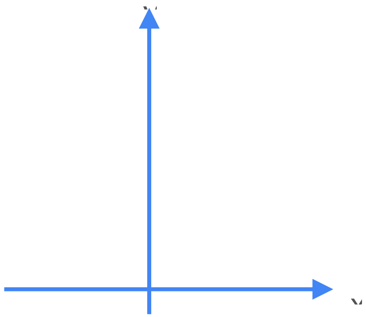
f#(0) > 0



Decreasing

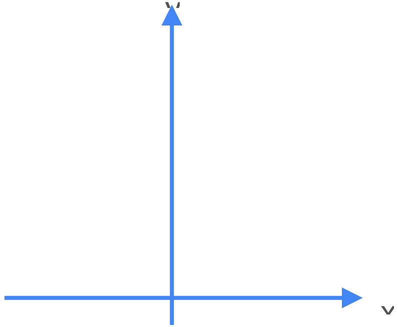
f#(0) < 0

Second derivative



Concave up

f##(0) > 0



Concave down

f##(0) < 0

Optimization in Neural Networks and NewtonÕs Method

The Hessian

Second Derivative

Second Derivative

	1 variable	2 variables

Second Derivative

	1 variable	2 variables
Function	$f(x)$	

Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$

Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f\#(x)$ Rate of change of $f(x)$	

Second Derivative

	1 variable		2 variables	
Function	$f(x)$		$f(x, y)$	
First derivative	$f'(x)$	Rate of change of $f(x)$	$f_x(x, y)$	Rate of change w.r.t x

Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t x $f_y(x, y)$ Rate of change w.r.t y

Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	f#(x) Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t x $f_y(x, y)$ Rate of change w.r.t y $- f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$

Second Derivative

	1 variable		2 variables	
Function	$f(x)$		$f(x, y)$	
First derivative	f#(x)	Rate of change of $f(x)$	$f_x(x, y)$	Rate of change w.r.t x
			$f_y(x, y)$	Rate of change w.r.t y
			$- \ f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$	
Second derivative	f##(x)	Rate of change of the rate of change of $f(x)$		

Second Derivative

	1 variable		2 variables	
Function	$f(x)$		$f(x, y)$	
First derivative	f#(x)	Rate of change of $f(x)$	$f_x(x, y)$ $f_y(x, y)$ $- \; f = \begin{matrix} f_x(x, y) \\ [\; f_y(x, y)] \end{matrix}$	Rate of change w.r.t x Rate of change w.r.t y
Second derivative	f##(x)	Rate of change of the rate of change of $f(x)$	???	

Second Derivative

Second Derivative

$$f(x, y) =$$

$$2x^2 + 3y^2 \text{ " } xy$$

Second Derivative

$$f(x, y) =$$

$$2x^2 + 3y^2 \text{ " } xy \xrightarrow{x} 4x \text{ " } y$$

Second Derivative

$f(x, y) =$

$2x^2 + 3y^2 \text{ " } xy \xrightarrow{x} 4x \text{ " } y$

$2x^2 + 3y^2 \text{ " } xy \xrightarrow{y} 6y \text{ " } x$

Second Derivative

$$\begin{tikzcd}
 & & 4 \\
f(x, y) = 2x^2 + 3y^2 - xy \arrow[r, "x"] \arrow[dr, "y"'] & 4x - y \arrow[ur, "x"] & \\
 & 6y - x &
\end{tikzcd}$$

Rate of change of $f'_x(x, y)$ w.r.t x

Second Derivative

$$
\begin{tikzcd}
 & & 4 \\
 & 4x - y \arrow[ur, "x"] \arrow[dr, "y"'] & \\
f(x, y) = 2x^2 + 3y^2 - xy \arrow[ur, "x"] \arrow[dr, "y"'] & & -1 \\
 & 6y - x &
\end{tikzcd}
$$

Rate of change of $f'_x(x, y)$ w.r.t x

Rate of change of $f'_x(x, y)$ w.r.t y

Second Derivative

$f(x, y) =$

$$
\begin{tikzcd}[row sep=small]
 & & 4 \\
 & 4x - y \arrow[ur, "x"] \arrow[dr, "y"'] & \\
2x^2 + 3y^2 - xy \arrow[ur, "x"] \arrow[dr, "y"'] & & -1 \\
 & & -1 \\
 & 6y - x \arrow[ur, "x"] &
\end{tikzcd}
$$

Rate of change of $f'_x(x, y)$ w.r.t x

Rate of change of $f'_x(x, y)$ w.r.t y

Rate of change of $f'_y(x, y)$ w.r.t x

Second Derivative

$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$\begin{tikzcd}
 & 4x - y \arrow[r, "x"] \arrow[dr, "y"'] & 4 \\
2x^2 + 3y^2 - xy \arrow[ur, "x"] \arrow[dr, "y"'] & & -1 \\
 & 6y - x \arrow[ur, "x"] \arrow[r, "y"'] & -1 \\
 & & 6
\end{tikzcd}$$

4: Rate of change of $f_x(x, y)$ w.r.t x

-1: Rate of change of $f_x(x, y)$ w.r.t y

-1: Rate of change of $f_y(x, y)$ w.r.t x

6: Rate of change of $f_y(x, y)$ w.r.t y

Second Derivative

$f(x, y) =$

$2x^2 + 3y^2 - xy$

$$
\begin{tikzcd}[column sep=large]
 & 4x - y \arrow[r, "x"] \arrow[dr, "y"'] & 4 \\
2x^2 + 3y^2 - xy \arrow[ur, "x"] \arrow[dr, "y"'] & & \boxed{\begin{matrix} -1 \\ -1 \end{matrix}} \\
 & 6y - x \arrow[ur, "x"] \arrow[r, "y"'] & 6
\end{tikzcd}
$$

4 Rate of change of $f'_x(x, y)$ w.r.t x

-1 Rate of change of $f'_x(x, y)$ w.r.t y

-1 Rate of change of $f'_y(x, y)$ w.r.t x

6 Rate of change of $f'_y(x, y)$ w.r.t y

What Do These Mean?

Rate of change of $f_x(x, y)$ w.r.t x

Rate of change of $f_y(x, y)$ w.r.t y

Rate of change of $f_x(x, y)$ w.r.t y

Rate of change of $f_y(x, y)$ w.r.t x

What Do These Mean?

Rate of change of $f_x(x, y)$ w.r.t x

Rate of change of $f_y(x, y)$ w.r.t y

Change in the change in the function w.r.t tiny changes in x and y

Rate of change of $f_x(x, y)$ w.r.t y

Rate of change of $f_y(x, y)$ w.r.t x

What Do These Mean?

Rate of change of $f_x(x, y)$ w.r.t x

Rate of change of $f_y(x, y)$ w.r.t y

Change in the change in the function w.r.t tiny changes in x and y

Same idea as with one variable!

Rate of change of $f_x(x, y)$ w.r.t y

Rate of change of $f_y(x, y)$ w.r.t x

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis

What Do These Mean?

Rate of change of $f_x(x, y)$ w.r.t x

Rate of change of $f_y(x, y)$ w.r.t y

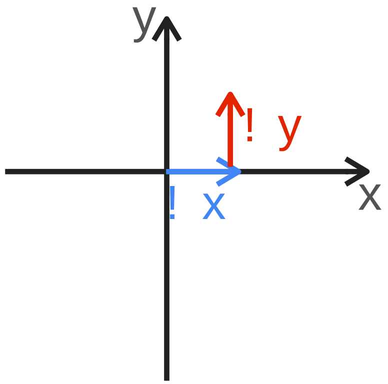
Change in the change in the function w.r.t tiny changes in x and y

Same idea as with one variable!

Rate of change of $f_x(x, y)$ w.r.t y

Rate of change of $f_y(x, y)$ w.r.t x

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis



What Do These Mean?

Rate of change of $f_x(x, y)$ w.r.t x

Rate of change of $f_y(x, y)$ w.r.t y

Change in the change in the function w.r.t tiny changes in x and y

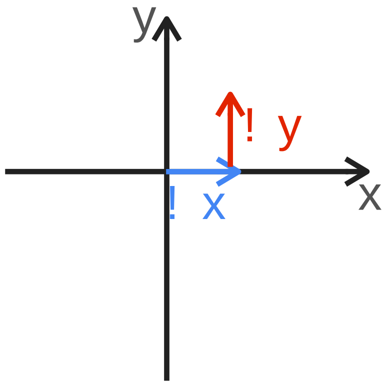
Same idea as with one variable!

Rate of change of $f_x(x, y)$ w.r.t y

Rate of change of $f_y(x, y)$ w.r.t x

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis

2. They are the same!



What Do These Meant?

Rate of change of $f_x(x, y)$ w.r.t x

Rate of change of $f_y(x, y)$ w.r.t y

Change in the change in the function w.r.t tiny changes in x and y

Same idea as with one variable!

