DeepLearning.AI

Optimization in Neural Networks and Newton's Method

Regression with a perceptron

Regression Problem Motivation

Regression Problem Motivation

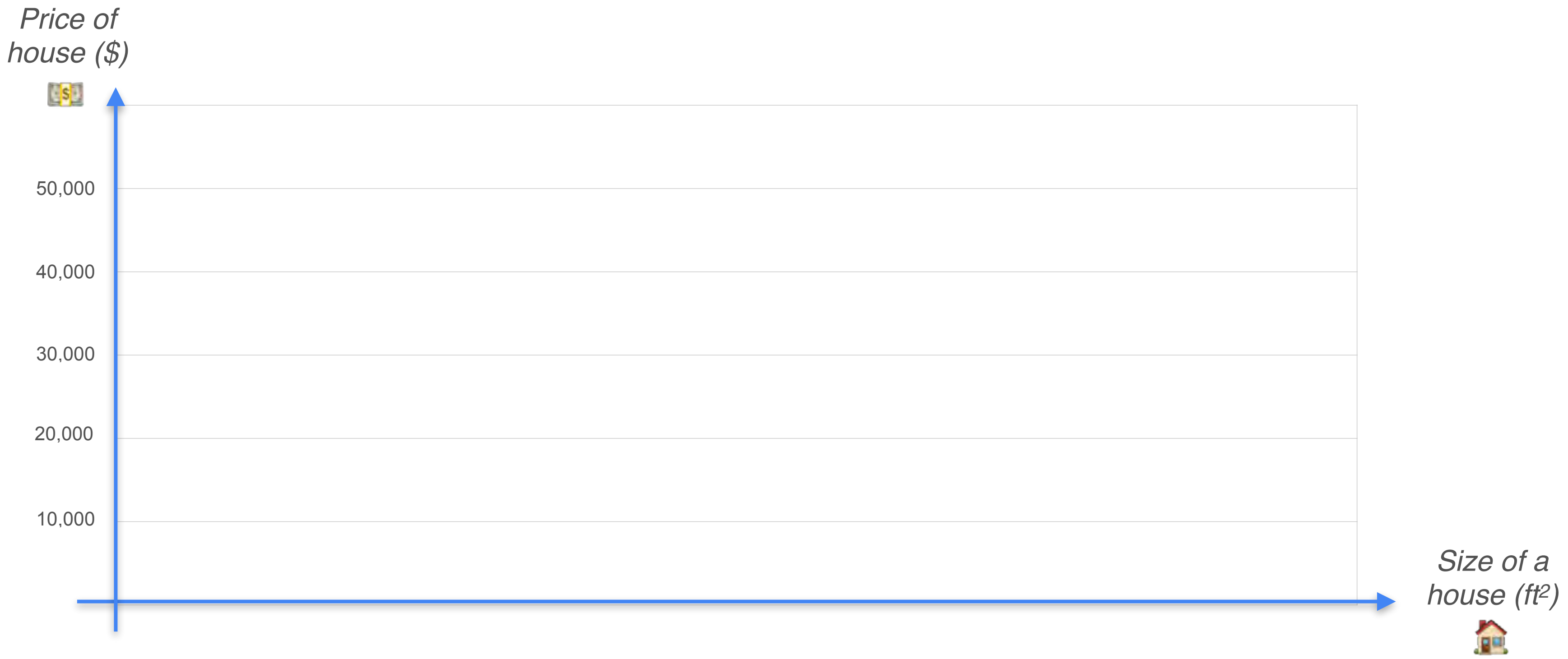
Predicting

the price of a house

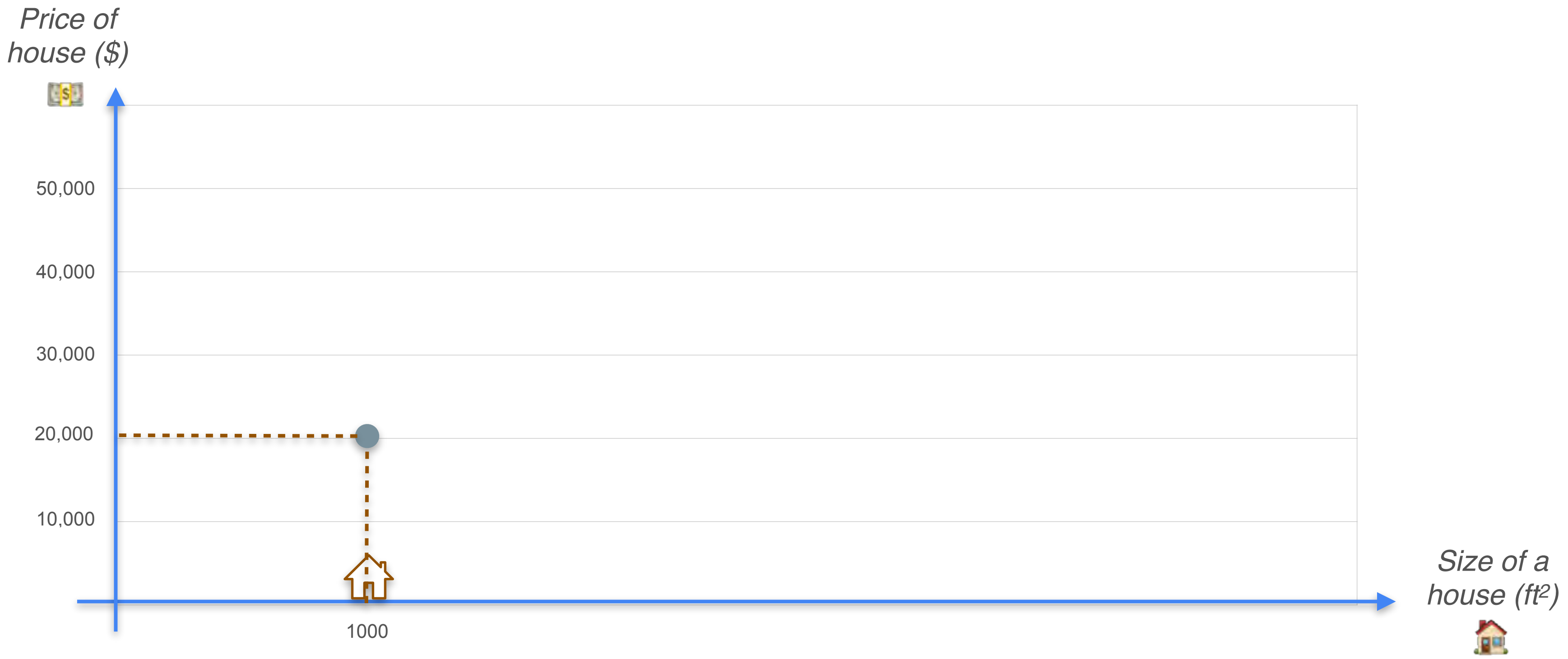
from

the size of the house

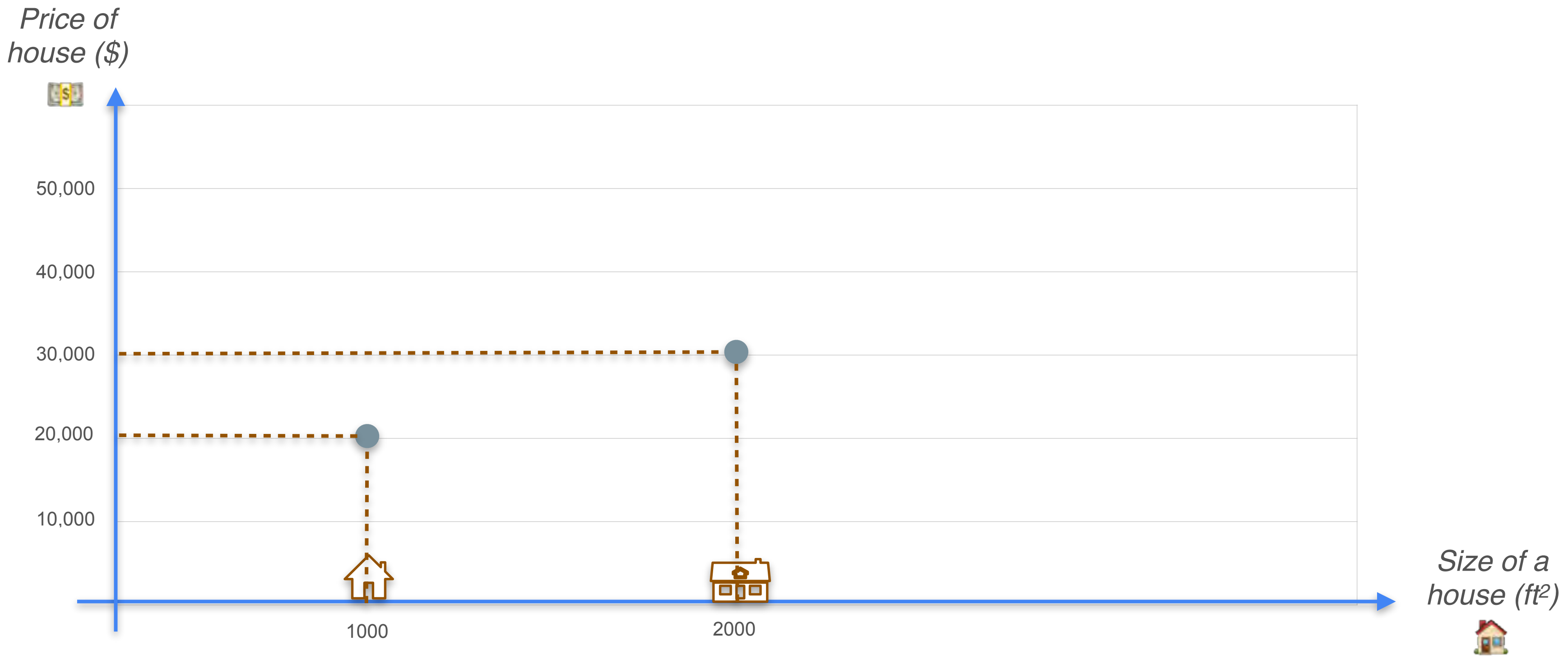
Regression Problem Motivation



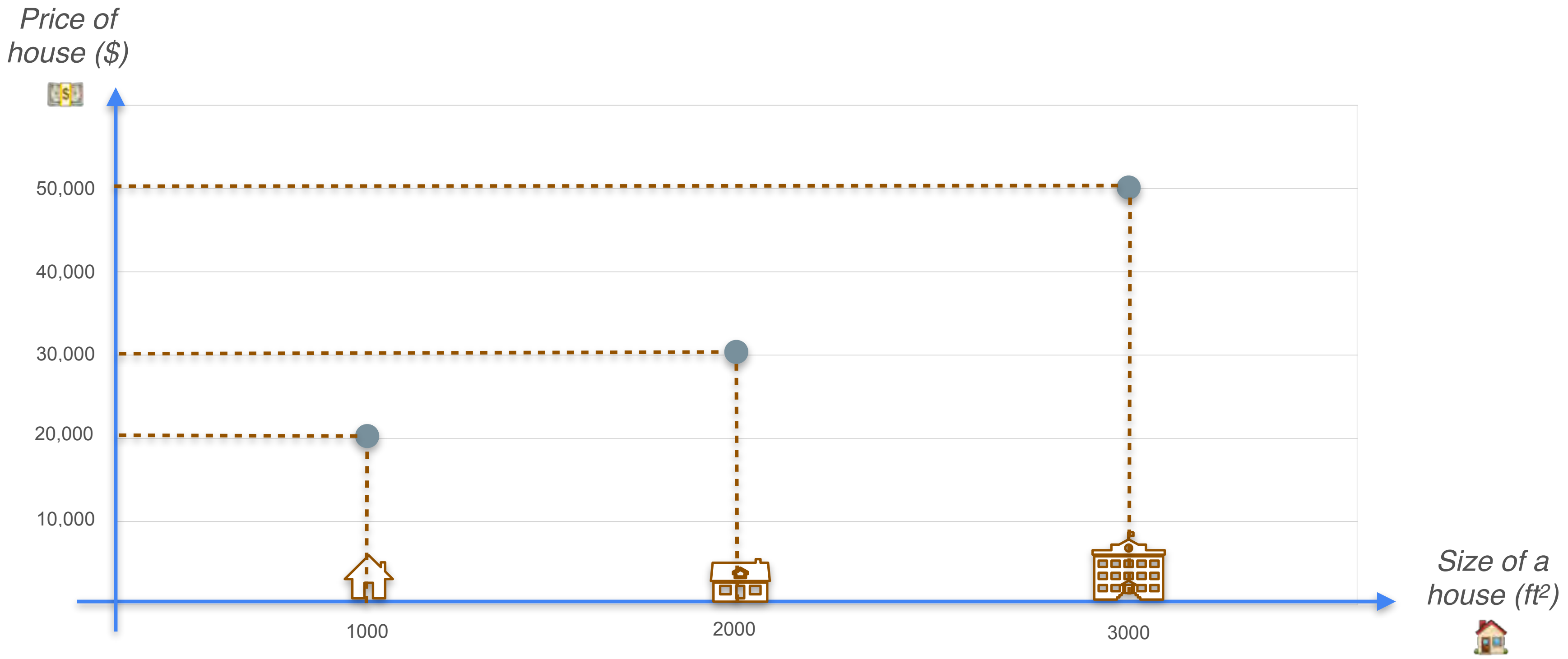
Regression Problem Motivation



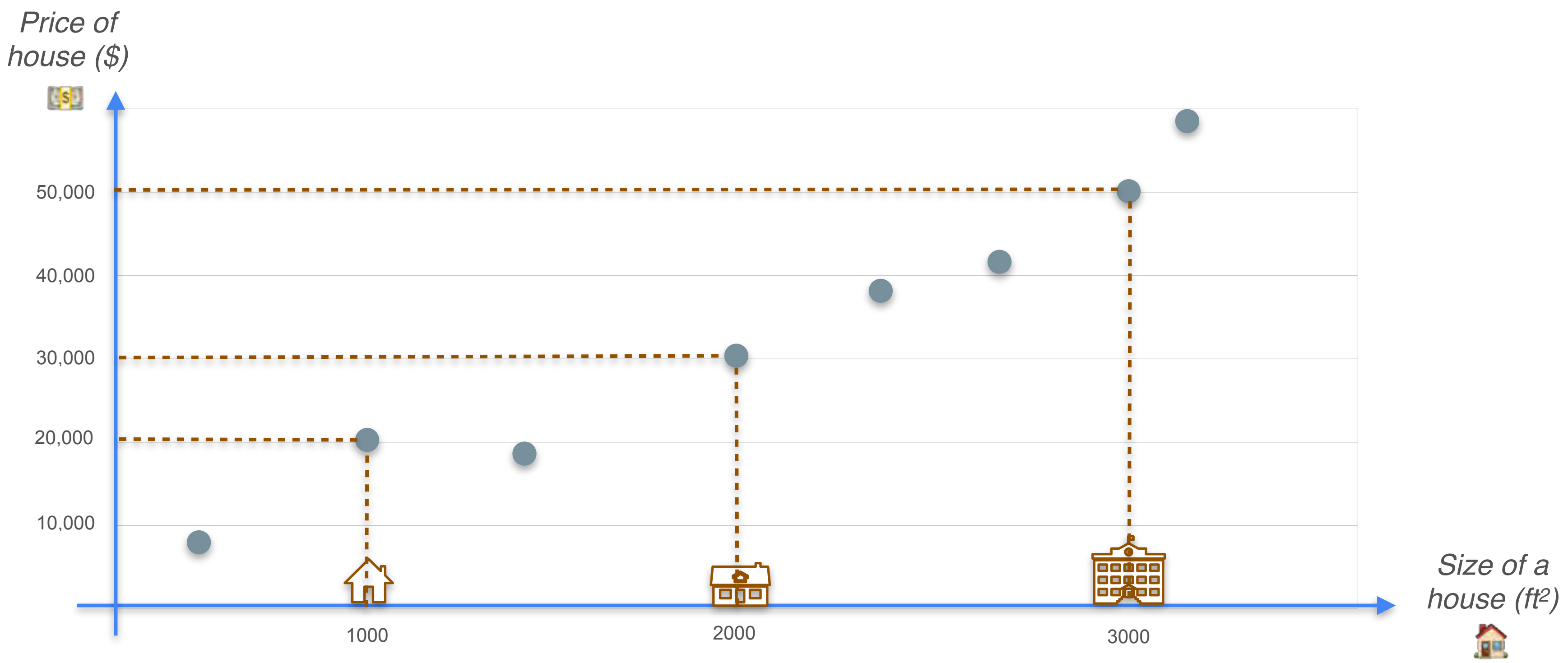
Regression Problem Motivation



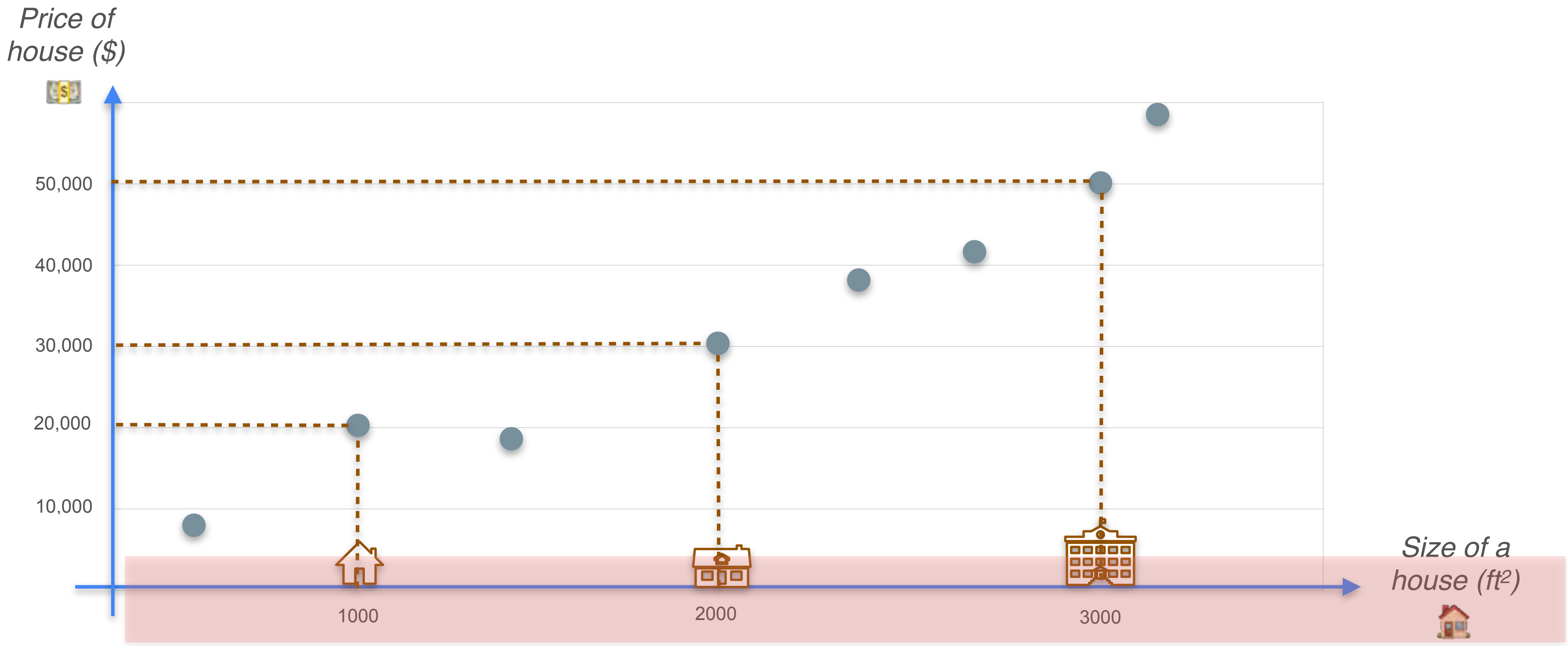
Regression Problem Motivation



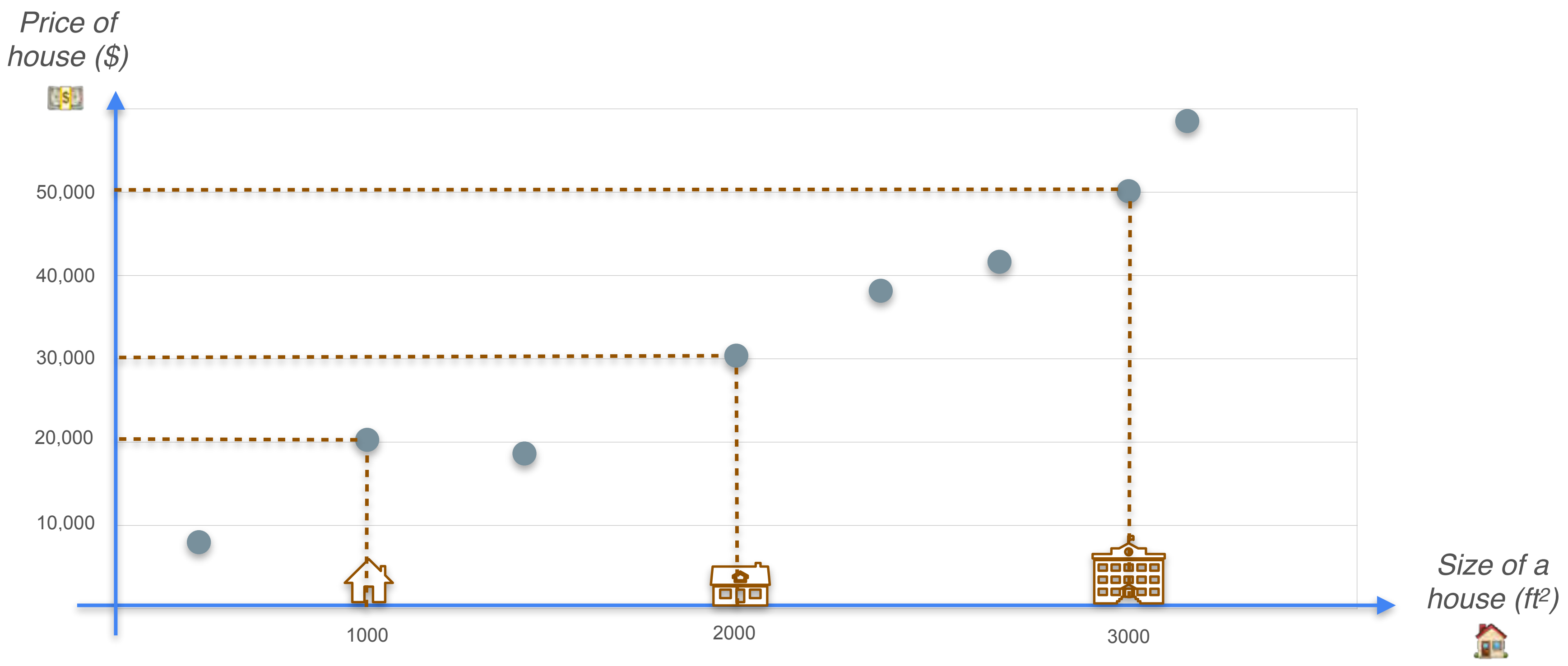
Regression Problem Motivation



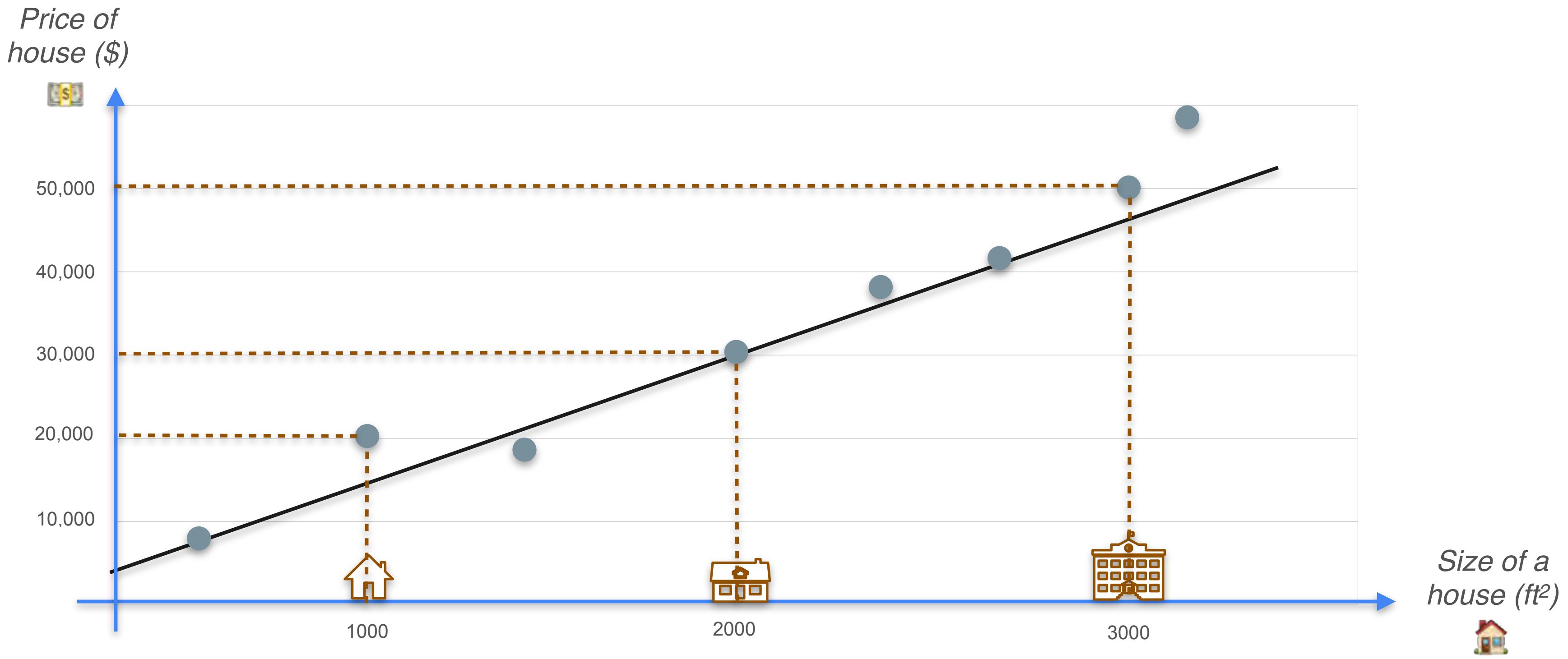
Regression Problem Motivation



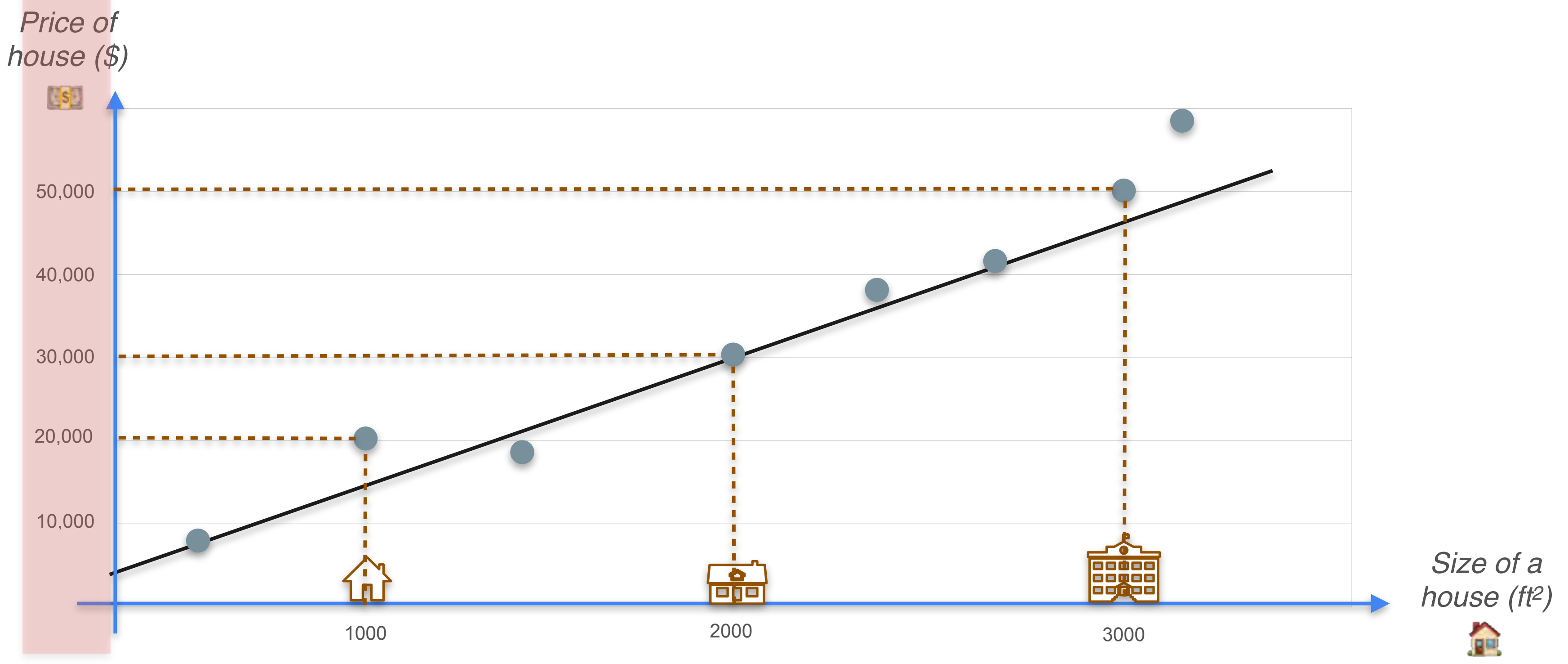
Regression Problem Motivation



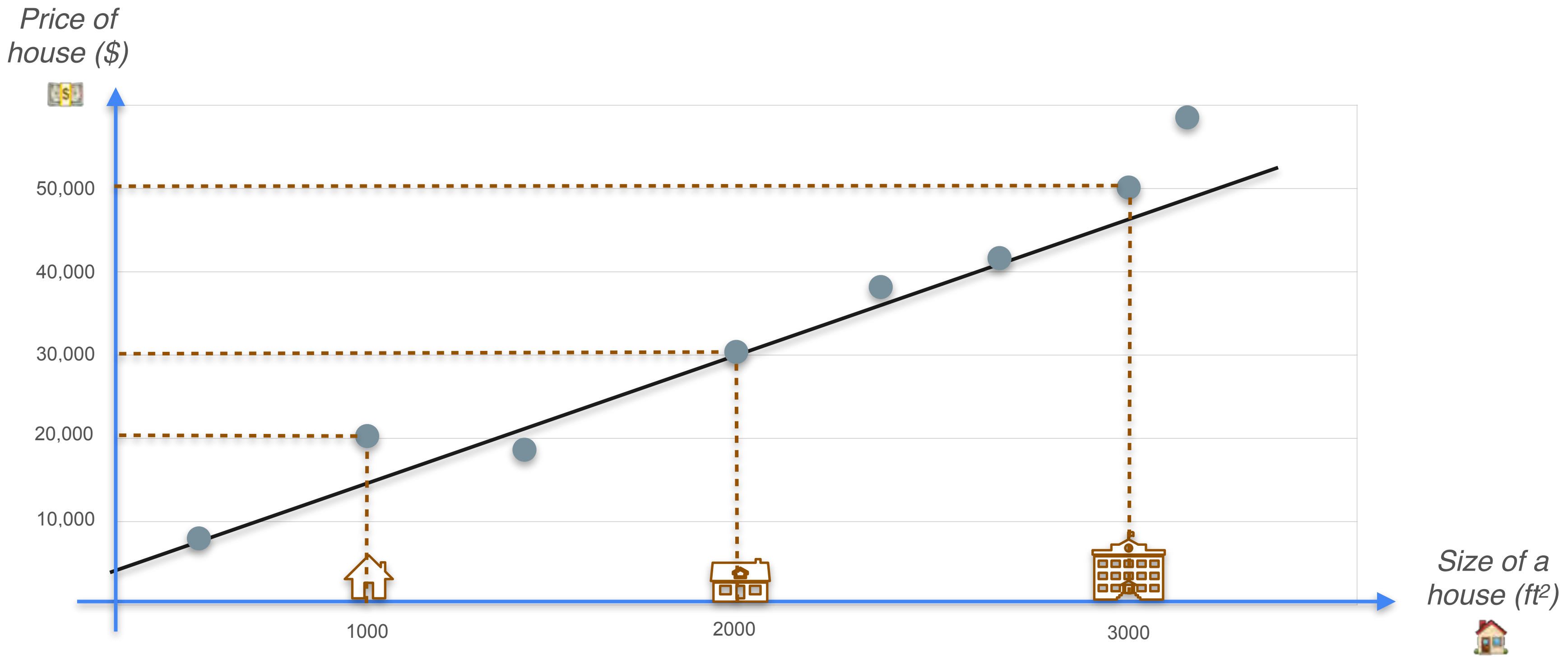
Regression Problem Motivation



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

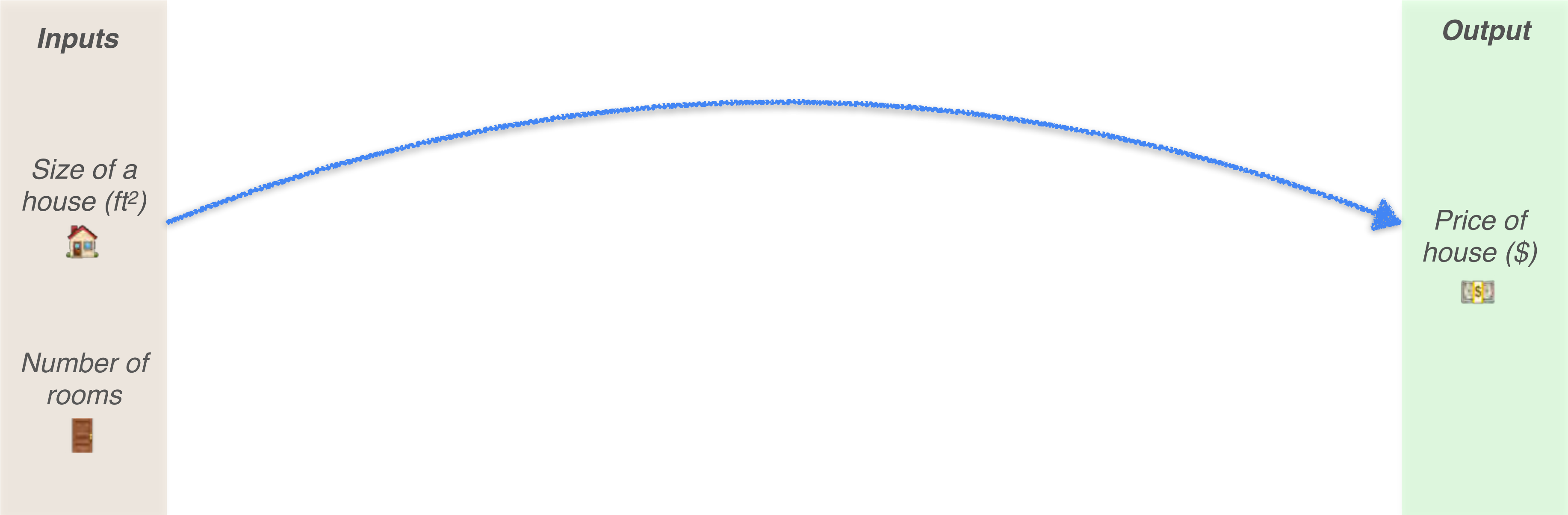
Size of a house (ft^2) 🏠

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 (ft²) 🏠

Number of rooms 🚪

Output

Price of house ($) 💵

Regression With a Perceptron

Single Layer Neural Network Perceptron

Inputs

Size of a house (ft^2)

Number of rooms

x_1

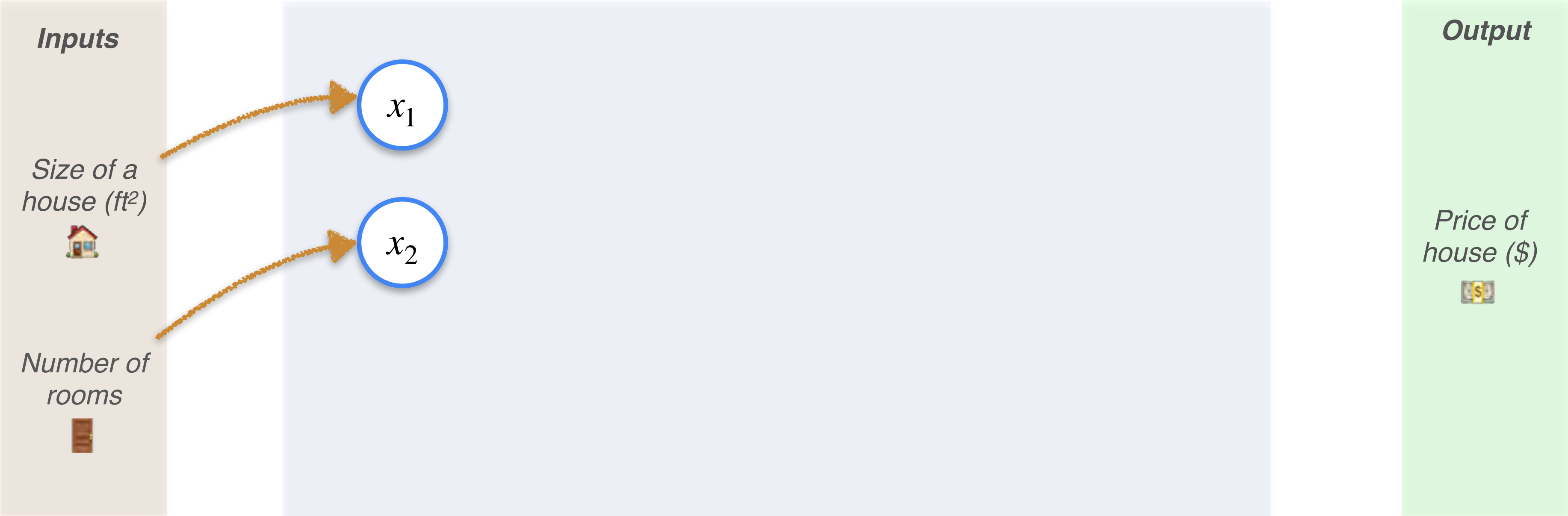
x_2

Output

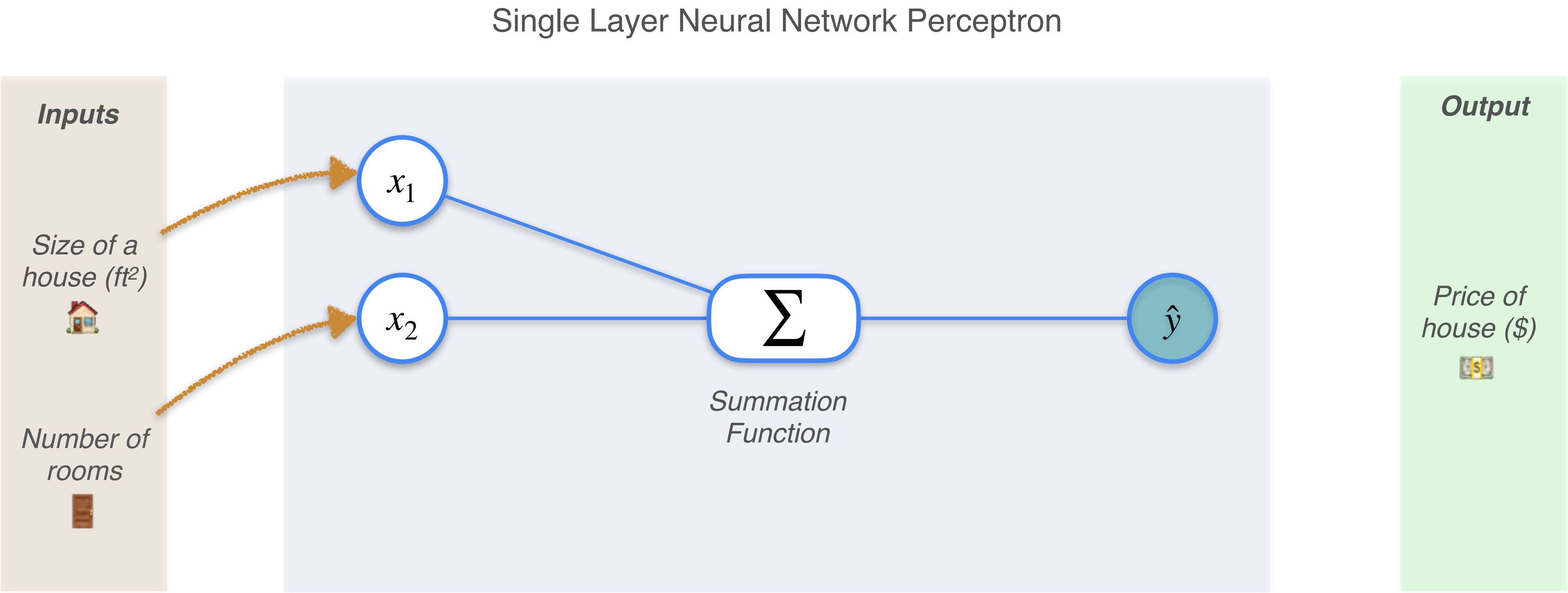
Price of house ($)