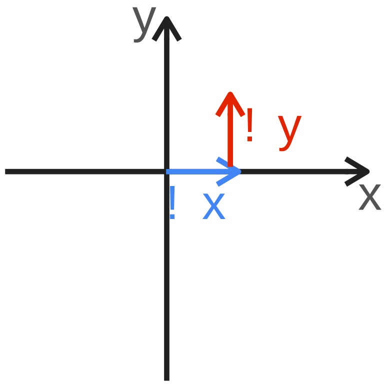
Rate of change of $f_x(x, y)$ w.r.t y

Rate of change of $f_y(x, y)$ w.r.t x

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis

2. They are the same!

(In most cases)



Notation

Rate of change of
$f\#_{x}(x, y)$ w.r.t x

Rate of change of
$f\#_{y}(x, y)$ w.r.t y

Rate of change of
$f\#_{x}(x, y)$ w.r.t y

Rate of change of
$f\#_{y}(x, y)$ w.r.t x

Notation

Leibniz's notation

Rate of change of $f_x(x, y)$ w.r.t x

Rate of change of $f_y(x, y)$ w.r.t y

Rate of change of $f_x(x, y)$ w.r.t y

Rate of change of $f_y(x, y)$ w.r.t x

Notation

	Leibniz's notation
Rate of change of $f\#_x(x, y)$ w.r.t x	$\frac{,\ ^2f}{,\ x^2}$
Rate of change of $f\#_y(x, y)$ w.r.t y	$\frac{,\ ^2f}{,\ y^2}$
Rate of change of $f\#_x(x, y)$ w.r.t y	
Rate of change of $f\#_y(x, y)$ w.r.t x	

Notation

	Leibniz's notation
Rate of change of $f\#_{x}(x, y)$ w.r.t x	$\frac{\partial^2 f}{\partial x^2}$
Rate of change of $f\#_{y}(x, y)$ w.r.t y	$\frac{\partial^2 f}{\partial y^2}$
Rate of change of $f\#_{x}(x, y)$ w.r.t y	$\frac{\partial^2 f}{\partial x, y}$
Rate of change of $f\#_{y}(x, y)$ w.r.t x	$\frac{\partial^2 f}{\partial y, x}$

Notation

	Leibniz's notation	LagrangeÕs notation
Rate of change of $f_x\#(x, y)$ w.r.t x	$\frac{,\ ^2f}{,\ x^2}$	
Rate of change of $f_y\#(x, y)$ w.r.t y	$\frac{,\ ^2f}{,\ y^2}$	
Rate of change of $f_x\#(x, y)$ w.r.t y	$\frac{,\ ^2f}{,\ x,\ y}$	
Rate of change of $f_y\#(x, y)$ w.r.t x	$\frac{,\ ^2f}{,\ y,\ x}$	

Notation

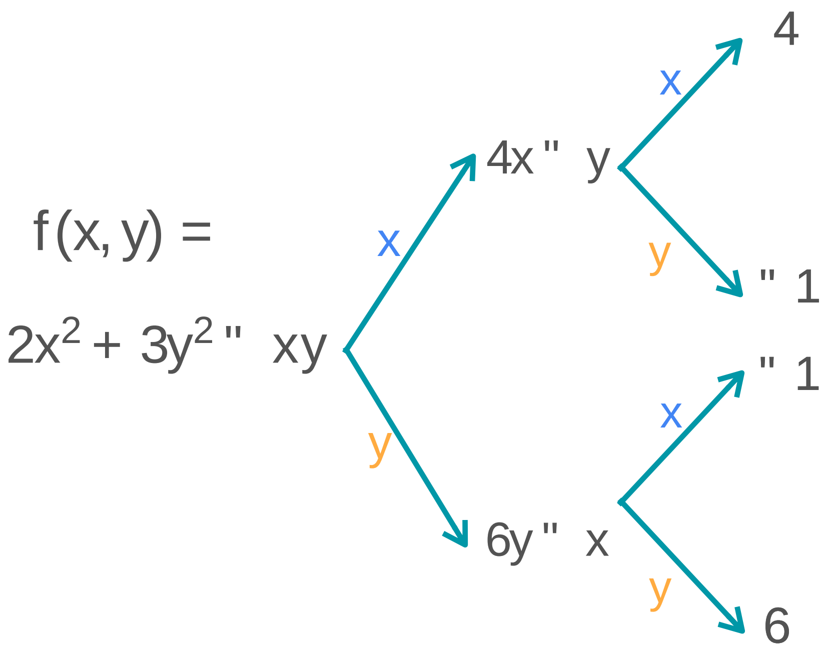
	Leibniz's notation	LagrangeÕs notation
Rate of change of f_x#(x, y) w.r.t x	$\frac{,\,^2f}{,\,x^2}$	$f_{xx}(x, y)$
Rate of change of f_y#(x, y) w.r.t y	$\frac{,\,^2f}{,\,y^2}$	$f_{yy}(x, y)$
Rate of change of f_x#(x, y) w.r.t y	$\frac{,\,^2f}{,\,x,\,y}$	
Rate of change of f_y#(x, y) w.r.t x	$\frac{,\,^2f}{,\,y,\,x}$	

Notation

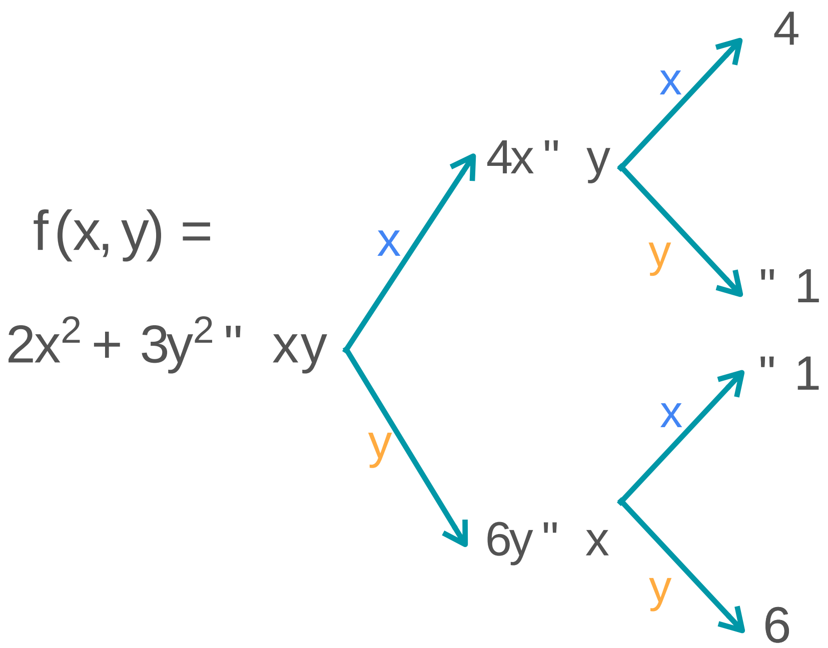
	Leibniz's notation	LagrangeÕs notation
Rate of change of f_x#(x, y) w.r.t x	$\frac{,^2 f}{, x^2}$	$f_{xx}(x, y)$
Rate of change of f_y#(x, y) w.r.t y	$\frac{,^2 f}{, y^2}$	$f_{yy}(x, y)$
Rate of change of f_x#(x, y) w.r.t y	$\frac{,^2 f}{, x, y}$	$f_{xy}(x, y)$
Rate of change of f_y#(x, y) w.r.t x	$\frac{,^2 f}{, y, x}$	$f_{yx}(x, y)$

Hessian Matrix

Hessian Matrix



Hessian Matrix



4

" 1

Hessian Matrix

$$f(x, y) =$$

$$\begin{tikzcd}[column sep=large]
 & & 4 \\
 & 4x - y \arrow[ur, "x"] \arrow[dr, "y"'] & \\
 & & -1 \\
2x^2 + 3y^2 - xy \arrow[uur, "x"] \arrow[ddr, "y"'] & & \\
 & & -1 \\
 & 6y - x \arrow[ur, "x"] \arrow[dr, "y"'] & \\
 & & 6
\end{tikzcd}$$

$$\nabla f_x = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

Hessian Matrix

$$f(x, y) =$$

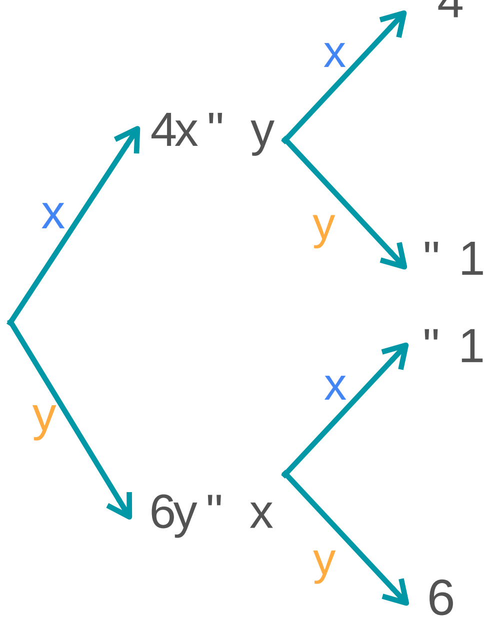
$$\begin{tikzcd}[column sep=large]
 & & 4 \\
 & 4x \text{ " } y \arrow[ur, "x"] \arrow[dr, "y"'] & \\
2x^2 + 3y^2 \text{ " } xy \arrow[ur, "x"] \arrow[dr, "y"'] & & \text{" } 1 \\
 & & \text{" } 1 \\
 & 6y \text{ " } x \arrow[ur, "x"] \arrow[dr, "y"'] & \\
 & & 6
\end{tikzcd}$$

$$- \; f_x = \begin{matrix} 4 \\ [\text{ " } 1] \end{matrix} \qquad \begin{matrix} \text{" } 1 \\ 6 \end{matrix}$$

Hessian Matrix

$f(x, y) =$

$2x^2 + 3y^2 \text{ " } xy$

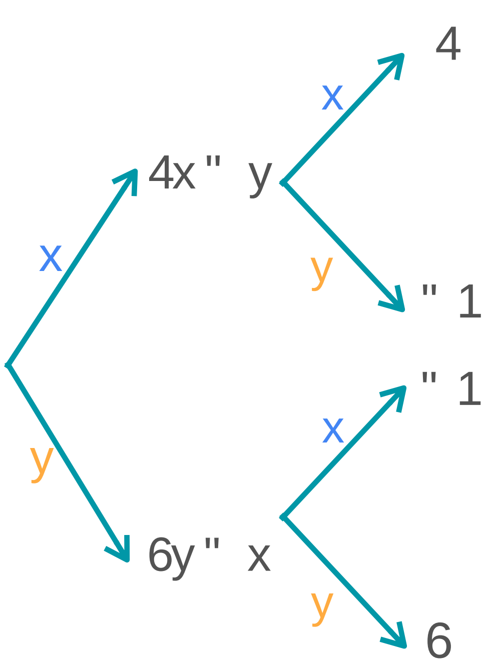


- $f_x = \begin{bmatrix} 4 \\ \text{"}\ 1 \end{bmatrix}$
- $f_y = \begin{bmatrix} \text{"}\ 1 \\ 6 \end{bmatrix}$

Hessian Matrix

$$f(x, y) =$$

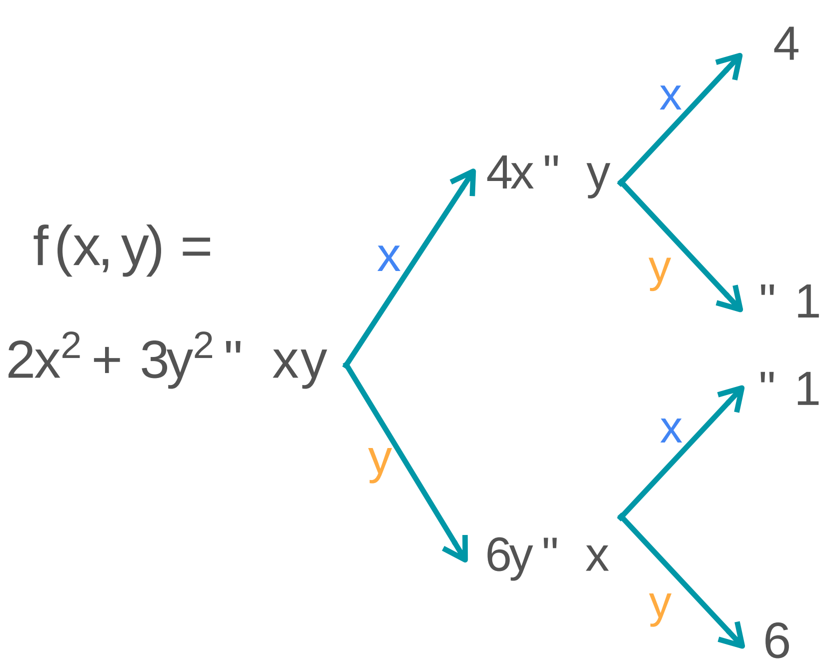
$$2x^2 + 3y^2 - xy$$



- $f_x = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$
- $f_y = \begin{bmatrix} -1 \\ 6 \end{bmatrix}$

$$\begin{matrix} 4 & -1 \\ -1 & 6 \end{matrix}$$

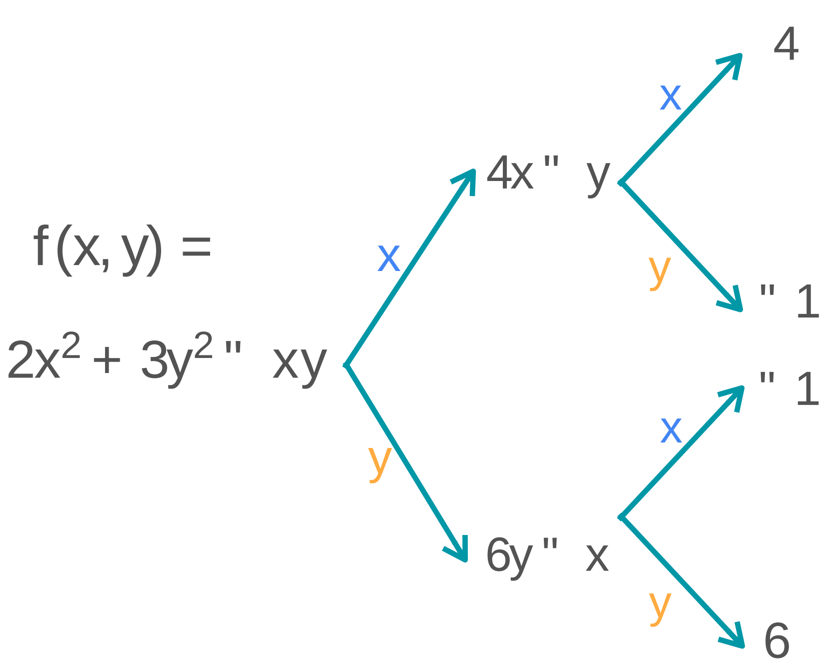
Hessian Matrix



$$- f_x = \begin{bmatrix} 4 \\ \text{"}\ 1 \end{bmatrix} \qquad - f_y = \begin{bmatrix} \text{"}\ 1 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 4 & \text{"}\ 1 \\ \text{"}\ 1 & 6 \end{bmatrix} = \begin{bmatrix} - f_x^T \\ - f_y^T \end{bmatrix}$$

Hessian Matrix



$$- f_x = \begin{bmatrix} 4 \\ " 1 \end{bmatrix} \qquad - f_y = \begin{bmatrix} " 1 \\ 6 \end{bmatrix}$$

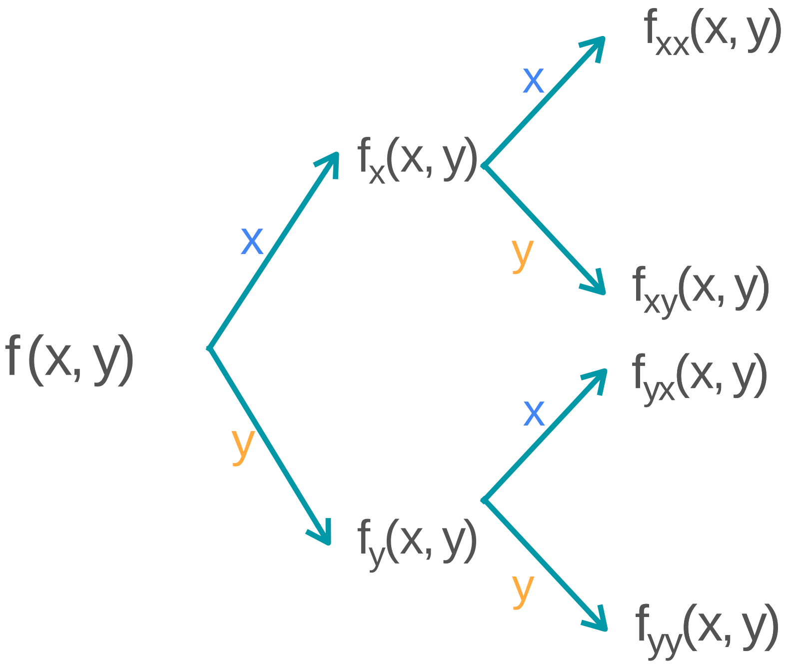
$$H = \begin{bmatrix} 4 & " 1 \\ " 1 & 6 \end{bmatrix} = \begin{bmatrix} - f_x^T \\ - f_y^T \end{bmatrix}$$

Hessian matrix

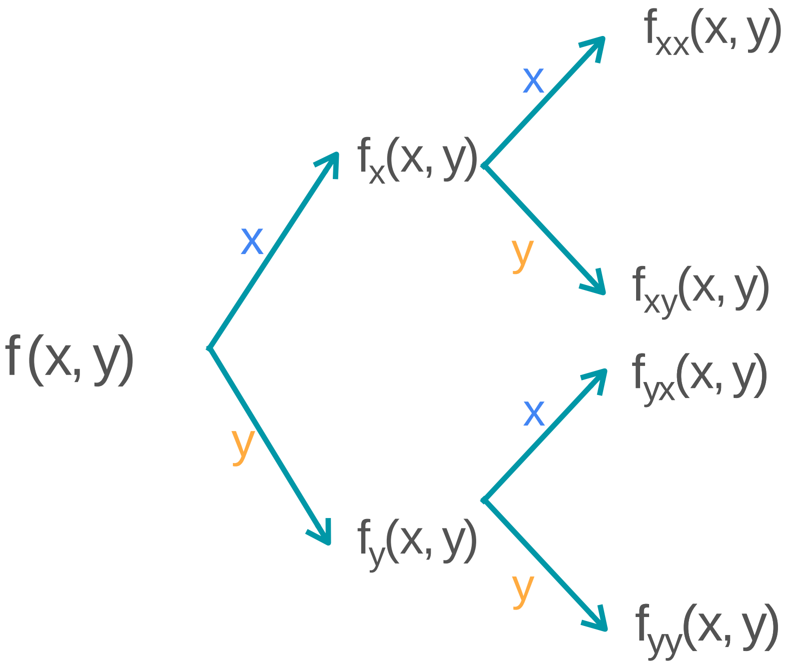
All information about second derivatives