Regression With a Perceptron

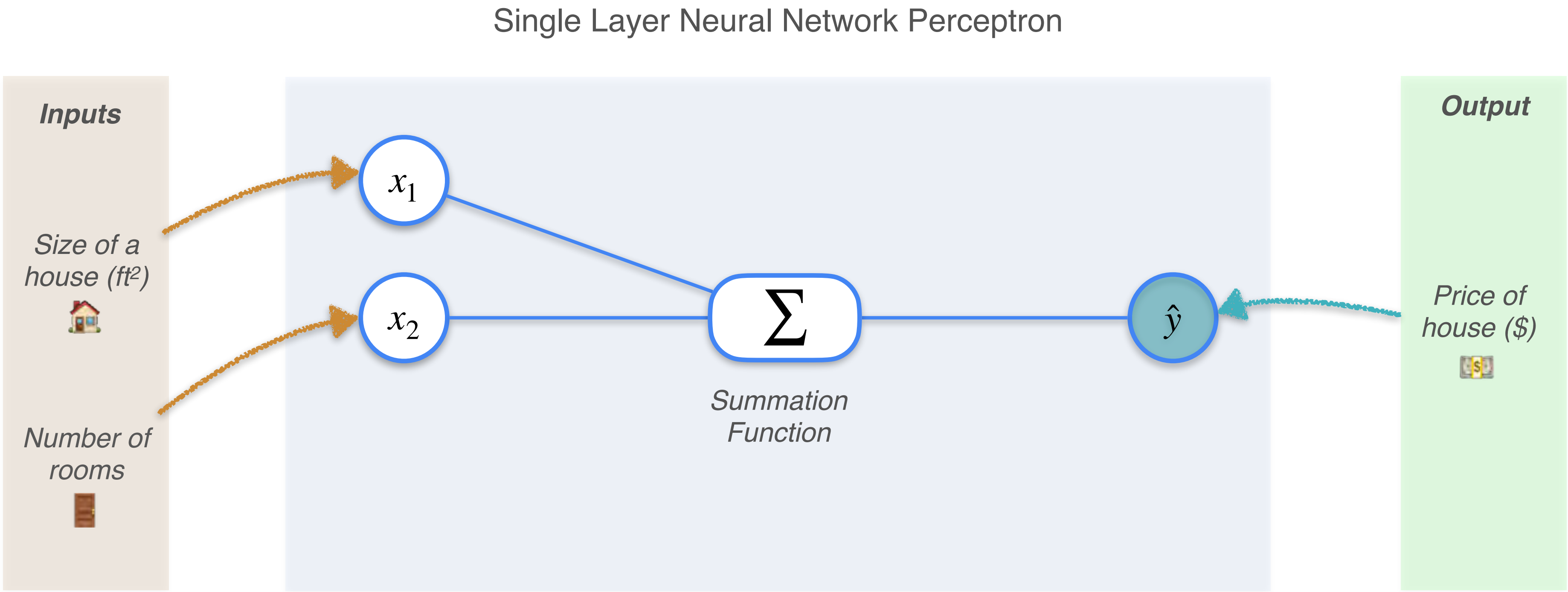
Single Layer Neural Network Perceptron



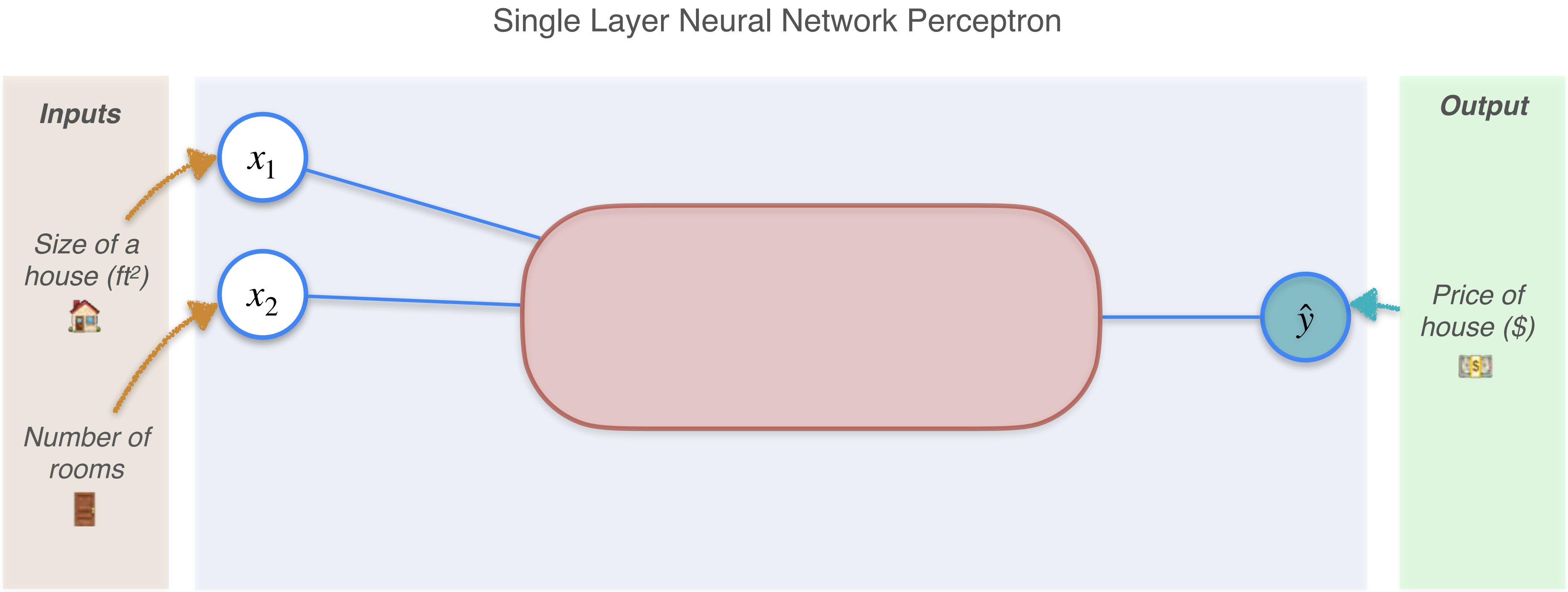
Regression With a Perceptron



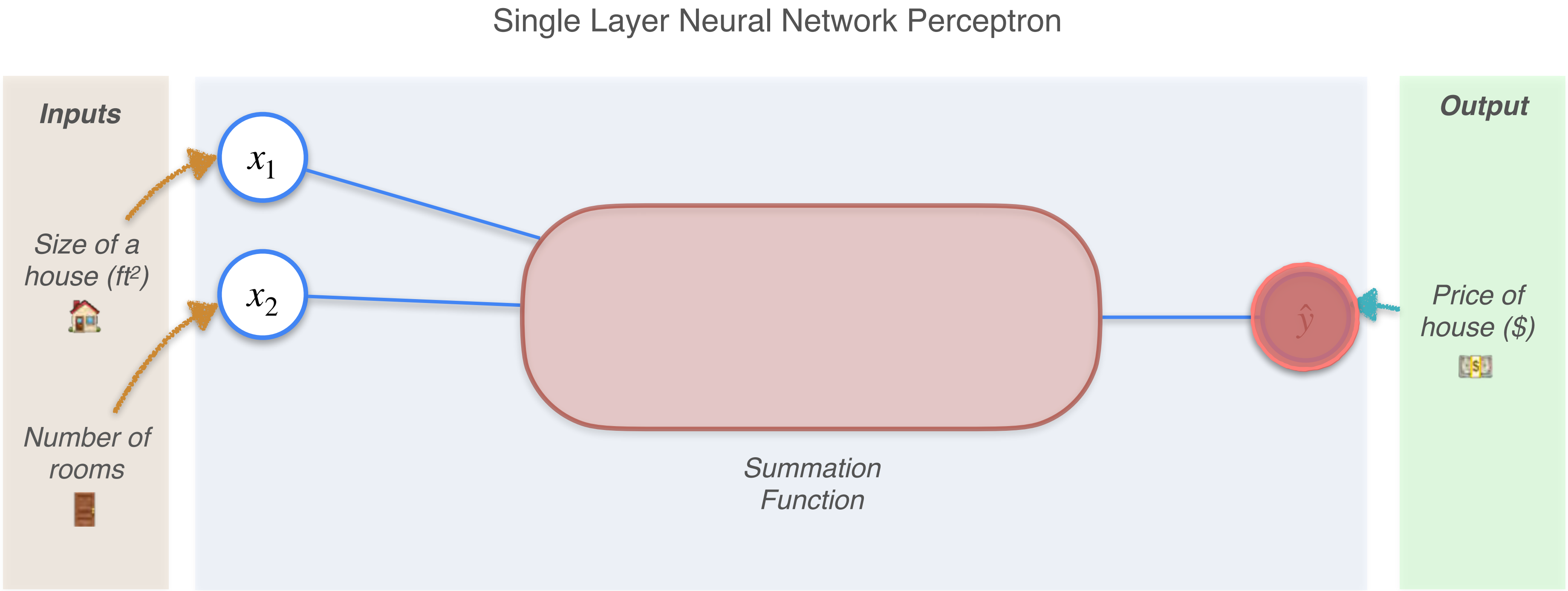
Regression With a Perceptron



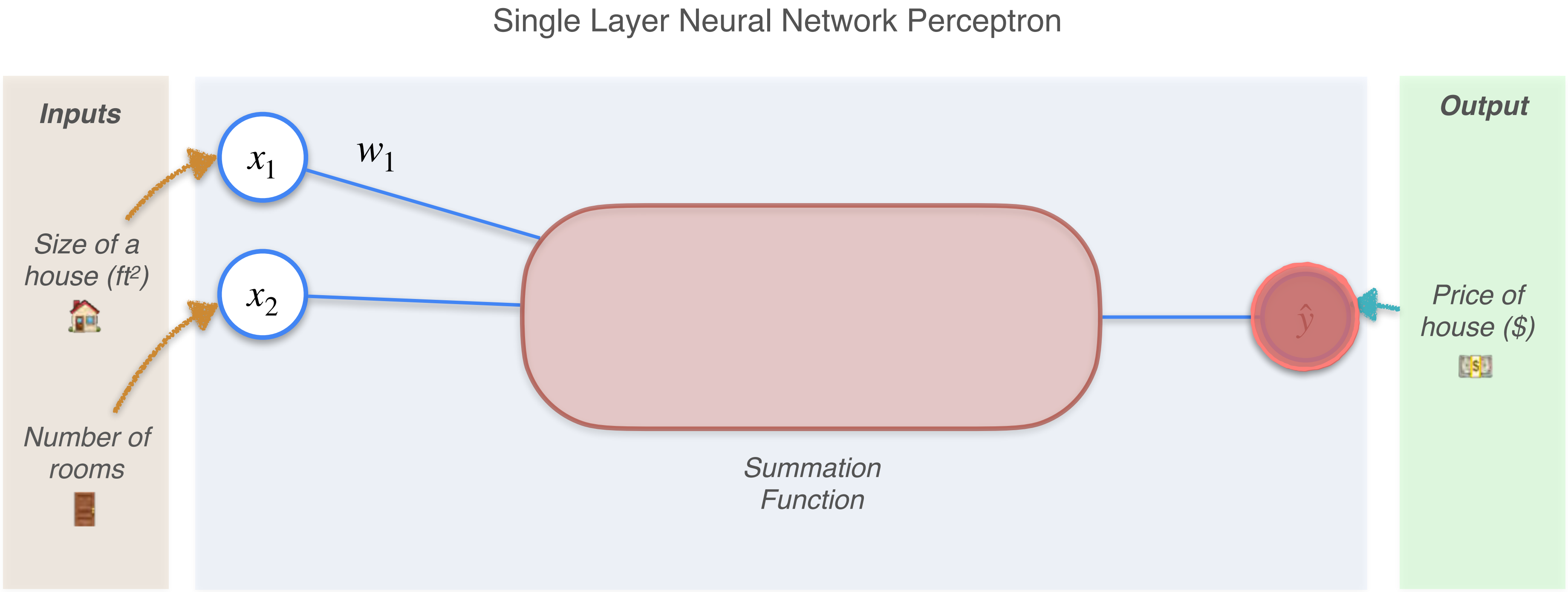
Regression With a Perceptron



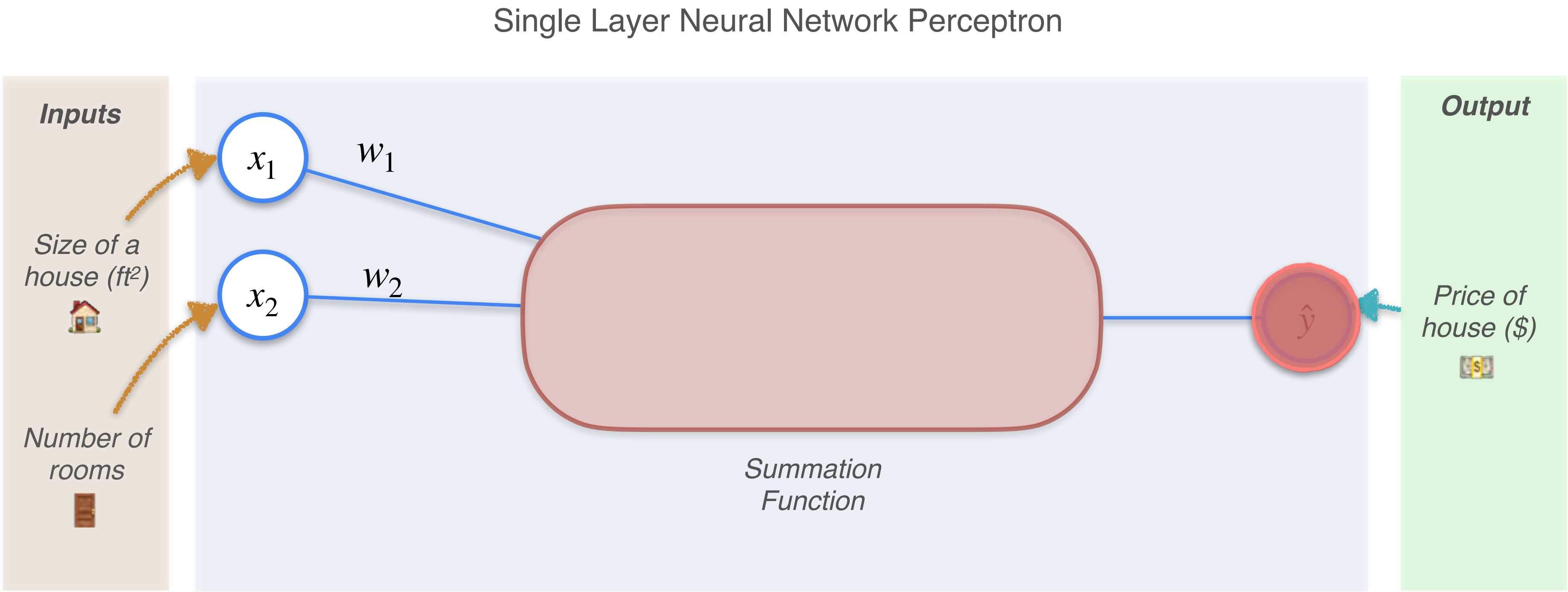
Regression With a Perceptron



Regression With a Perceptron

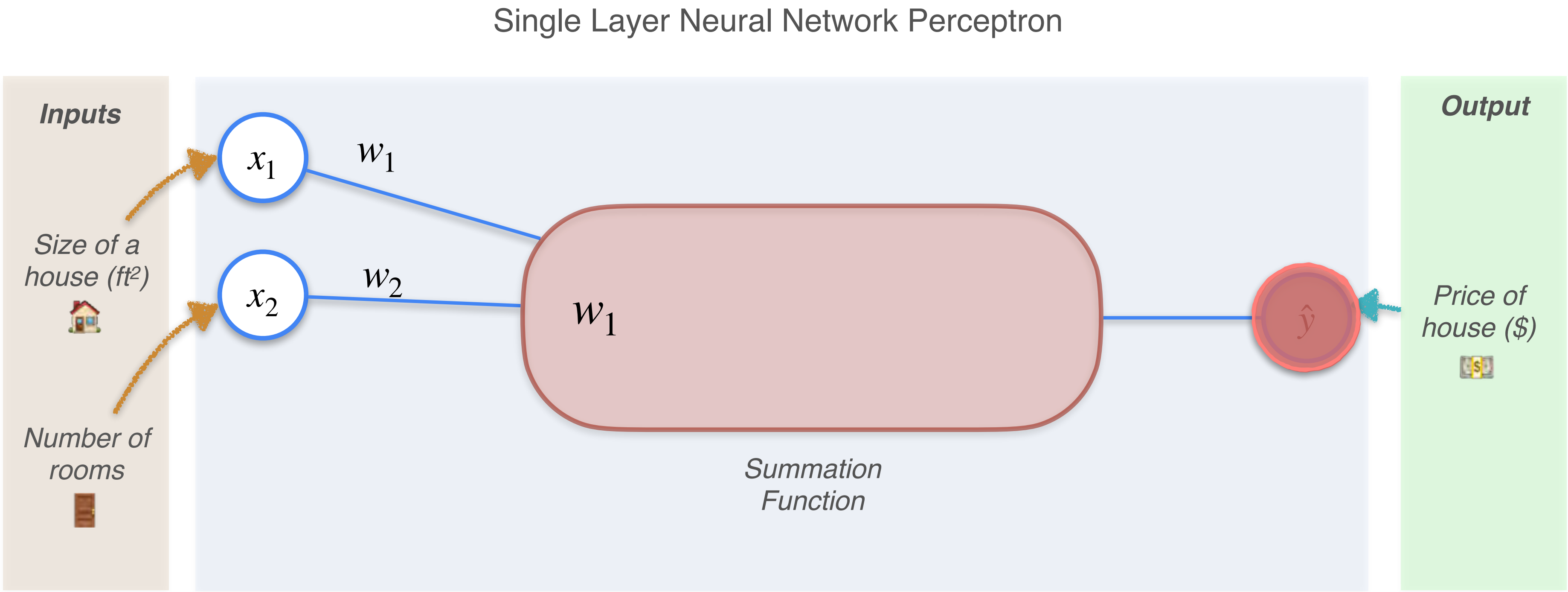


Regression With a Perceptron



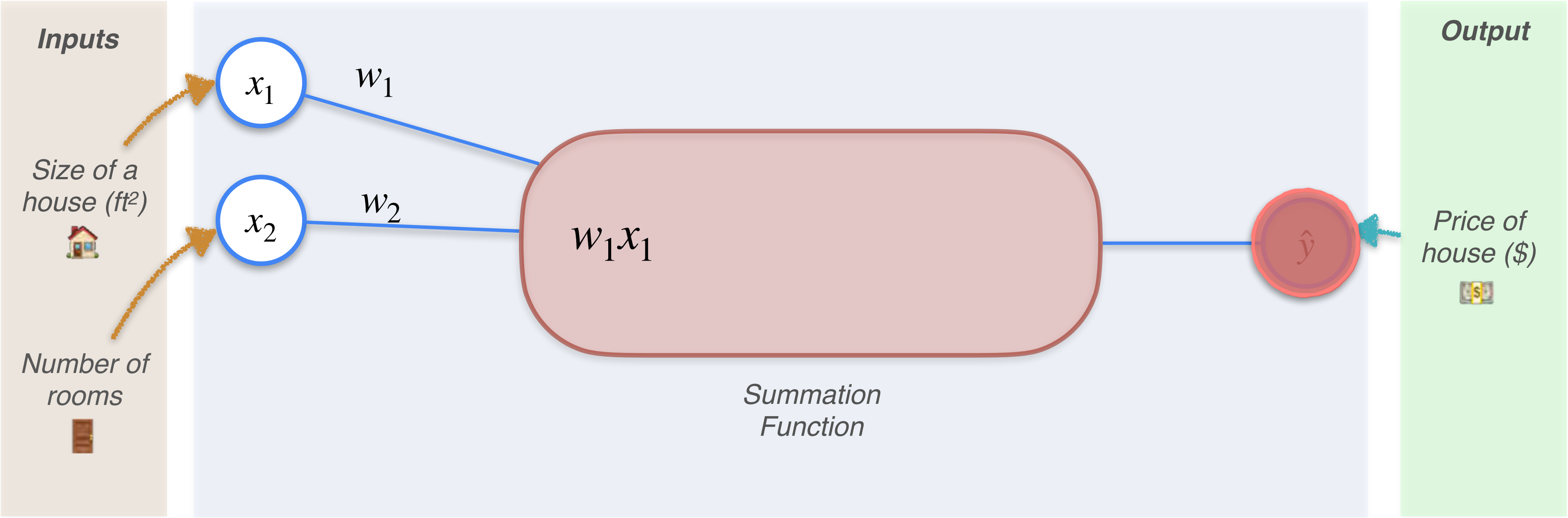
DeepLearning.AI

Regression With a Perceptron

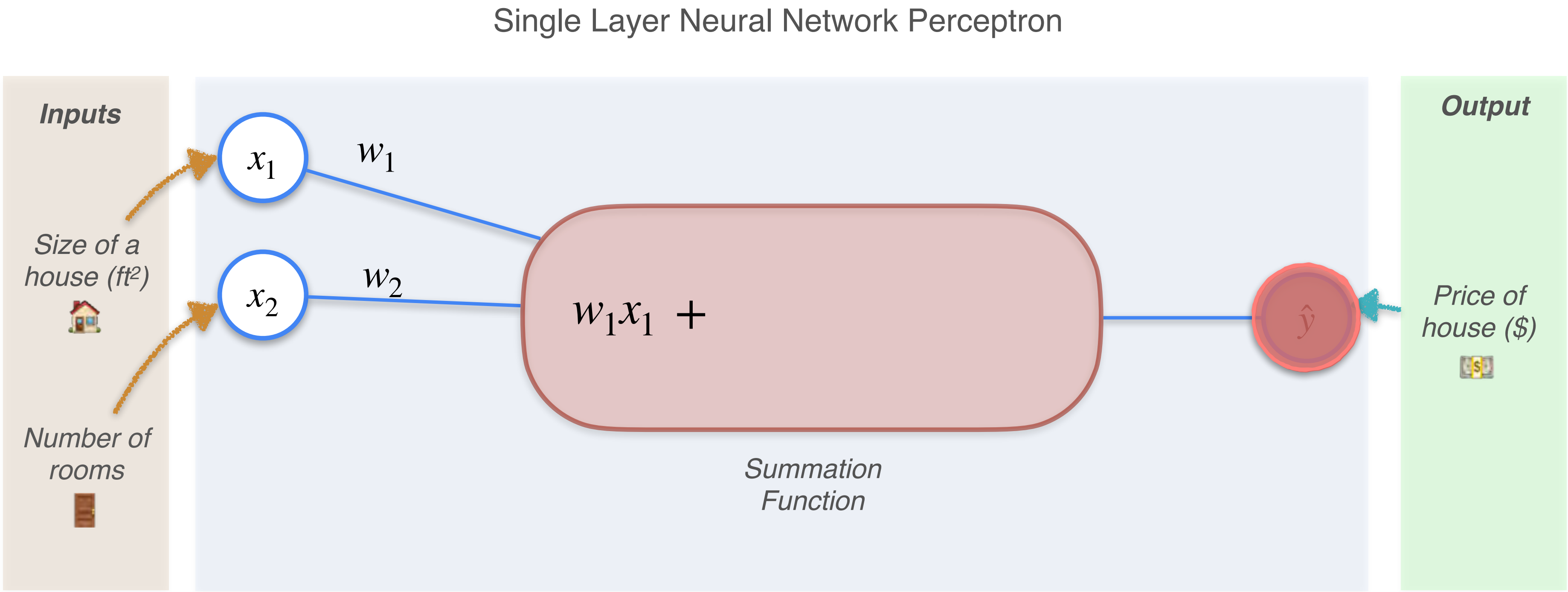


Regression With a Perceptron

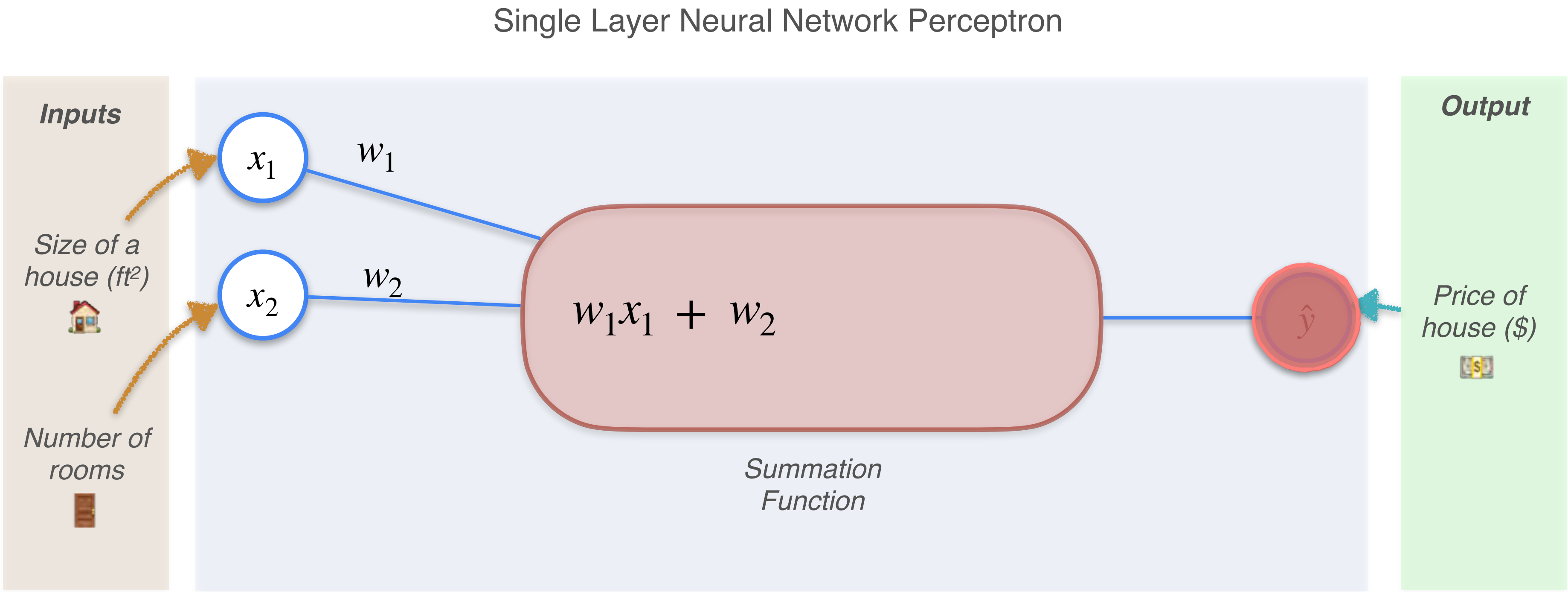
Single Layer Neural Network Perceptron



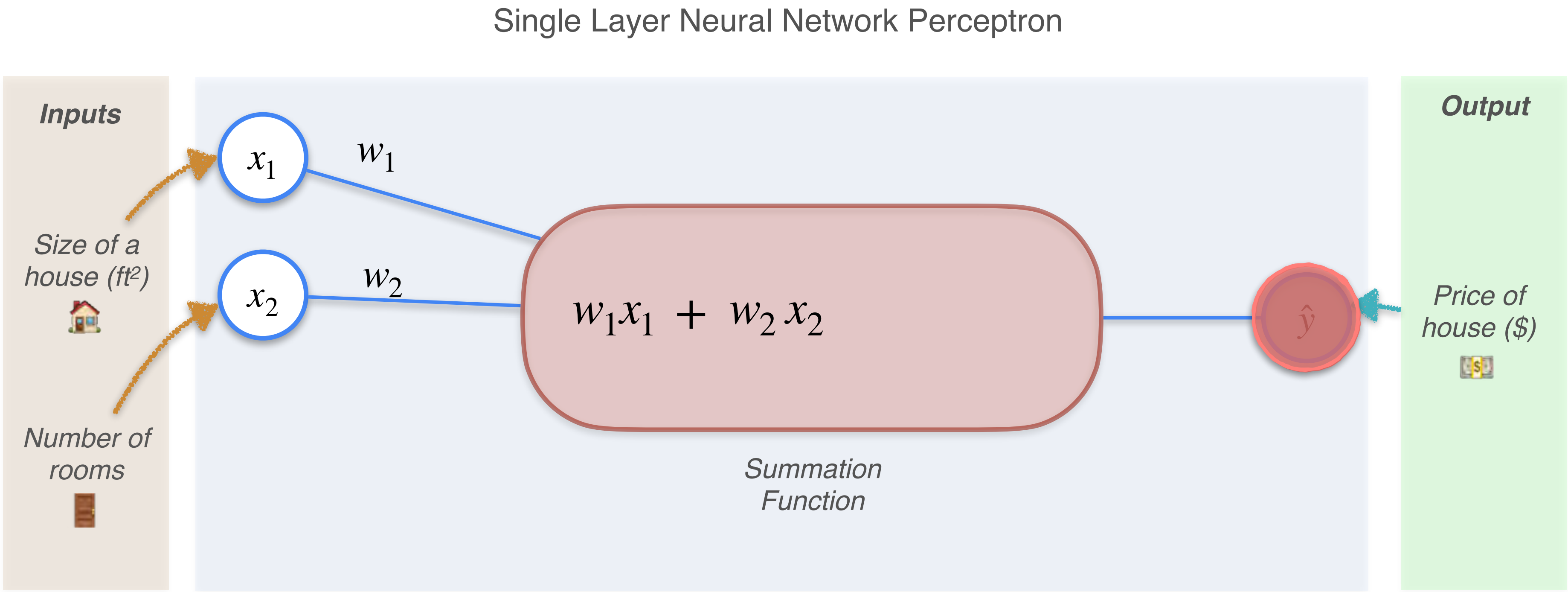
Regression With a Perceptron



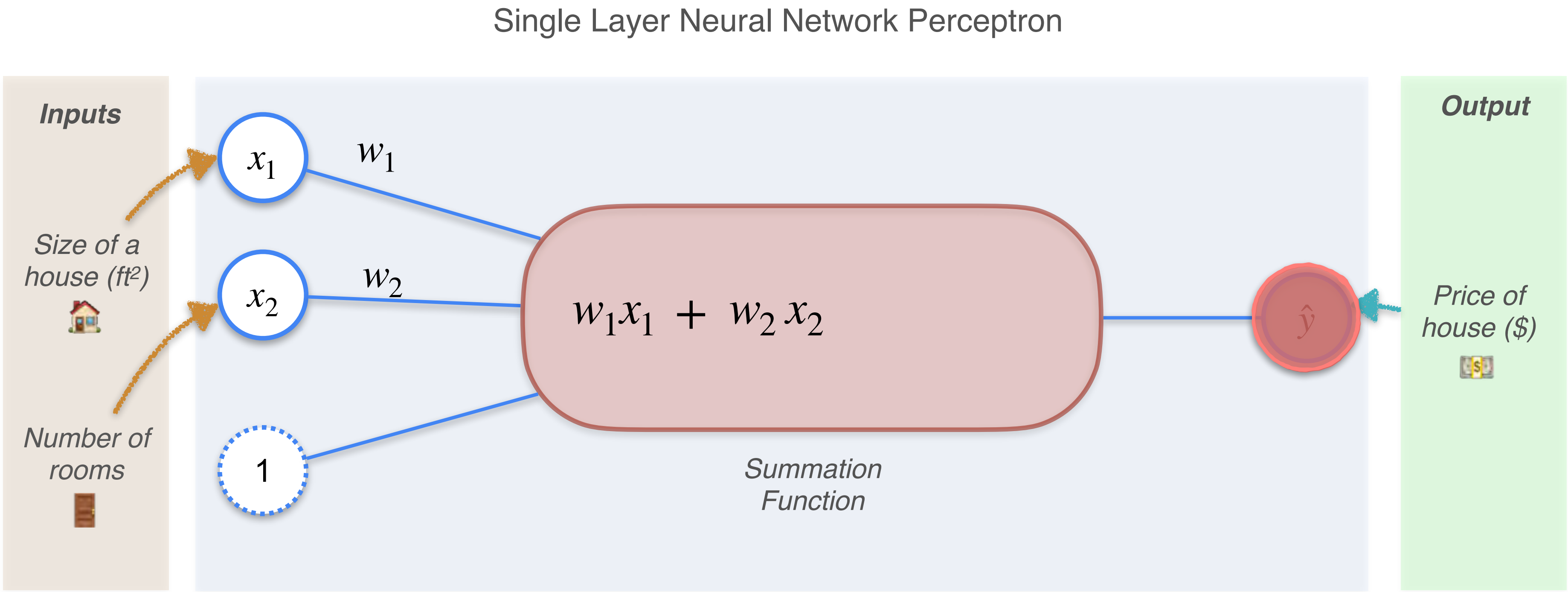
Regression With a Perceptron



Regression With a Perceptron

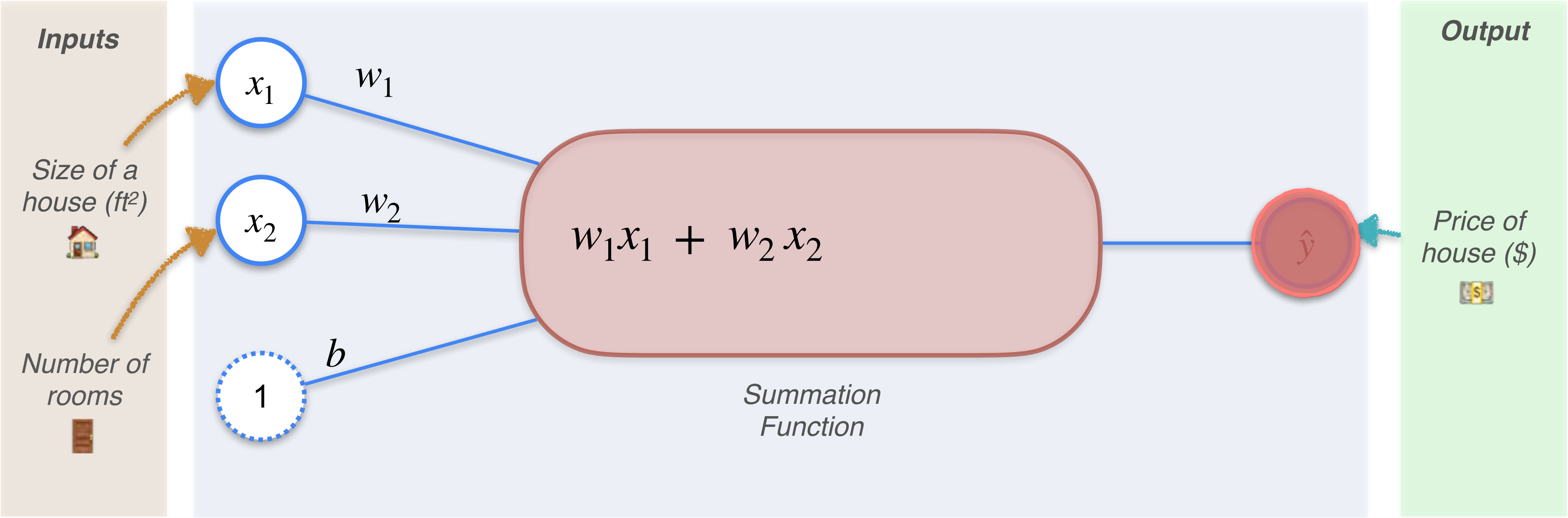


Regression With a Perceptron

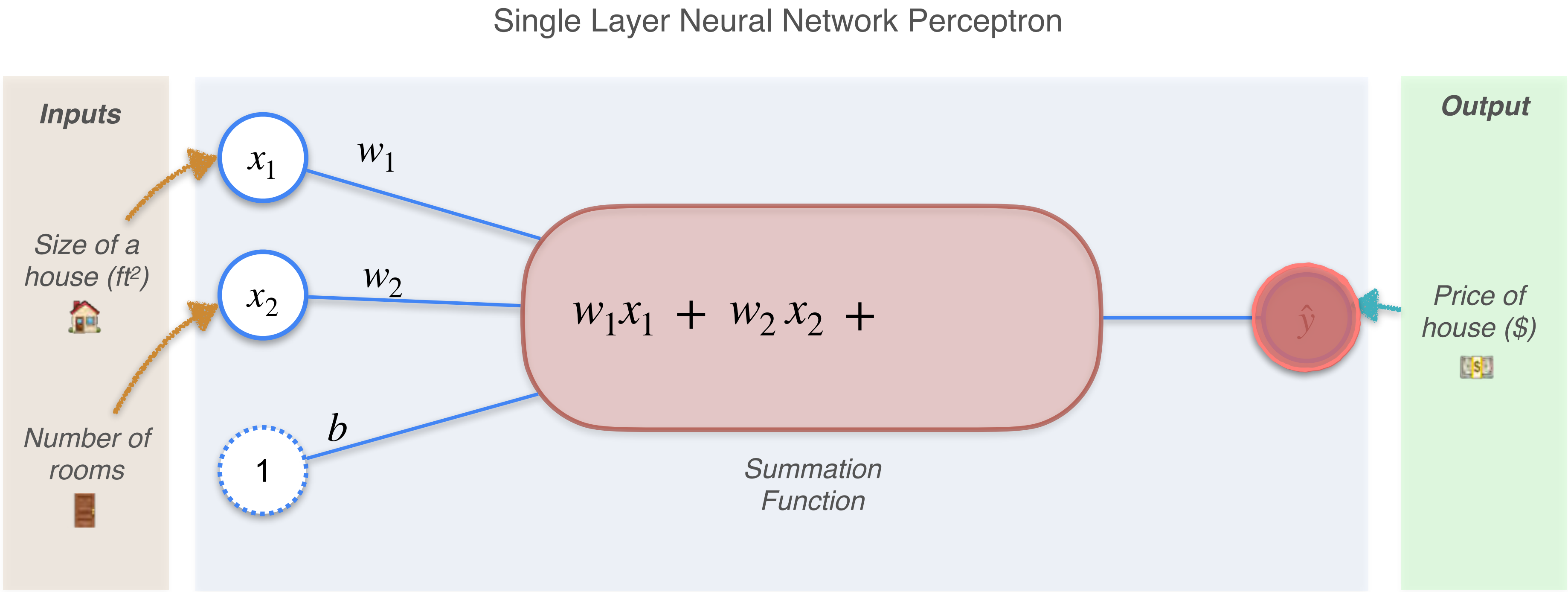


Regression With a Perceptron

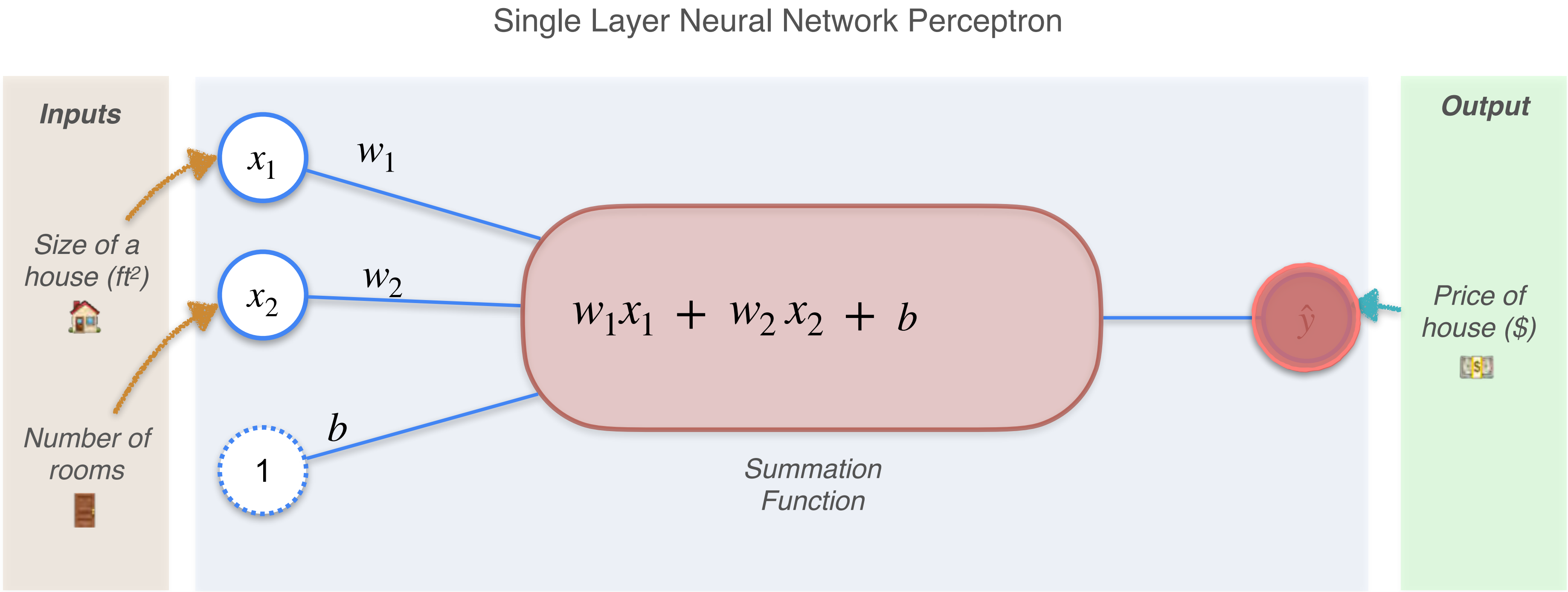
Single Layer Neural Network Perceptron



Regression With a Perceptron

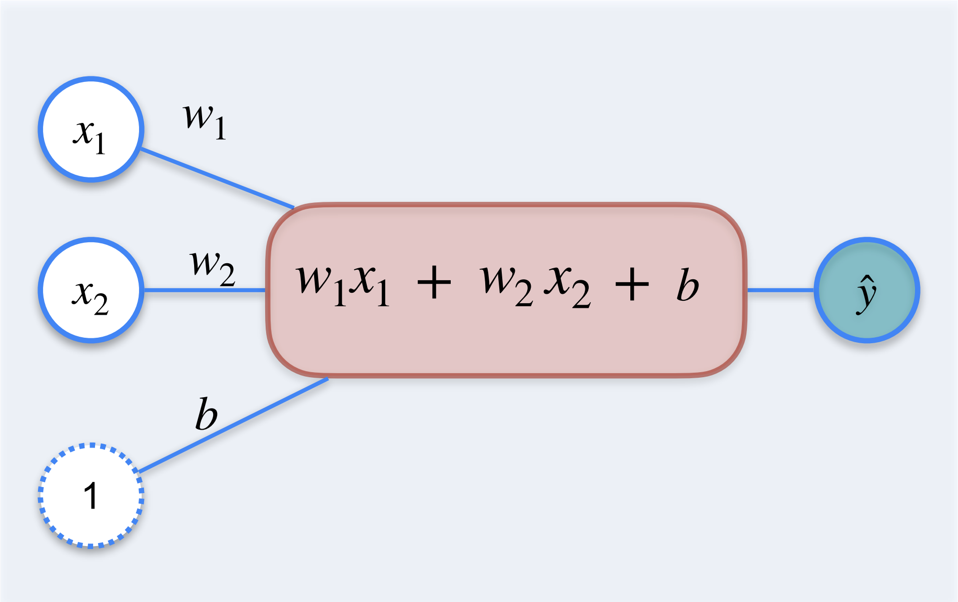


Regression With a Perceptron



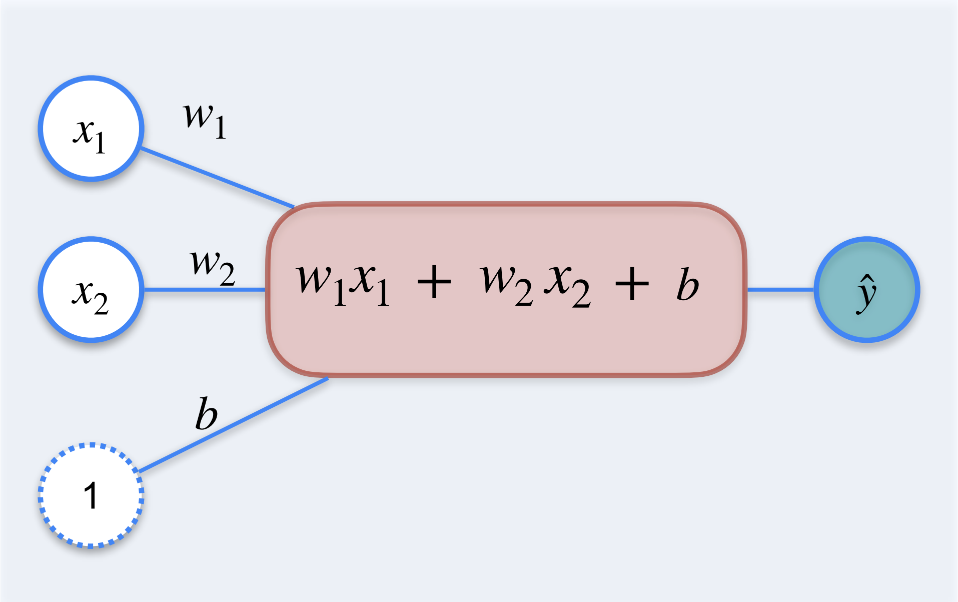
Regression With a Perceptron

Single Layer Neural Network Perceptron



Regression With a Perceptron

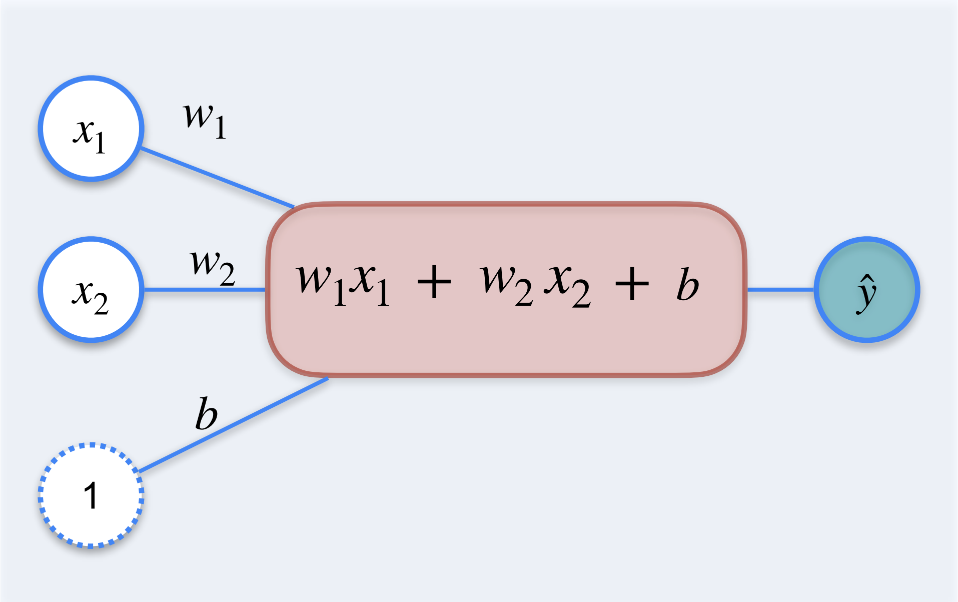
Single Layer Neural Network Perceptron



$\hat{y}$

Regression With a Perceptron

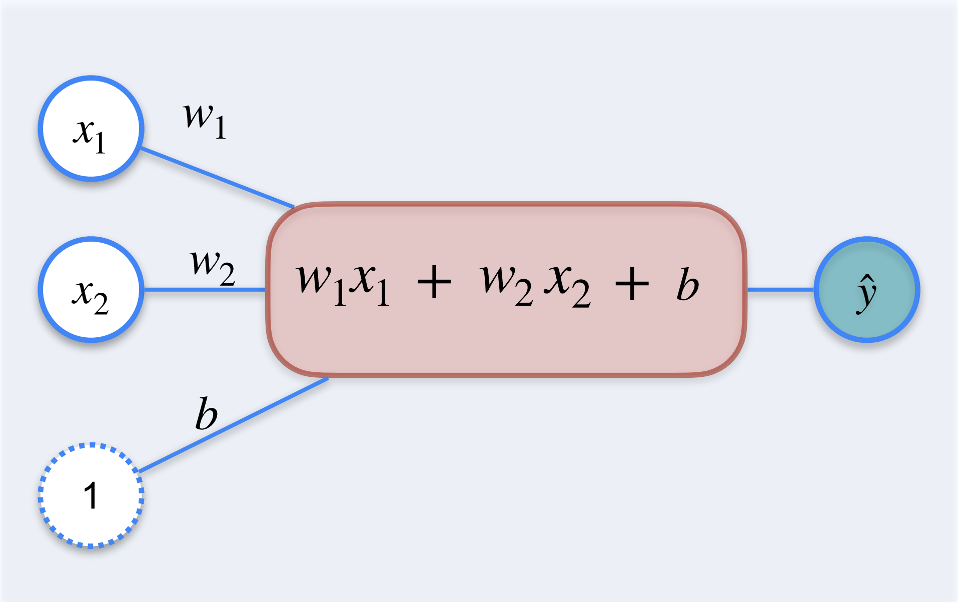
Single Layer Neural Network Perceptron



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

Regression With a Perceptron

Single Layer Neural Network Perceptron

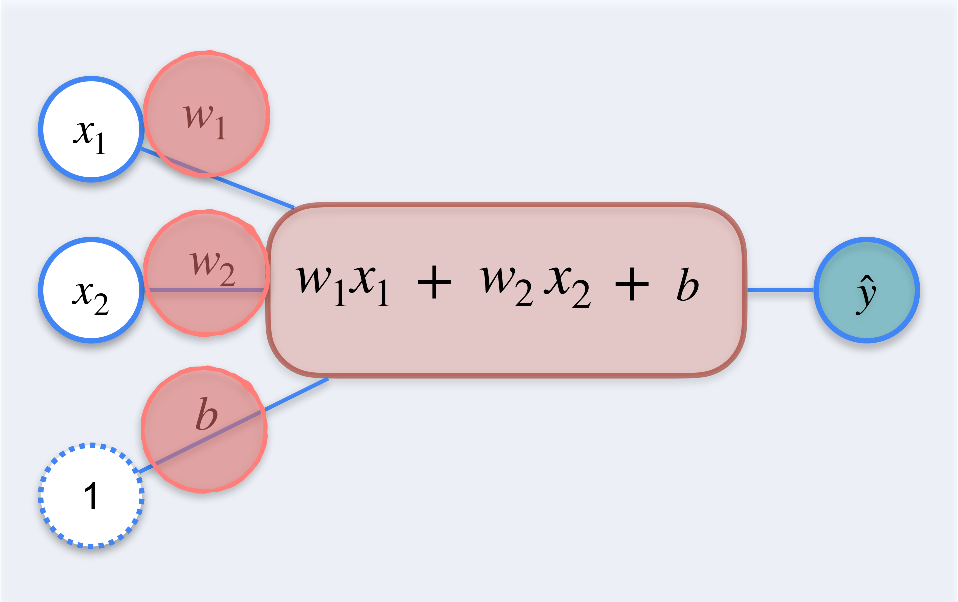


$$\hat{y} = w_1 x_1 + w_2 x_2 + b$$

Main Goal:

Regression With a Perceptron

Single Layer Neural Network Perceptron

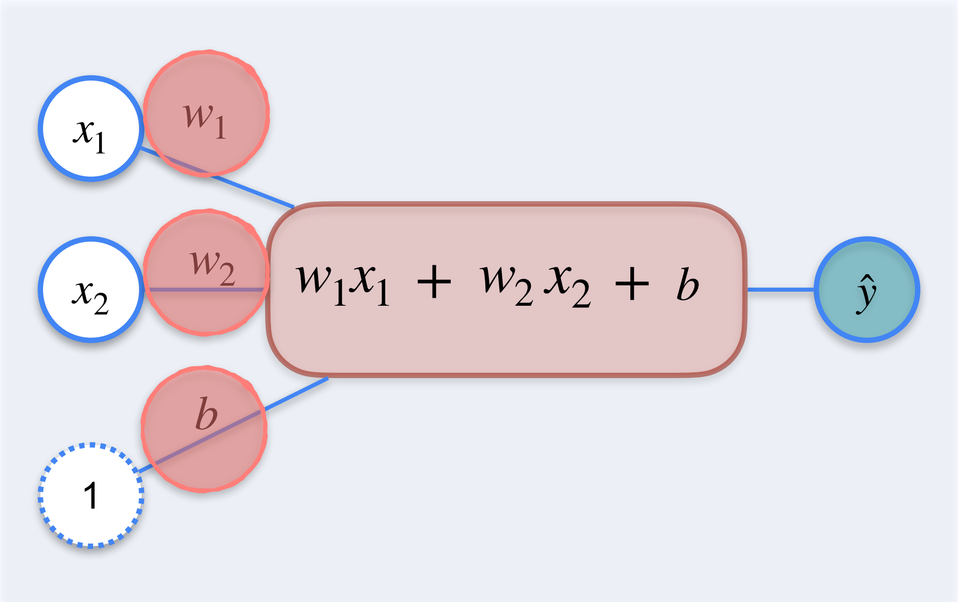


$$\hat{y} = w_1 x_1 + w_2 x_2 + b$$

Main Goal:

Regression With a Perceptron

Single Layer Neural Network Perceptron



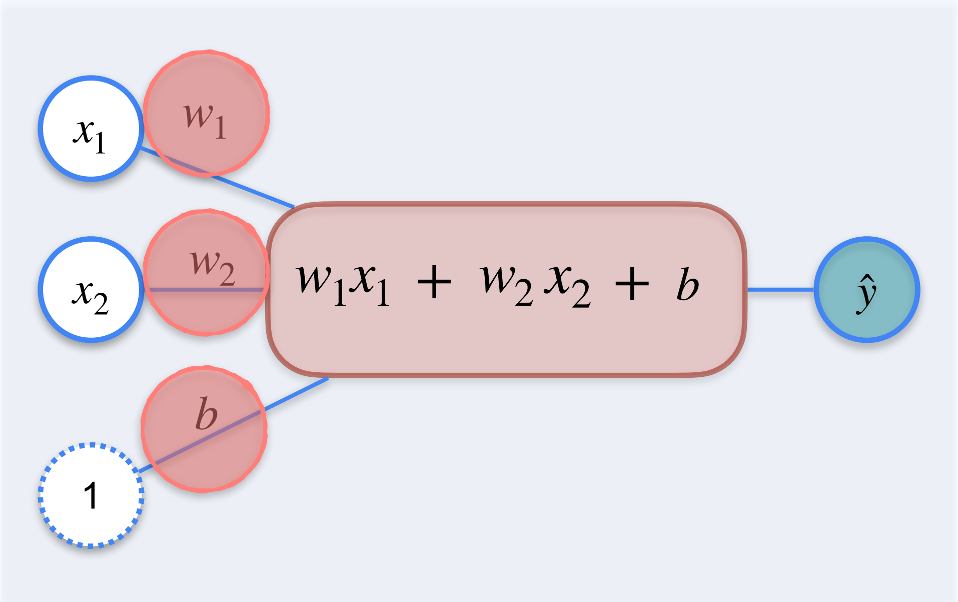
$$\hat{y} = w_1 x_1 + w_2 x_2 + b$$

Main Goal:

Find weights and bias that will optimise the predictions.