Hessian Matrix - General Case

Hessian Matrix - General Case



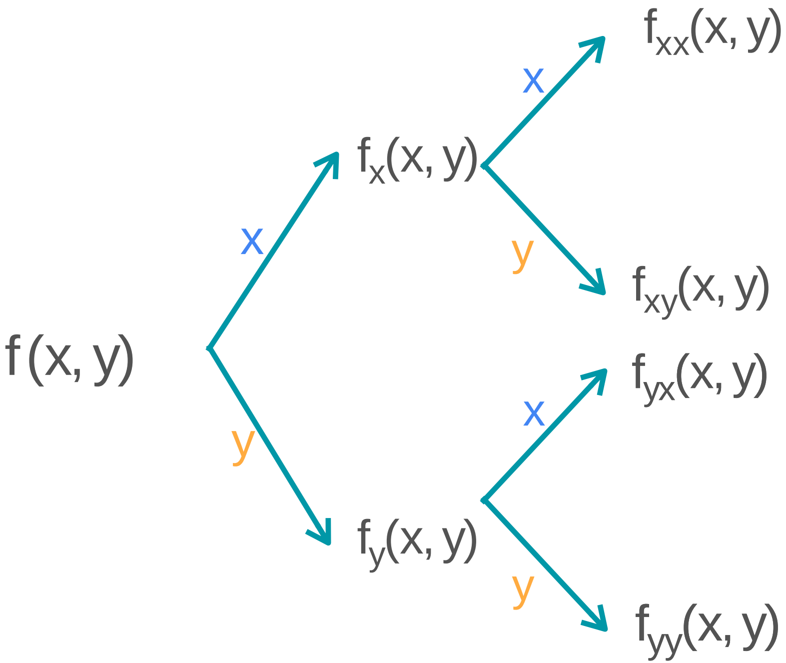
Hessian Matrix - General Case



$f_{xx}(x, y)$

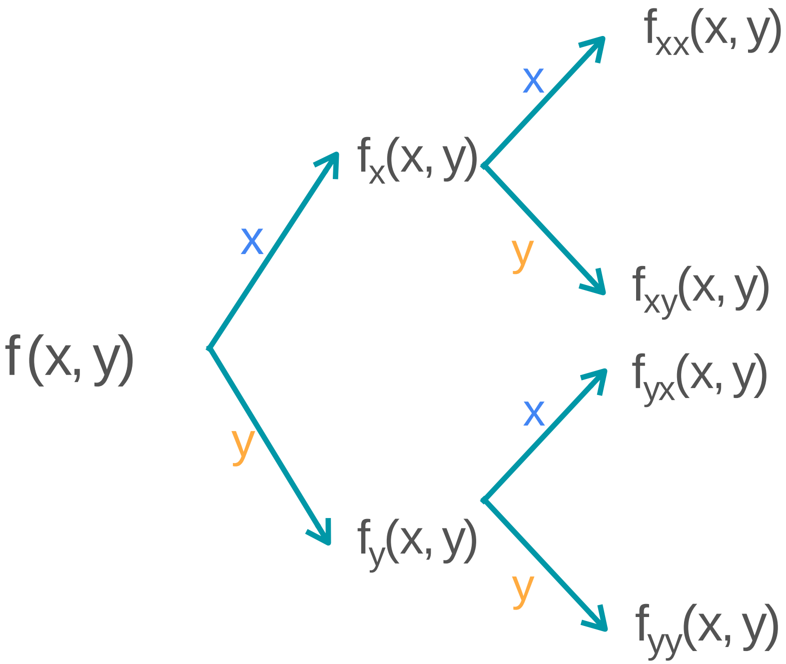
$f_{xy}(x, y)$

Hessian Matrix - General Case



- $f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix}$

Hessian Matrix - General Case

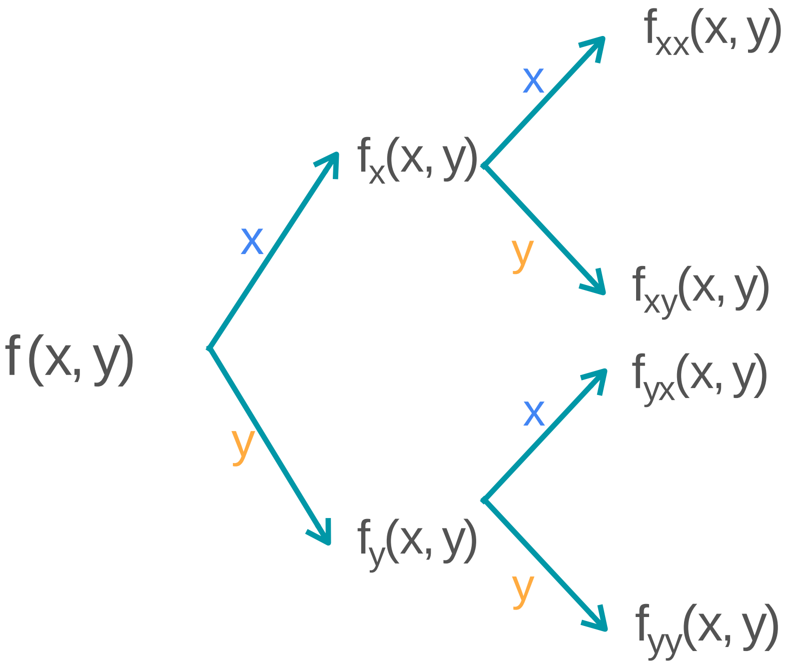


- $f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix}$

$f_{yx}(x, y)$

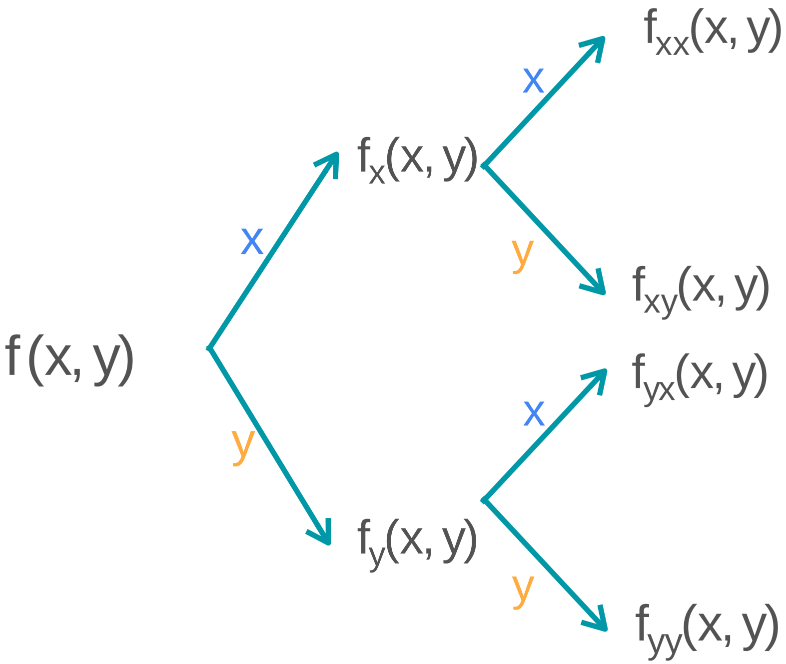
$f_{yy}(x, y)$

Hessian Matrix - General Case



- $f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix}$
- $f_y = \begin{bmatrix} f_{yx}(x, y) \\ f_{yy}(x, y) \end{bmatrix}$

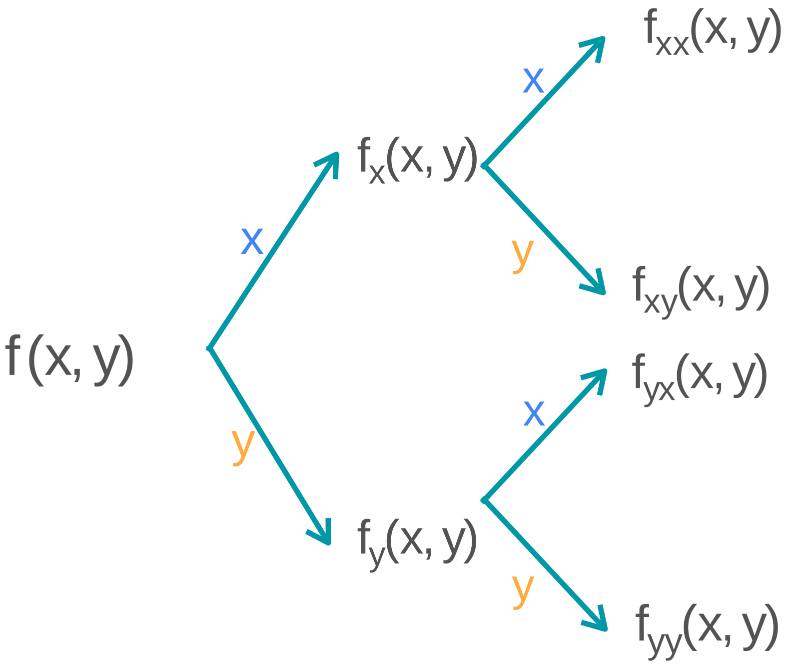
Hessian Matrix - General Case



- $f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix}$
- $f_y = \begin{bmatrix} f_{yx}(x, y) \\ f_{yy}(x, y) \end{bmatrix}$

$$\begin{matrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{matrix}$$

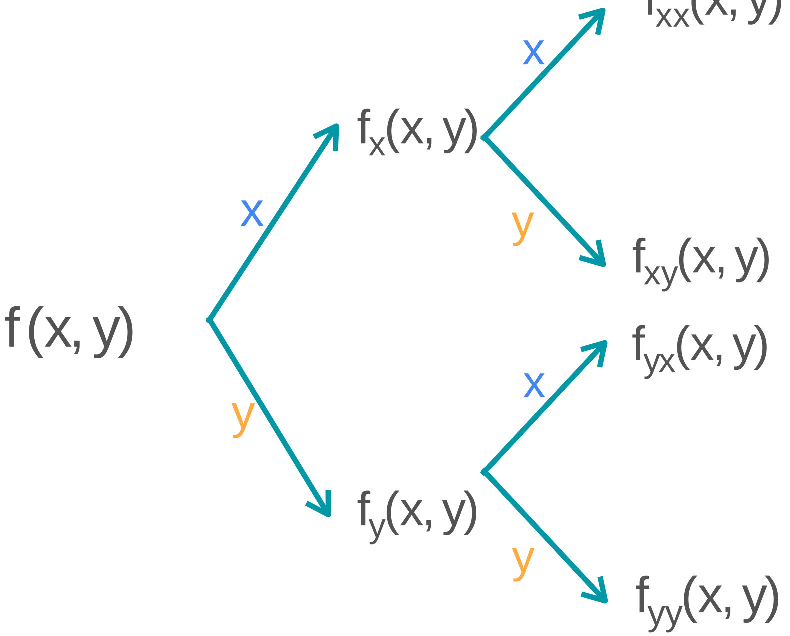
Hessian Matrix - General Case



$- \ f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix}$ $\qquad$ $- \ f_y = \begin{bmatrix} f_{yx}(x, y) \\ f_{yy}(x, y) \end{bmatrix}$

$$\begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix} = \begin{bmatrix} - \ f_x^T \\ - \ f_y^T \end{bmatrix}$$

Hessian Matrix - General Case



$$- f_x = \begin{bmatrix} f_{xx}(x,y) \\ f_{xy}(x,y) \end{bmatrix} \qquad - f_y = \begin{bmatrix} f_{yx}(x,y) \\ f_{yy}(x,y) \end{bmatrix}$$

$$H = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix} = \begin{bmatrix} - f_x^T \\ - f_y^T \end{bmatrix}$$

Hessian matrix

All information about second derivatives

Second Derivative

Second Derivative

	1 variable		2 variables	
Function	$f(x)$		$f(x, y)$	
First derivative	$f'(x)$	Rate of change of $f(x)$	$f_x(x, y)$	Rate of change w.r.t x
			$f_y(x, y)$	Rate of change w.r.t y
			$\nabla f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$	
Second derivative	$f''(x)$	Rate of change of the rate of change of $f(x)$		

Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	f#(x) Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t x $f_y(x, y)$ Rate of change w.r.t y - f = [$f_x(x, y)$ $f_y(x, y)$]
Second derivative	f##(x) Rate of change of the rate of change of $f(x)$	H(x, y) = [$f_{xx}(x, y)$ $f_{xy}(x, y)$ $f_{yx}(x, y)$ $f_{yy}(x, y)$]

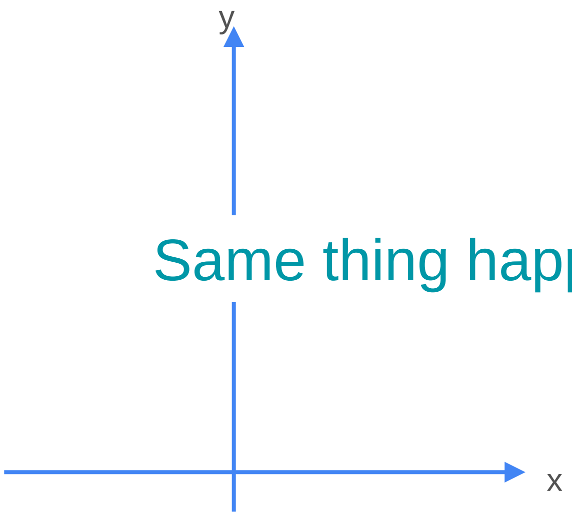
Optimization in Neural Networks and NewtonÕs Method

Hessians and concavity

RememberÉ

Same thing happens for many variables!

RememberÉ

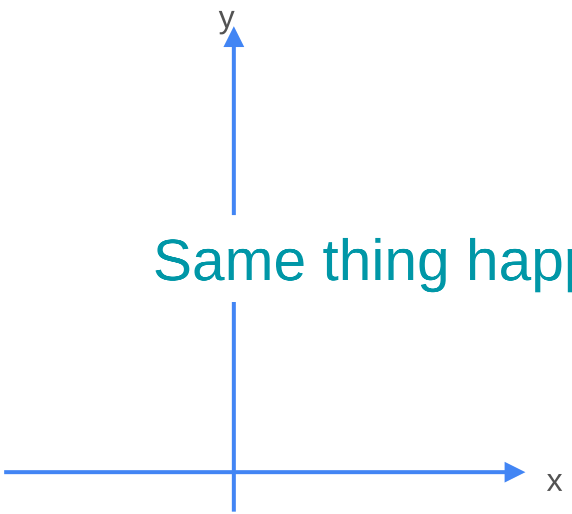


Same thing happens for many variables!

Concave up or convex

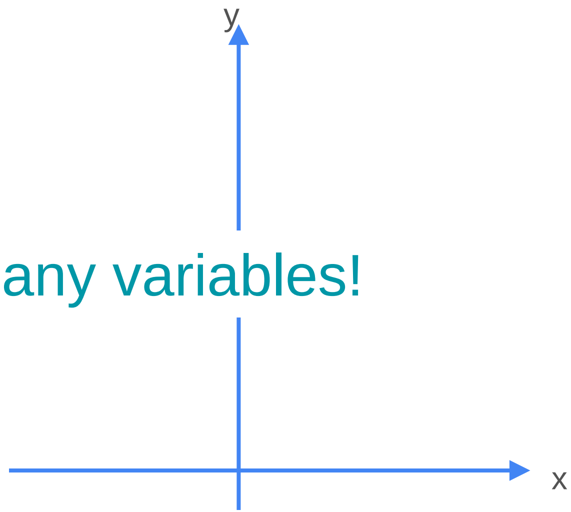
f#(0) > 0

RememberÉ



Concave up or convex

f#(0) > 0



Concave down

f#(0) < 0

Same thing happens for many variables!

Concave Up

Concave Up

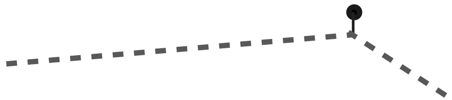
$$f(x, y) = 2x^2 + 3y^2 \text{ " } xy$$

Concave Up

$$f(x, y) = 2x^2 + 3y^2 - xy$$

Concave Up

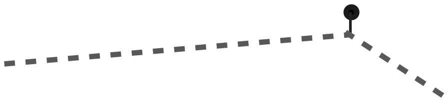
$$f(x, y) = 2x^2 + 3y^2 \text{ " } xy$$



Concave Up

$$f(x,y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

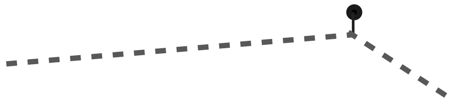


Concave Up

$$f(x,y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

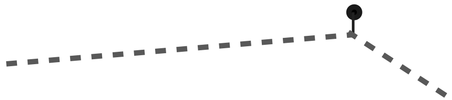


Concave Up

$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det\left(\begin{bmatrix} 4-\lambda & -1 \\ -1 & 6-\lambda \end{bmatrix}\right)$$



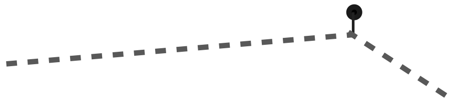
Concave Up

$$f(x,y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det\left(\begin{bmatrix} 4-\lambda & -1 \\ -1 & 6-\lambda \end{bmatrix}\right)$$

$$= (4-\lambda)(6-\lambda) - (-1)(-1)$$



Concave Up

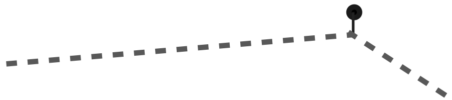
$$f(x,y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det\left(\begin{bmatrix} 4-\lambda & -1 \\ -1 & 6-\lambda \end{bmatrix}\right)$$

$$= (4-\lambda)(6-\lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23$$



Concave Up

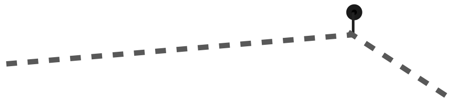
$f(x, y) = 2x^2 + 3y^2$ " xy

H(0,0) = [4 " 1 ; " 1 6]

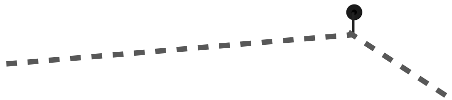
det(H(0,0) " %I) = det([4 " % " 1 ; " 1 6 " %])

= (4 " %)(6 " %) " (" 1)(" 1)