Regression With a Perceptron

Single Layer Neural Network Perceptron



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

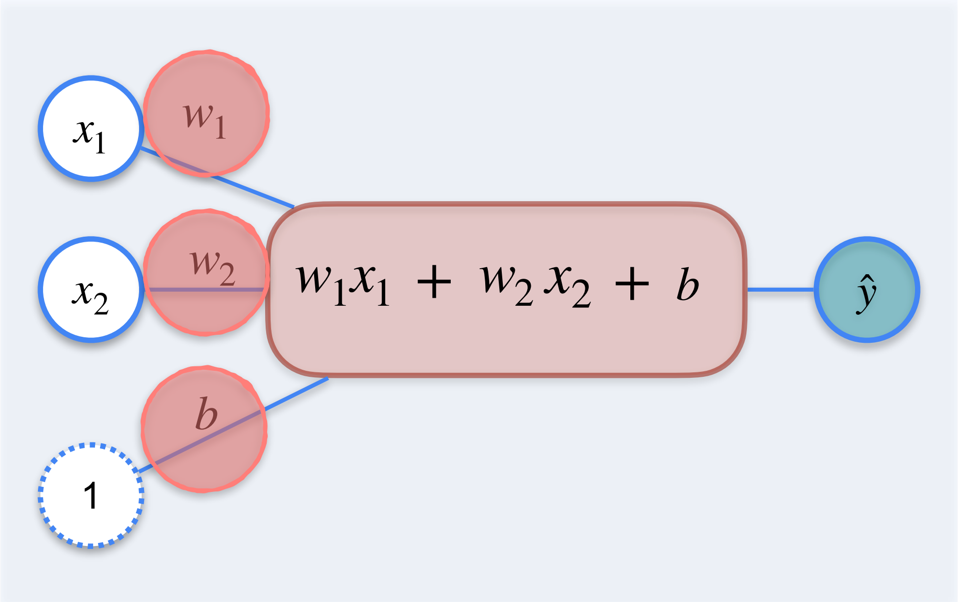
Main Goal:

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



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

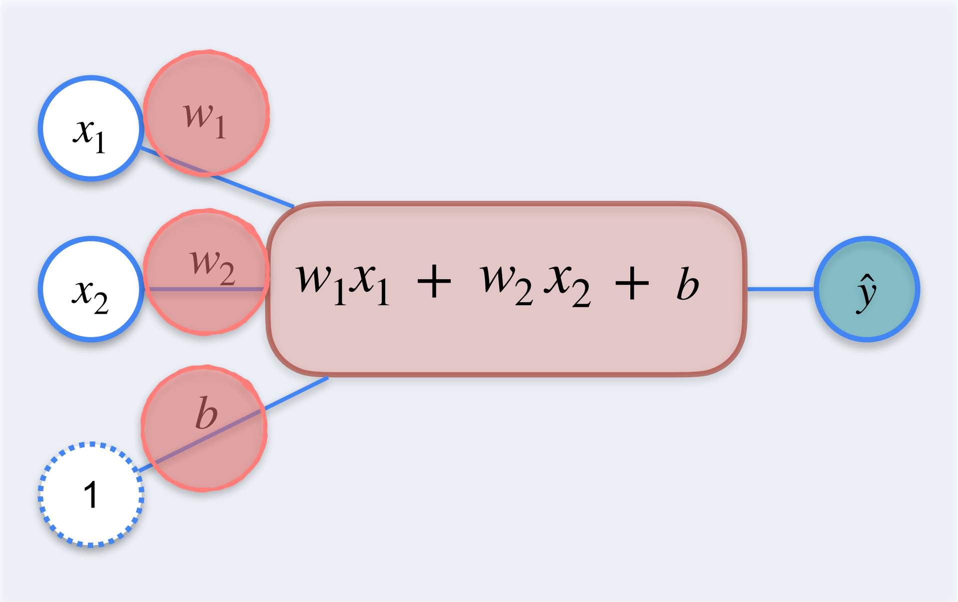
Main Goal:

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



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

Main Goal:

Find weights and bias that will optimise the predictions.

ie. Reduce the errors in the predictions

The Loss Function

DeepLearning.AI

Optimization in Neural Networks and Newton's Method

Regression with a perceptron: Loss function

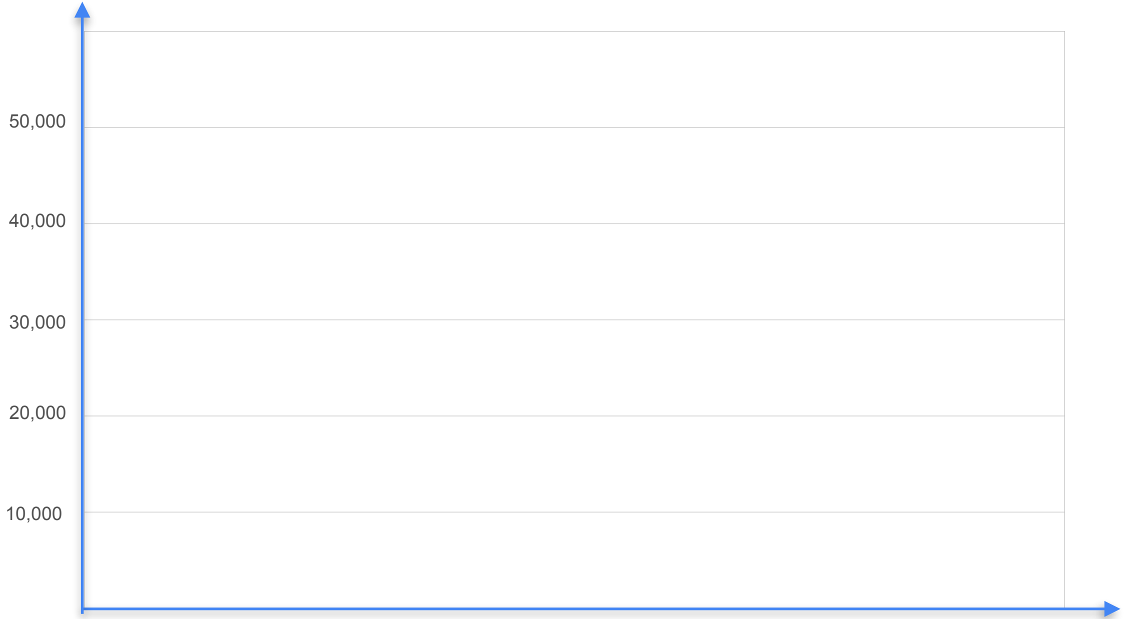
Mean Squared Error

Mean Squared Error

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

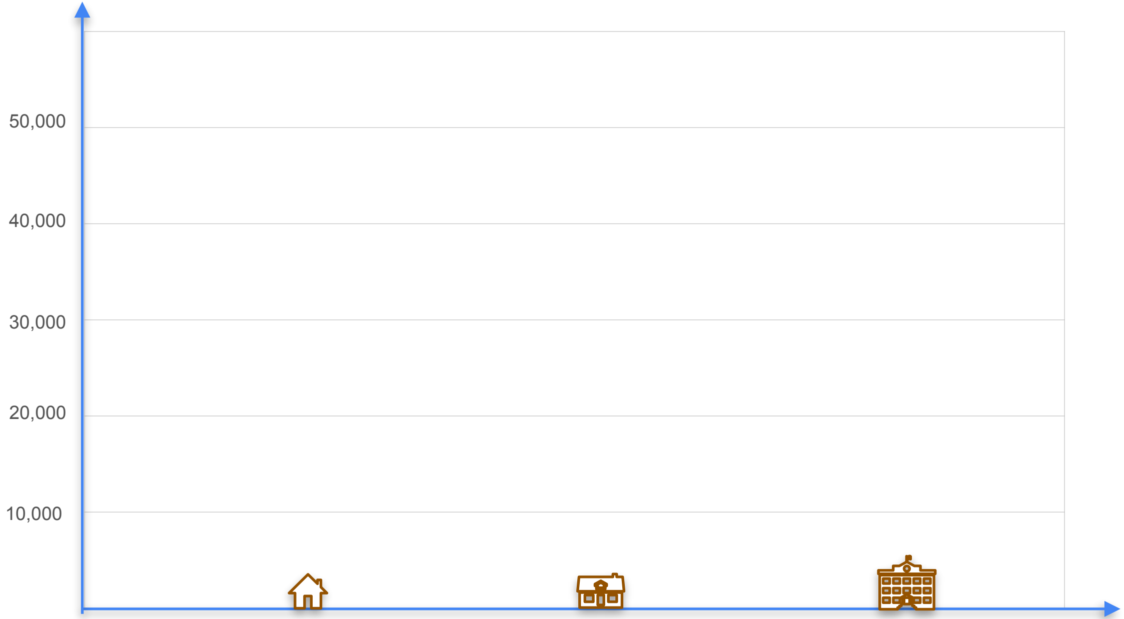
Mean Squared Error

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



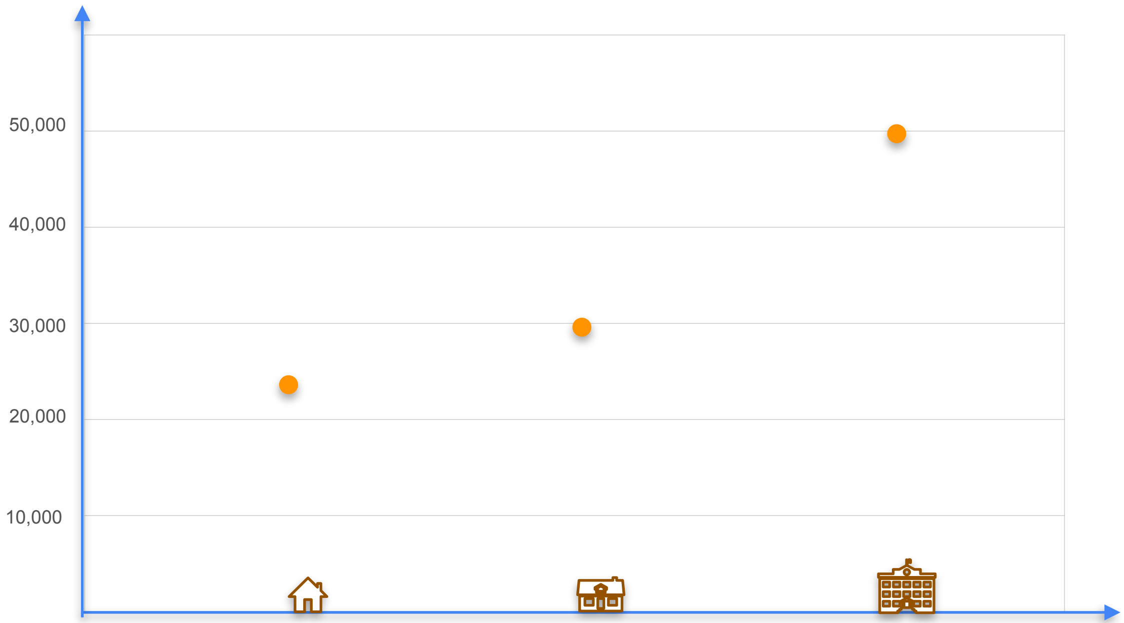
Mean Squared Error

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



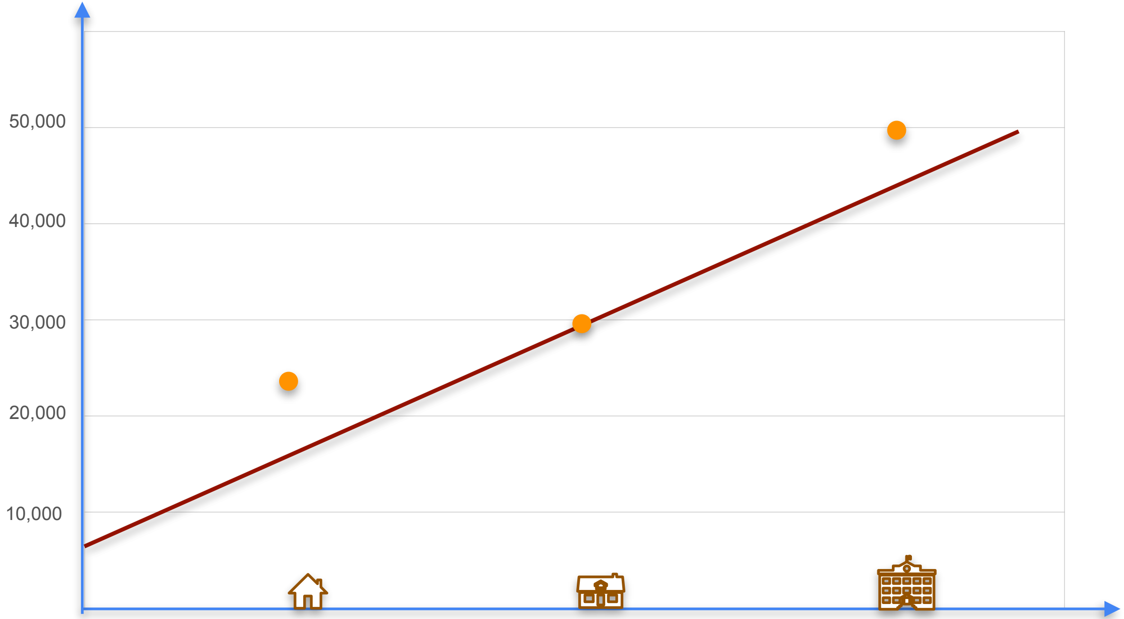
Mean Squared Error

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



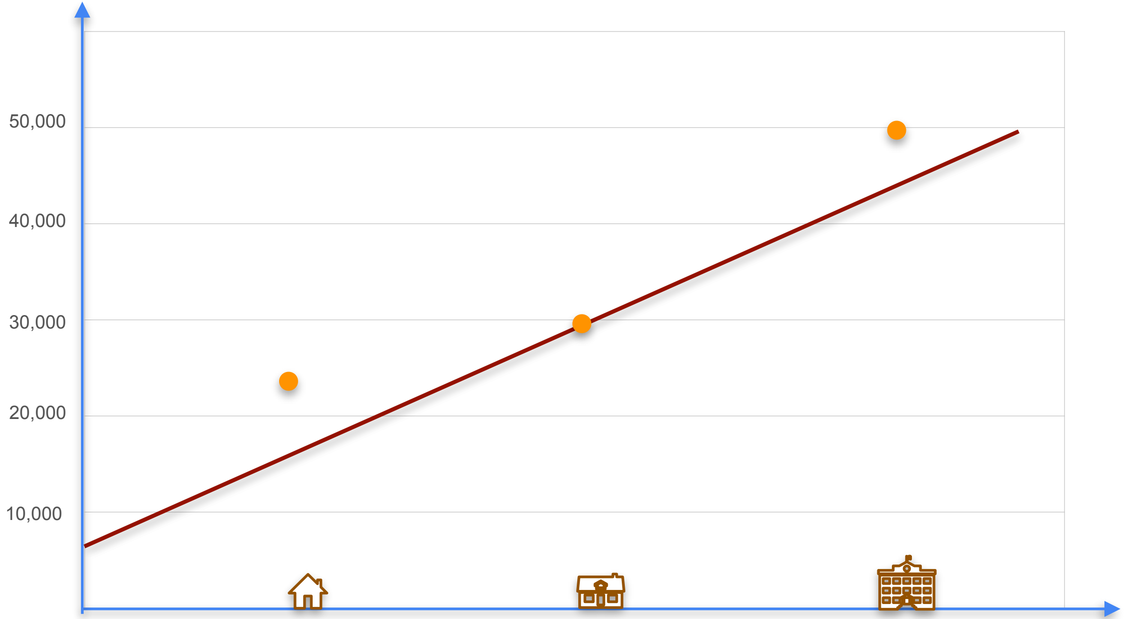
Mean Squared Error

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



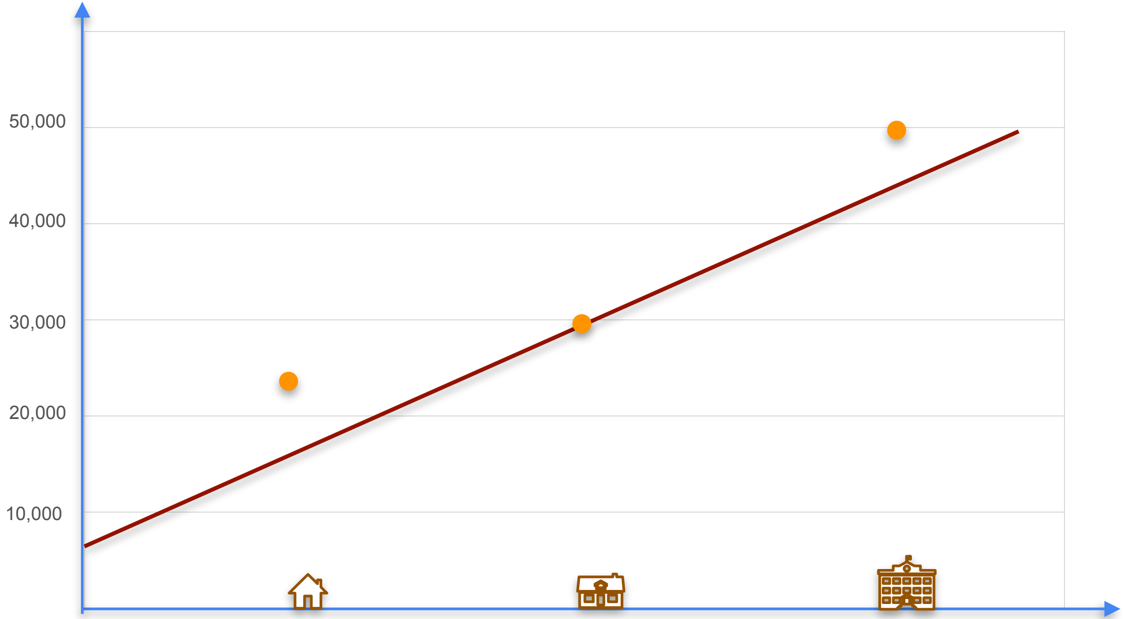
Mean Squared Error

	y	$\hat{y}$	
	$20,000	$15,000	
	$30,000	$30,000	
	$50,000	$45,000	



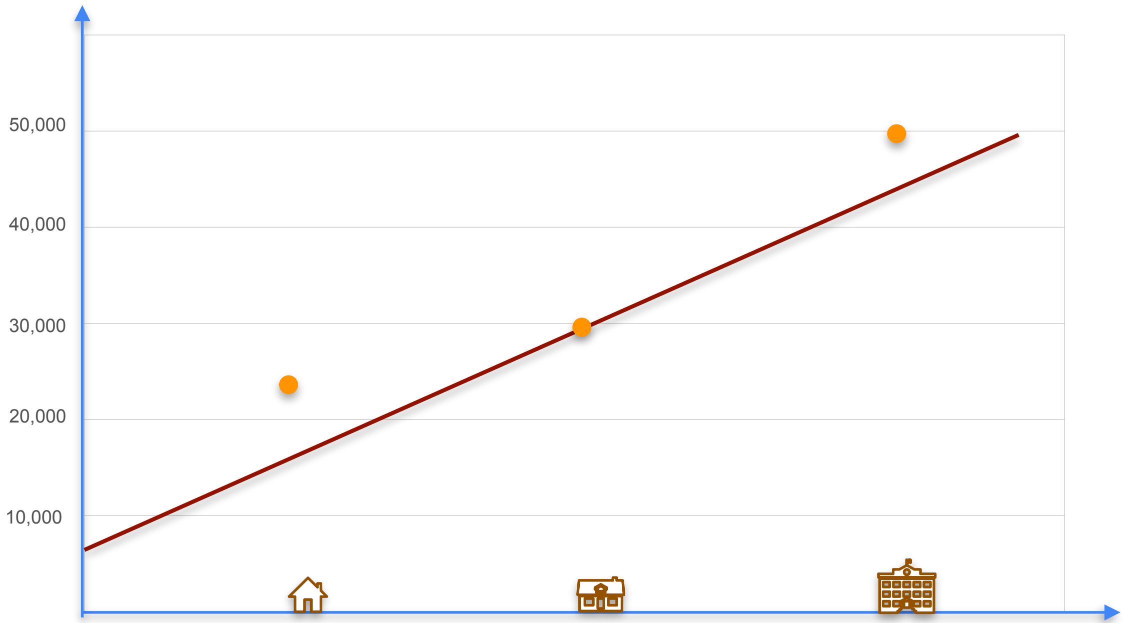
Mean Squared Error

	y	$\hat{y}$	$y - \hat{y}$
	$20,000	$15,000	
	$30,000	$30,000	
	$50,000	$45,000	



Mean Squared Error

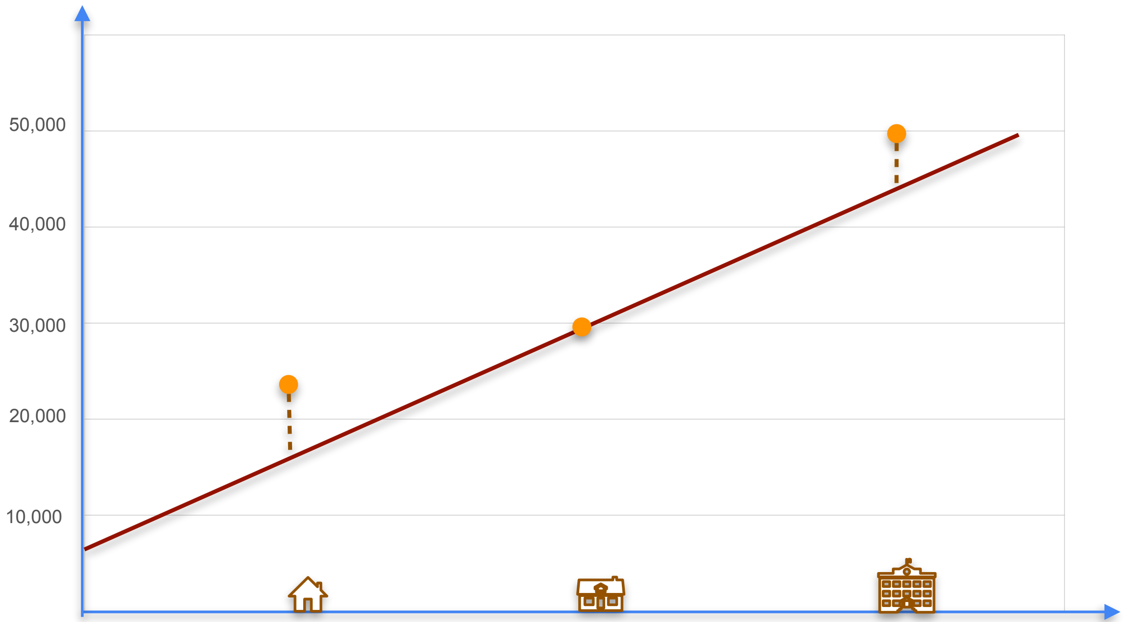
	y	$\hat{y}$	$y - \hat{y}$
	$20,000	$15,000	Error
	$30,000	$30,000	Error
	$50,000	$45,000	Error



DeepLearning.AI

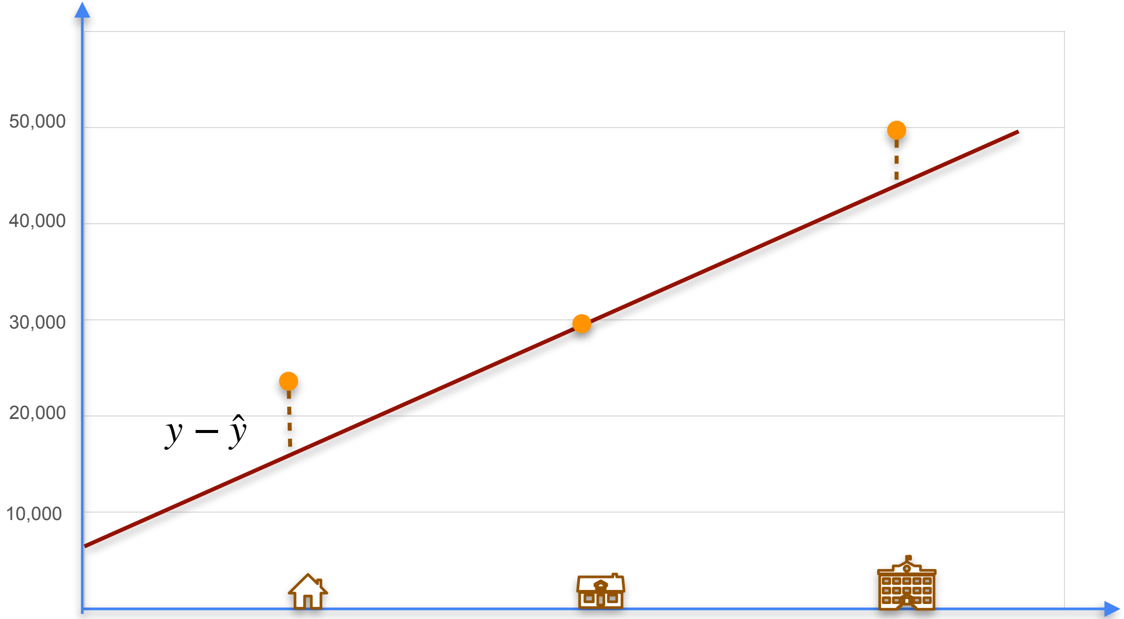
Mean Squared Error

	y	$\hat{y}$	$y - \hat{y}$
	$20,000	$15,000	**Error**
	$30,000	$30,000	**Error**
	$50,000	$45,000	**Error**



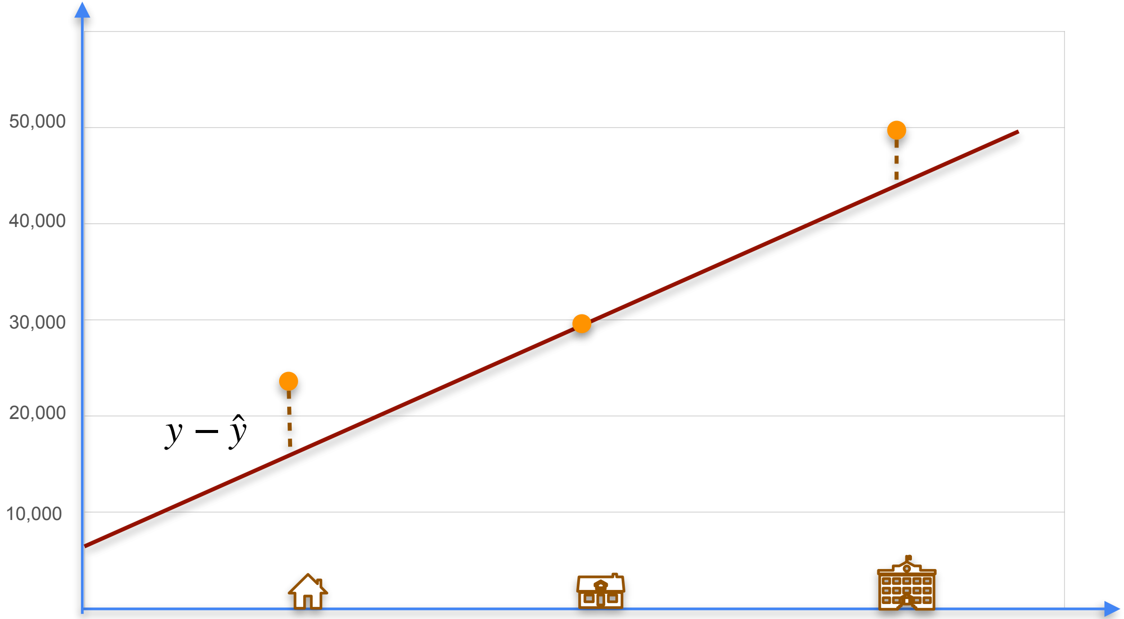
Mean Squared Error

	y	$\hat{y}$	$y - \hat{y}$
	$20,000	*$15,000*	***Error***
	$30,000	*$30,000*	***Error***
	$50,000	*$45,000*	***Error***



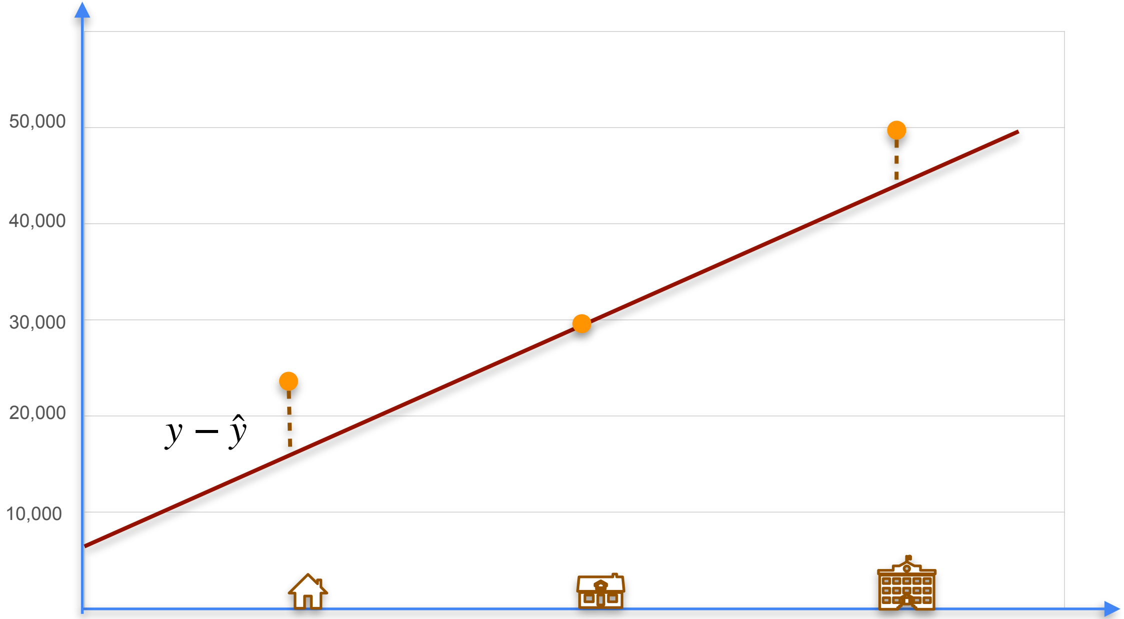
Mean Squared Error

	y	$\hat{y}$	$(y - \hat{y})^2$
	\$20,000	*\$15,000*	***Error***
	\$30,000	*\$30,000*	***Error***
	\$50,000	*\$45,000*	***Error***



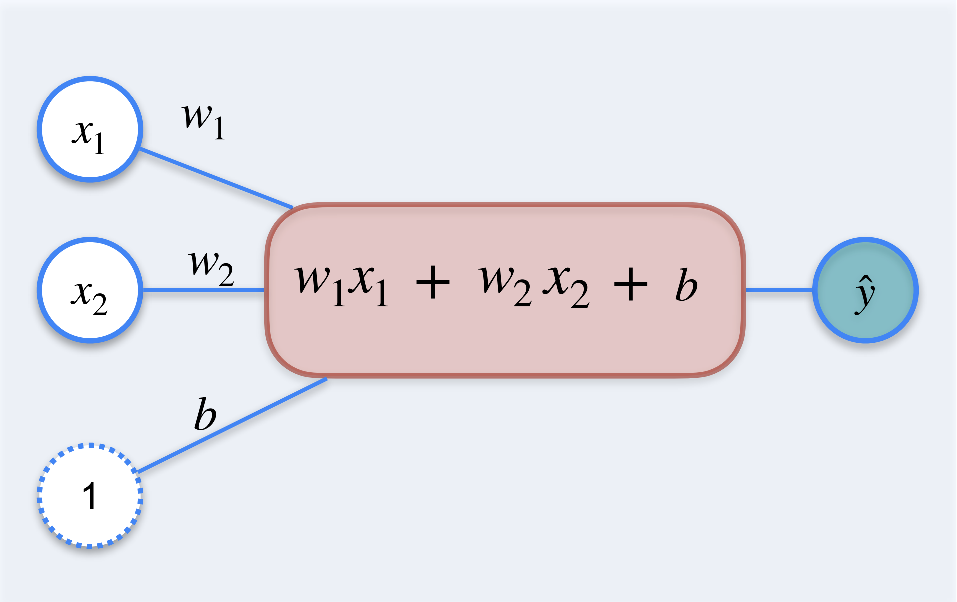
Mean Squared Error

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



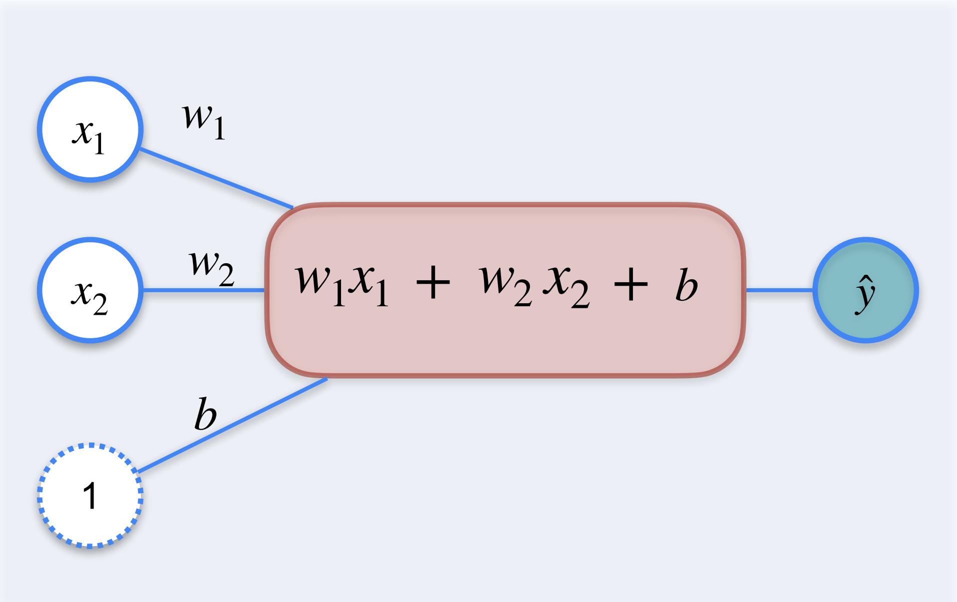
Regression With a Perceptron

Single Layer Neural Network Perceptron



Regression With a Perceptron

Single Layer Neural Network Perceptron

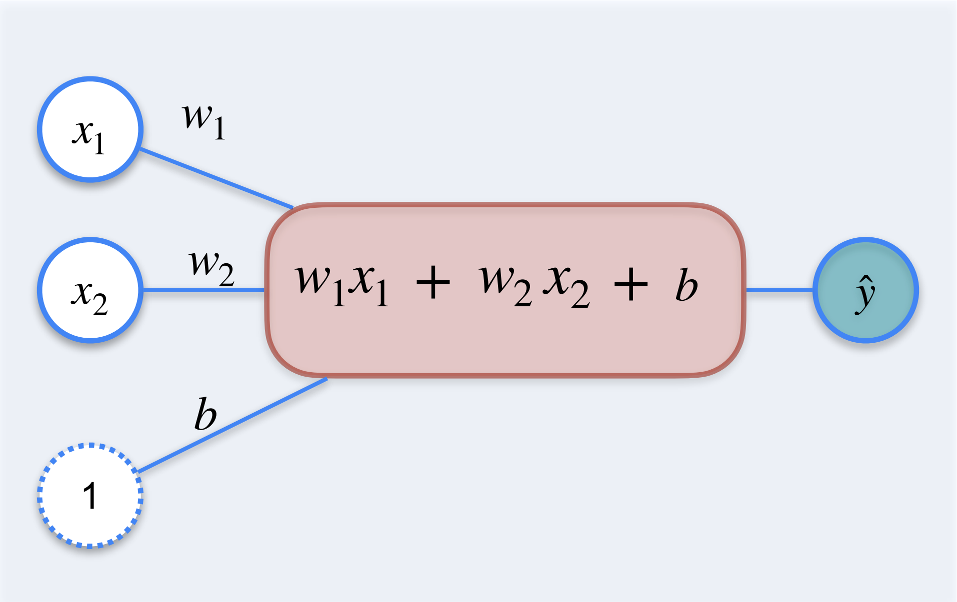


Prediction Function:

$\hat{y}$

Regression With a Perceptron

Single Layer Neural Network Perceptron

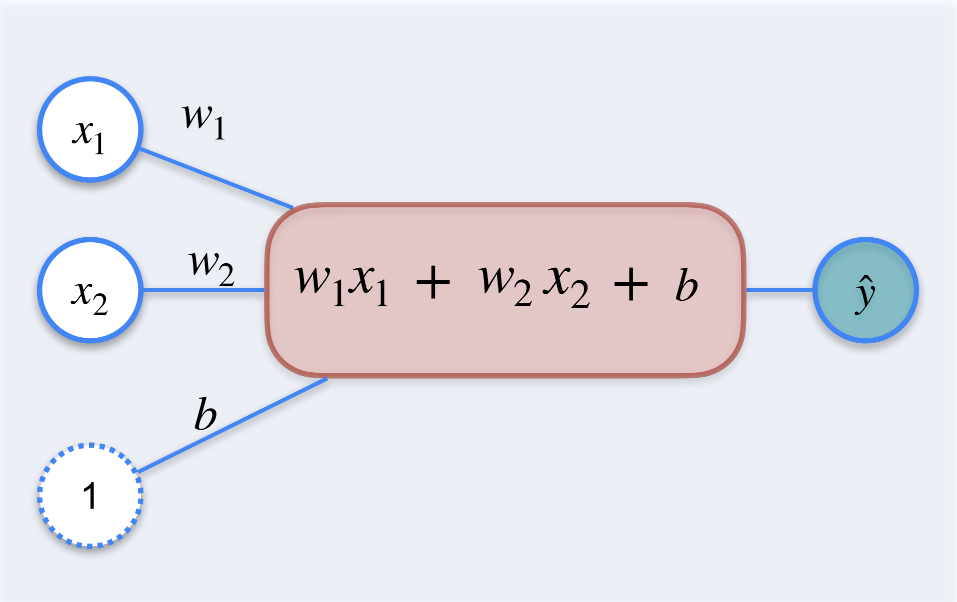


Prediction Function:

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

Regression With a Perceptron

Single Layer Neural Network Perceptron



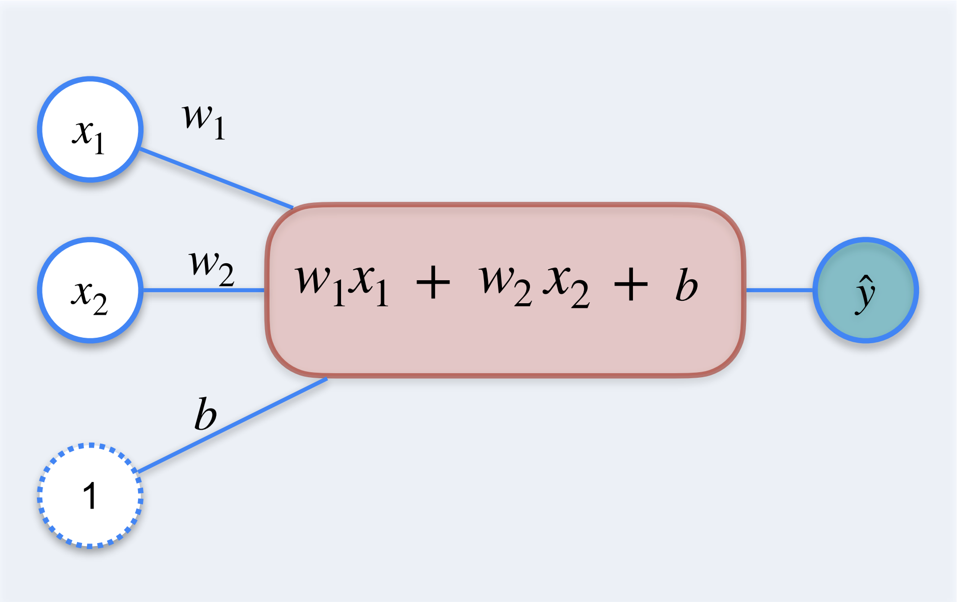
Prediction Function:

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

Loss Function:

Regression With a Perceptron

Single Layer Neural Network Perceptron



Prediction Function:

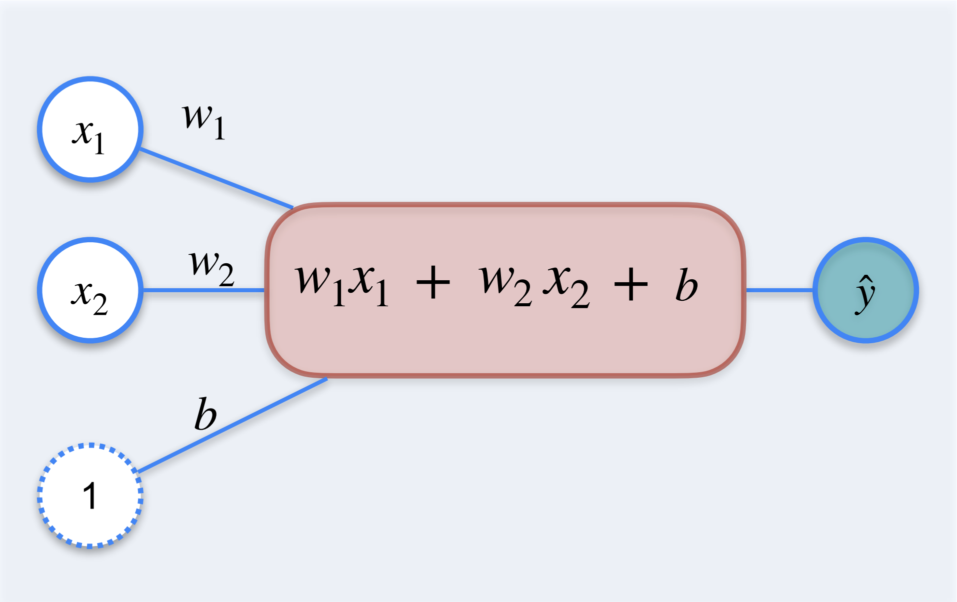
$$\hat{y} = w_1x_1 + w_2x_2 + b$$

Loss Function:

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

Regression With a Perceptron

Single Layer Neural Network Perceptron



Prediction Function:

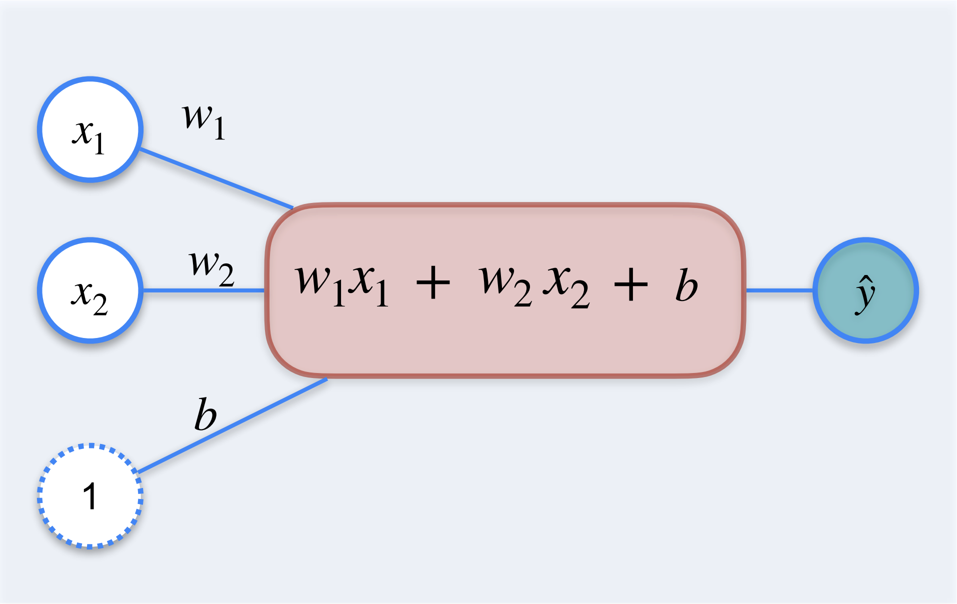
$$\hat{y} = w_1x_1 + w_2x_2 + b$$

Loss Function:

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

Regression With a Perceptron

Single Layer Neural Network Perceptron



Prediction Function:

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

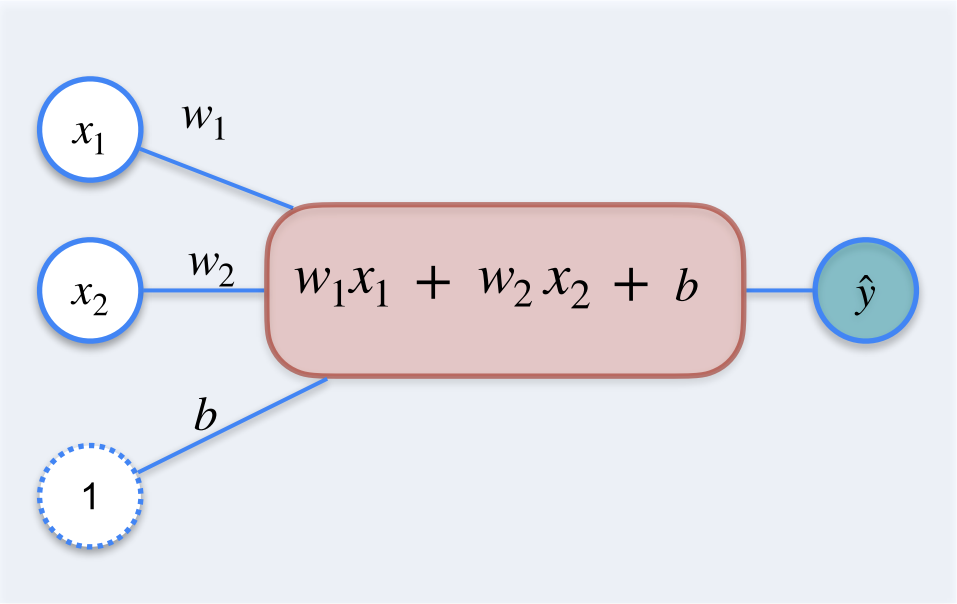
Loss Function:

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

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 $\hat{y}$ with the least error

Optimization in Neural Networks and Newton's Method

DeepLearning.AI

Regression with a perceptron: Gradient Descent

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

Main Goal:

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

Main Goal:

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

Main Goal:

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

To find optimal values for:

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

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:

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

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

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

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

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

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

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

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

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

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

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

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

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

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

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

Some initial starting values

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

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

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

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

Some initial starting values

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

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

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

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

Some initial starting values

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

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

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

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

Some initial starting values

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

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

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

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

Regression With a Perceptron

Prediction Function:

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

Loss Function:

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

Some initial starting values

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

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

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

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

SUB-TASK

Find the following partial derivatives

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$$\frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

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

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

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

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

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

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

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

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

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

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \partial \hat{y}$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

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

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

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

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \partial \hat{y}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} =$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} =$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

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

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \partial \hat{y}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

Using chain rule:

$$\frac{\partial L}{\partial b} = \boxed{\frac{\partial L}{\partial \hat{y}}} \cdot \boxed{\frac{\partial \hat{y}}{\partial b}}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \boxed{\frac{\partial \hat{y}}{\partial w_1}}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \boxed{\frac{\partial \hat{y}}{\partial w_2}}$$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= (y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= (y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

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

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	$= x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	$= x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	$= x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	$= x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	$= x_2$

Regression With Perceptron

Prediction Function:

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

Loss Function:

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

$\frac{\partial L}{\partial \hat{y}}$	$= -(y - \hat{y})$
$\frac{\partial \hat{y}}{\partial b}$	$= 1$
$\frac{\partial \hat{y}}{\partial w_1}$	$= x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	$= x_2$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \text{-}(y - \hat{y})$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

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

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

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

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \qquad = \text{-}(y - \hat{y})$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \qquad = \text{-}(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \qquad = \text{-}(y - \hat{y})$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = -(y - \hat{y})\, x_2$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = -(y - \hat{y})\, x_2$$

Regression With Perceptron

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

$$\frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial \hat{y}}{\partial w_1} = x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y - \hat{y})\, x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = -(y - \hat{y})\, x_2$$

Regression With a Perceptron

Main Goal:

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

Regression With a Perceptron

Main Goal:

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

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

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

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 - \alpha$$

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 - \alpha(-x_1(y - \hat{y}))$$

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

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

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

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

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

$$w_2 = w_2 - \alpha$$

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 = w_2 - \alpha(-x_2(y - \hat{y}))$$

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 = w_2 - \alpha(-x_2(y - \hat{y}))$$

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

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 = w_2 - \alpha(-x_2(y - \hat{y}))$$

$$b = b - \alpha$$

Regression With a Perceptron

Main Goal:

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

ie. optimal values for:
w_1 , w_2 , b

Perform Gradient Descent

$$w_1 = w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 = w_2 - \alpha(-x_2(y - \hat{y}))$$

$$b = b - \alpha(-(y - \hat{y}))$$

Optimization in Neural Networks and Newton's Method

DeepLearning.AI

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* 😊

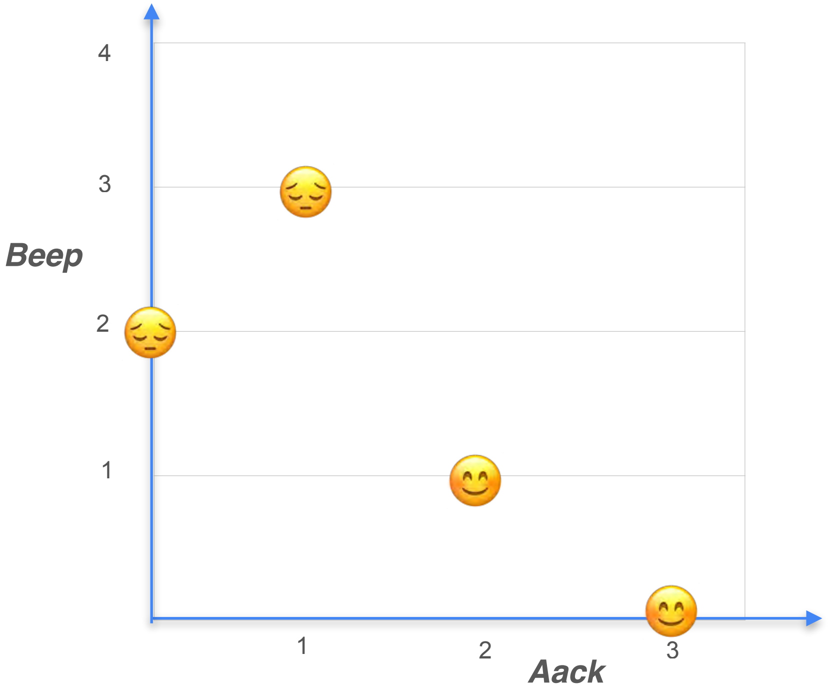
Classification Problem Motivation

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 😊

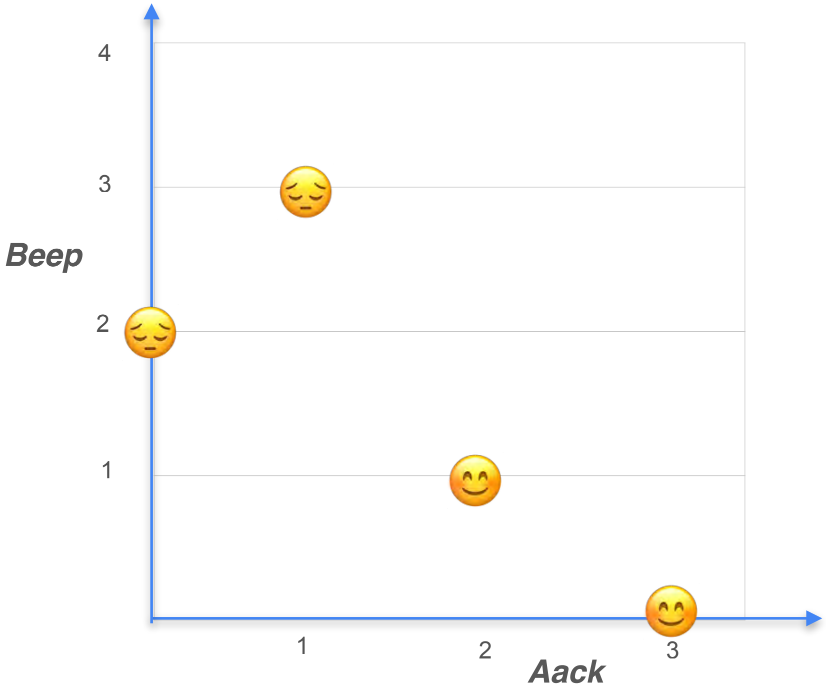
Classification Problem Motivation

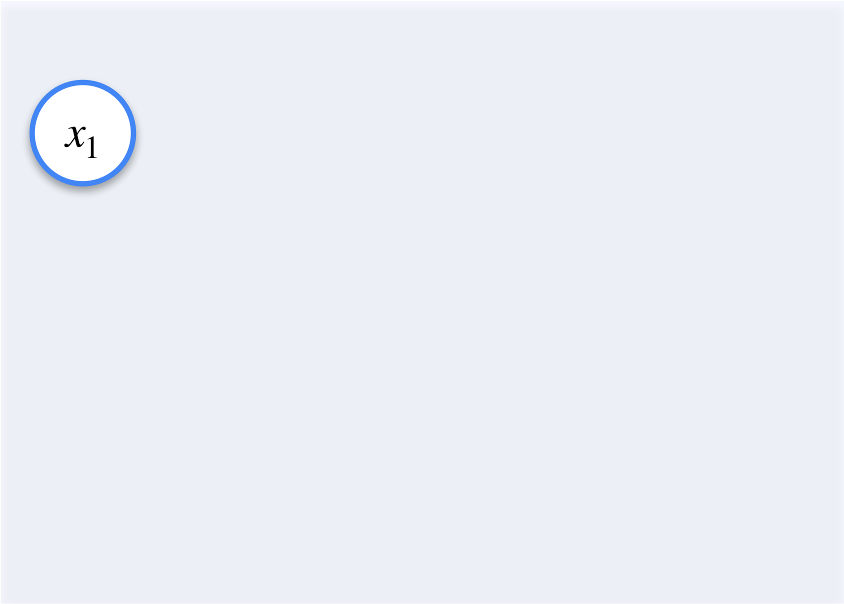
Sentence	*Aack*	*Beep*	*Mood*
Aack aack aack!	*3*	*0*	*Happy* 😊
Beep beep!	*0*	*2*	*Sad* 😔
Aack beep beep beep!	*1*	*3*	*Sad* 😔
Aack beep aack!	*2*	*1*	*Happy* 😊