= $\%^2$ " 10% + 23 → $\%_1$ = 6.41



Concave Up



$$f(x,y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det\left(\begin{bmatrix} 4-\lambda & -1 \\ -1 & 6-\lambda \end{bmatrix}\right)$$

$$= (4-\lambda)(6-\lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23 \rightarrow \begin{array}{l} \lambda_1 = 6.41 \\ \lambda_2 = 3.59 \end{array}$$

Concave Up

$$f(x,y) = 2x^2 + 3y^2 \text{"}\ xy$$

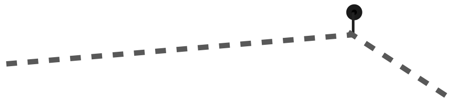
$$H(0,0) = \begin{bmatrix} 4 & \text{"}\ 1 \\ \text{"}\ 1 & 6 \end{bmatrix}$$

$$\det(H(0,0)\ \text{"}\ \%I) = \det\left(\begin{bmatrix} 4\ \text{"}\ \% & \text{"}\ 1 \\ \text{"}\ 1 & 6\ \text{"}\ \% \end{bmatrix}\right)$$

$$= (4\ \text{"}\ \%)(6\ \text{"}\ \%)\ \text{"}\ (\text{"}\ 1)(\text{"}\ 1)$$

$$= \%^2\ \text{"}\ 10\% + 23 \quad \boxed{\begin{matrix} \%_1 = 6.41 \\ \%_2 = 3.59 \end{matrix}}$$

$$> 0$$



Concave Up

$$f(x,y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

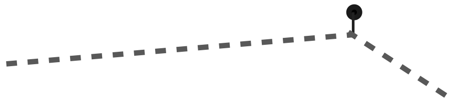
$$\det(H(0,0) - \lambda I) = \det\left(\begin{bmatrix} 4-\lambda & -1 \\ -1 & 6-\lambda \end{bmatrix}\right)$$

$$= (4-\lambda)(6-\lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23 \quad \boxed{\begin{matrix} \lambda_1 = 6.41 \\ \lambda_2 = 3.59 \end{matrix}}$$

$(0,0)$ is a minimum!

> 0



Concave Down

Concave Down

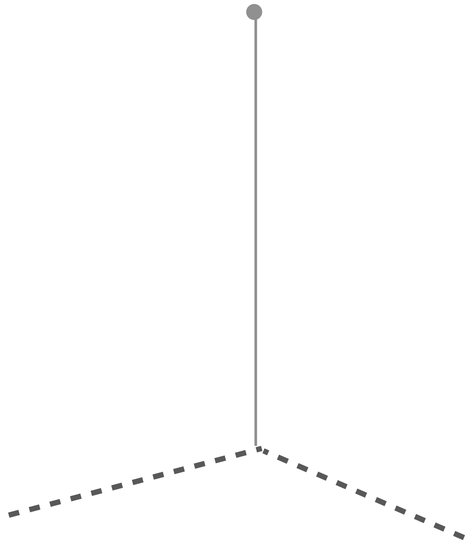
$$f(x, y) = -2x^2 + -3y^2 - xy + 15$$

Concave Down

$$f(x, y) = \text{"}\ 2x^2 + \text{"}\ 3y^2\ \text{"}\ xy + 15$$

Concave Down

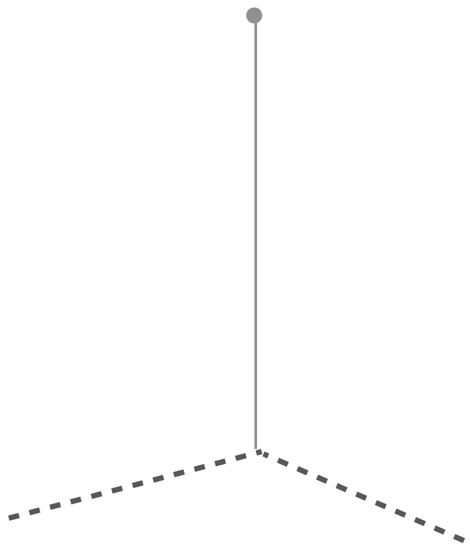
$$f(x, y) = -2x^2 + -3y^2 - xy + 15$$



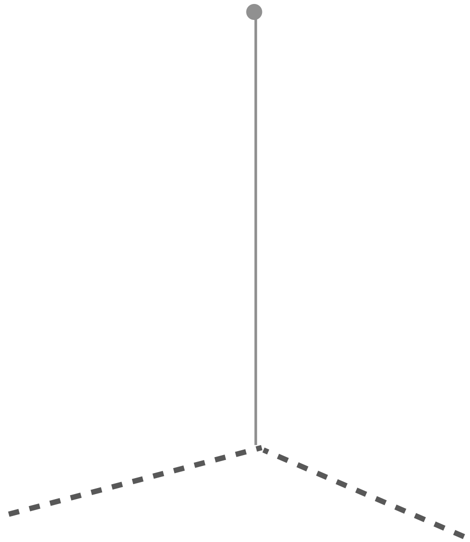
Concave Down

$$f(x,y) = \text{"}\ 2x^2 + \text{"}\ 3y^2 \text{"}\ \ xy + 15$$

$$-\ f(x,y) = \begin{matrix} \text{"}\ 4x \text{"}\ \ y \\ [\text{"}\ x \text{"}\ \ 6y] \end{matrix}$$



Concave Down

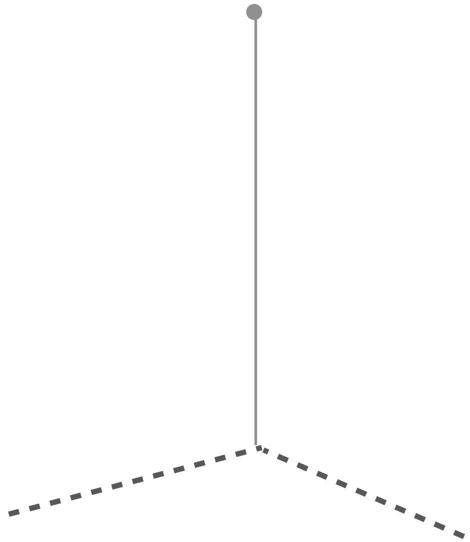


$$f(x,y) = -2x^2 + -3y^2 - xy + 15$$

$$\nabla f(x,y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

Concave Down



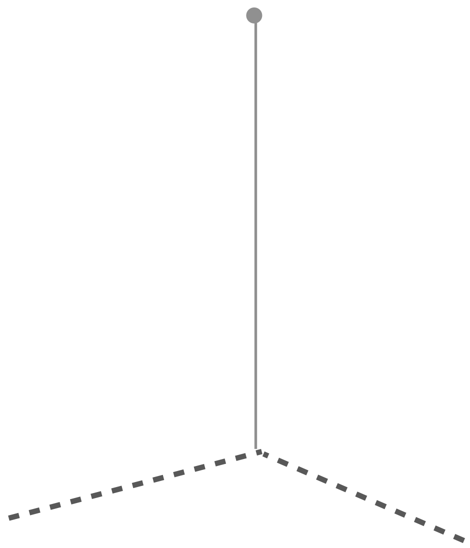
$$f(x,y) = -2x^2 + -3y^2 - xy + 15$$

$$\nabla f(x,y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

Concave Down



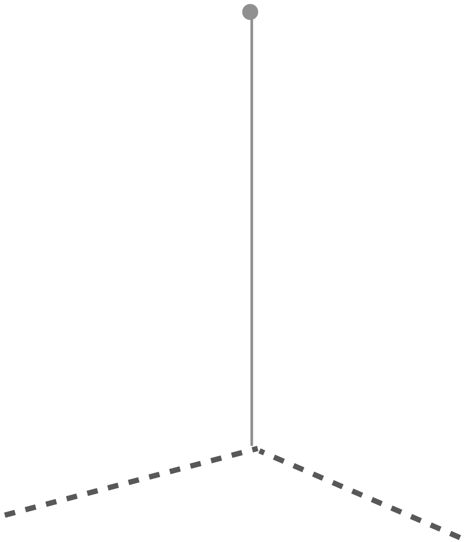
$$f(x,y) = -2x^2 + -3y^2 - xy + 15$$

$$\nabla f(x,y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = (-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

Concave Down



$$f(x,y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x,y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

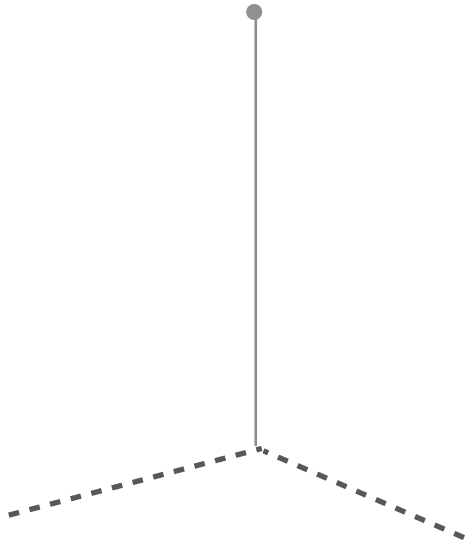
$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 + 10\lambda + 23$$

Concave Down



$$f(x,y) = -2x^2 + -3y^2 - xy + 15$$

$$\nabla f(x,y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

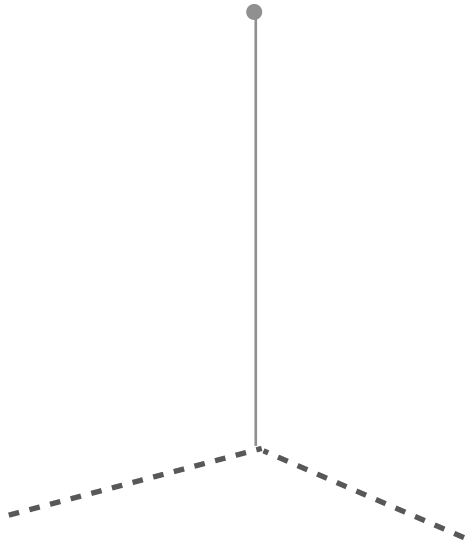
$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 + 10\lambda + 23 \rightarrow \lambda_1 = -3.49$$

Concave Down



$$f(x,y) = -2x^2 + -3y^2 - xy + 15$$

$$\nabla f(x,y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

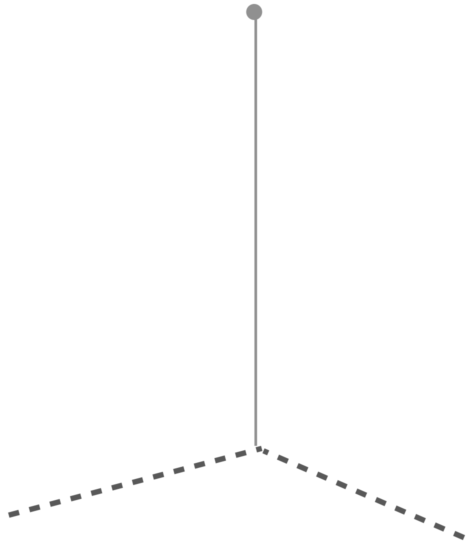
$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 + 10\lambda + 23 \rightarrow \begin{matrix} \lambda_1 = -3.49 \\ \lambda_2 = -6.41 \end{matrix}$$

Concave Down



$$f(x,y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x,y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = (-4-\lambda)(-6-\lambda) - (-1)(-1)$$

$$= \lambda^2 + 10\lambda + 23 \rightarrow \boxed{\begin{matrix} \lambda_1 = -3.49 \\ \lambda_2 = -6.41 \end{matrix}}$$

< 0

$(0,0)$ is a maximum!

Saddle Point

Saddle Point

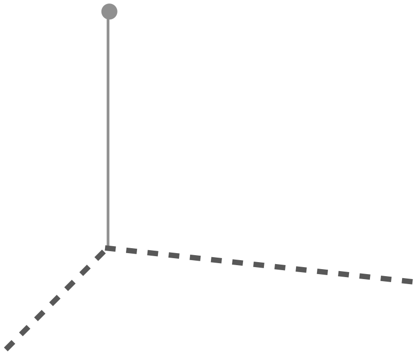
$$f(x, y) = 2x^2 - 2y^2$$

Saddle Point

$$f(x, y) = 2x^2 - 2y^2$$

Saddle Point

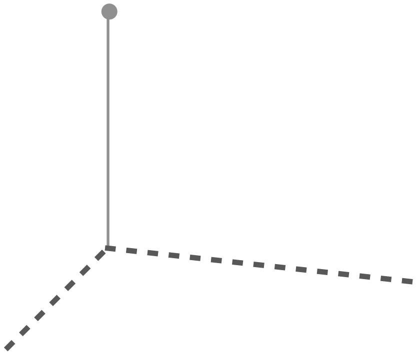
$$f(x, y) = 2x^2 - 2y^2$$



Saddle Point

$$f(x, y) = 2x^2 \text{ " } 2y^2$$

$$- \ f(x, y) = \begin{bmatrix} 4x \\ \text{"} \ 4y \end{bmatrix}$$

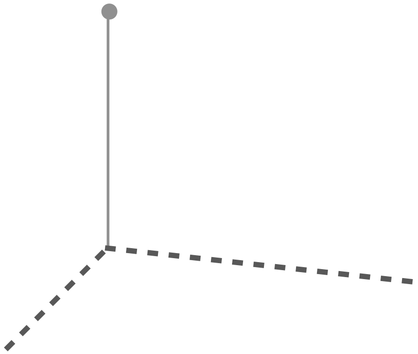


Saddle Point

$$f(x,y) = 2x^2 - 2y^2$$

$$\nabla f(x,y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$



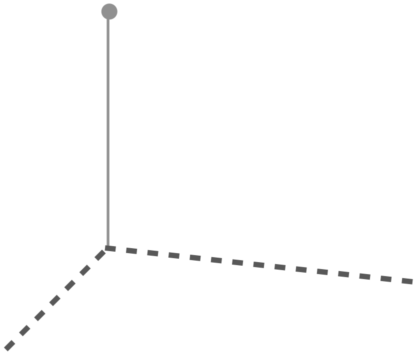
Saddle Point

$$f(x,y) = 2x^2 - 2y^2$$

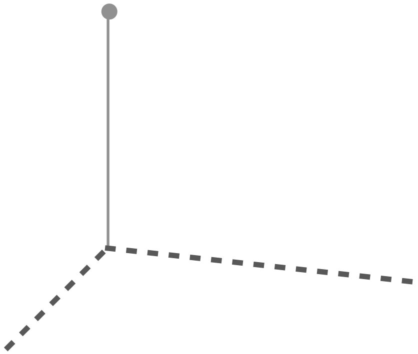
$$\nabla f(x,y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$



Saddle Point



$$f(x,y) = 2x^2 - 2y^2$$

$$\nabla f(x,y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

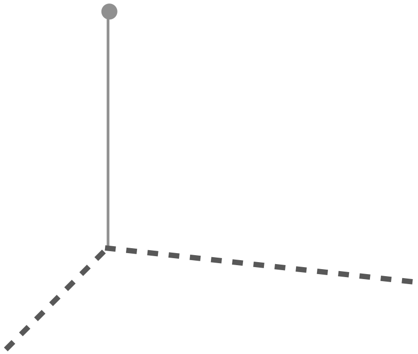
$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(4-\lambda)(-4-\lambda) - 0$$

$$\lambda_1 = -4$$

Saddle Point



f (x, y) = 2x² " 2y²

- f (x, y) = [4x ; " 4y]

H(0,0) = [4 0 ; 0 " 4]

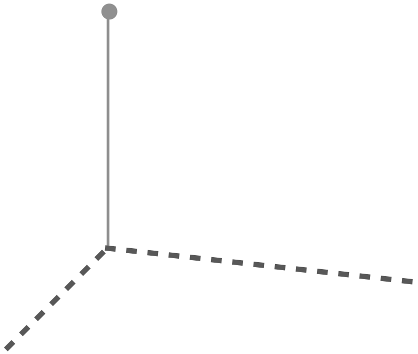
det(H(0,0) " %I) =

(4 " %)(" 4 " %) " 0

%₁ = " 4

%₂ = 4

Saddle Point



$$f(x,y) = 2x^2 - 2y^2$$

$$\nabla f(x,y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

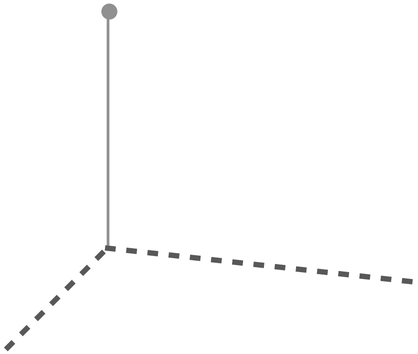
$$\det(H(0,0) - \lambda I) = (4-\lambda)(-4-\lambda) - 0 < 0$$

$$\boxed{\lambda_1 = -4}$$

$$\lambda_2 = 4$$

(0,0) is saddle point

Saddle Point



$$f(x, y) = 2x^2 - 2y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(4 - \lambda)(-4 - \lambda) - 0$$

< 0

$\boxed{\lambda_1 = -4}$

$\boxed{\lambda_2 = 4}$

> 0

(0,0) is saddle point

Summary

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \text{É}, x_n)$

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \ldots, x_n)$
(Local) minima	Happy face $f''(x) > 0$		

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \ldots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0$ & $\lambda_2 > 0$	

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \ldots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0$ & $\lambda_2 > 0$	All $\lambda_i > 0$

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \ldots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0$ & $\lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$		

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2,$ É $, x_n)$
(Local) minima	Happy face f#(x) > 0	Upper paraboloid $\%_1 > 0$ & $\%_2 > 0$	All $\%_i > 0$
(Local) maxima	Sad face f#(x) < 0	Down paraboloid $\%_1 < 0$ & $\%_2 < 0$	

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, É\ , x_n)$
(Local) minima	Happy face $f\#(x) > 0$	Upper paraboloid $\%\rho_1 > 0\ \&\ \%\rho_2 > 0$	All $\%\rho_i > 0$
(Local) maxima	Sad face $f\#(x) < 0$	Down paraboloid $\%\rho_1 < 0\ \&\ \%\rho_2 < 0$	All $\%\rho_i < 0$

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \text{É}, x_n)$
(Local) minima	Happy face $f\#(x) > 0$	Upper paraboloid $\%_1 > 0$ & $\%_2 > 0$	All $\%_i > 0$
(Local) maxima	Sad face $f\#(x) < 0$	Down paraboloid $\%_1 < 0$ & $\%_2 < 0$	All $\%_i < 0$
Need more information	$f\#(x) = 0$		