Classification Problem Motivation

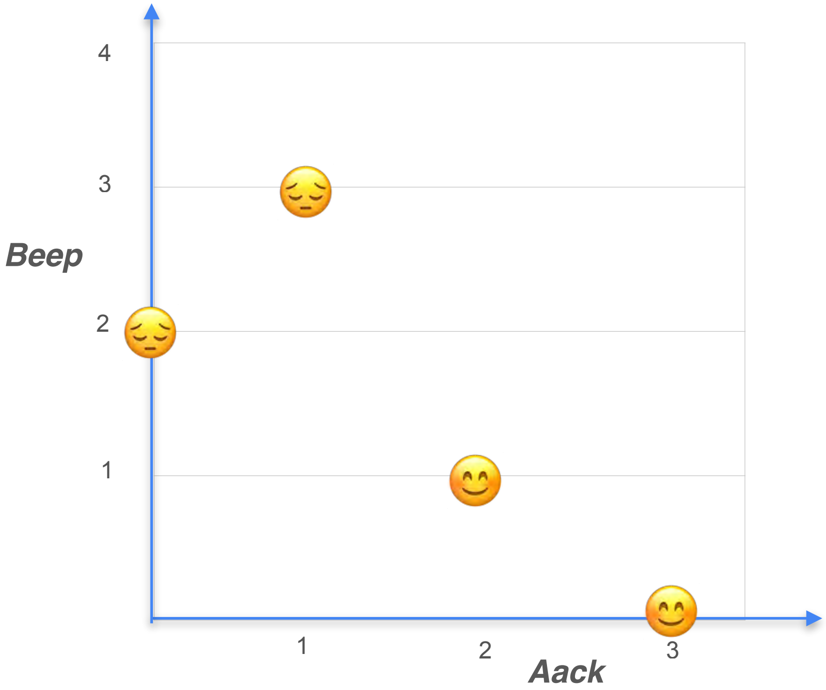


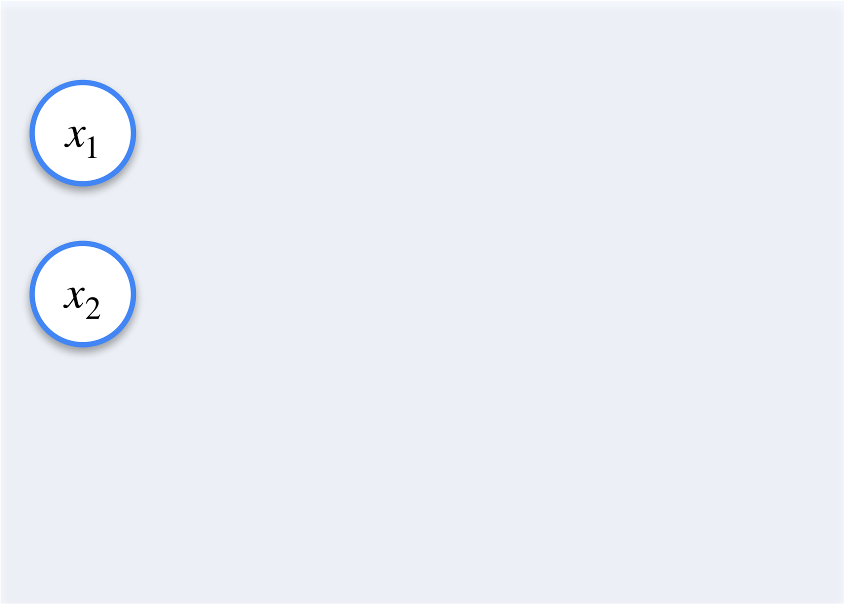
Classification Problem Motivation



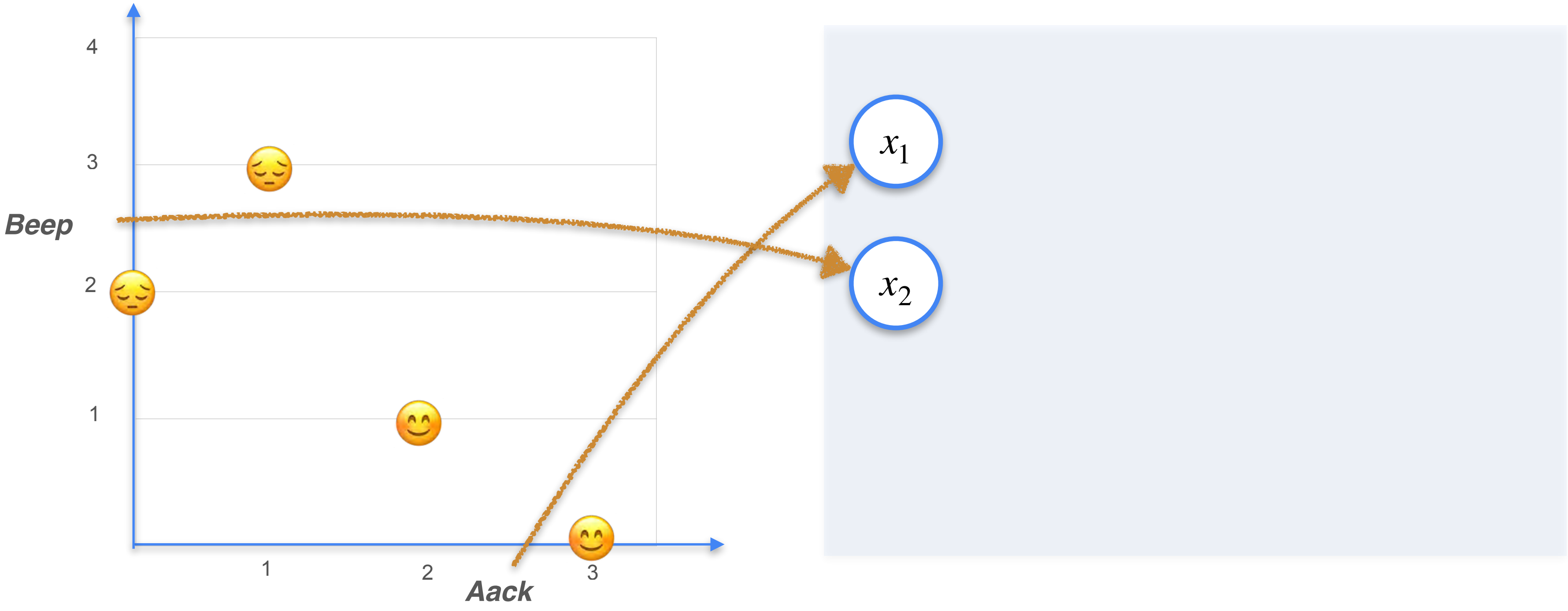


Classification Problem Motivation

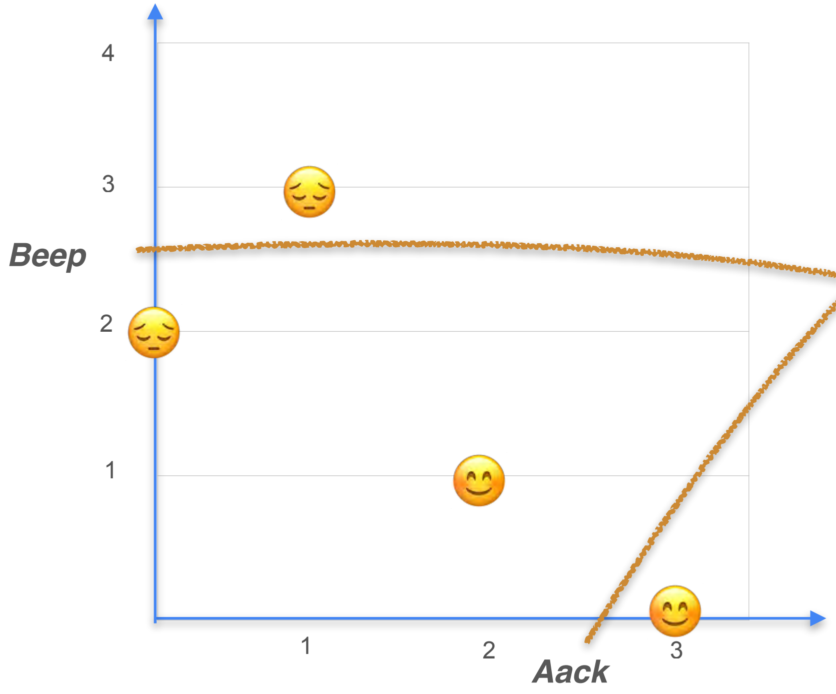


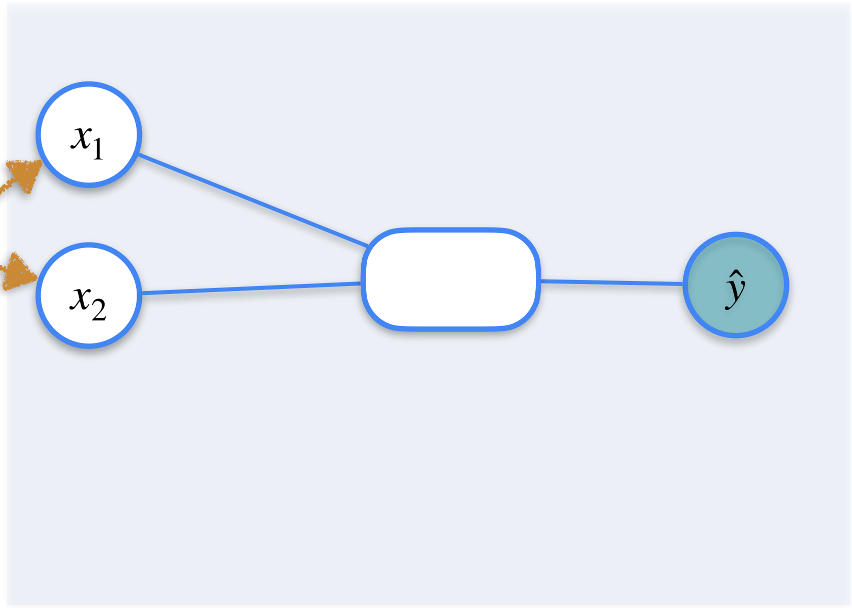


Classification Problem Motivation

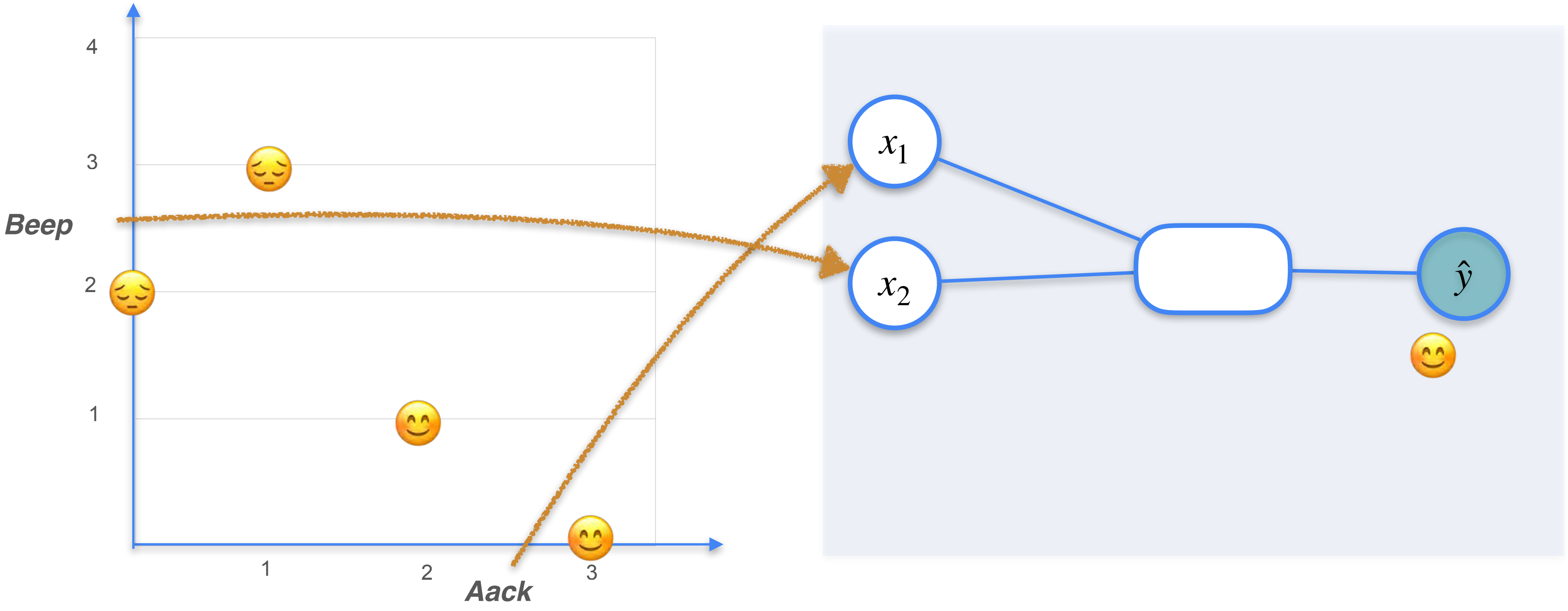


Classification Problem Motivation

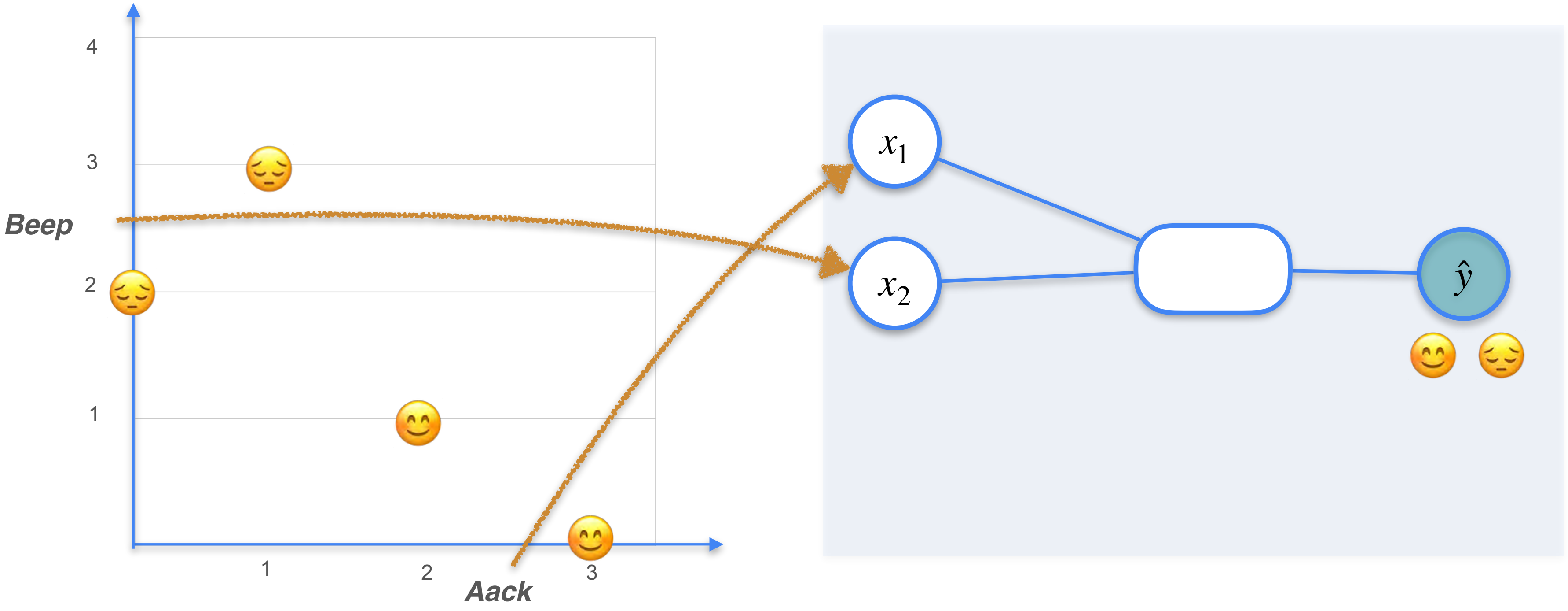




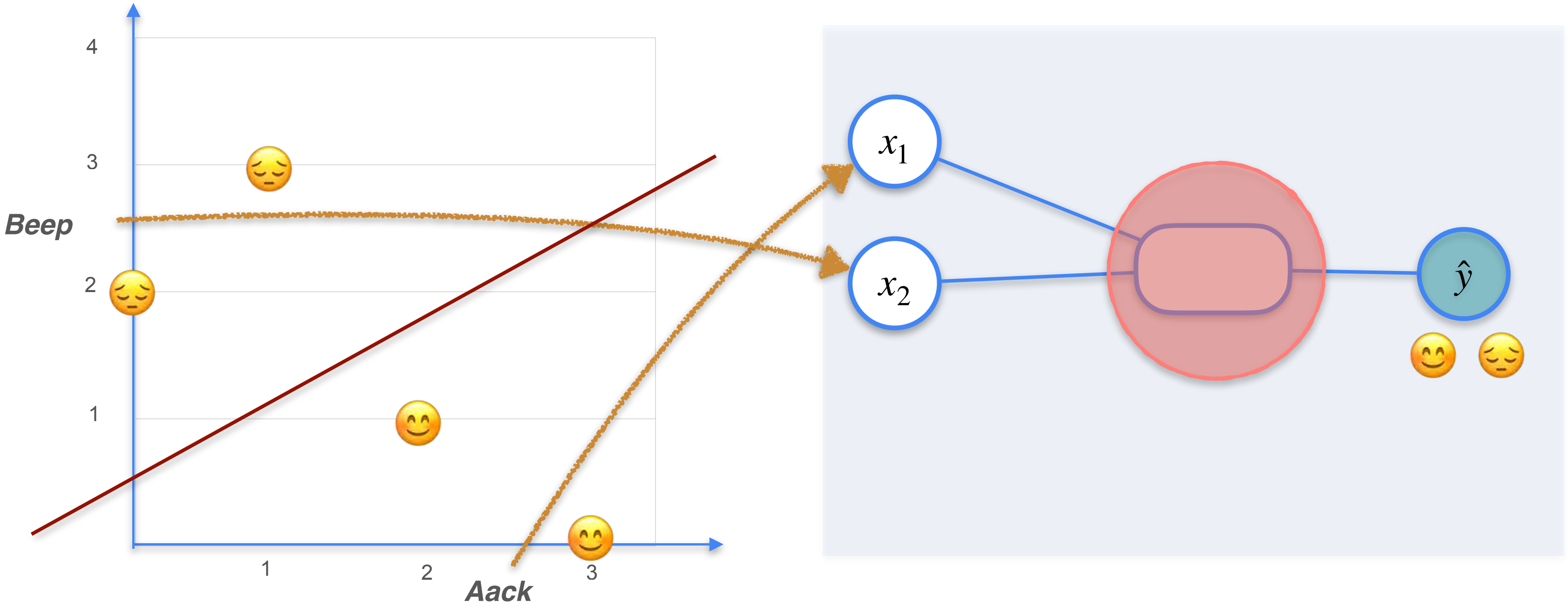
Classification Problem Motivation



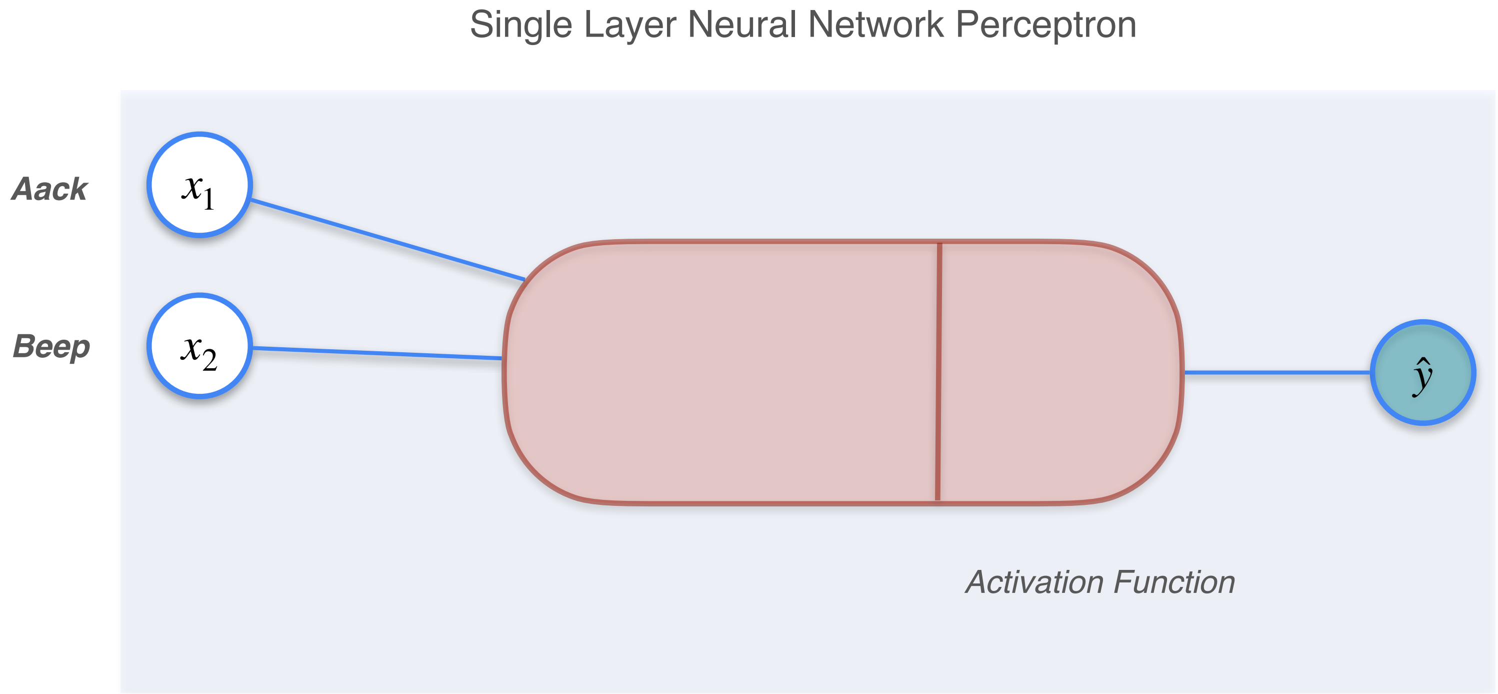
Classification Problem Motivation



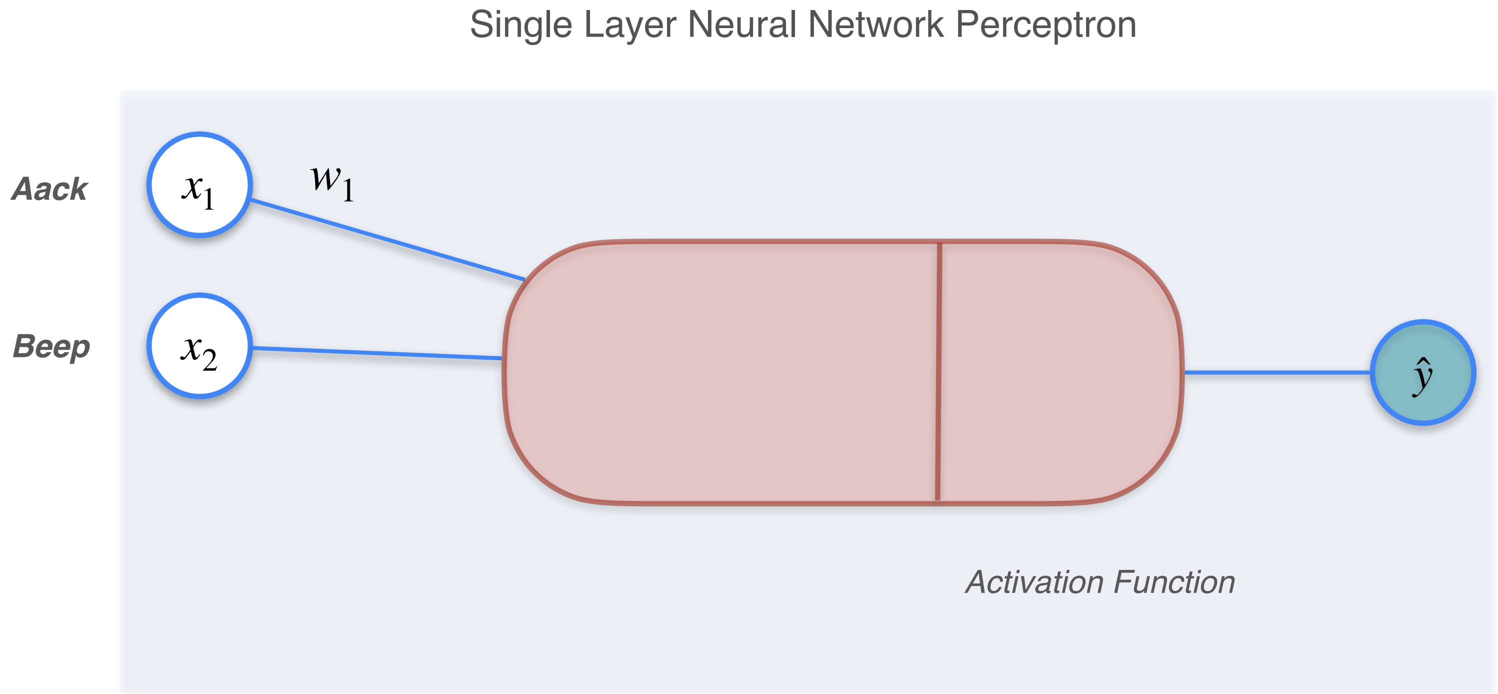
Classification Problem Motivation



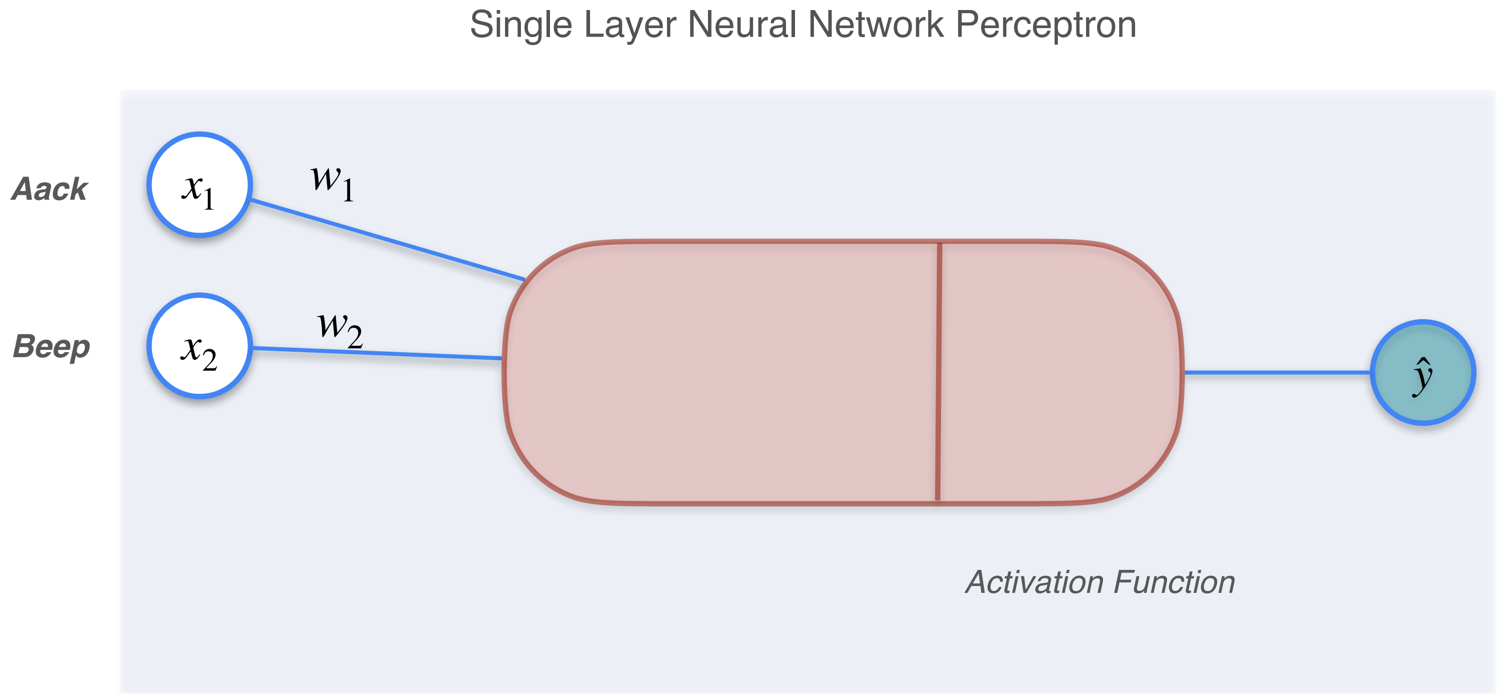
Classification With a Perceptron



Classification With a Perceptron

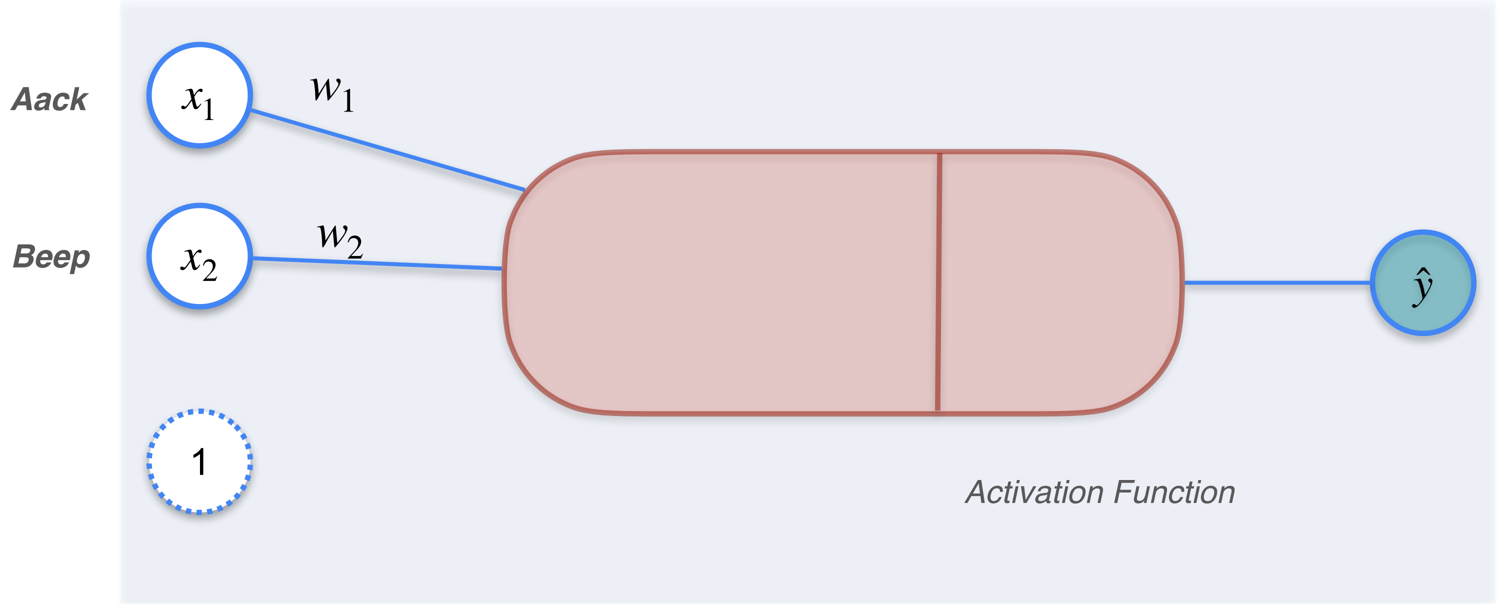


Classification With a Perceptron

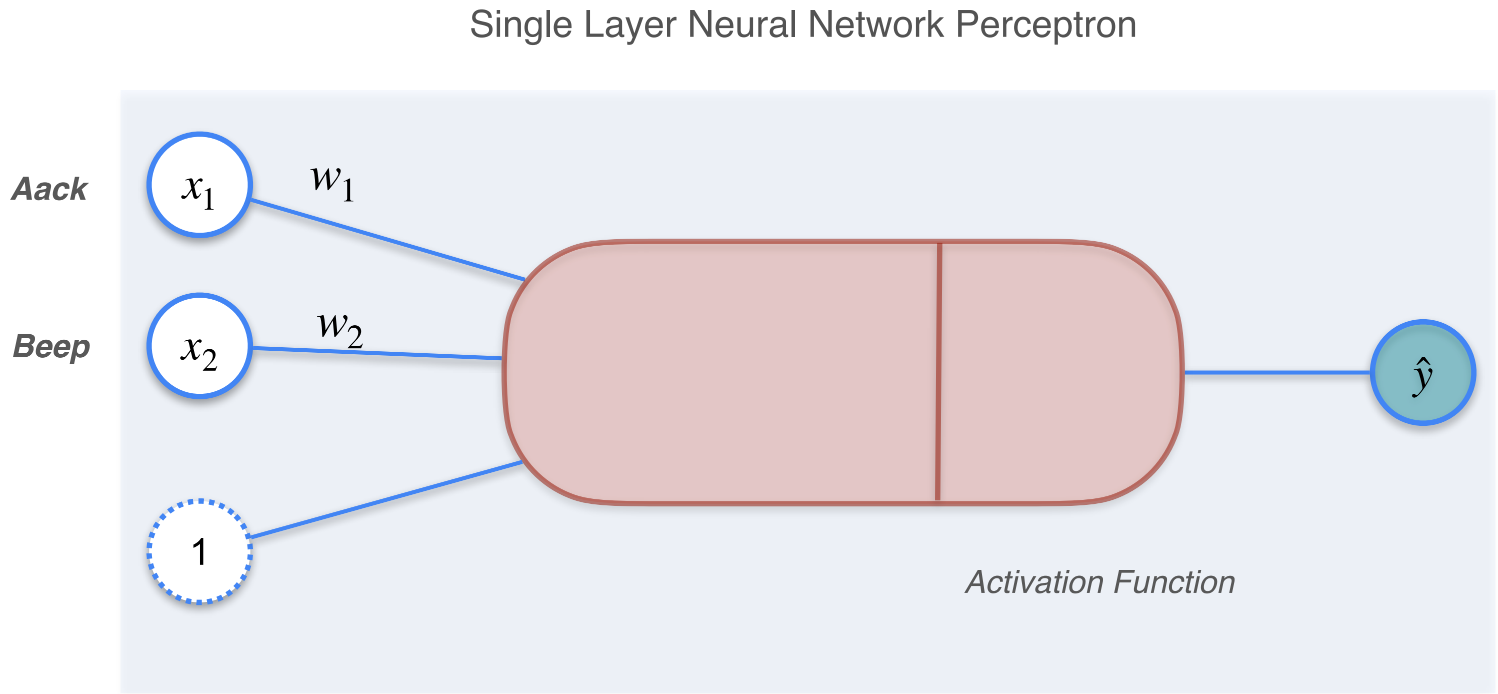


Classification With a Perceptron

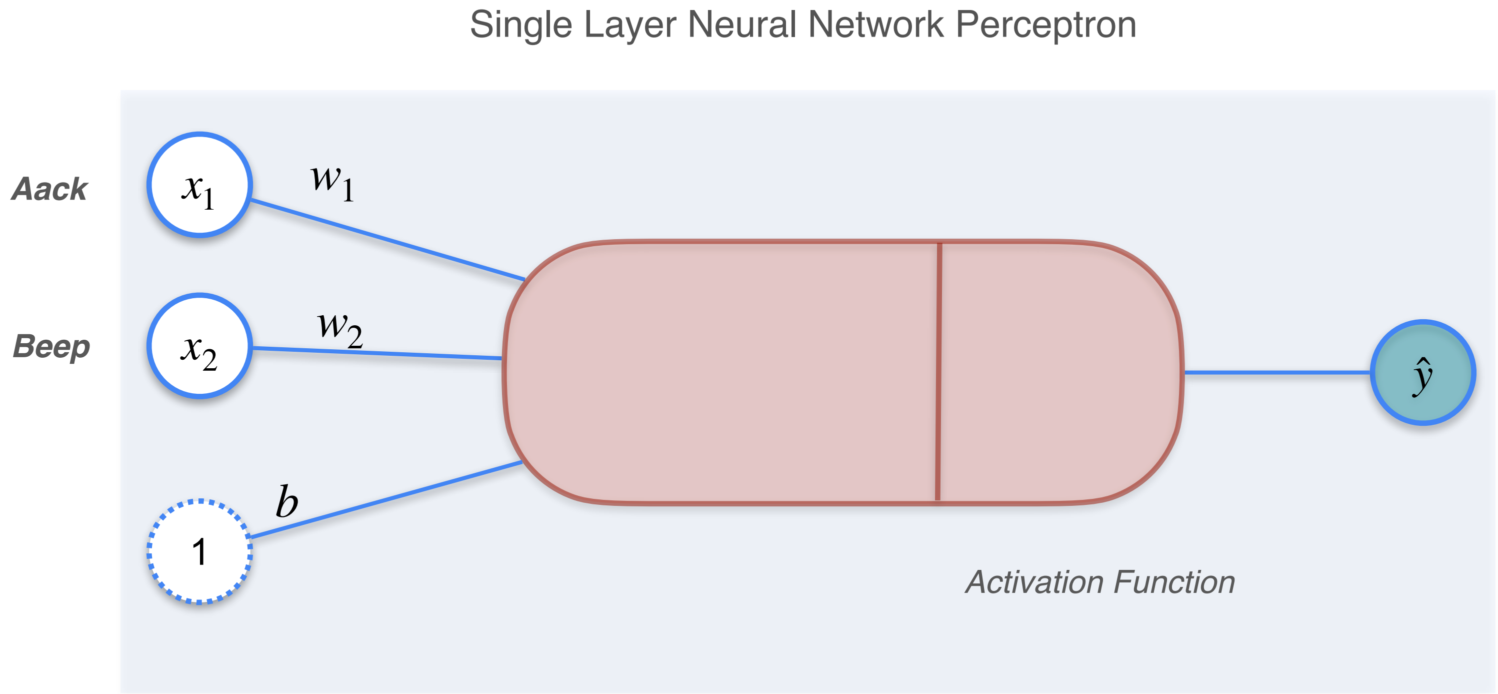
Single Layer Neural Network Perceptron



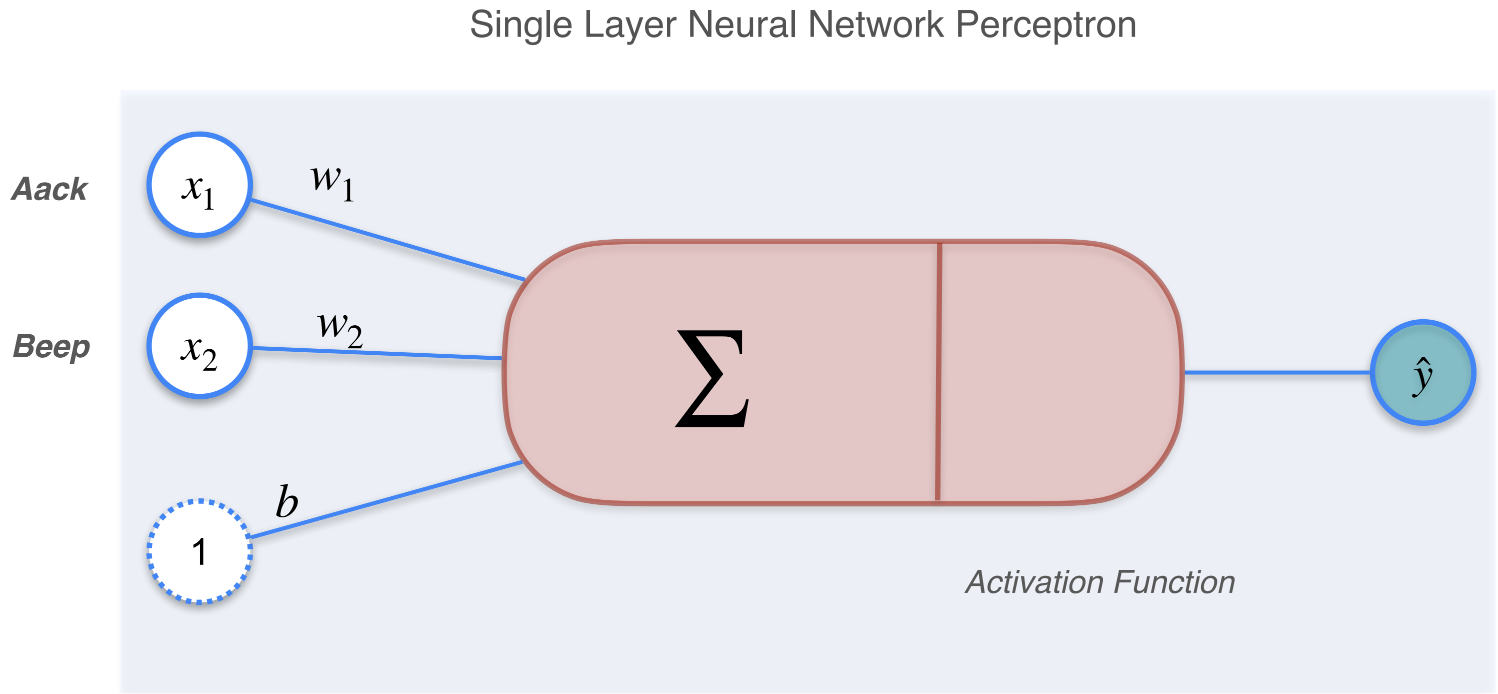
Classification With a Perceptron



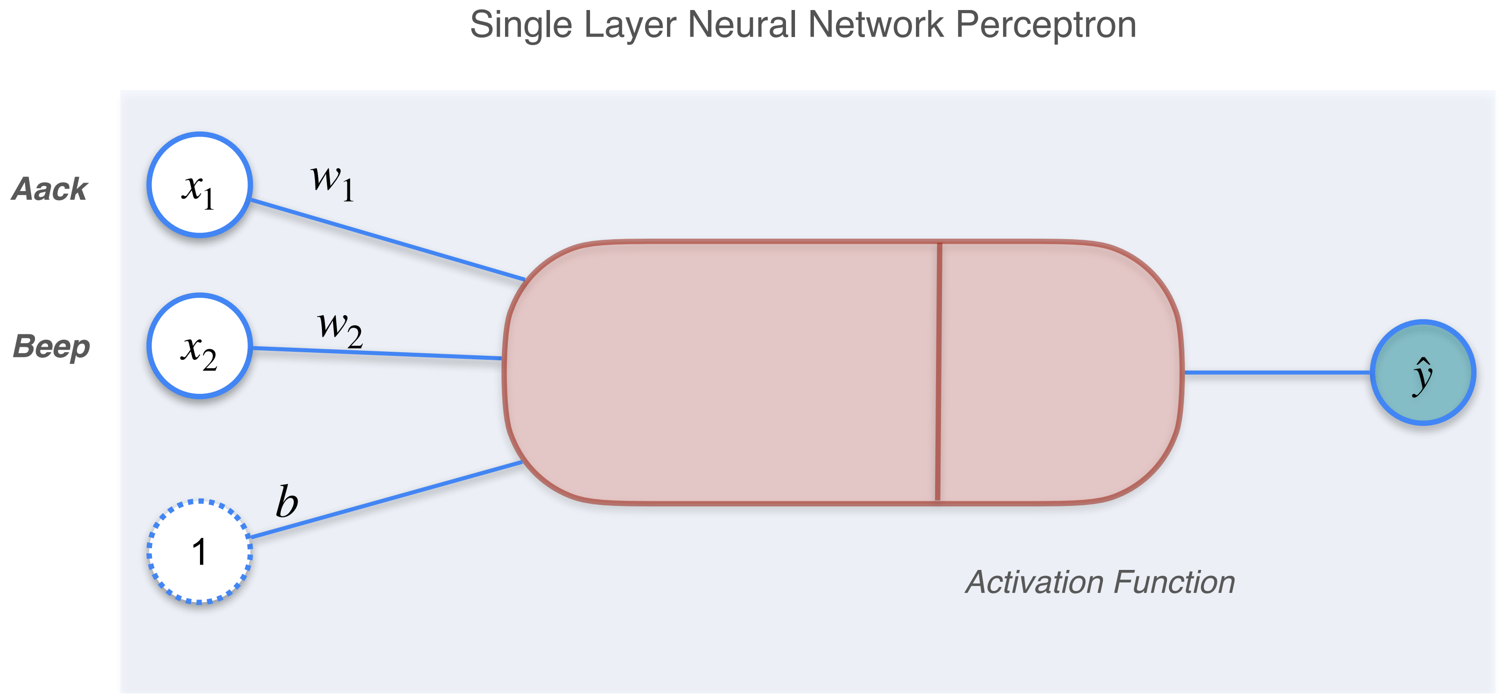
Classification With a Perceptron



Classification With a Perceptron

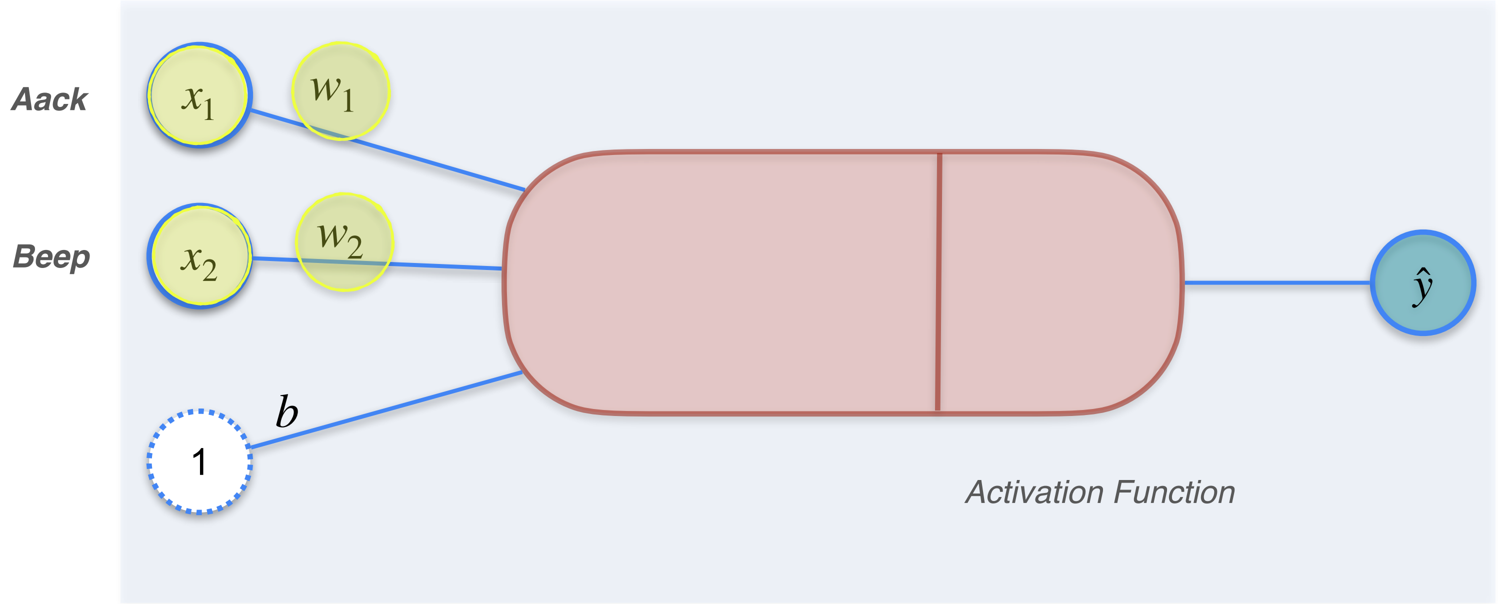


Classification With a Perceptron

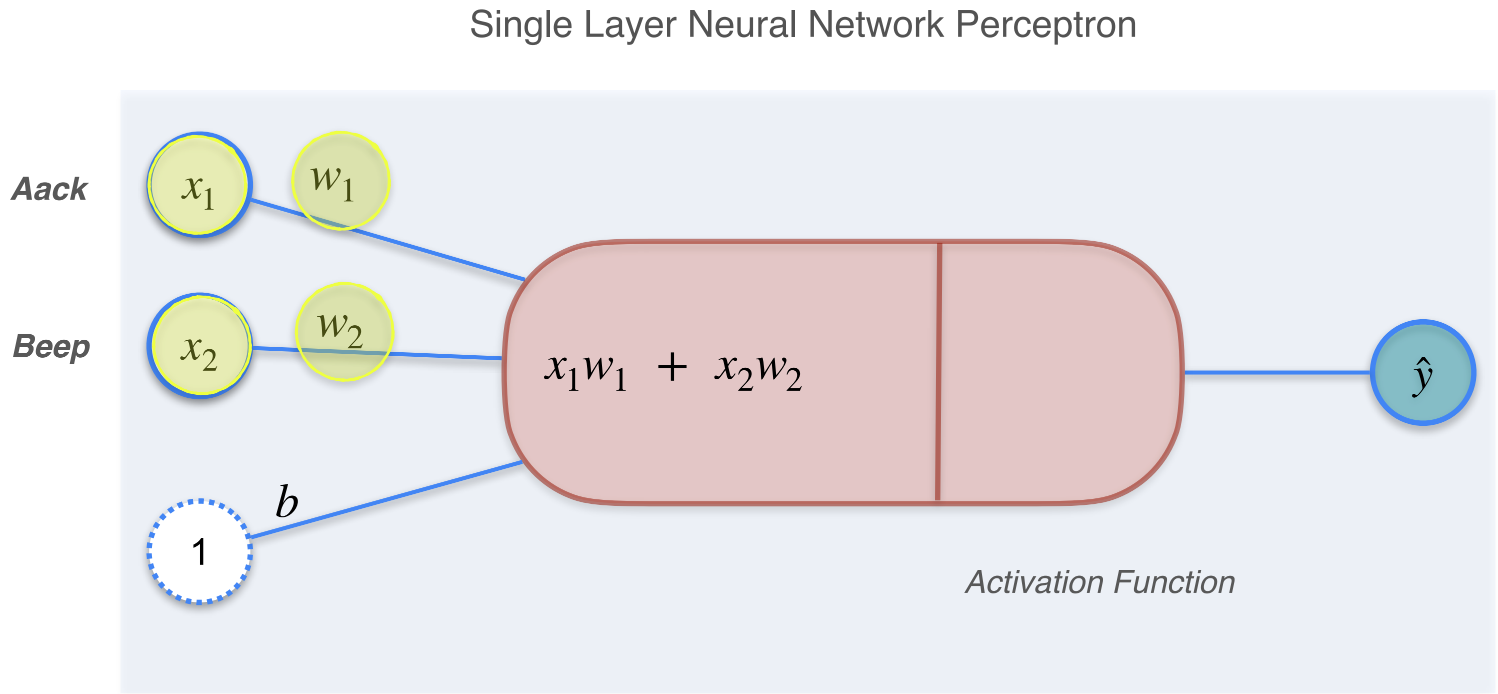


Classification With a Perceptron

Single Layer Neural Network Perceptron

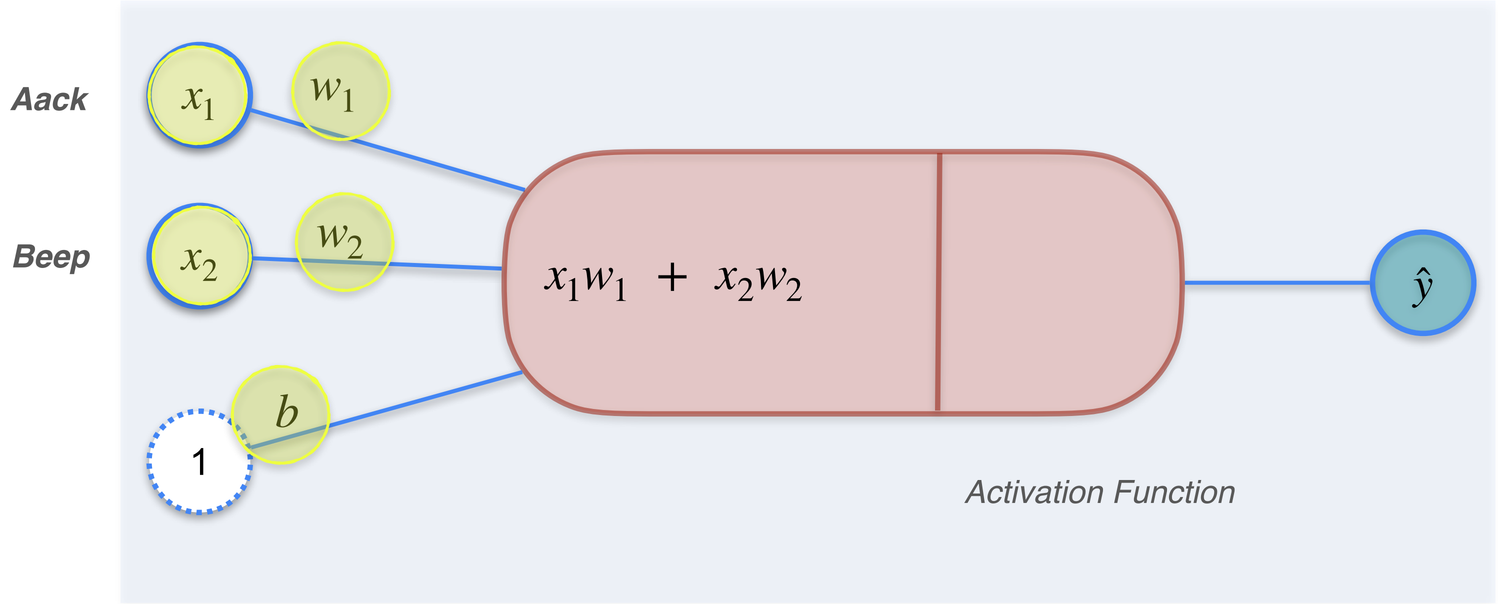


Classification With a Perceptron

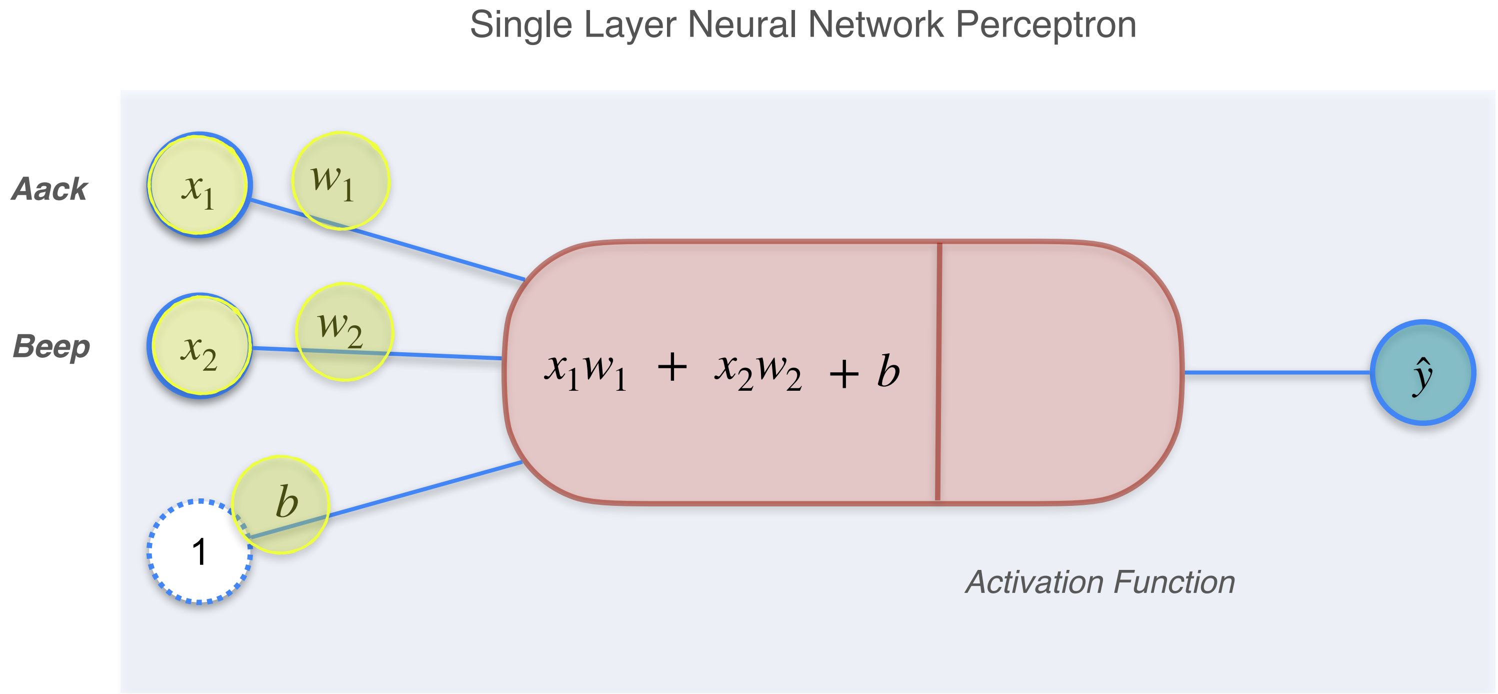


Classification With a Perceptron

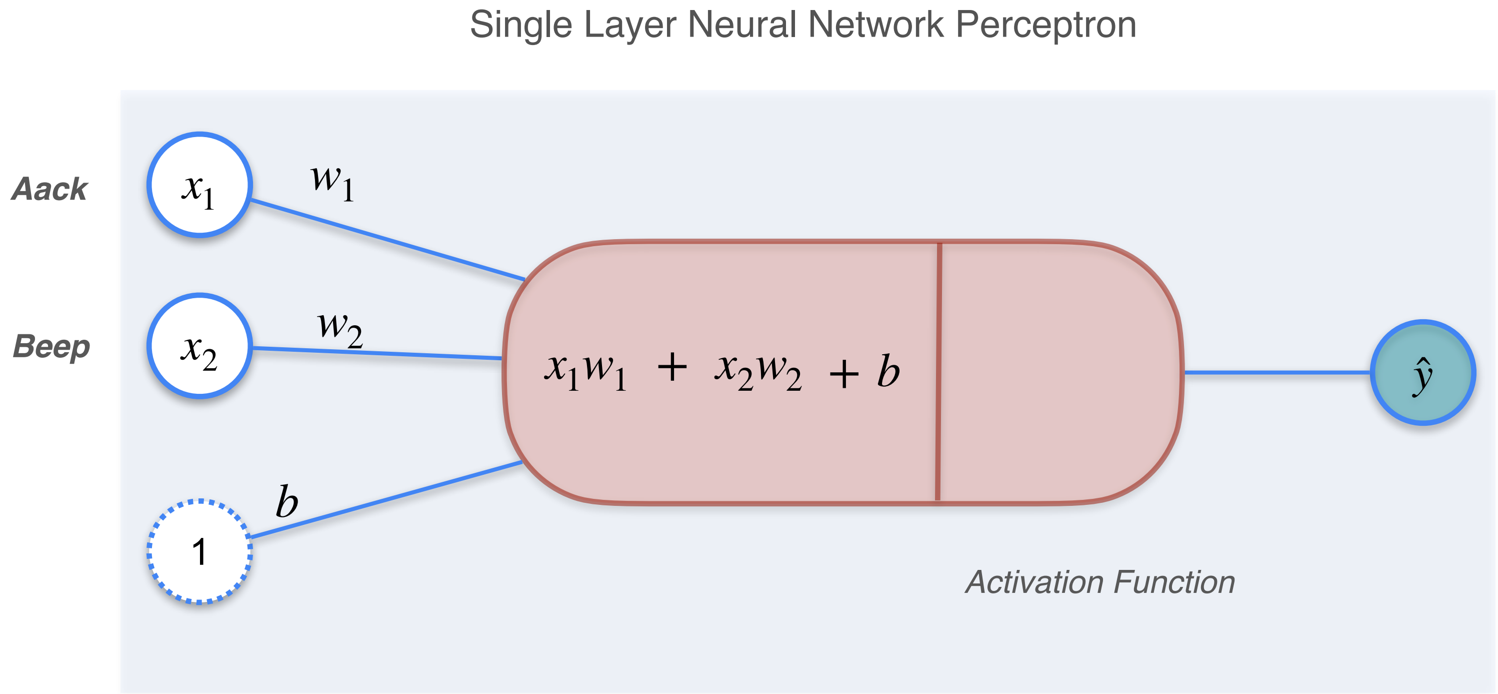
Single Layer Neural Network Perceptron



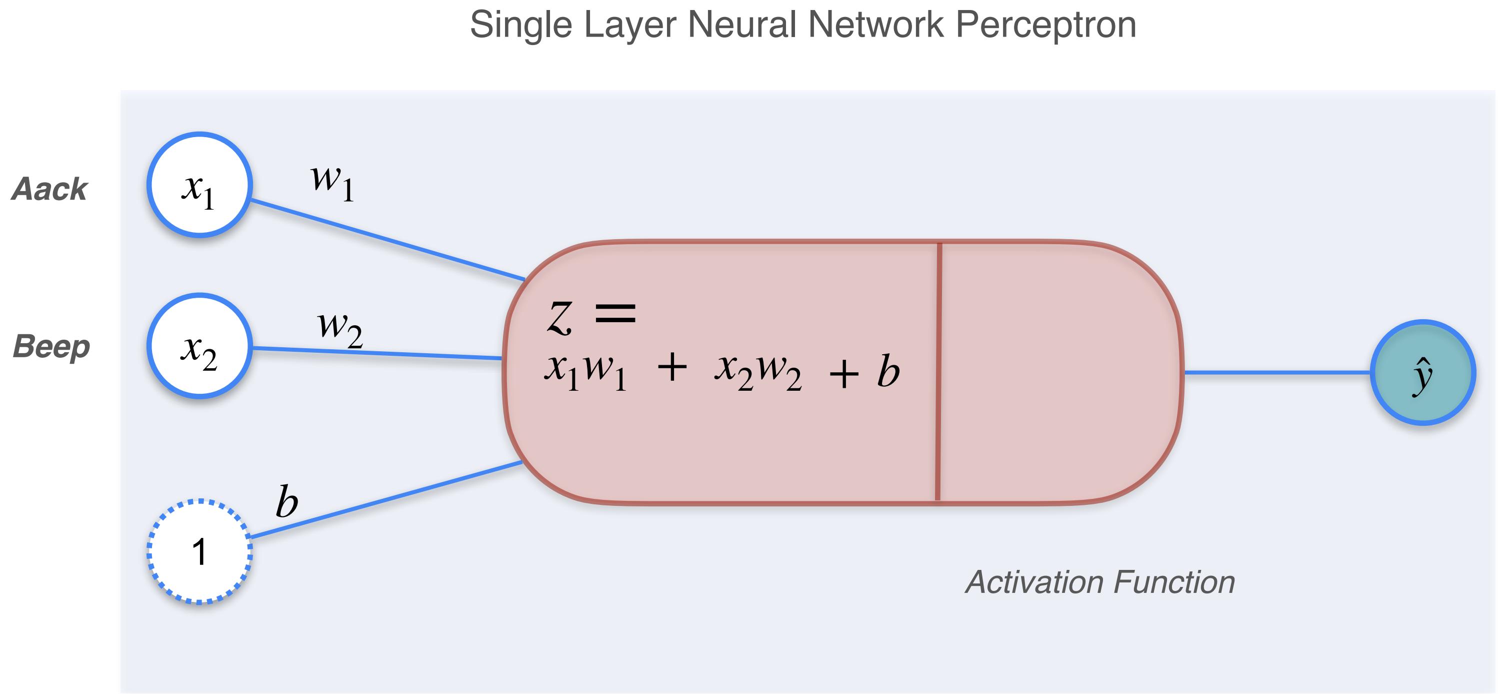
Classification With a Perceptron



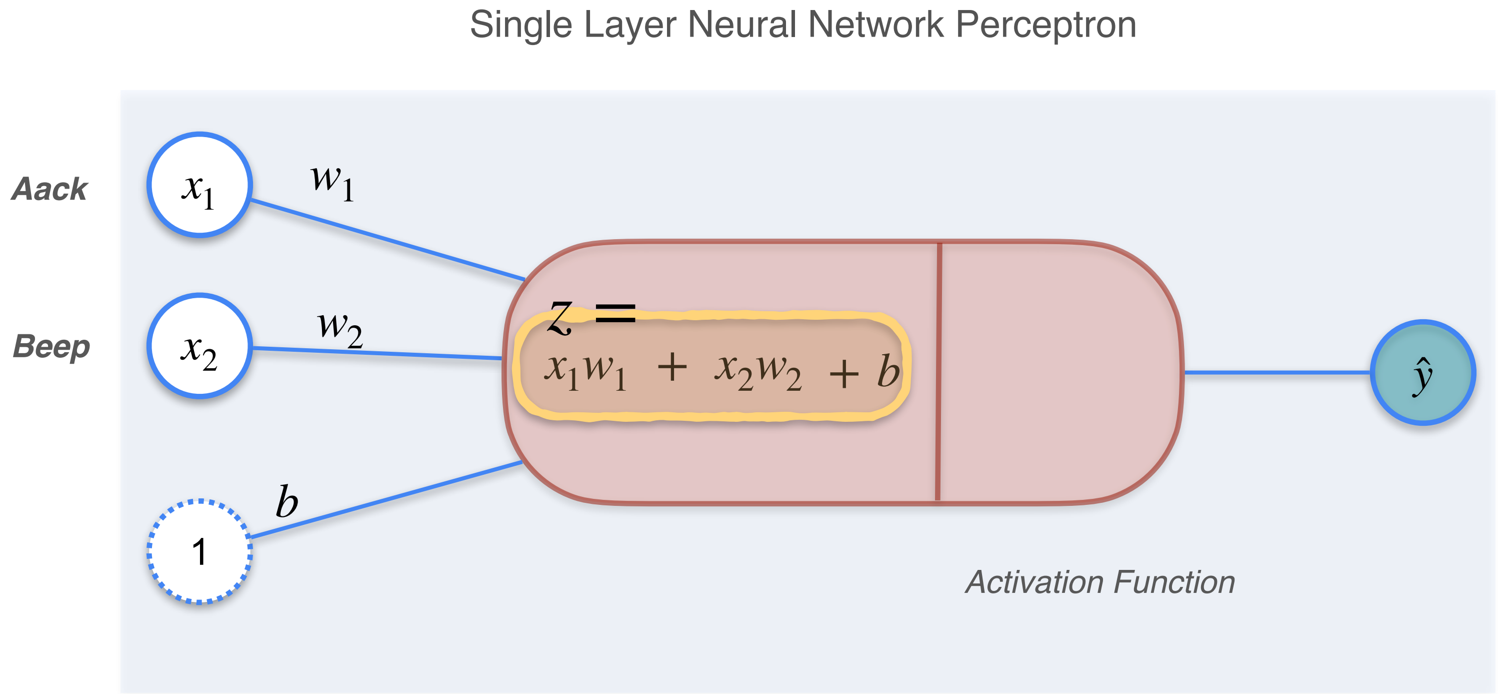
Classification With a Perceptron



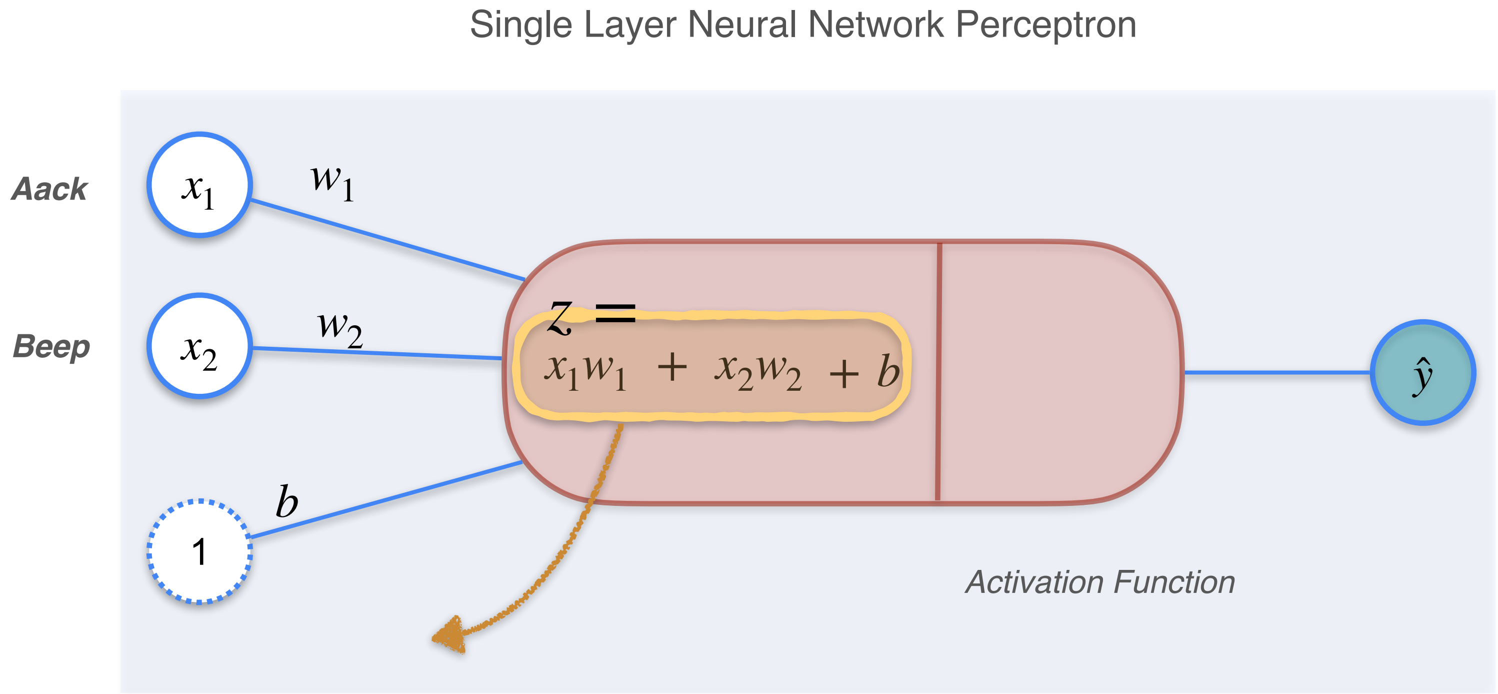
Classification With a Perceptron



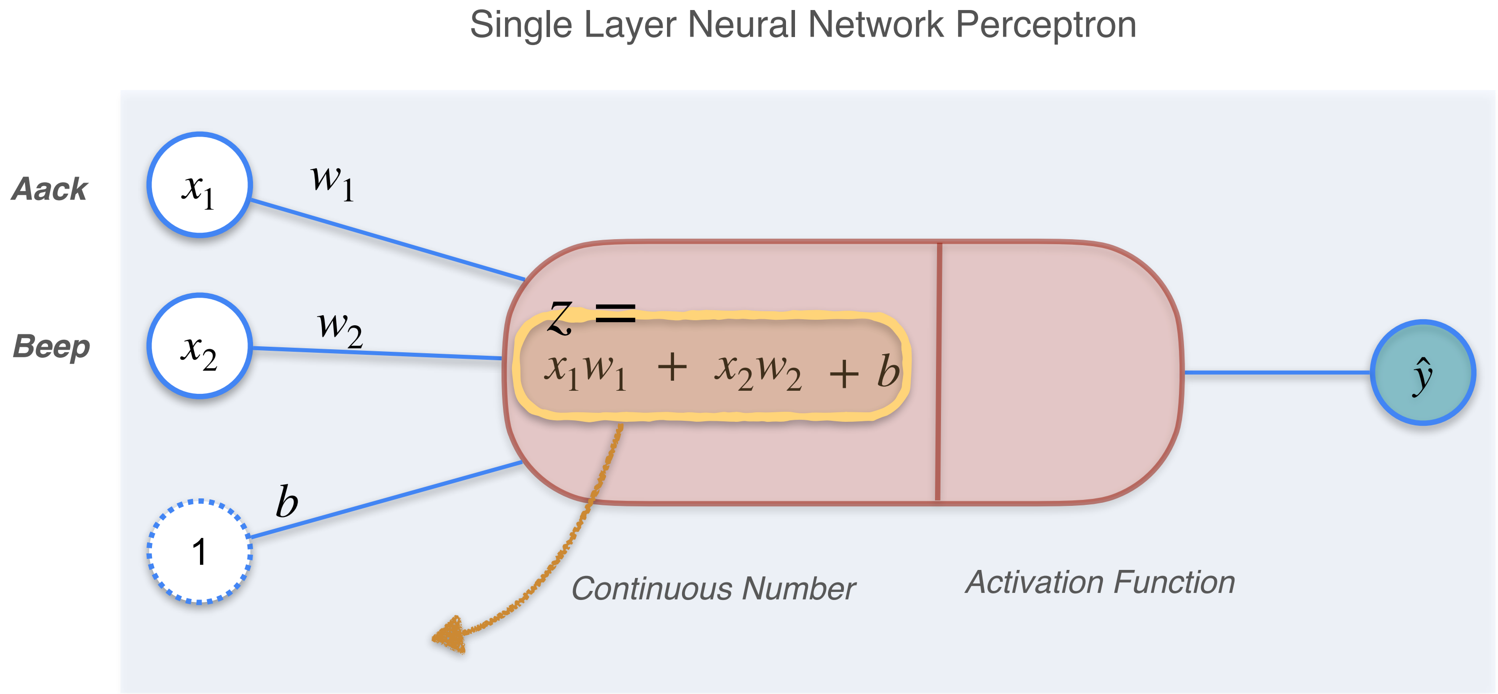
Classification With a Perceptron



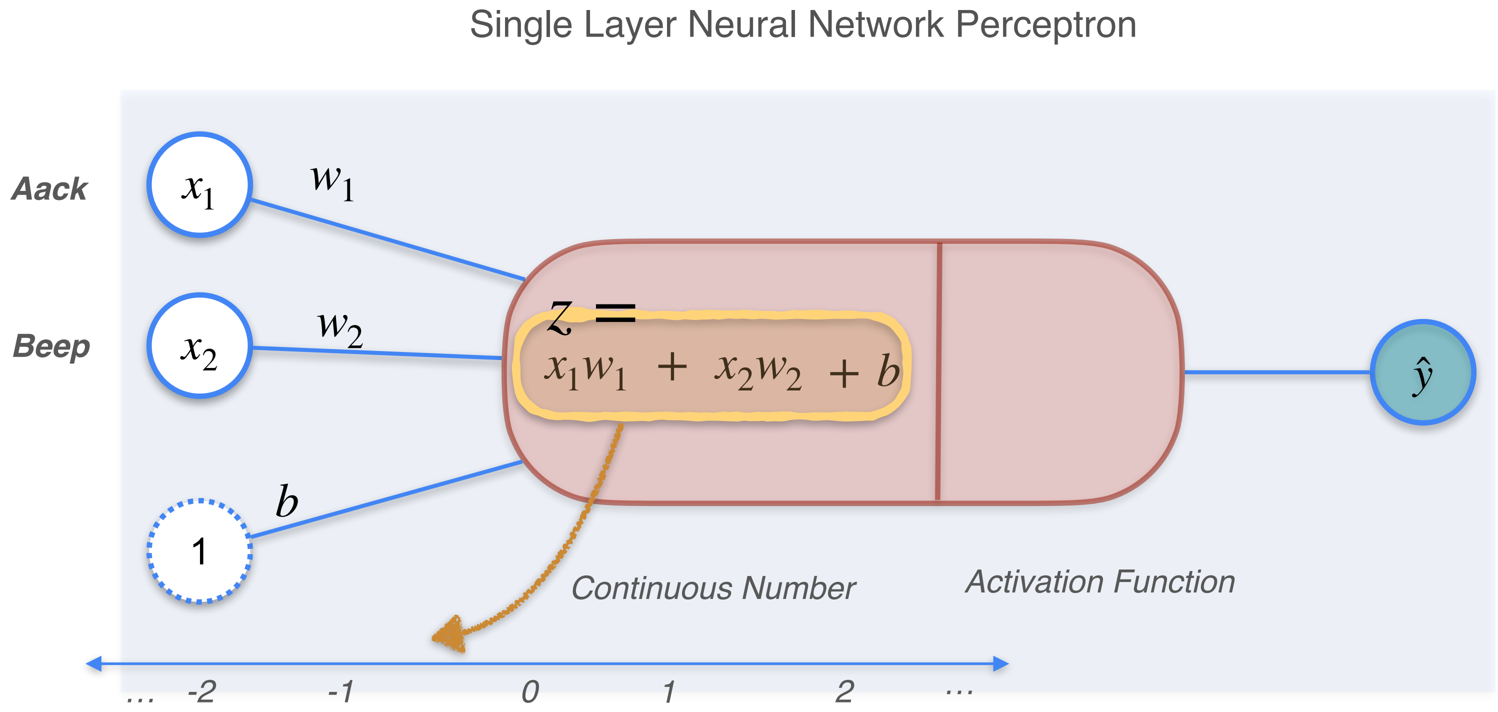
Classification With a Perceptron



Classification With a Perceptron

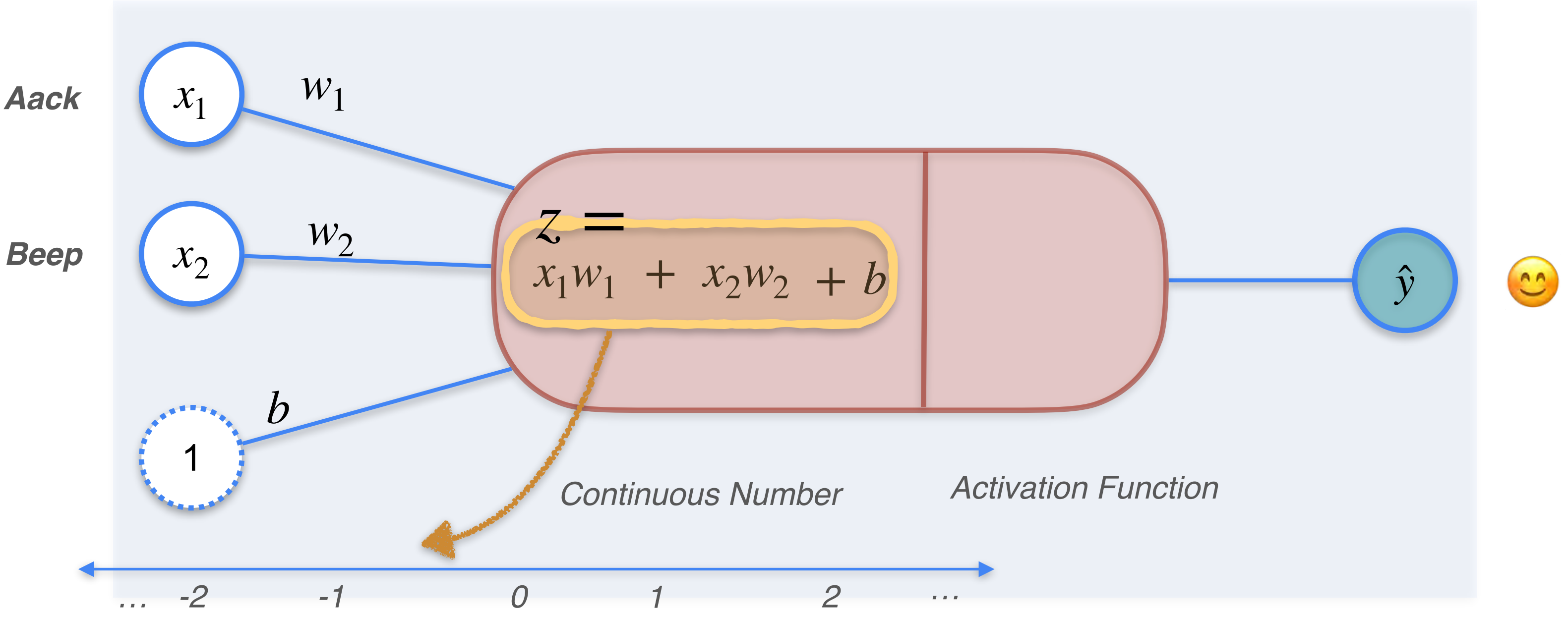


Classification With a Perceptron



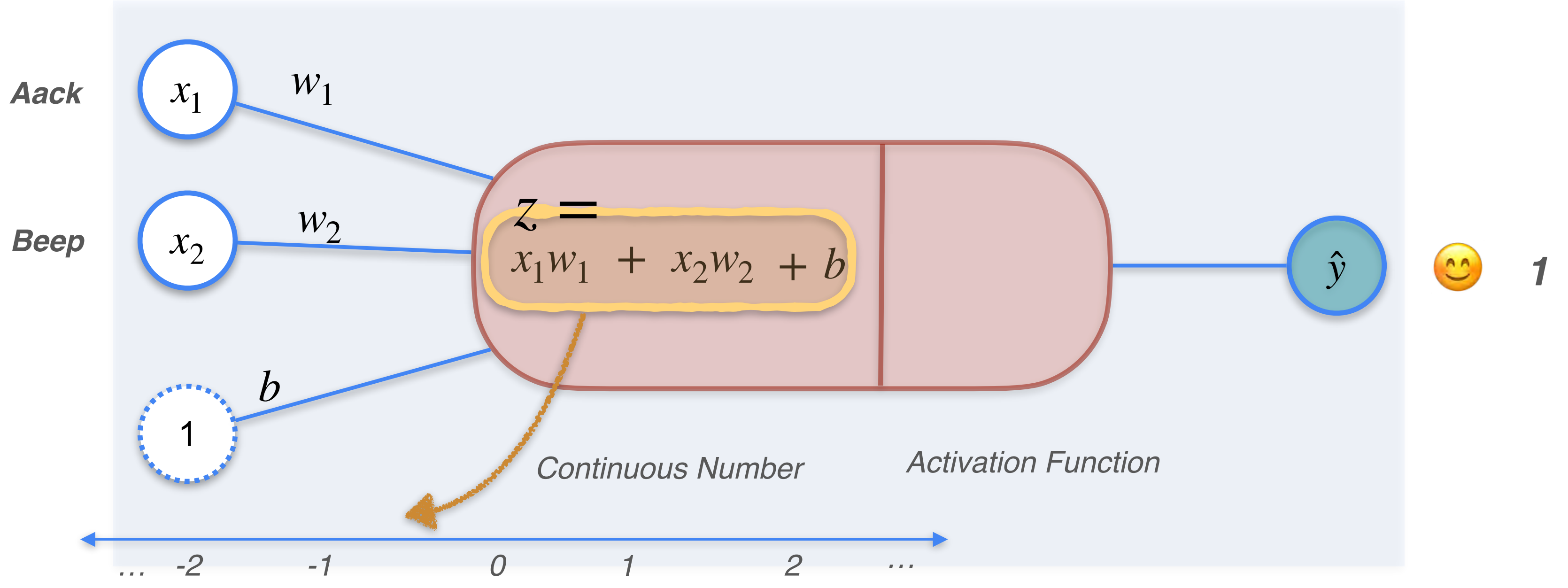
Classification With a Perceptron

Single Layer Neural Network Perceptron



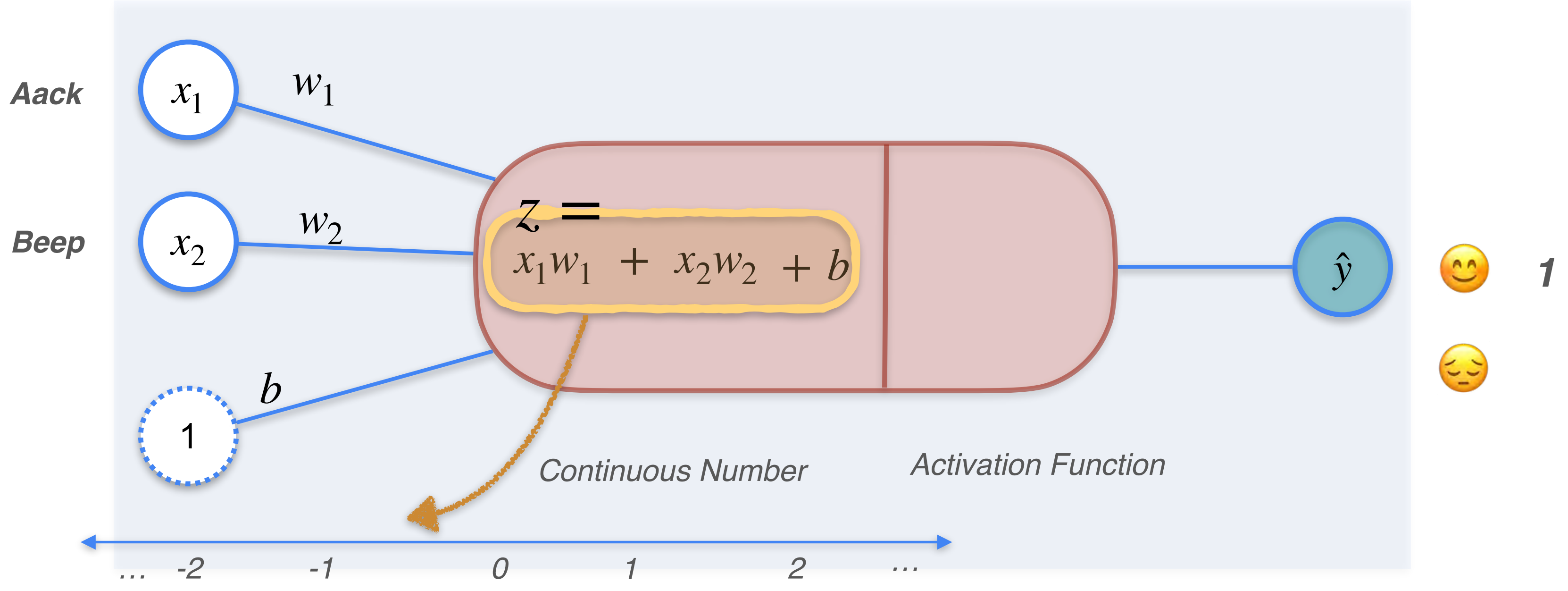
Classification With a Perceptron

Single Layer Neural Network Perceptron



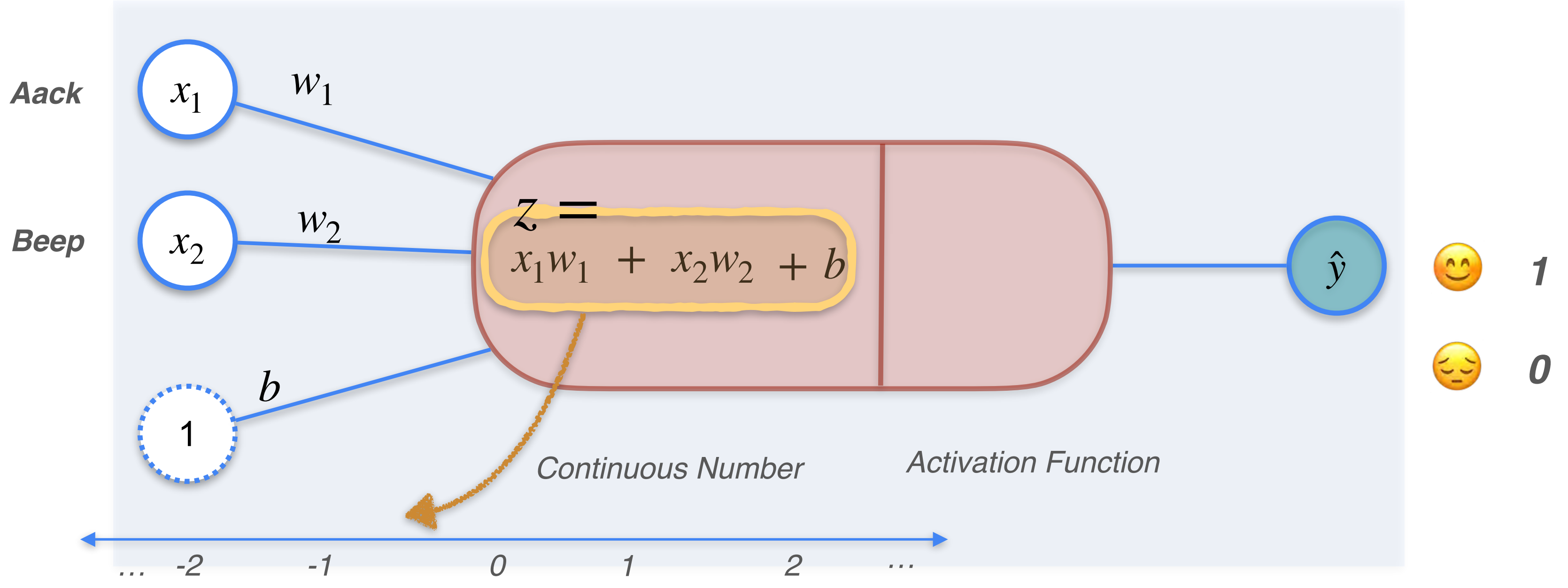
Classification With a Perceptron

Single Layer Neural Network Perceptron

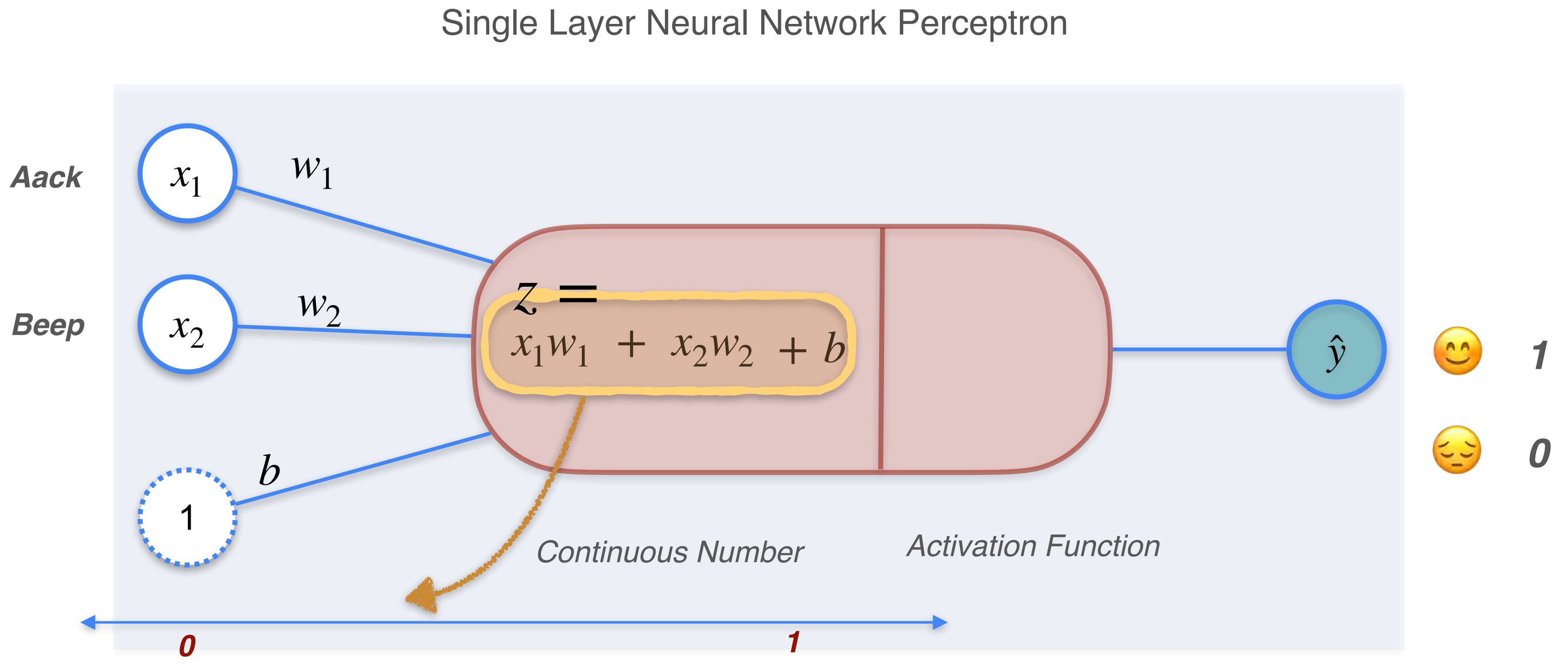


Classification With a Perceptron

Single Layer Neural Network Perceptron

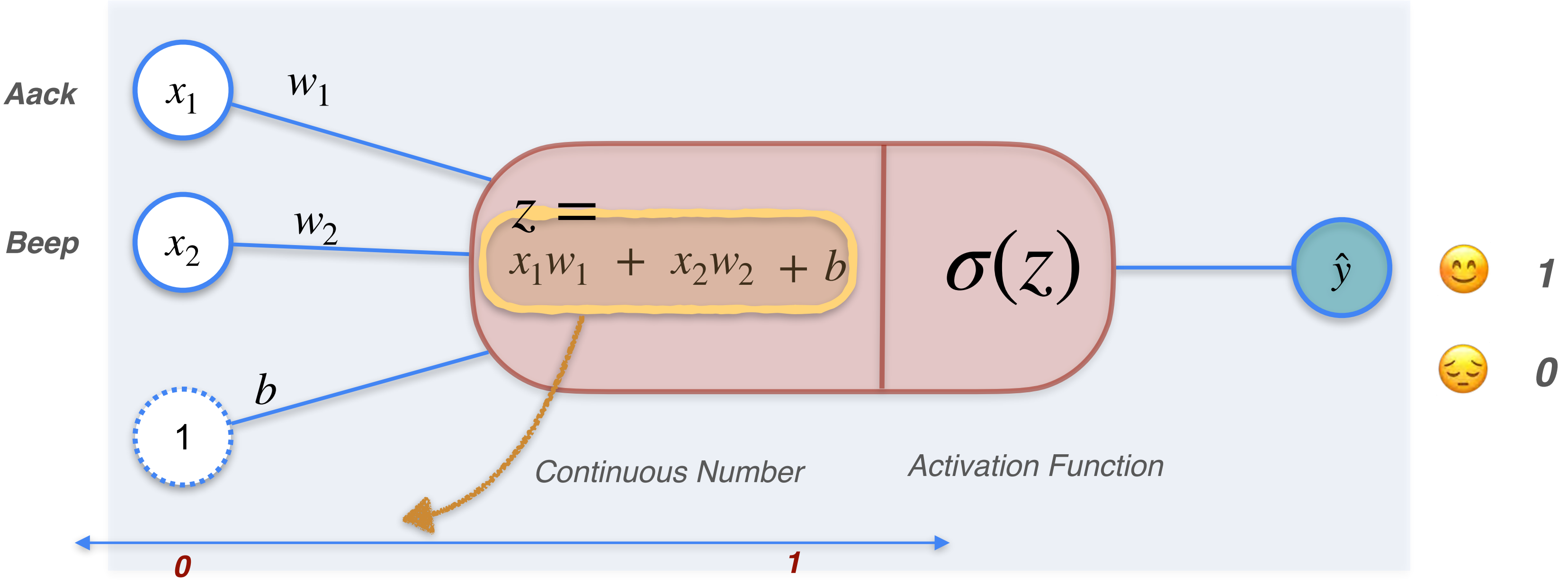


Classification With a Perceptron



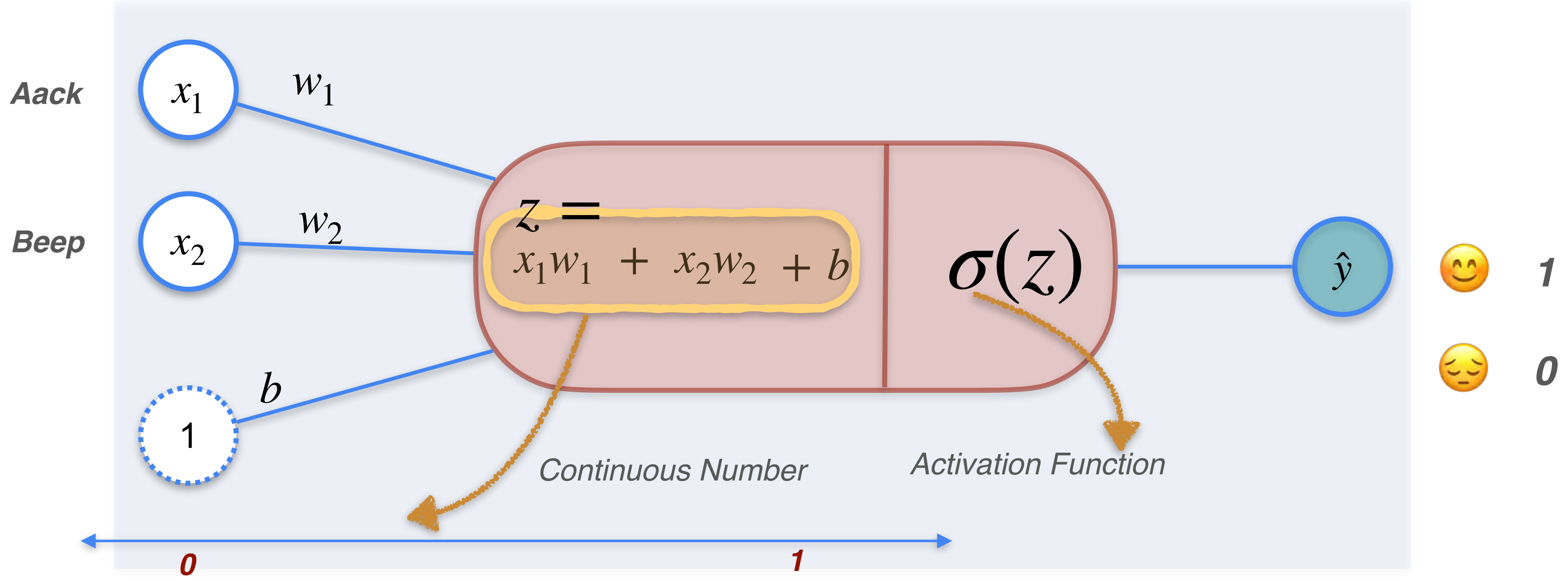
Classification With a Perceptron

Single Layer Neural Network Perceptron



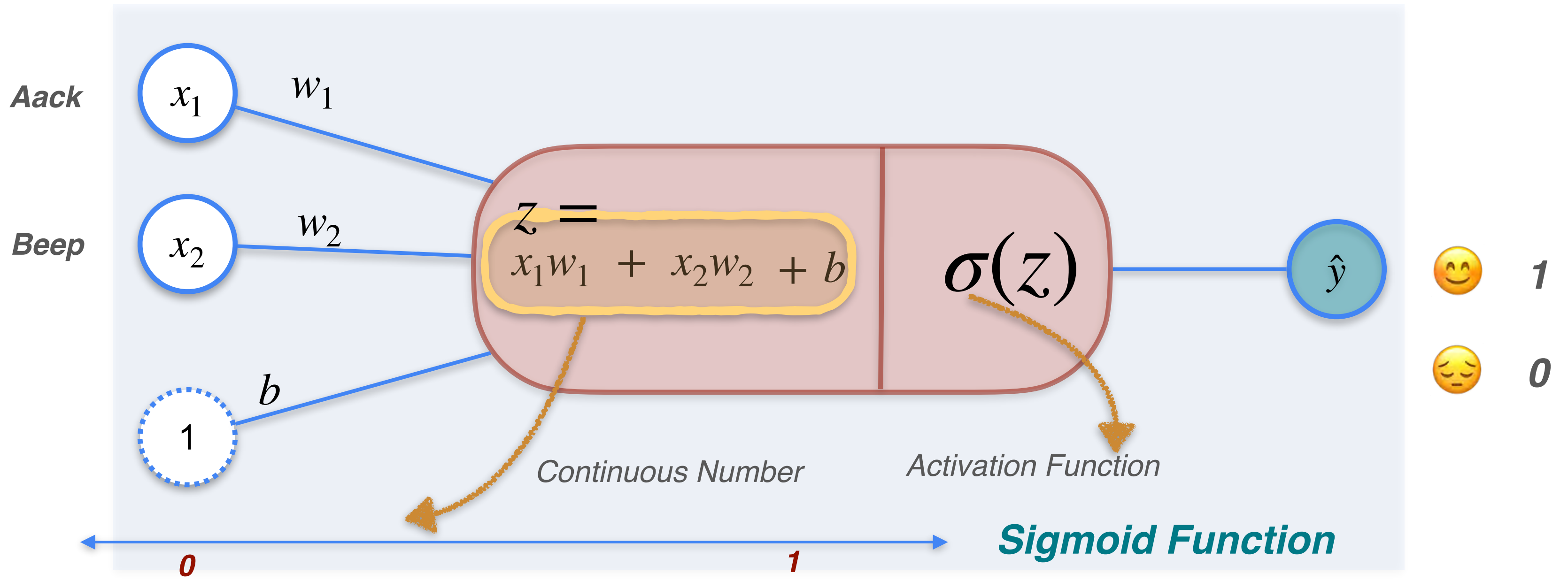
Classification With a Perceptron

Single Layer Neural Network Perceptron

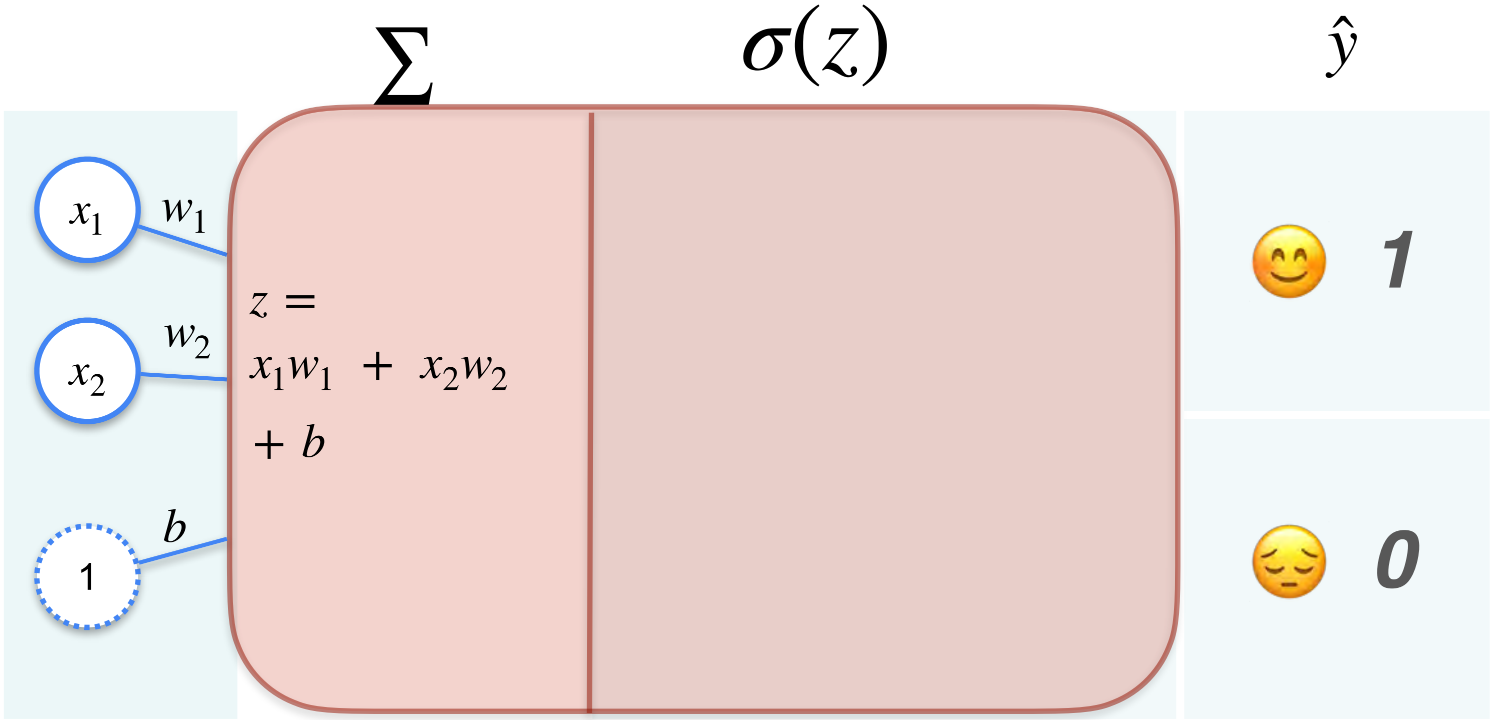


Classification With a Perceptron

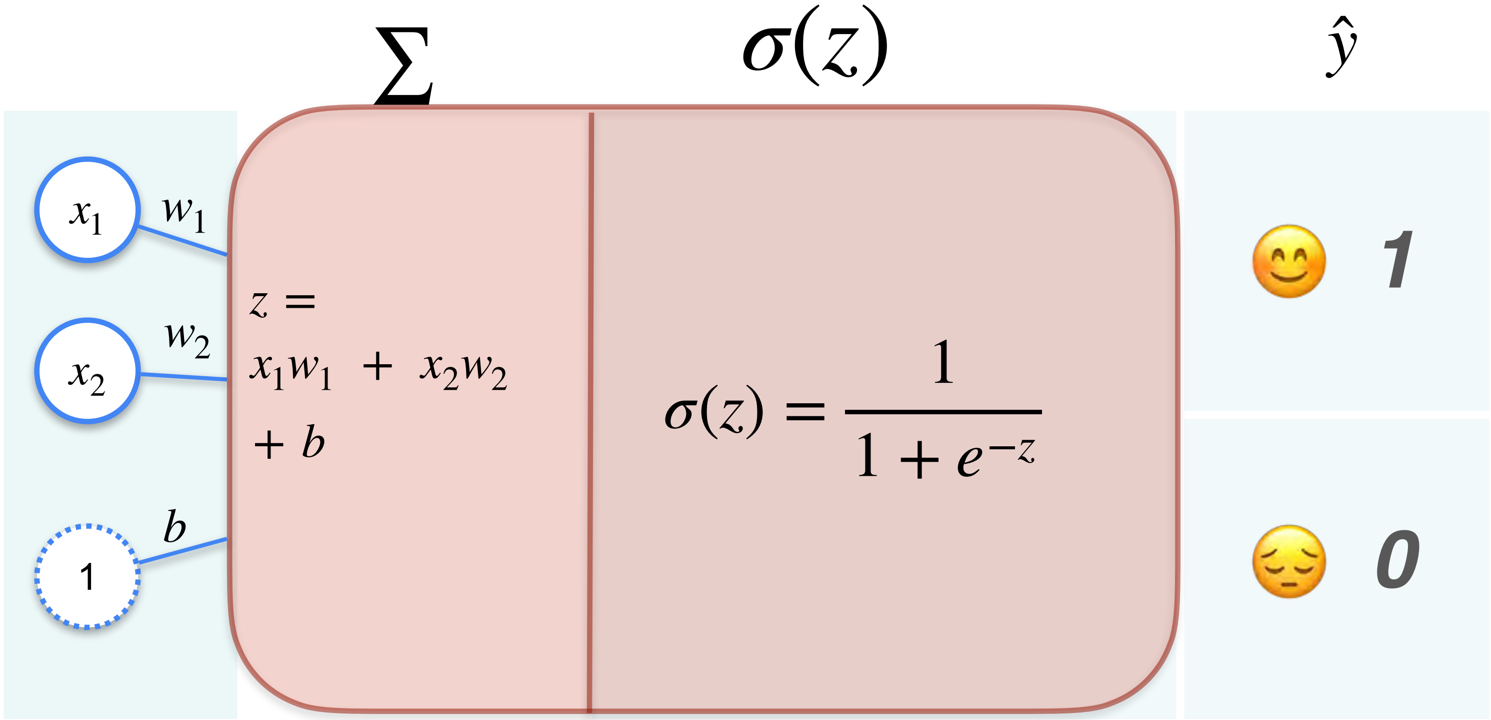
Single Layer Neural Network Perceptron



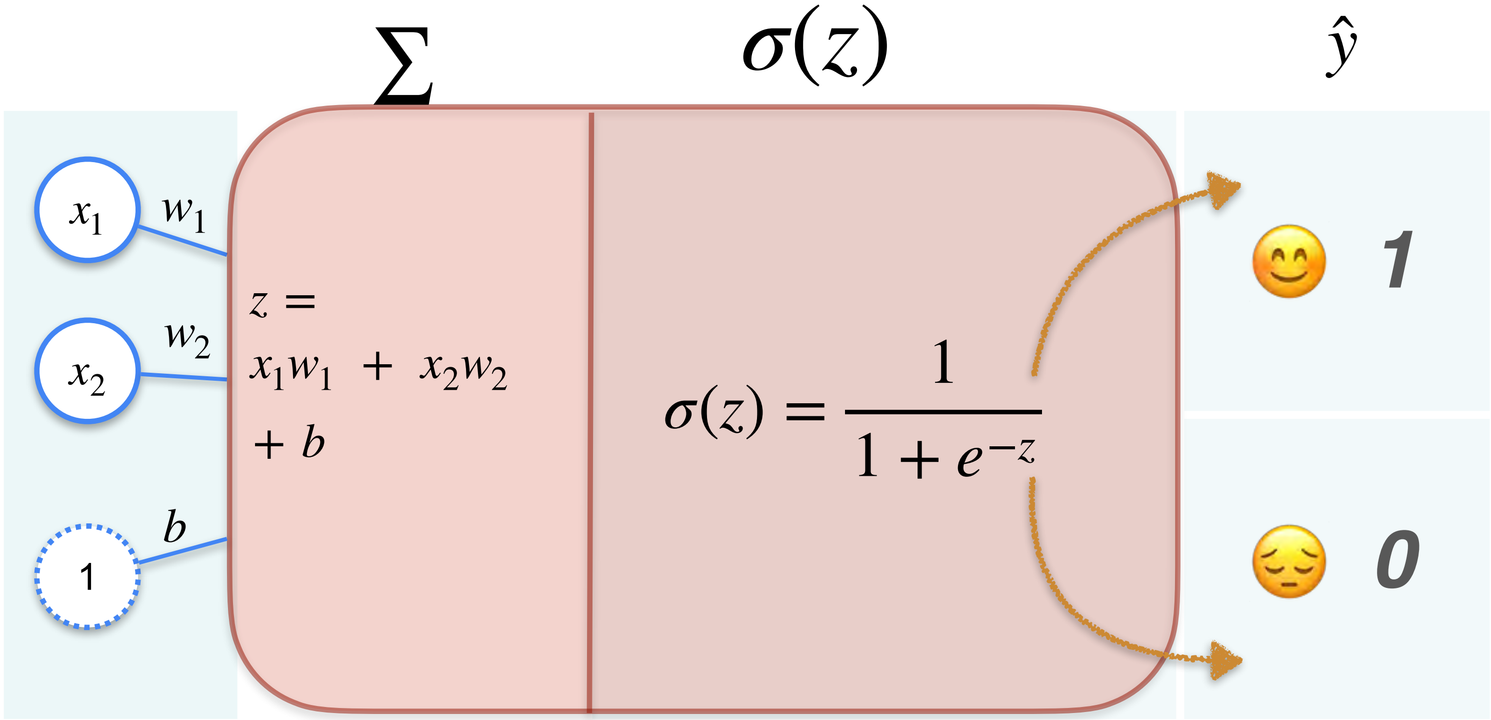
Sigmoid Function



Sigmoid Function



Sigmoid Function



Optimization in Neural Networks and Newton's Method

DeepLearning.AI

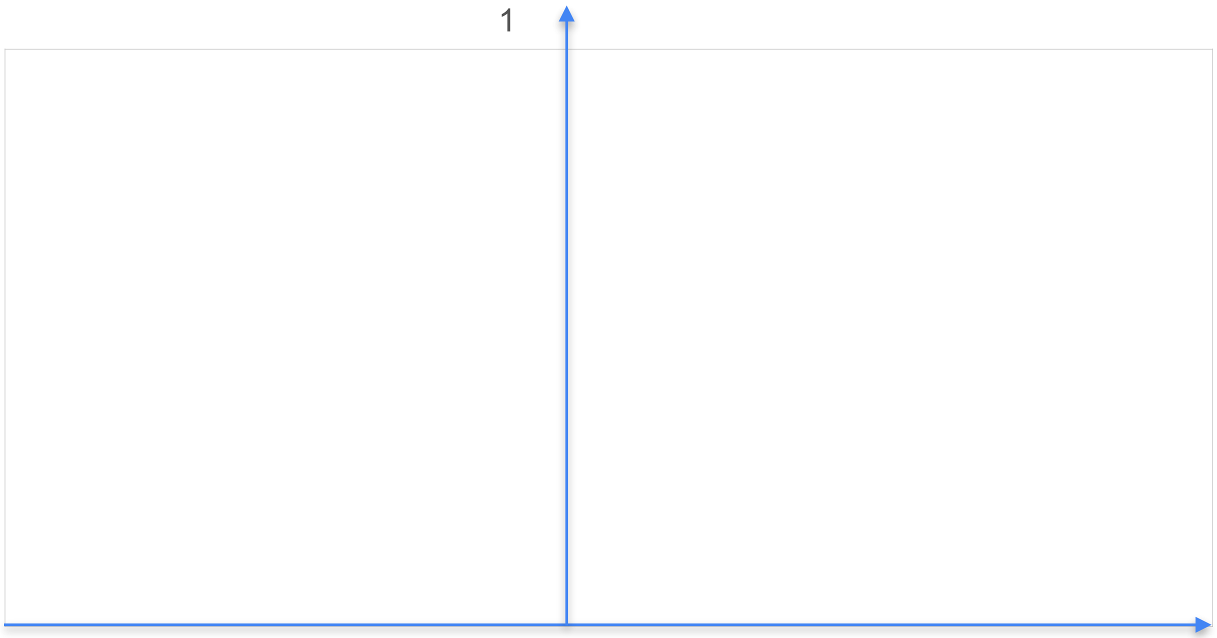
Classification with a perceptron: The sigmoid function

Sigmoid Function

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