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \ldots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0$ & $\lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Down paraboloid $\lambda_1 < 0$ & $\lambda_2 < 0$	All $\lambda_i < 0$
Need more information	$f''(x) = 0$	Saddle point $\lambda_1 > 0$ & $\lambda_2 < 0$ $\lambda_1 < 0$ & $\lambda_2 > 0$	

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, É\ , x_n)$
(Local) minima	Happy face $f\#(x) > 0$	Upper paraboloid $\%_1 > 0$ & $\%_2 > 0$	All $\%_i > 0$
(Local) maxima	Sad face $f\#(x) < 0$	Down paraboloid $\%_1 < 0$ & $\%_2 < 0$	All $\%_i < 0$
Need more information	$f\#(x) = 0$	Saddle point $\%_1 > 0$ & $\%_2 < 0$ $\%_1 < 0$ & $\%_2 > 0$ Or some $\%_i = 0$	

Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, É, x_n)$
(Local) minima	Happy face $f\#(x) > 0$	Upper paraboloid $\%o_1 > 0$ & $\%o_2 > 0$	All $\%o_i > 0$
(Local) maxima	Sad face $f\#(x) < 0$	Down paraboloid $\%o_1 < 0$ & $\%o_2 < 0$	All $\%o_i < 0$
Need more information	$f\#(x) = 0$	Saddle point $\%o_1 > 0$ & $\%o_2 < 0$ $\%o_1 < 0$ & $\%o_2 > 0$ Or some $\%o_i = 0$	Some $\%o_i > 0$ and some $\%o_j < 0$ OR At least one $\%o_i = 0$

Optimization in Neural Networks and NewtonÕs Method

NewtonÕs method for two variables

NewtonÕs Method

NewtonÕs Method

1 variable

$$x_{k+1} = x_k \,\text{"}\, \frac{f\#(x_k)}{f\#\#(x_k)}$$

NewtonÕs Method

1 variable

$$x_{k+1} = \quad " \; \frac{f\#(x_k)}{f\#\#(x_k)}$$

NewtonÕs Method

1 variable

$$x_{k+1} = x_k - \frac{\boxed{}}{f'(x_k)}$$

NewtonÕs Method

1 variable

$$x_{k+1} = x_k " \frac{f\#(x_k)}{f\#\#(x_k)}$$

NewtonÕs Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

NewtonÕs Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

2 variables

NewtonÕs Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix}$$

NewtonÕs Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} -$$

NewtonÕs Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - \boxed{f''(x_k)}^{-1} f'(x_k)$$

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k)$$

NewtonÕs Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - \boxed{f''(x_k)^{-1}}\,\boxed{f'(x_k)}$$

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k) - f(x_k, y_k)$$

NewtonÕs Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k) \nabla f(x_k, y_k)$$

NewtonÕs Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} \text{"} \; H^{\text{"}1}(x_k, y_k) \; - \; f(x_k, y_k)$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} \text{"} \; - \; f(x_k, y_k) \; H^{\text{"}1}(x_k, y_k)$$

NewtonÕs Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k) \; \nabla f(x_k, y_k)$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \cancel{\nabla f(x_k, y_k) \; H^{-1}(x_k, y_k)}$$

NewtonÕs Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \underbrace{H^{-1}(x_k, y_k)}_{2\times 2} \underbrace{\nabla f(x_k, y_k)}_{2\times 1}$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \cancel{\nabla f(x_k, y_k)\, H^{-1}(x_k, y_k)}$$

NewtonÕs Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} \text{"} \underbrace{H^{\text{"}1}(x_k, y_k)}_{2\%2} \underbrace{- f(x_k, y_k)}_{2\%1}$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} \text{"} \cancel{- f(x_k, y_k) H^{\text{"}1}(x_k, y_k)}$$

When working with 2 variables the order is crucial!

An Example

An Example

$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

An Example

$$f(x,y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$f(x,y) \xrightarrow{x} 4x^3 + 8x - y - 0.4xy$$

An Example

$$f(x,y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\begin{tikzcd}
 & 4x^3 + 8x - y - 0.4xy \\
f(x,y) \arrow[ur, "x"] \arrow[dr, "y"'] & \\
 & 3.2y^3 + 4y - x - 0.2x^2
\end{tikzcd}$$

An Example

$$f(x,y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\begin{tikzcd}
 & & 12x^2 + 8 - 0.4y \\
 & 4x^3 + 8x - y - 0.4xy \arrow[ur, "x"] & \\
f(x,y) \arrow[ur, "x"] \arrow[dr, "y"'] & & \\
 & 3.2y^3 + 4y - x - 0.2x^2 &
\end{tikzcd}$$

An Example

$$f(x,y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\begin{tikzcd}
 & & 12x^2 + 8 - 0.4y \\
 & 4x^3 + 8x - y - 0.4xy \arrow[ur, "x"] \arrow[dr, "y"'] & \\
f(x,y) \arrow[ur, "x"] \arrow[dr, "y"'] & & -1 - 0.4x \\
 & 3.2y^3 + 4y - x - 0.2x^2 &
\end{tikzcd}$$

An Example

$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\begin{tikzcd}
 & & 12x^2 + 8 - 0.4y \\
 & 4x^3 + 8x - y - 0.4xy \arrow[ur, "x"] \arrow[dr, "y"'] & \\
f(x, y) \arrow[ur, "x"] \arrow[dr, "y"'] & & -1 - 0.4x \\
 & 3.2y^3 + 4y - x - 0.2x^2 \arrow[r, "x"] & -1 - 0.4x
\end{tikzcd}$$

An Example

$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\begin{tikzcd}
 & & 12x^2 + 8 - 0.4y \\
 & 4x^3 + 8x - y - 0.4xy \arrow[ur, "x"] \arrow[dr, "y"'] & \\
 & & -1 - 0.4x \\
f(x, y) \arrow[uur, "x"] \arrow[ddr, "y"'] & & \\
 & & -1 - 0.4x \\
 & 3.2y^3 + 4y - x - 0.2x^2 \arrow[ur, "x"] \arrow[dr, "y"'] & \\
 & & 9.6y^2 + 4
\end{tikzcd}$$

An Example

$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

An Example

$$f(x,y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\nabla f(x,y) = \begin{bmatrix} 4x^3 + 8x - y - 0.4xy \\ 3.2y^3 + 4y - x - 0.2x^2 \end{bmatrix}$$

An Example

$$f(x,y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\nabla f(x,y) = \begin{bmatrix} 4x^3 + 8x - y - 0.4xy \\ 3.2y^3 + 4y - x - 0.2x^2 \end{bmatrix}$$

$$H(x,y) = \begin{bmatrix} 12x^2 + 8 - 0.4y & -1 - 0.4x \\ -1 - 0.4x & 9.6y^2 + 4 \end{bmatrix}$$

An Example

An Example

Start at some point $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$\begin{bmatrix} 4 \\ 4 \end{bmatrix}$

An Example

Start at some point $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$\begin{bmatrix} 4 \\ 4 \end{bmatrix}$

- $f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$

An Example

Start at some point $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$\begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$- f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix} \qquad H(4,4) = \begin{bmatrix} 198.4 & \text{"}\ 2.6 \\ \text{"}\ 2.6 & 157.6 \end{bmatrix}$

An Example

Start at some point $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$\begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$- f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$ $\quad H(4,4) = \begin{bmatrix} 198.4 & \text{"}\ 2.6 \\ \text{"}\ 2.6 & 157.6 \end{bmatrix}$

$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \text{"}$

An Example

$$\begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

Start at some point $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$$\nabla f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix} \qquad H(4,4) = \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}^{-1} \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$$

An Example

$\begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$

Start at some point $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$$- f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix} \qquad H(4,4) = \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}^{-1} \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$$

$$= \begin{bmatrix} 2.58 \\ 2.62 \end{bmatrix}$$

An Example

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

An Example

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$- f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

$$\begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

An Example

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$- f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} \quad H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

$$\begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

An Example

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$- f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} \quad H(2.58, 2.61) = \begin{bmatrix} 86.83 & " 2.032 \\ " 2.032 & 69.39 \end{bmatrix}$$

$$\begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix} "$$

An Example

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\nabla f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} \quad H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

$$\begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix} - \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}^{-1} \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

An Example

$$\begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\begin{bmatrix} 1.59 \\ 1.67 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$- f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} \quad H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix} - \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}^{-1} \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

$$= \begin{bmatrix} 1.59 \\ 1.67 \end{bmatrix}$$

An Example

1.59
[1.67]

An Example

Repeat until you are close enough to the actual zero!

1.59
[1.67]

An Example

Repeat until you are close enough to the actual zero!

Needed $k = 8$ steps

1.59

[1.67]

An Example

$$\begin{bmatrix} 1.59 \\ 1.67 \end{bmatrix}$$

Repeat until you are close enough to the actual zero!

Needed $k = 8$ steps

$$\begin{bmatrix} x_8 \\ y_8 \end{bmatrix} = \begin{bmatrix} 4.15 \cdot 10^{-17} \\ -2.05 \cdot 10^{-17} \end{bmatrix}$$

An Example

$$\begin{bmatrix} 1.59 \\ 1.67 \end{bmatrix}$$

Repeat until you are close enough to the actual zero!

Needed $k = 8$ steps

$$\begin{bmatrix} x_8 \\ y_8 \end{bmatrix} = \begin{bmatrix} 4.15 \times 10^{-17} \\ -2.05 \times 10^{-17} \end{bmatrix}$$

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Optimization in Neural Networks and NewtonÕs Method

Conclusion