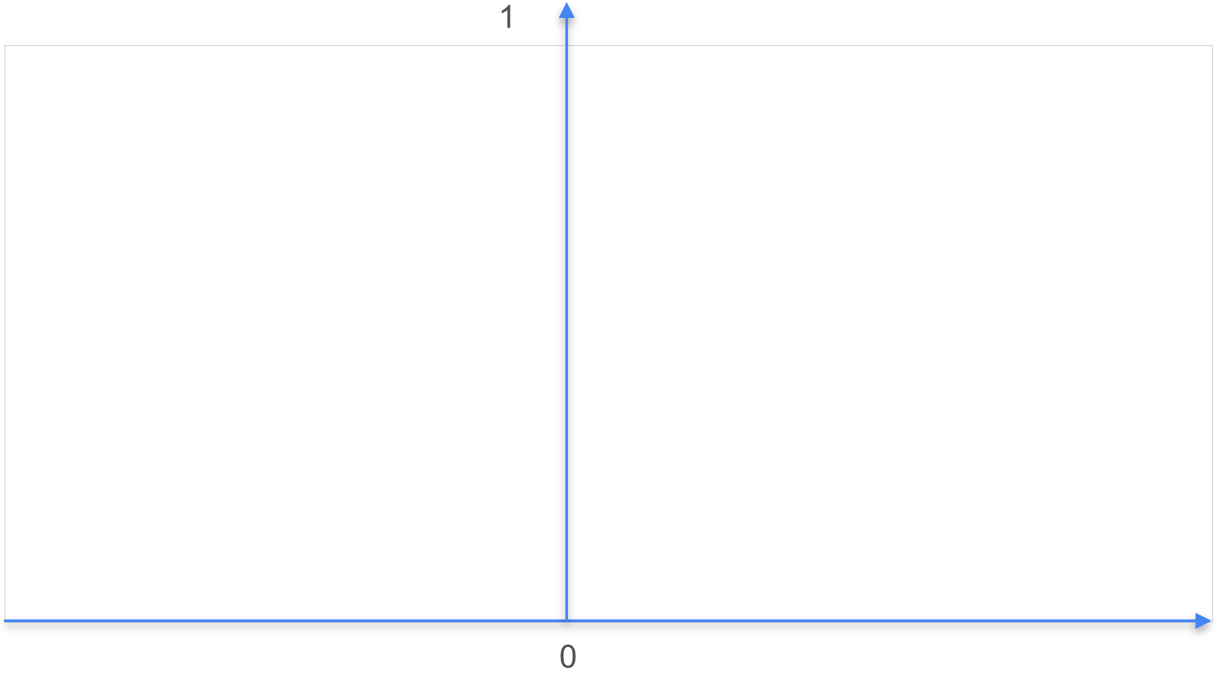
Sigmoid Function

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



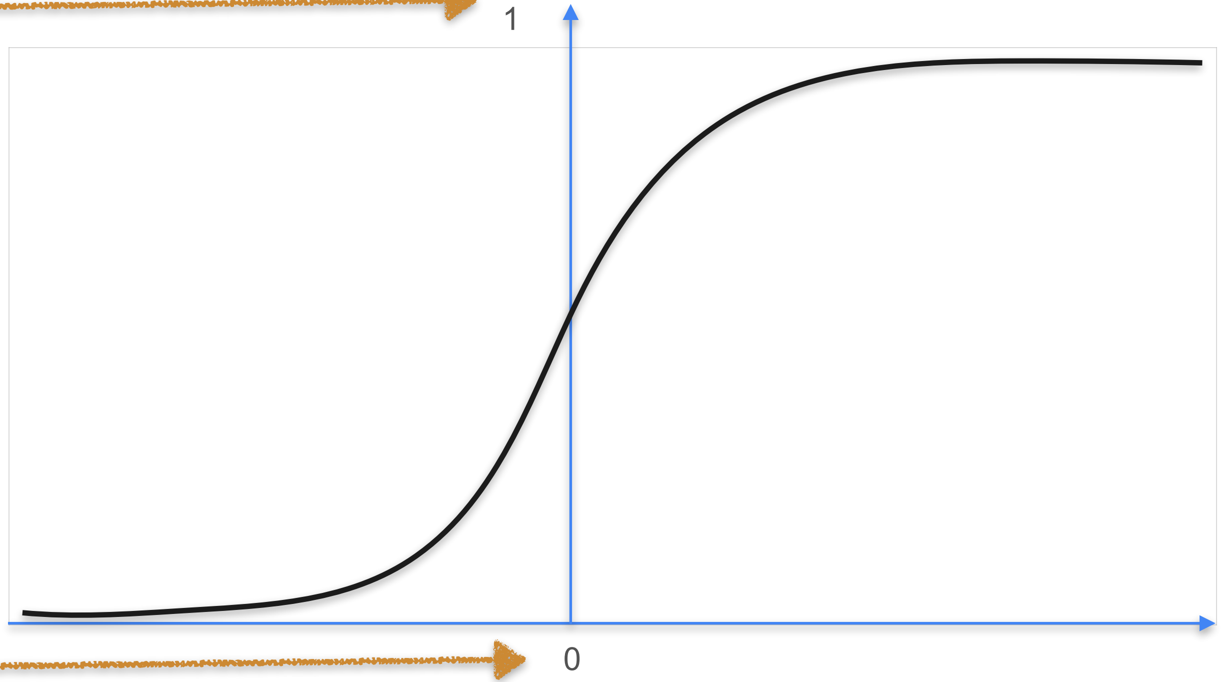
Sigmoid Function

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



Sigmoid Function

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



Derivative of a Sigmoid Function

Derivative of a Sigmoid Function

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

Derivative of a Sigmoid Function

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

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

Derivative of a Sigmoid Function

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

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

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

Derivative of a Sigmoid Function

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

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

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

Derivative of a Sigmoid Function

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

$$= -1$$

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

$$= -1$$

Derivative of a Sigmoid Function

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

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

DeepLearning.AI

Derivative of a Sigmoid Function

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

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

$$= -1$$

DeepLearning.AI

Derivative of a Sigmoid Function

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

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

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

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

DeepLearning.AI

Derivative of a Sigmoid Function

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

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

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

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

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

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

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

Derivative of a Sigmoid Function

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

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

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

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

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

Derivative of a Sigmoid Function

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

Derivative of a Sigmoid Function

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

Derivative of a Sigmoid Function

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

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

Derivative of a Sigmoid Function

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

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

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

Derivative of a Sigmoid Function

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

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

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

Derivative of a Sigmoid Function

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

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

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

Derivative of a Sigmoid Function

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

Recall that:

Derivative of a Sigmoid Function

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

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

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

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

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

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

Recall that: $\sigma(z) = \dfrac{1}{1+e^{-z}}$

Derivative of a Sigmoid Function

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

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

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

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

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

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

Recall that: $\sigma(z) = \dfrac{1}{1+e^{-z}}$

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

Recall that: $\sigma(z) = \dfrac{1}{1+e^{-z}}$

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

Derivative of a Sigmoid Function

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

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

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

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

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

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

Recall that: $\sigma(z) = \dfrac{1}{1+e^{-z}}$

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

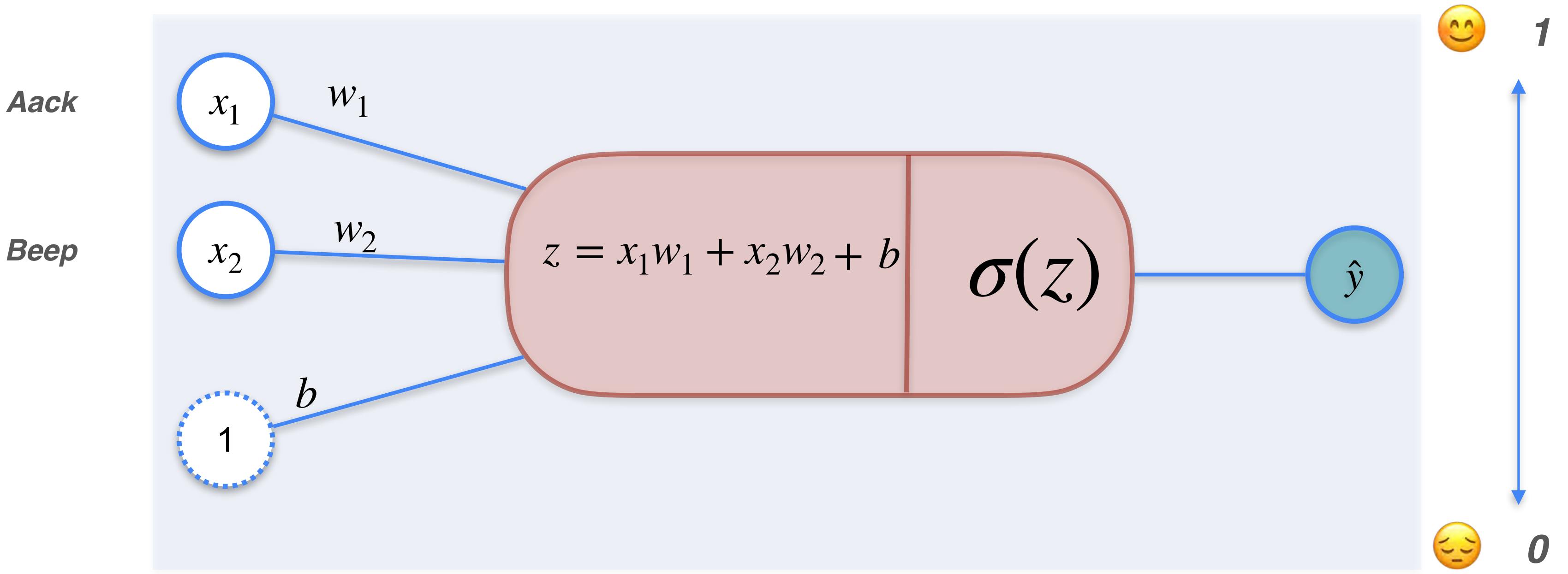
DeepLearning.AI

Optimization in Neural Networks and Newton's Method

Classification with a perceptron: Gradient Descent

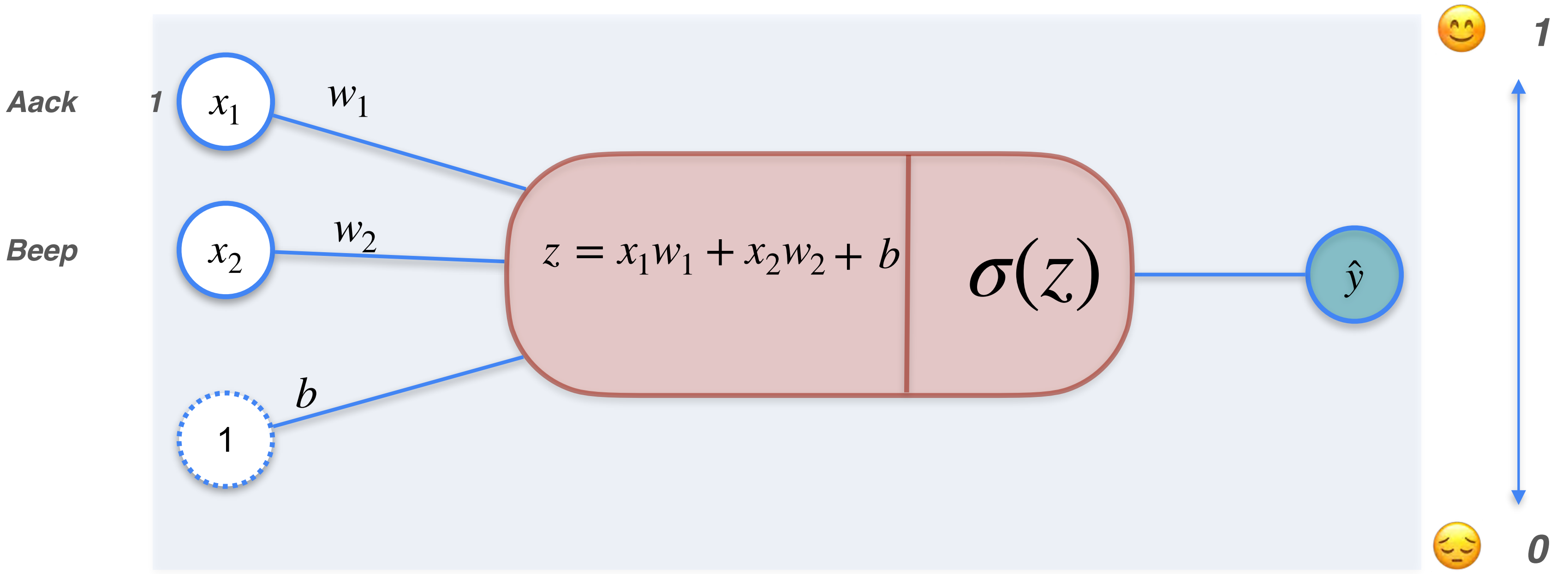
Classification With a Perceptron

Aack beep beep beep



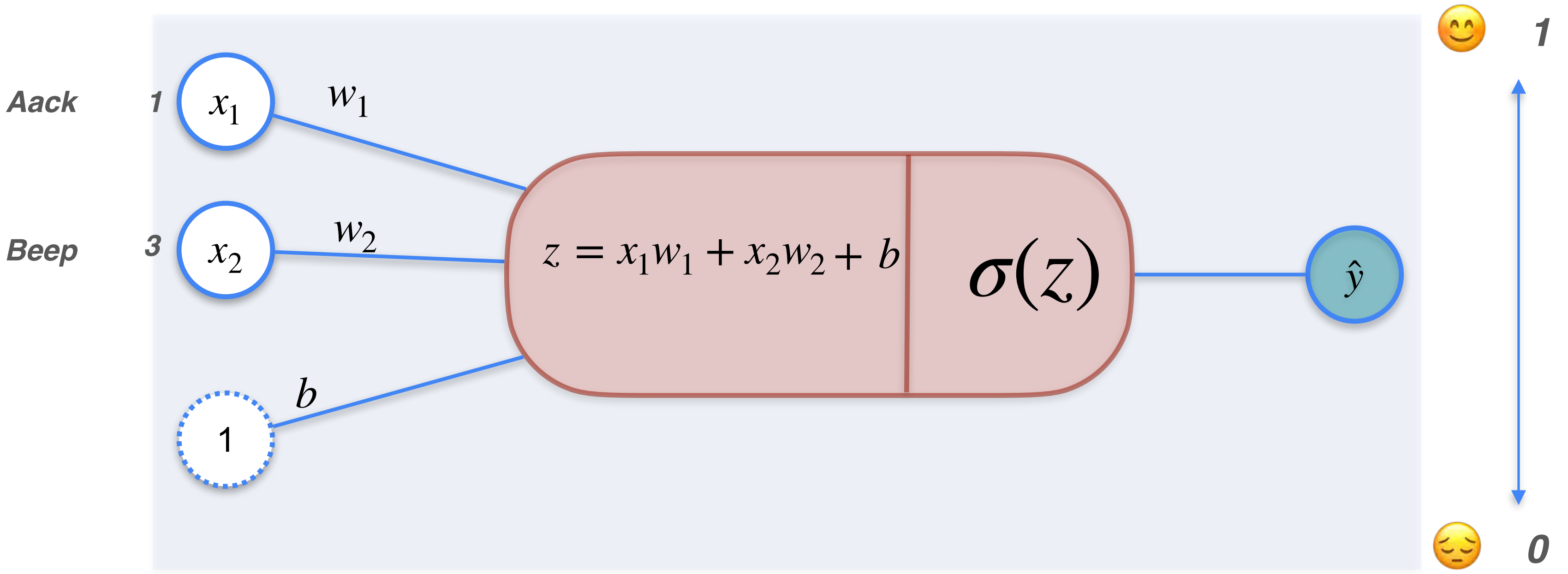
Classification With a Perceptron

Aack beep beep beep



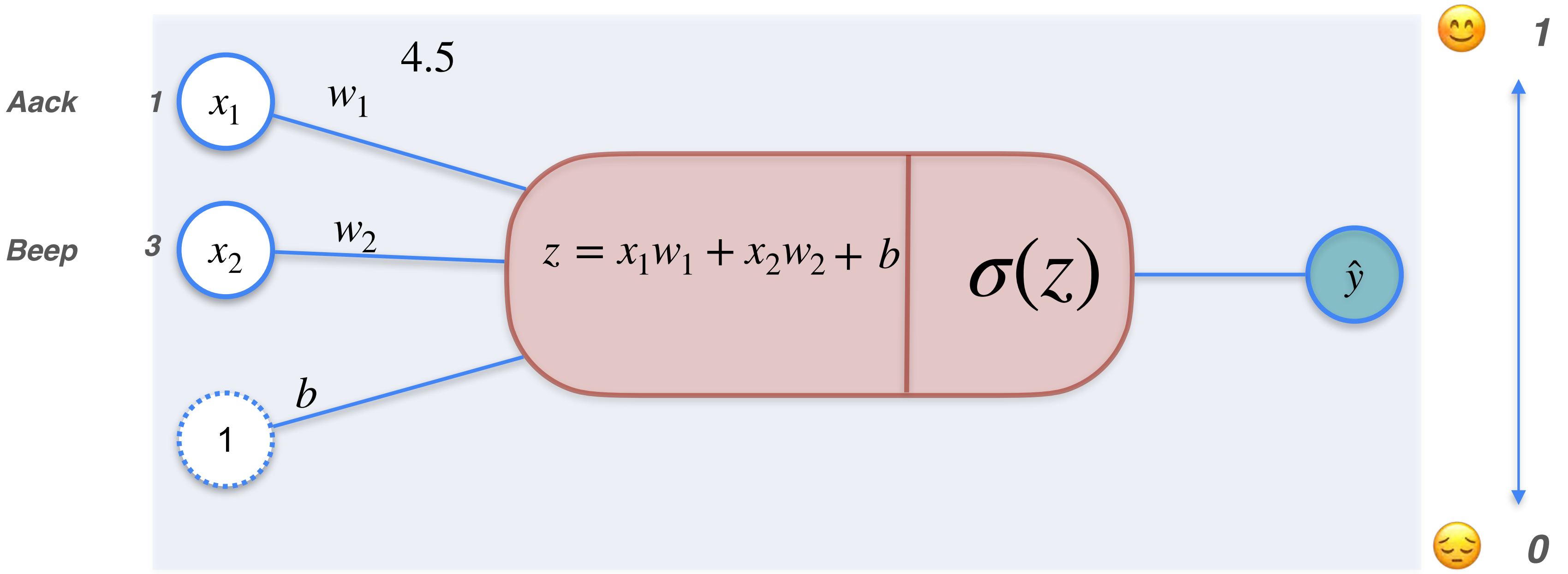
Classification With a Perceptron

Aack beep beep beep



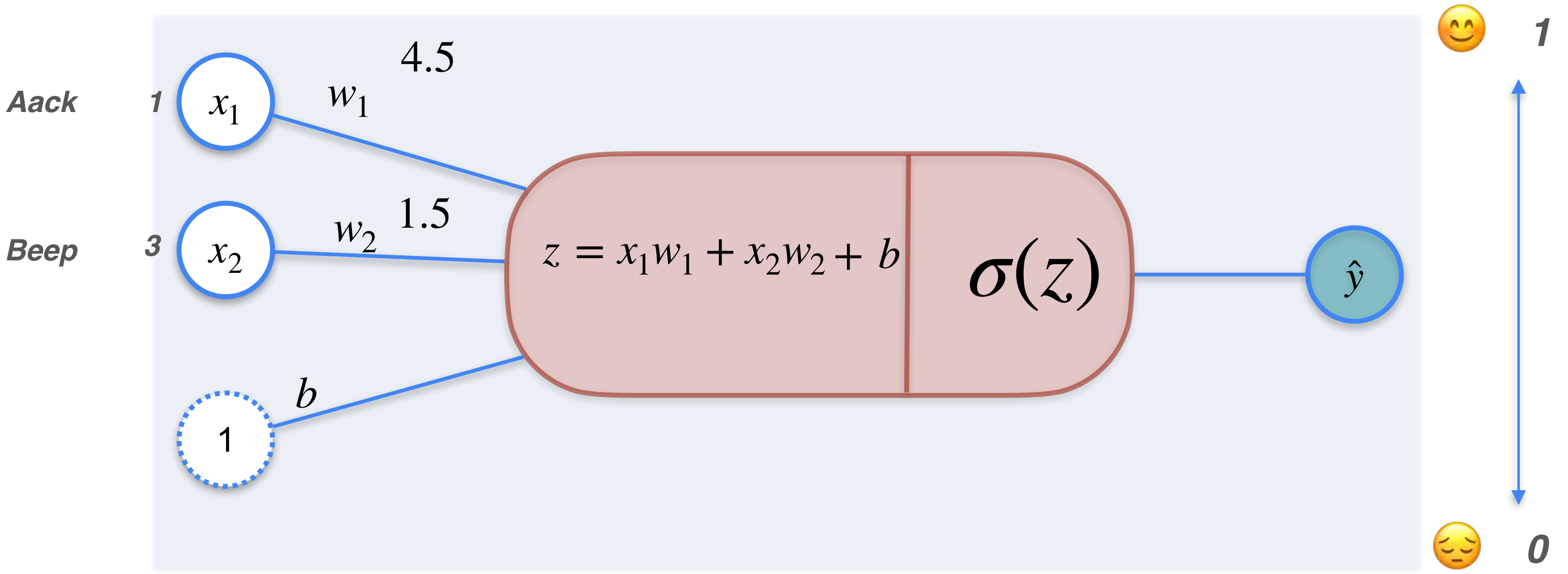
Classification With a Perceptron

Aack beep beep beep

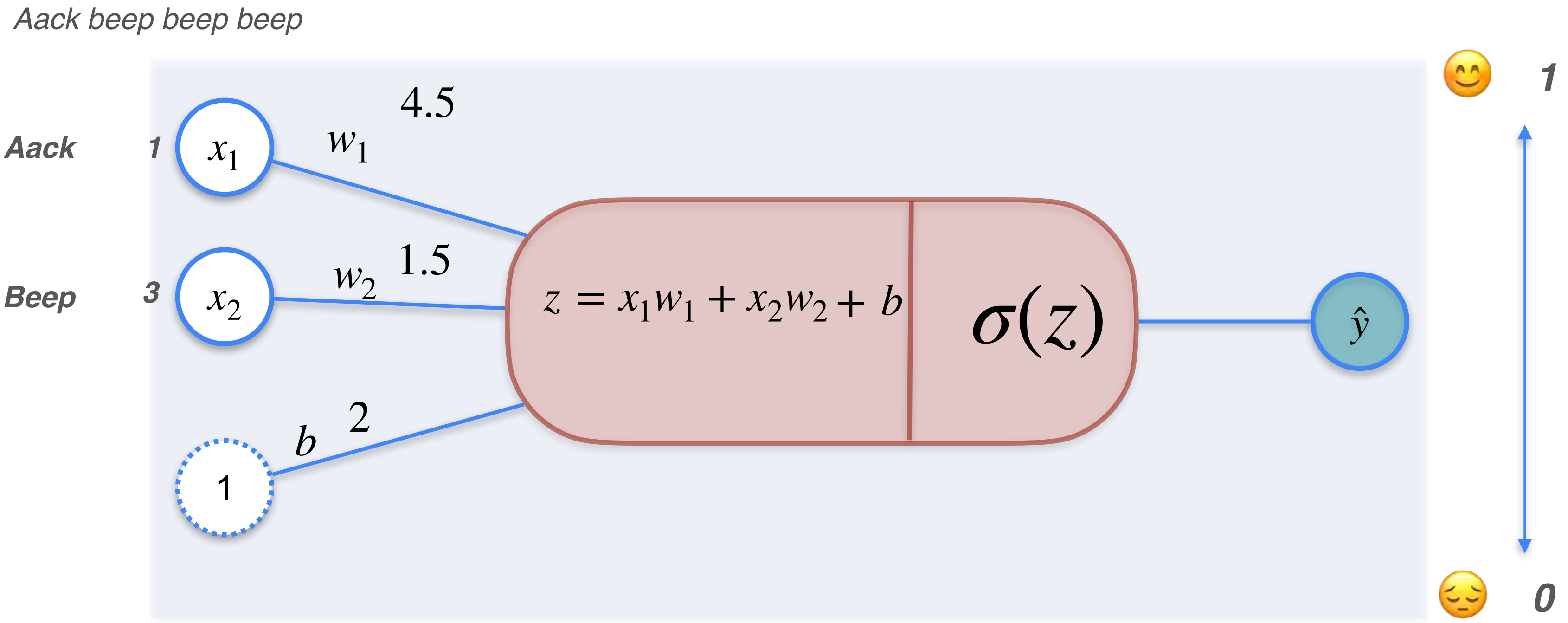


Classification With a Perceptron

Aack beep beep beep

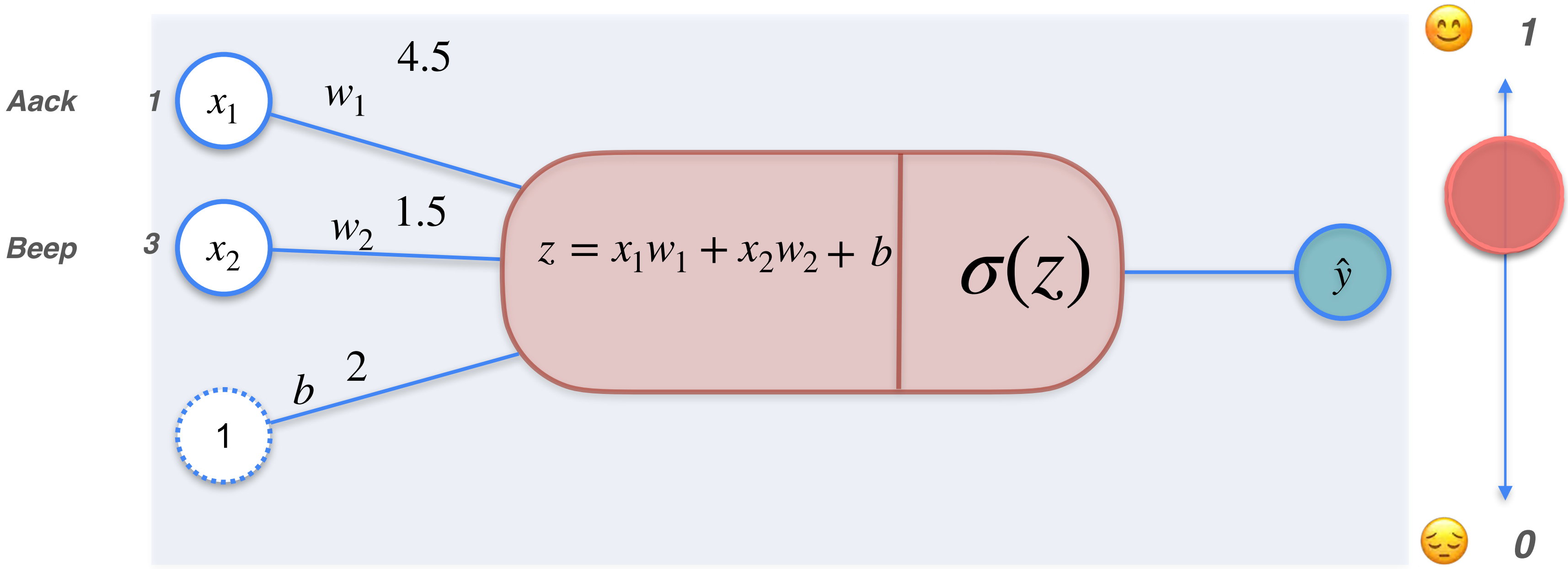


Classification With a Perceptron



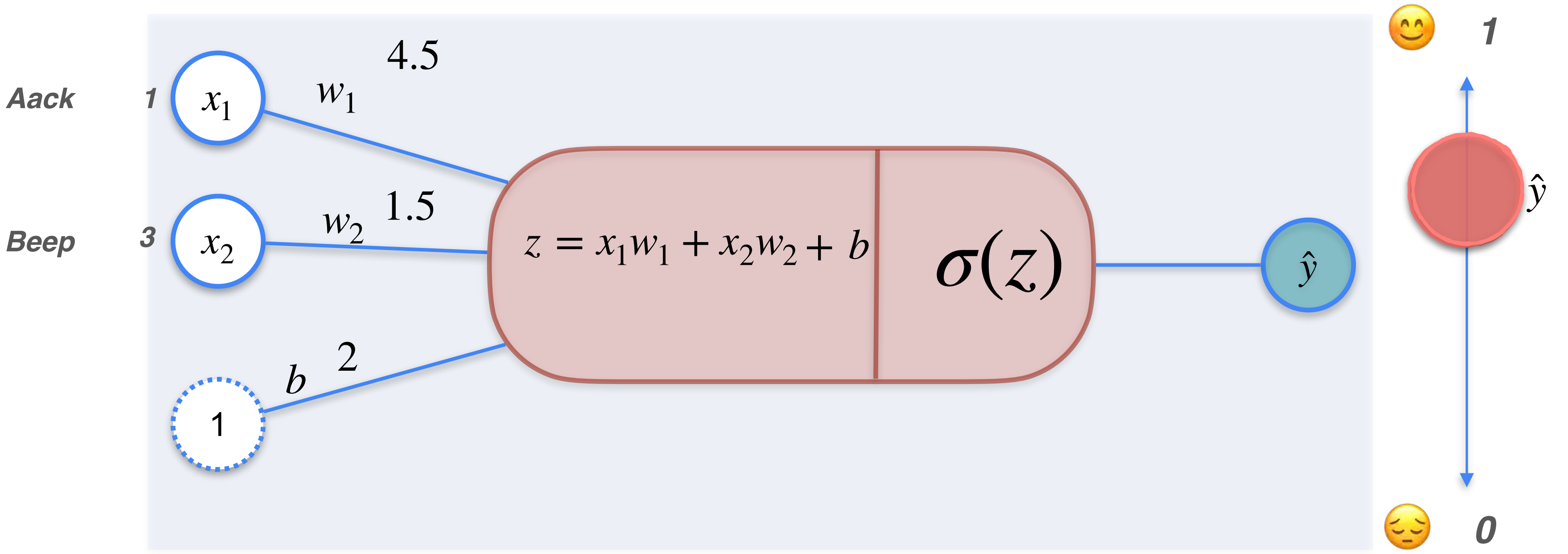
Classification With a Perceptron

Aack beep beep beep



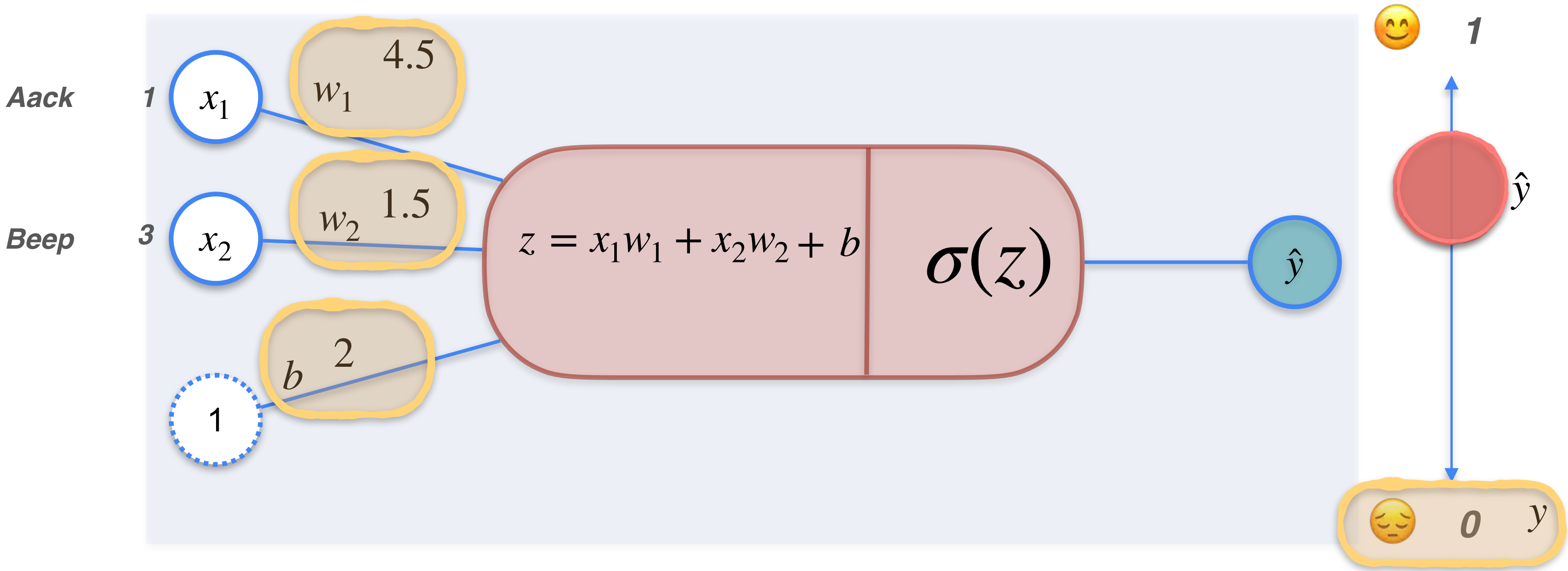
Classification With a Perceptron

Aack beep beep beep



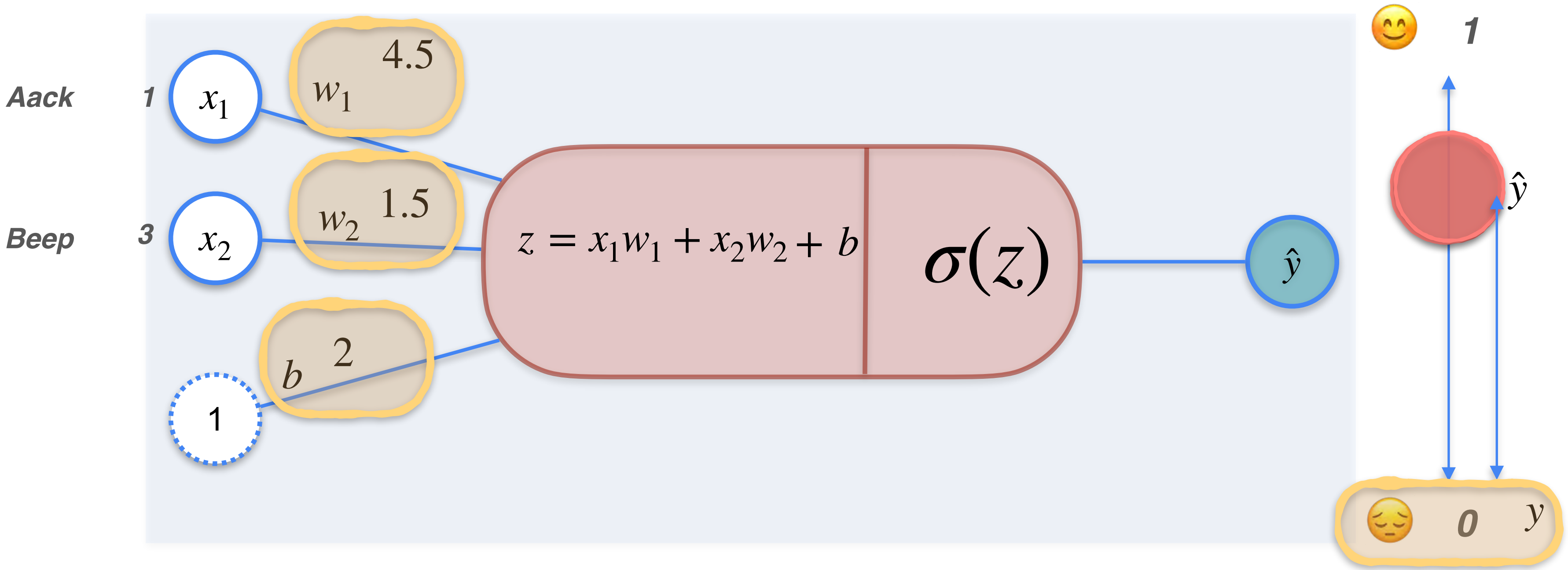
Classification With a Perceptron

Aack beep beep beep



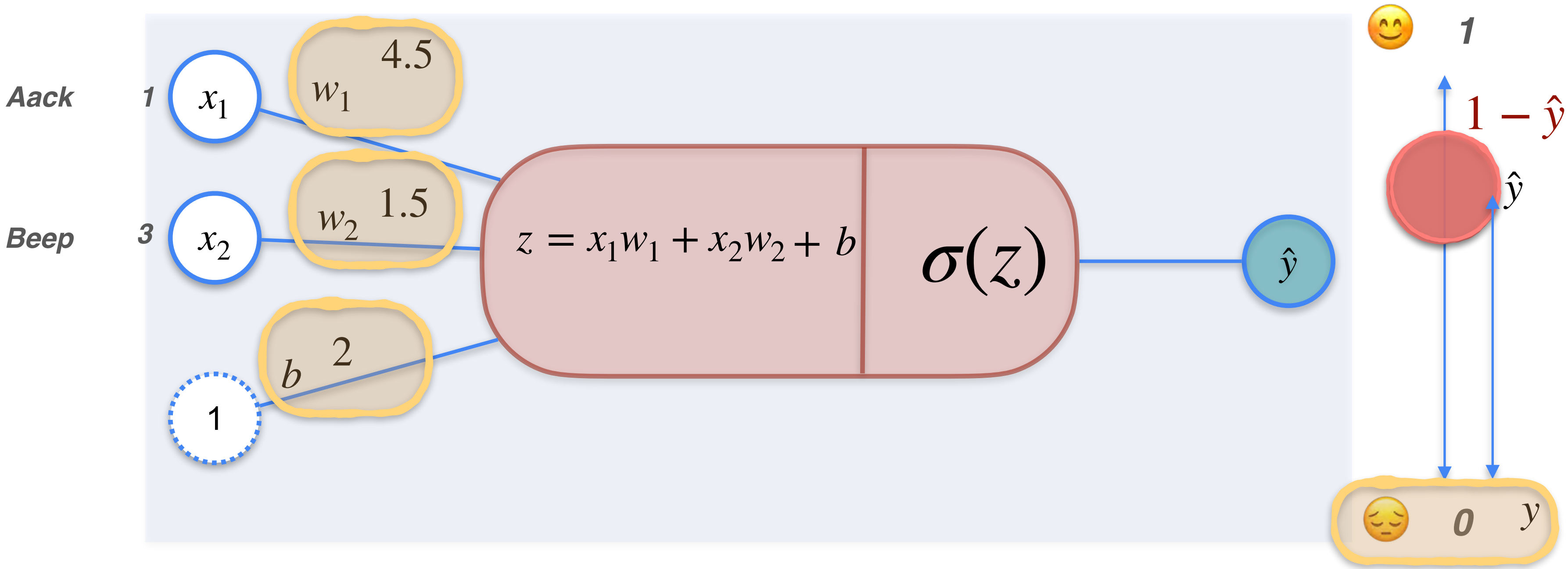
Classification With a Perceptron

Aack beep beep beep



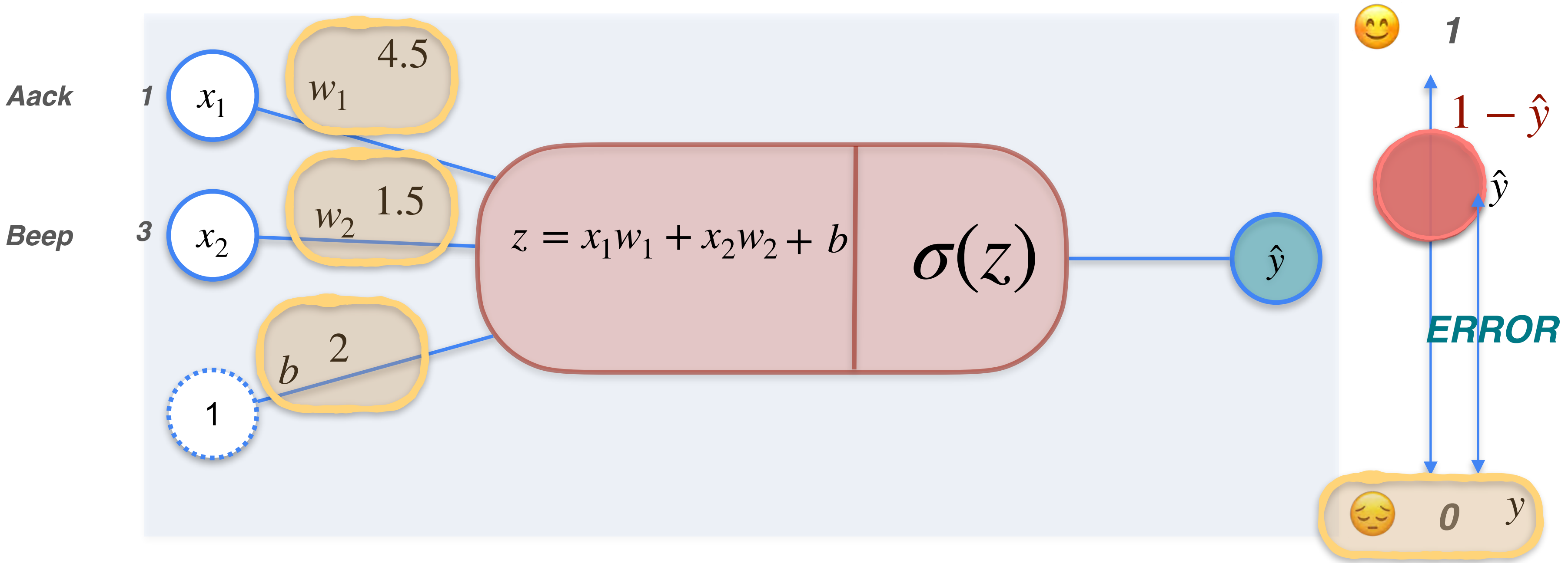
Classification With a Perceptron

Aack beep beep beep

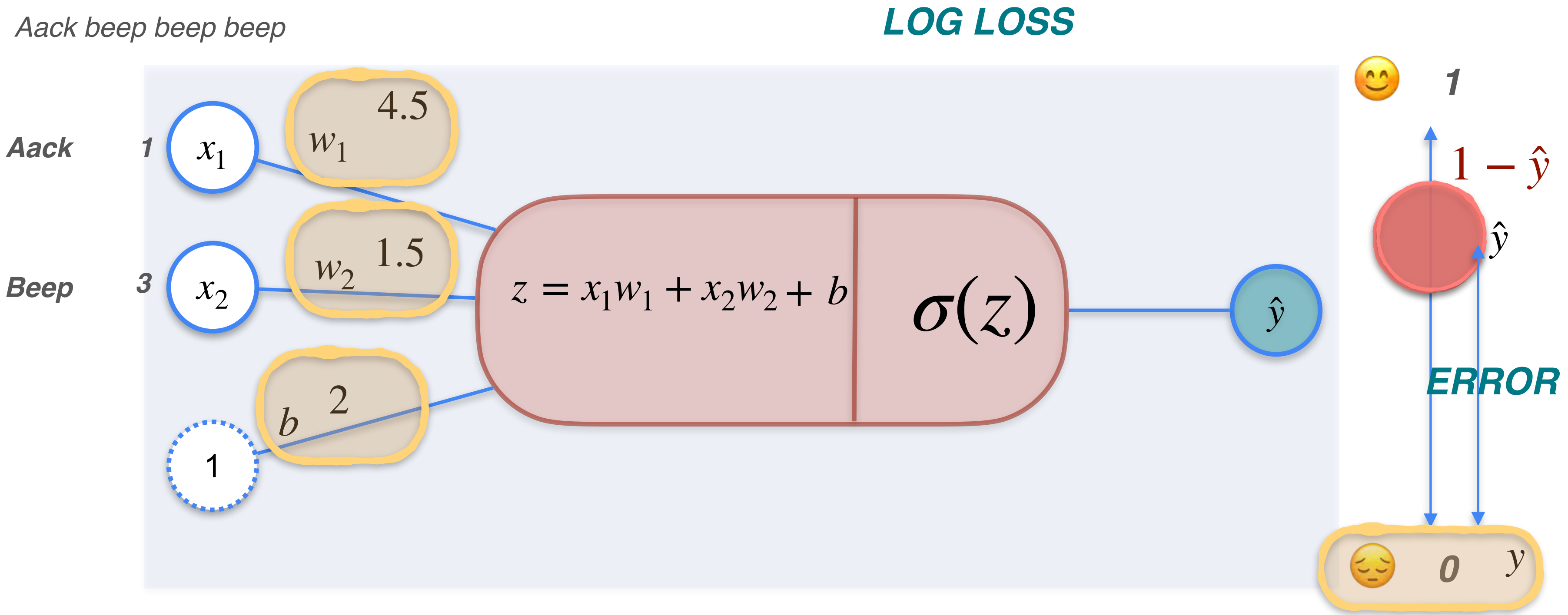


Classification With a Perceptron

Aack beep beep beep

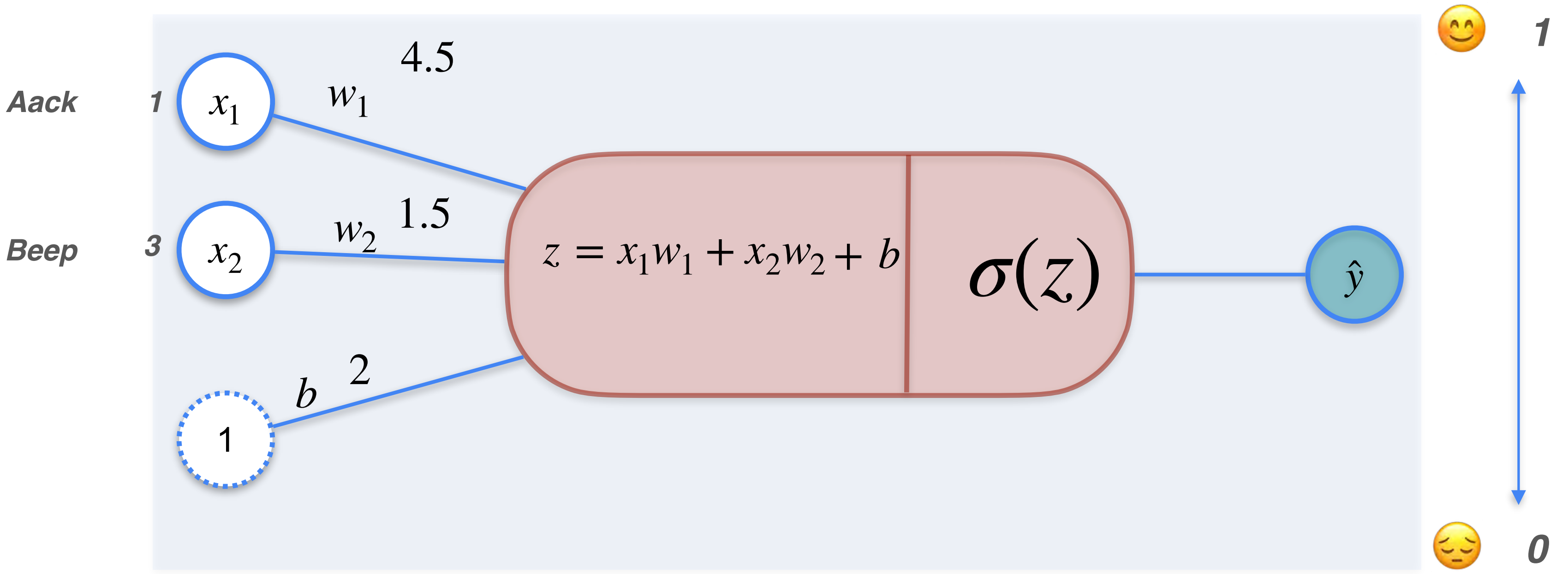


Classification With a Perceptron



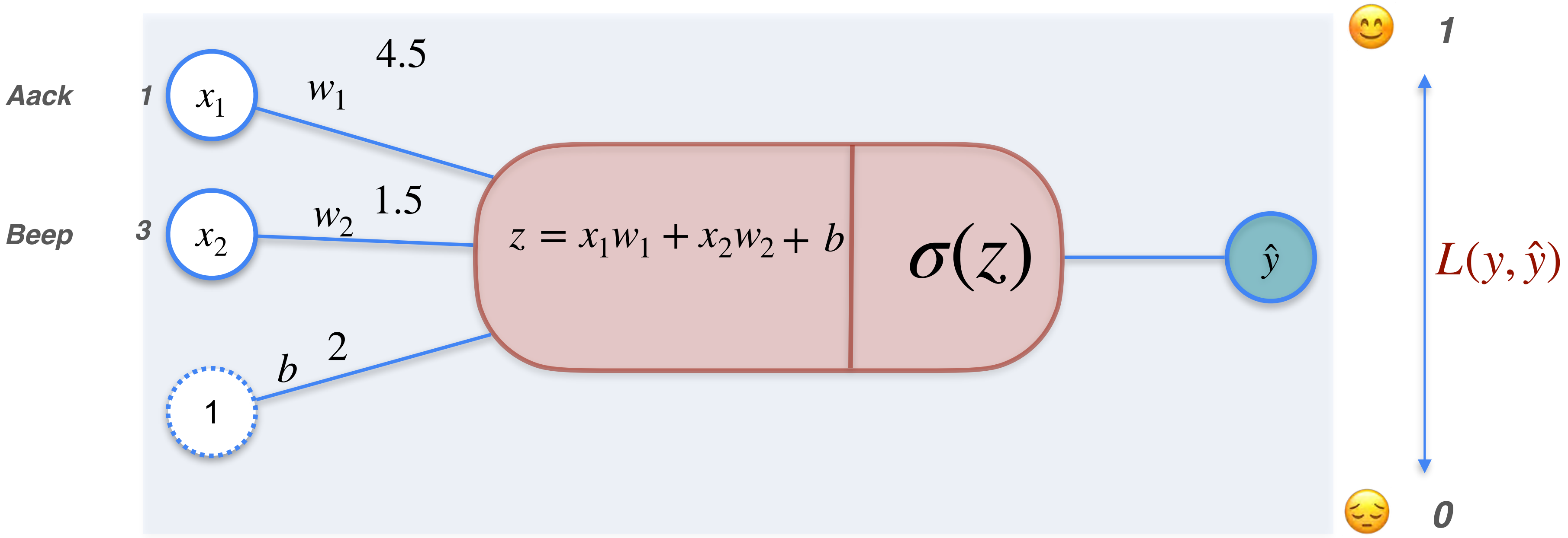
Classification With a Perceptron

Aack beep beep beep



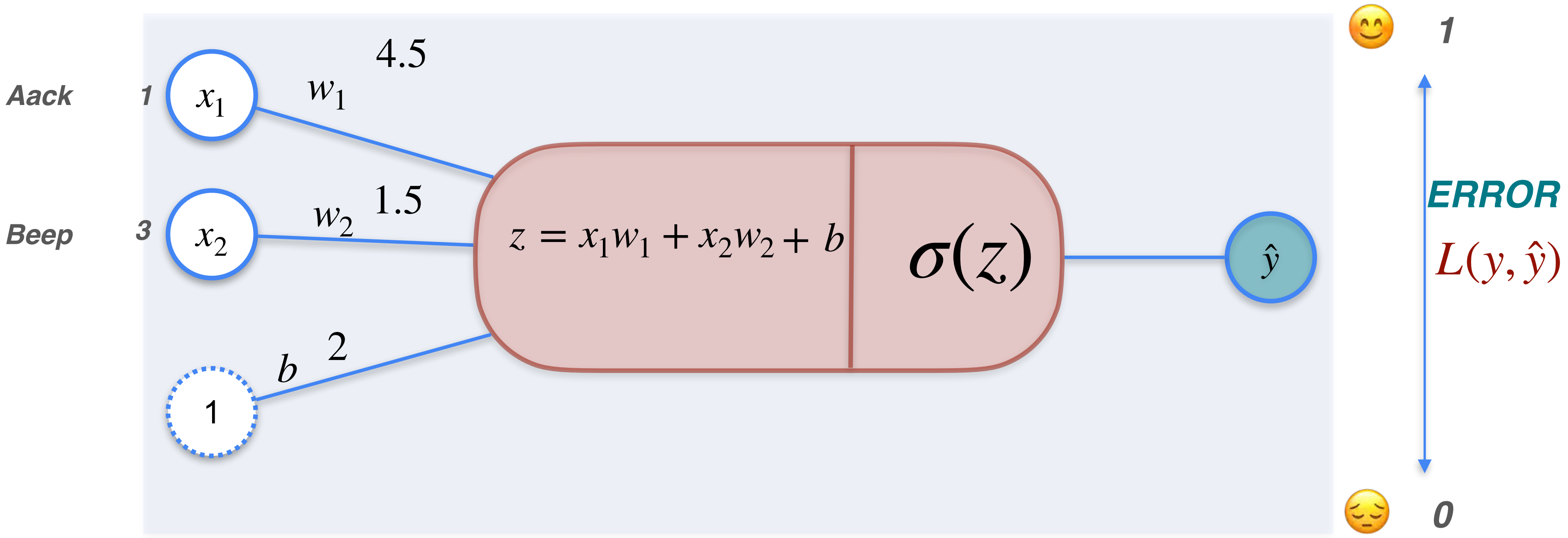
Classification With a Perceptron

Aack beep beep beep



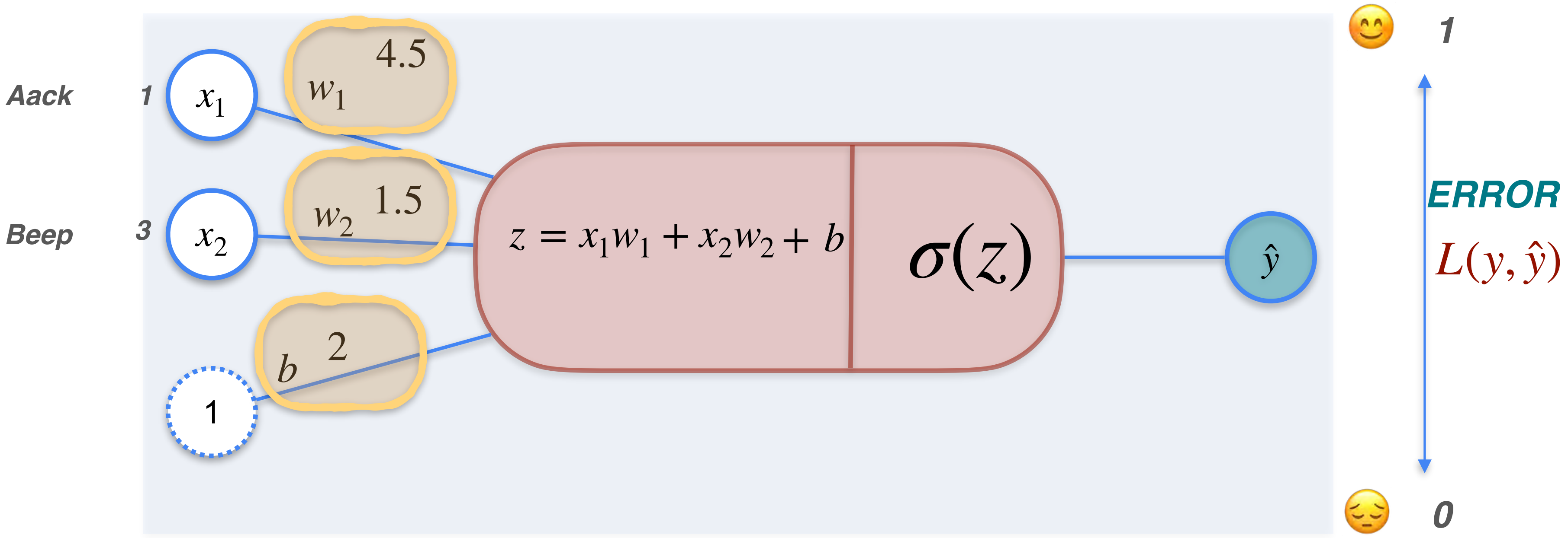
Classification With a Perceptron

Aack beep beep beep



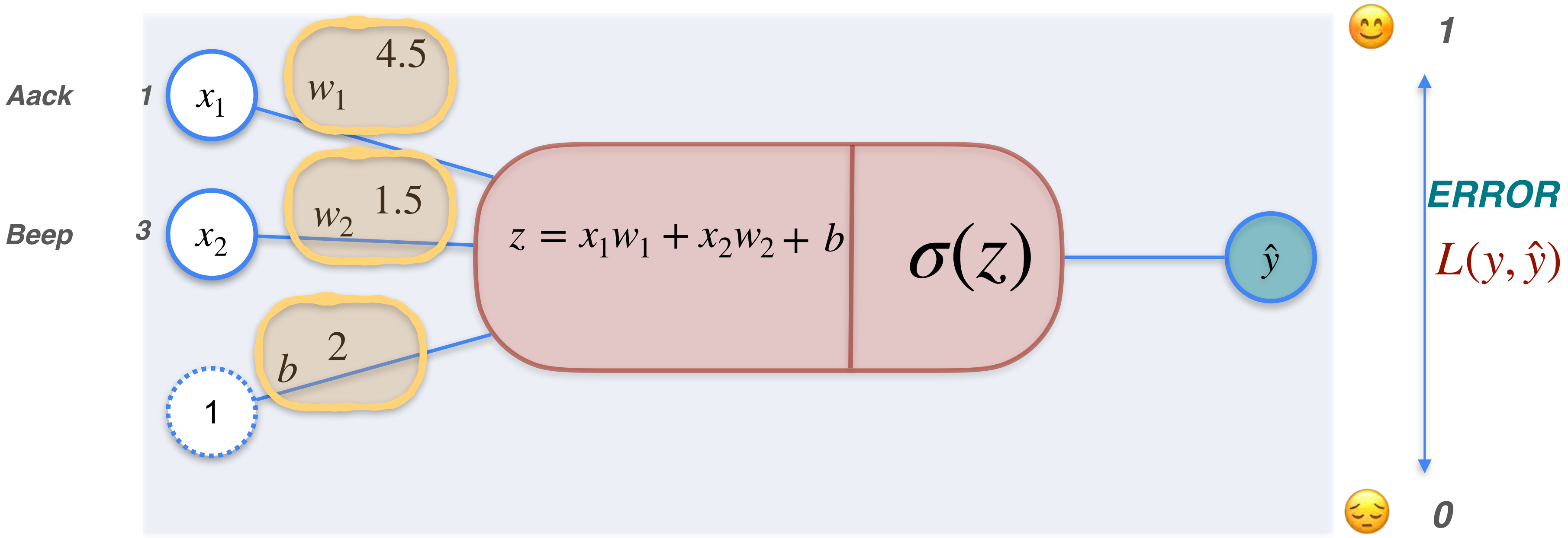
Classification With a Perceptron

Aack beep beep beep

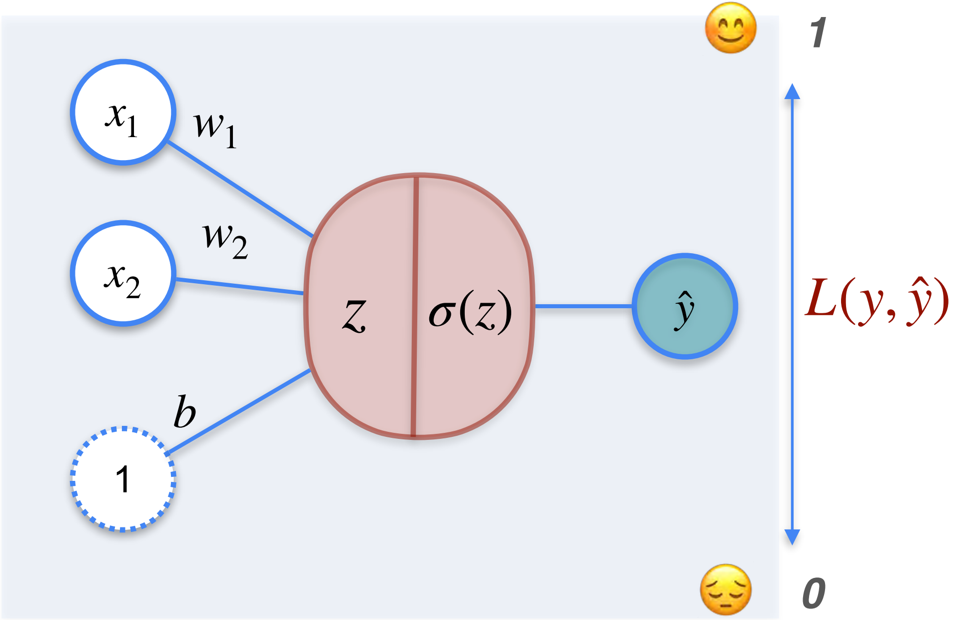


Classification With a Perceptron

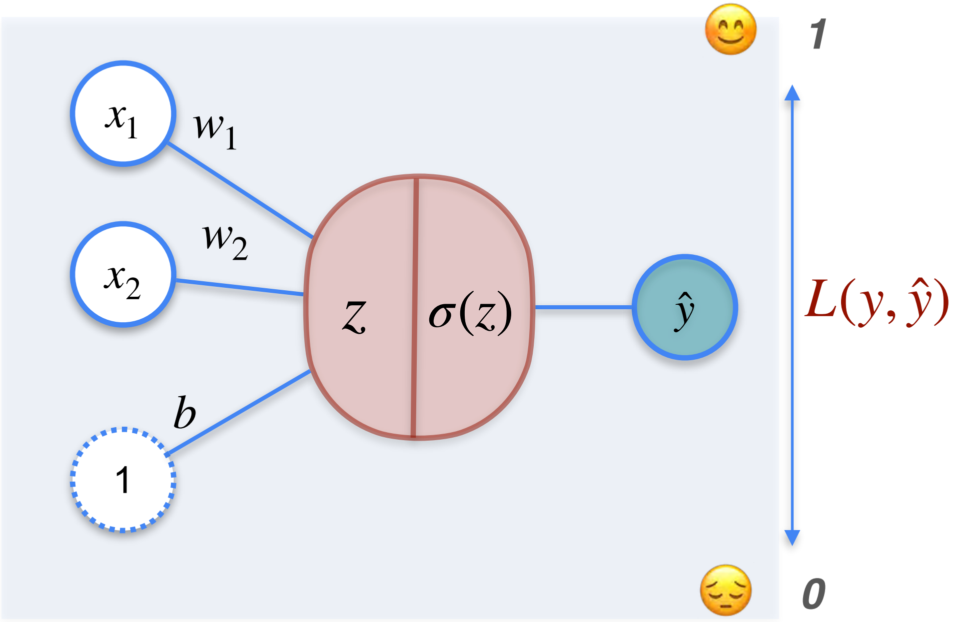
Aack beep beep beep



Classification With a Perceptron



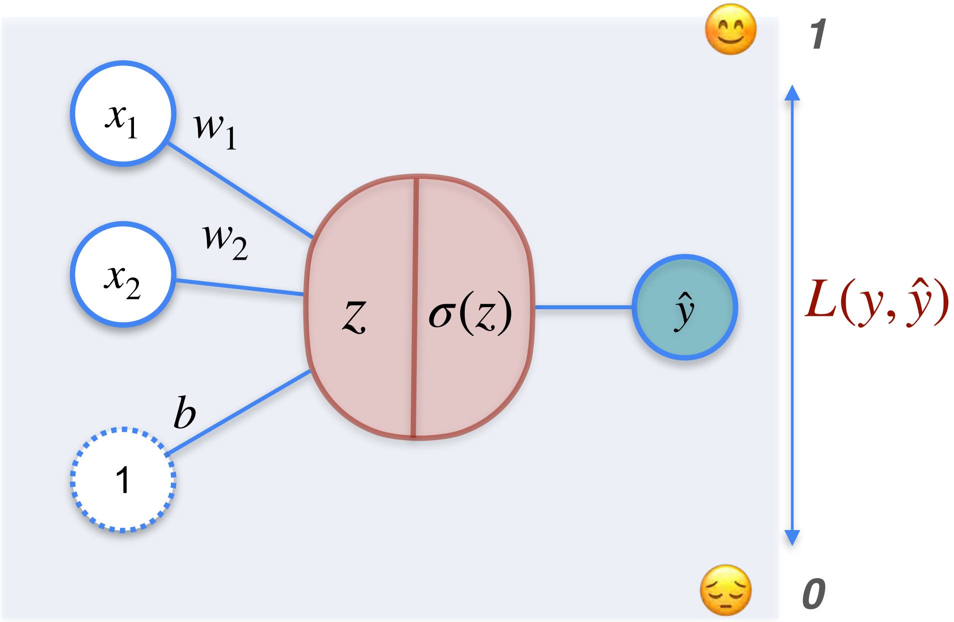
Classification With a Perceptron



Prediction Function:

$\hat{y}$

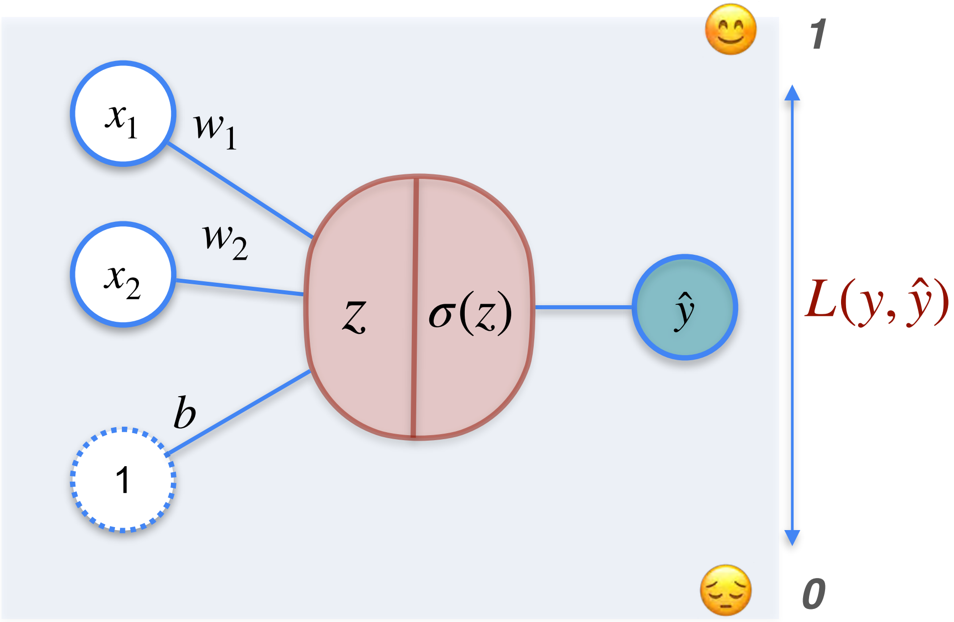
Classification With a Perceptron



Prediction Function:

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

Classification With a Perceptron

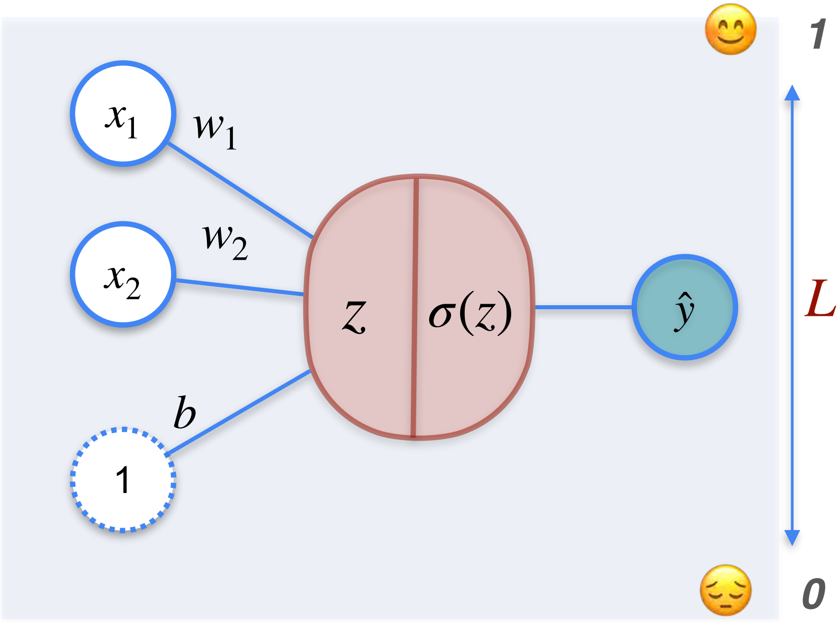


Prediction Function:

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

Loss Function:

Classification With a Perceptron



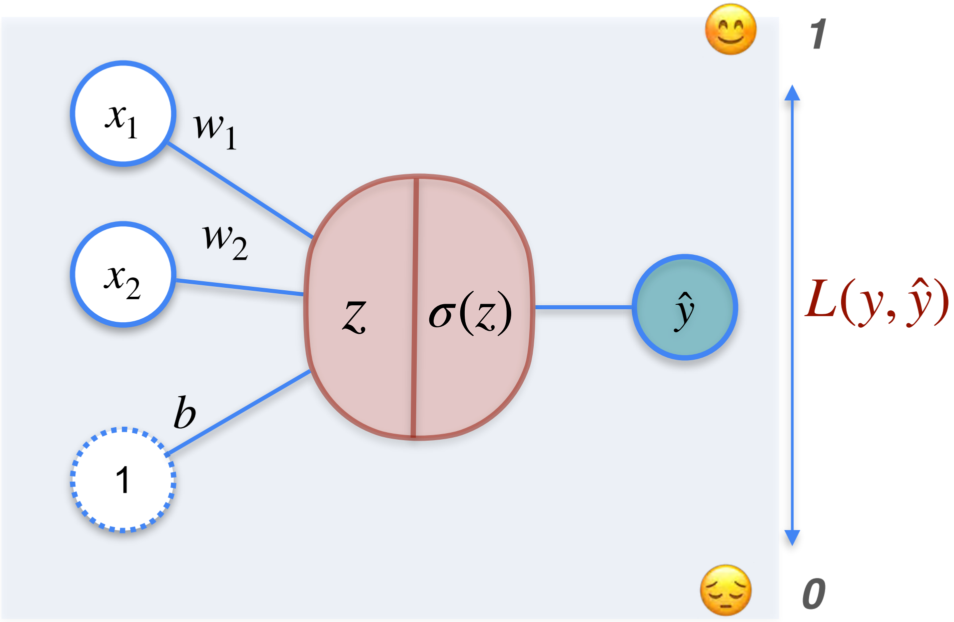
Prediction Function:

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

Loss Function:

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

Classification With a Perceptron



Prediction Function:

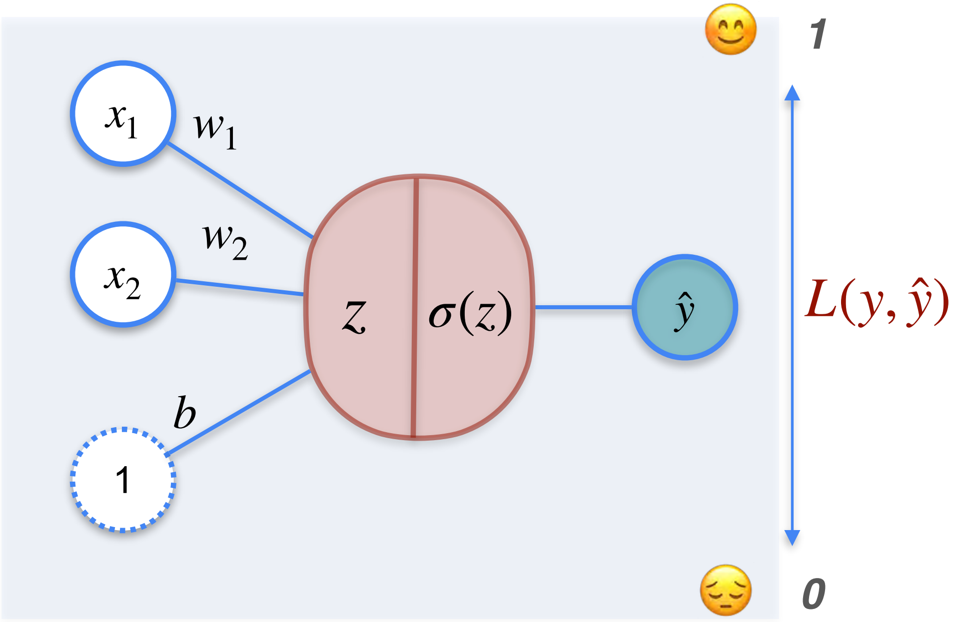
$$\hat{y} = \sigma(w_1 x_1 + w_2 x_2 + b)$$

Loss Function:

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

Main Goal:

Classification With a Perceptron



Prediction Function:

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

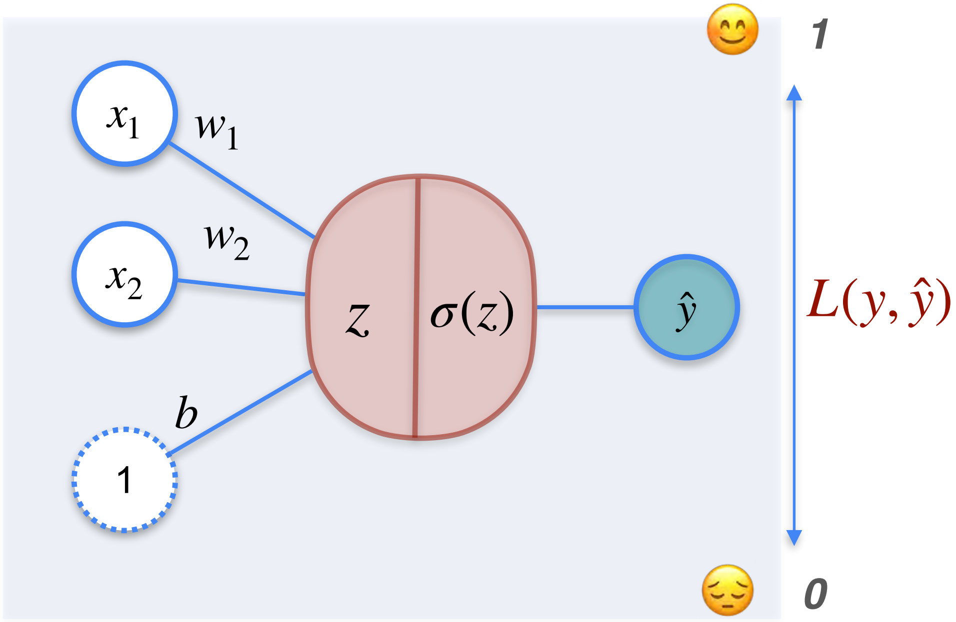
Loss Function:

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

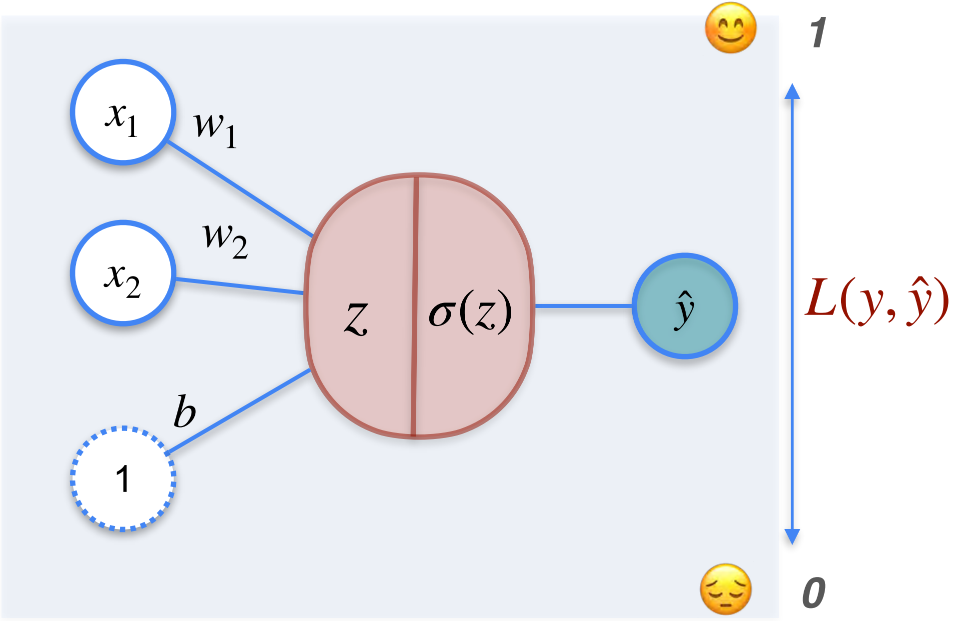
Main Goal:

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

Classification With a Perceptron

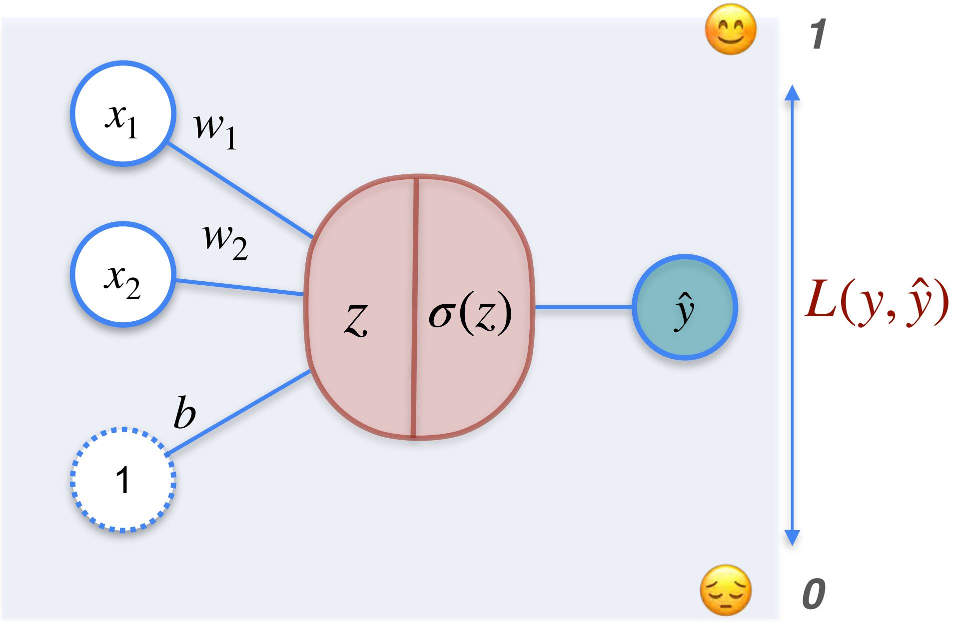


Classification With a Perceptron



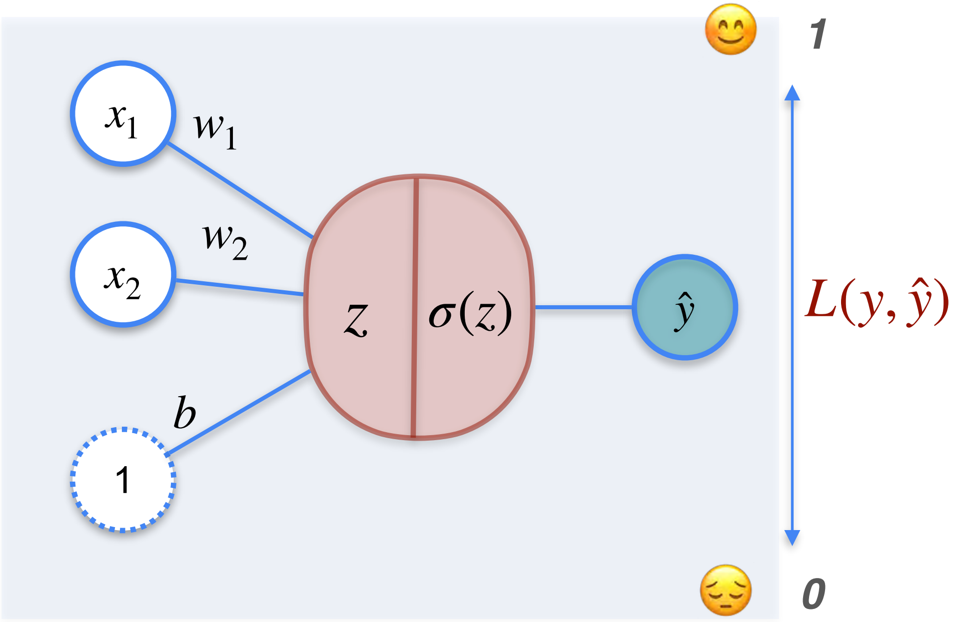
To find optimal values for:

Classification With a Perceptron



To find optimal values for:
w_1 , w_2 , b

Classification With a Perceptron

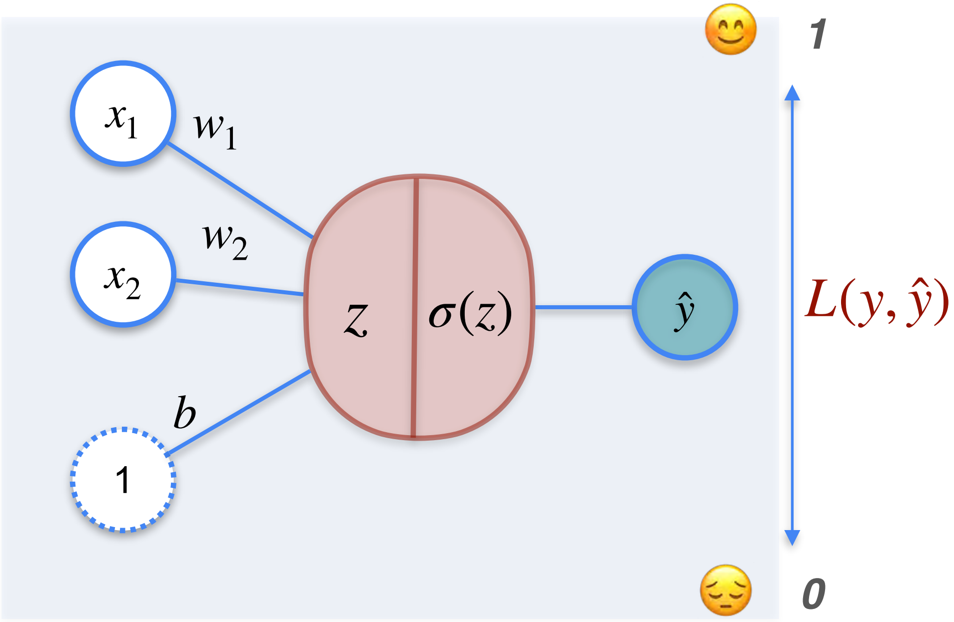


To find optimal values for:

w_1 , w_2 , b

You need gradient descent

Classification With a Perceptron



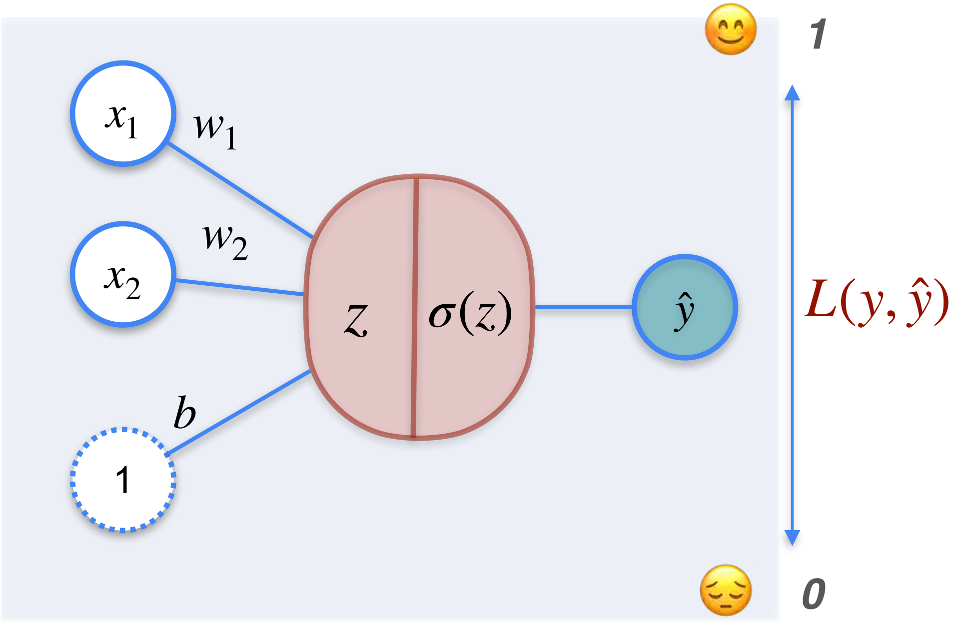
To find optimal values for:

w_1 , w_2 , b

You need gradient descent

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

Classification With a Perceptron



To find optimal values for:

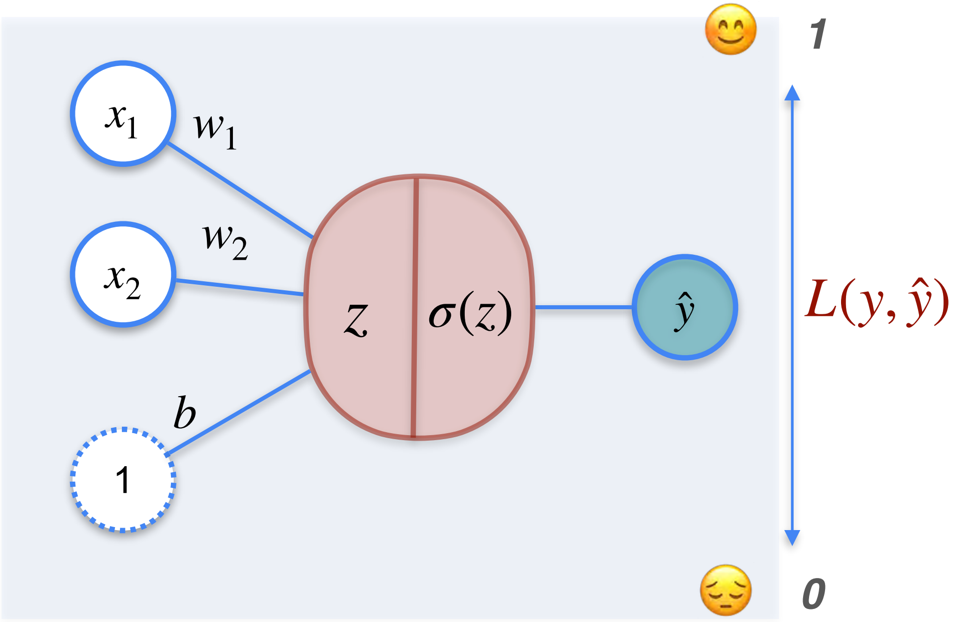
w_1 , w_2 , b

You need gradient descent

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

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

Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

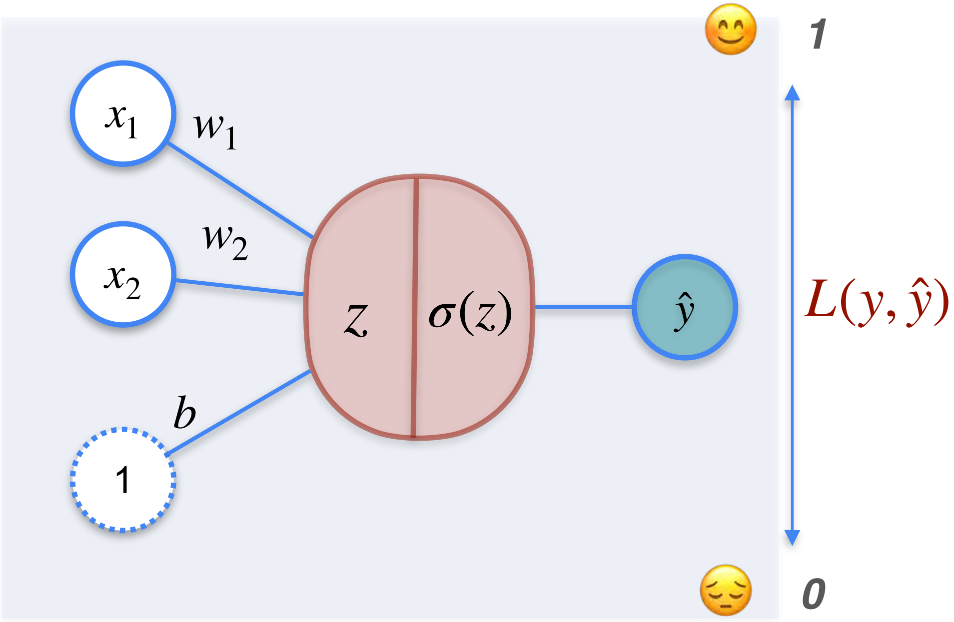
You need gradient descent

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

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

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

Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

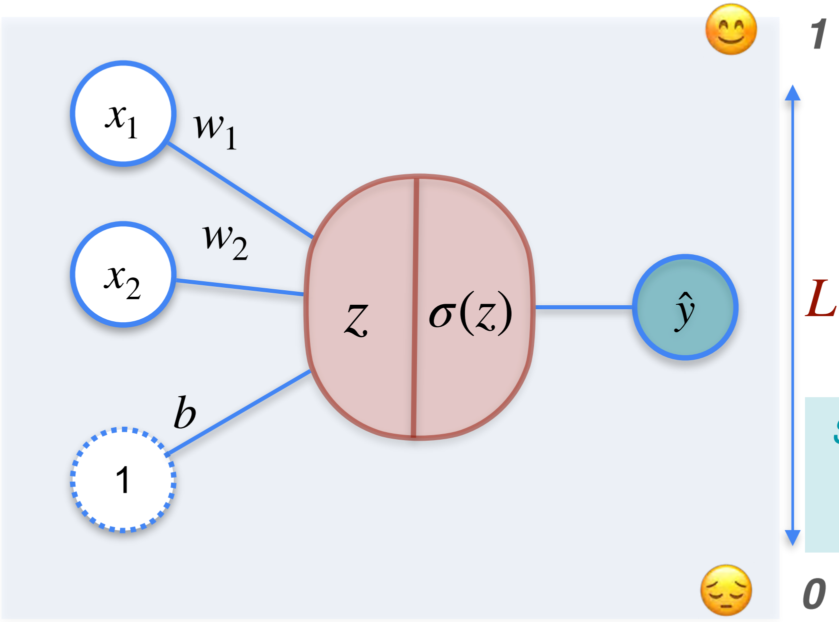
You need gradient descent

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

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

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

Classification With a Perceptron



$L(y, \hat{y})$

To find optimal values for:

$w_1\ ,\ w_2\ ,\ b$

You need gradient descent

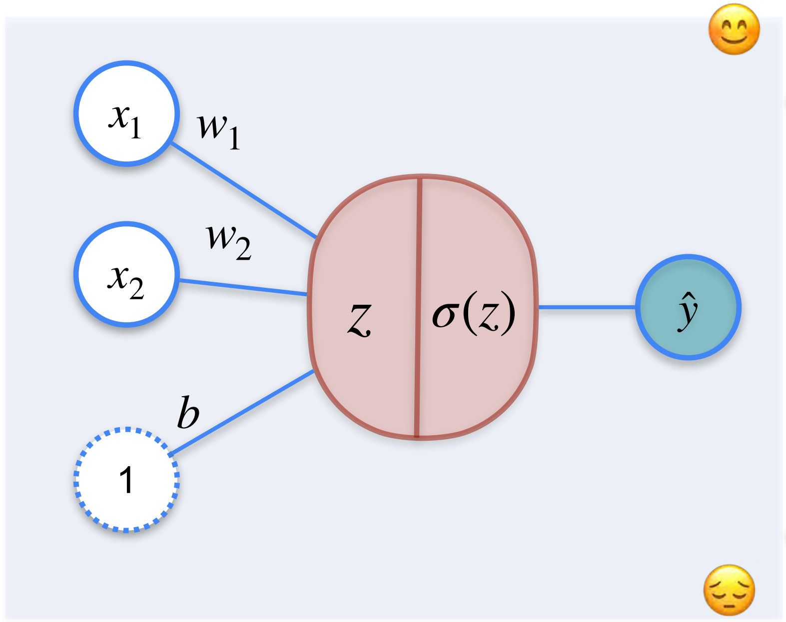
$$w_1 \rightarrow w_1 - \alpha \frac{\partial L}{\partial w_1}$$

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

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

Some initial starting values

Classification With a Perceptron



1

$L(y, \hat{y})$

0

To find optimal values for:

$w_1 \ , \ w_2 \ , \ b$

You need gradient descent

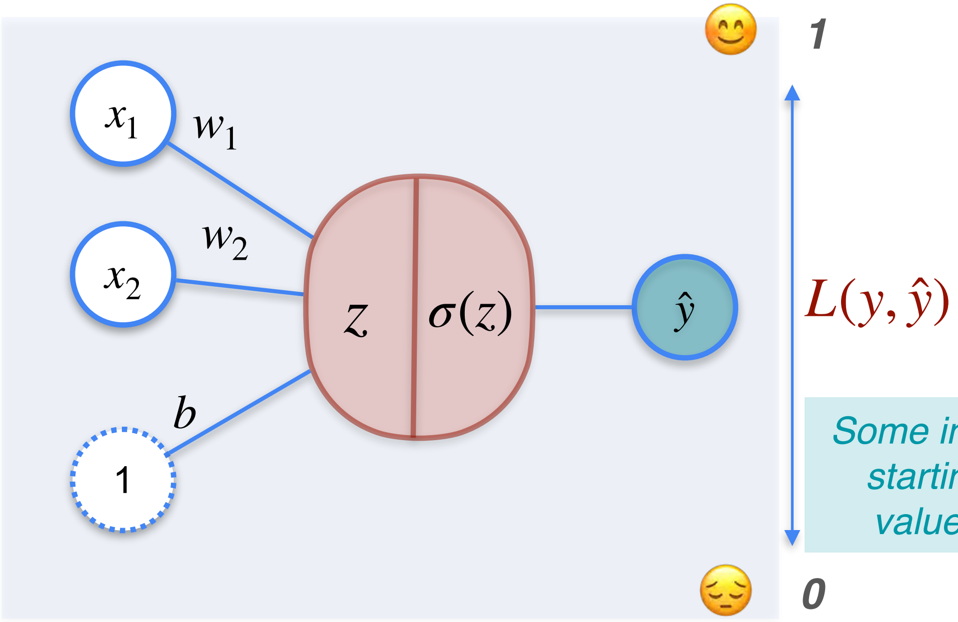
Some initial starting values

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

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

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

Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

You need gradient descent

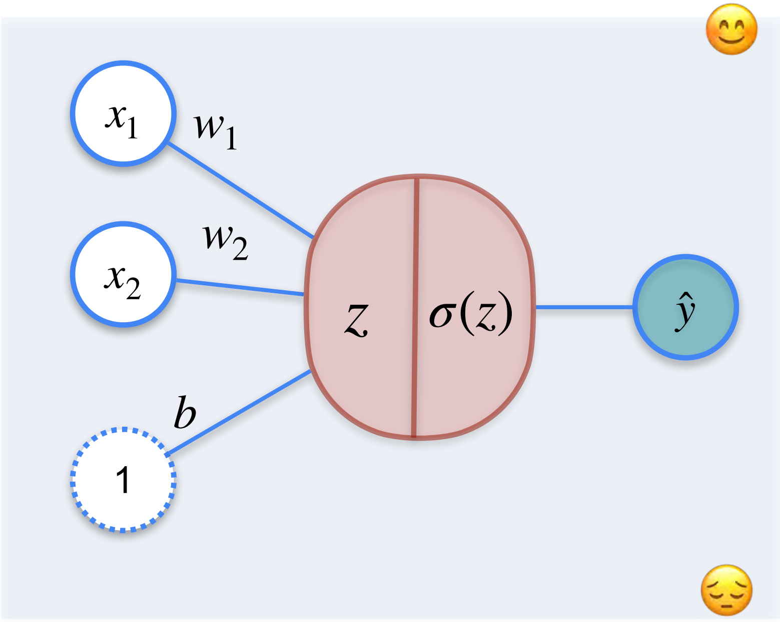
Some initial starting values

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

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

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

Classification With a Perceptron



1

$L(y, \hat{y})$

0

To find optimal values for:

w_1 , w_2 , b

You need gradient descent

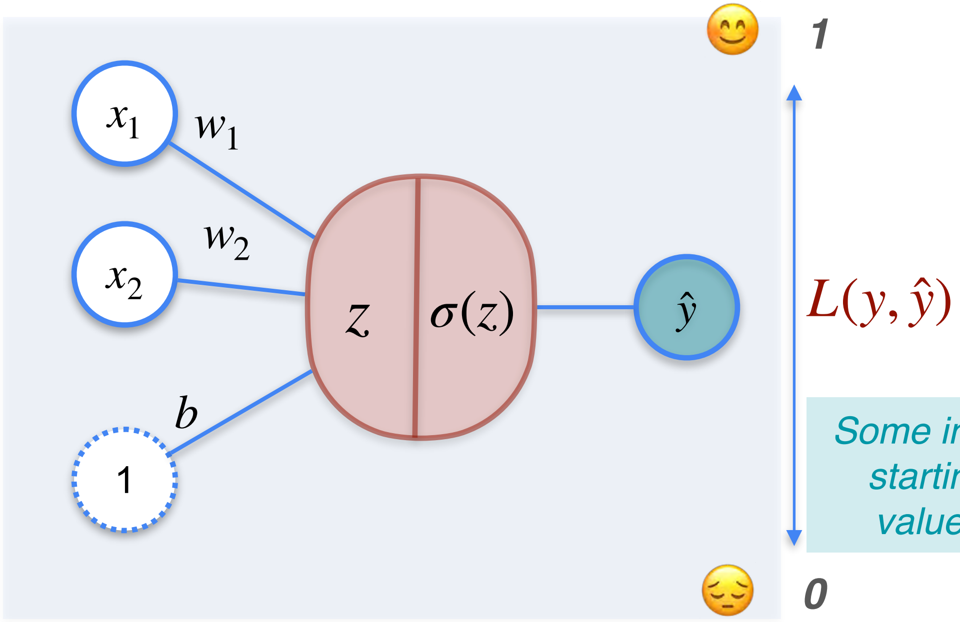
Some initial starting values

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

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

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

Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

You need gradient descent

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

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

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

Some initial starting values

SUB-TASK

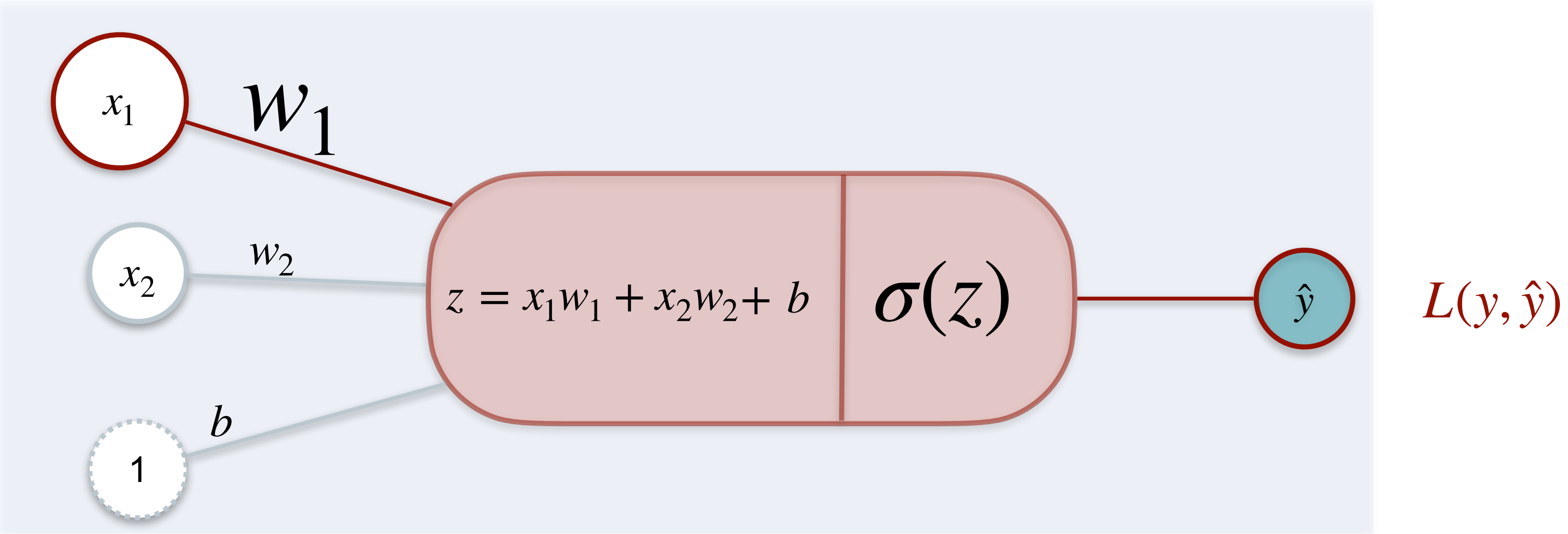
Find the following partial derivatives

DeepLearning.AI

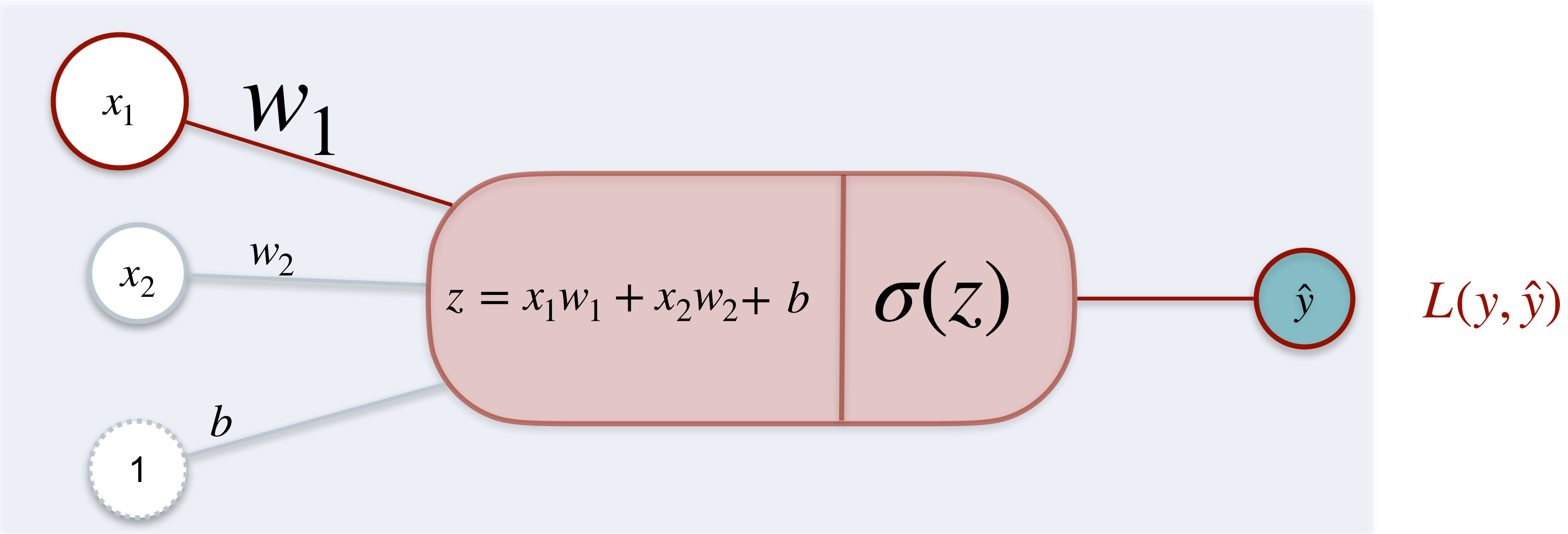
Optimization in Neural Networks and Newton's Method

Classification with a perceptron: Calculating the derivatives

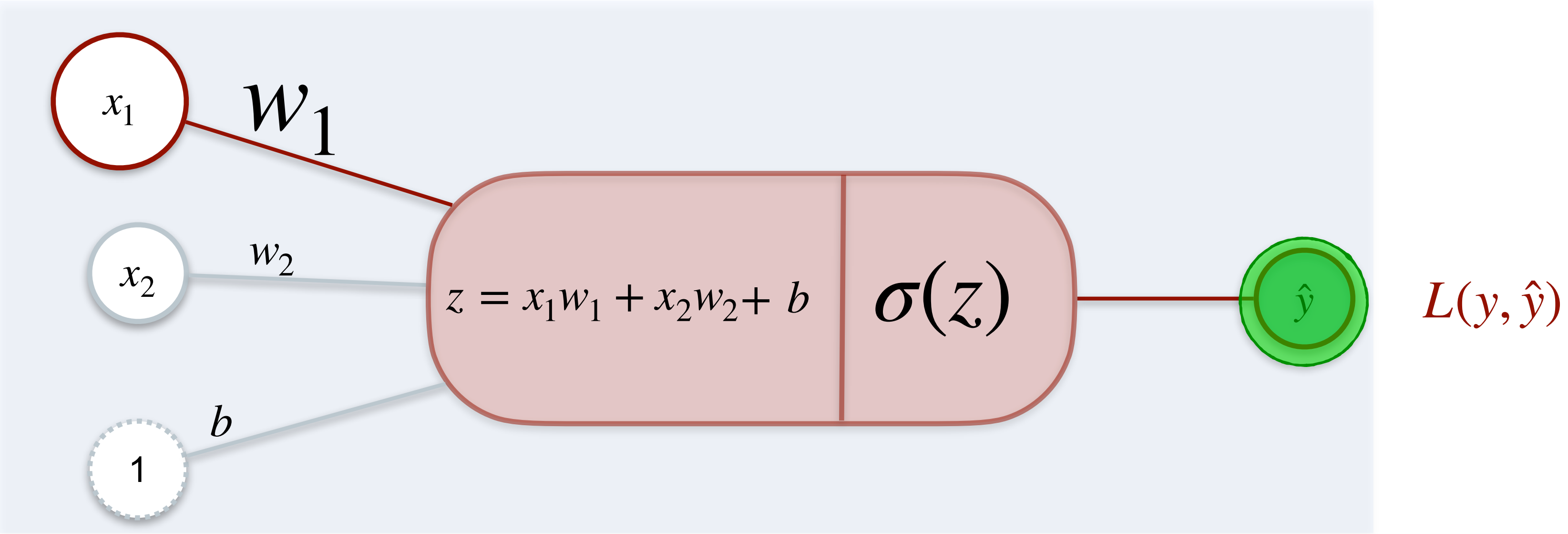
Classification With a Perceptron



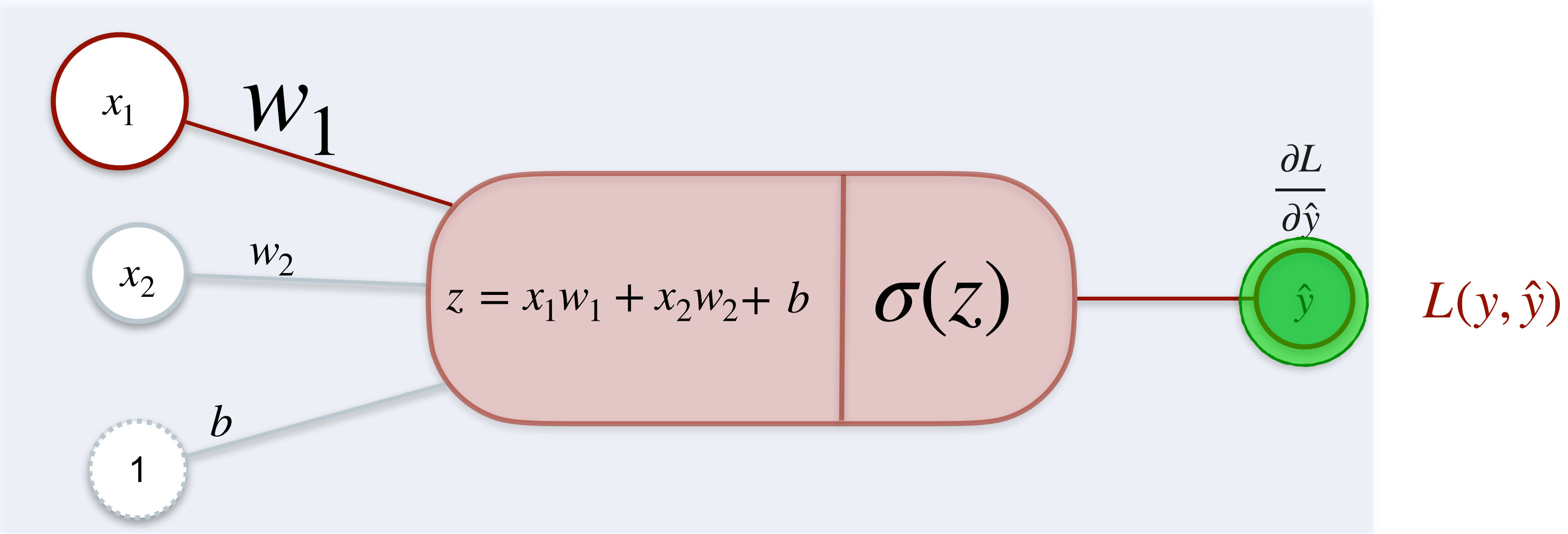
Classification With a Perceptron



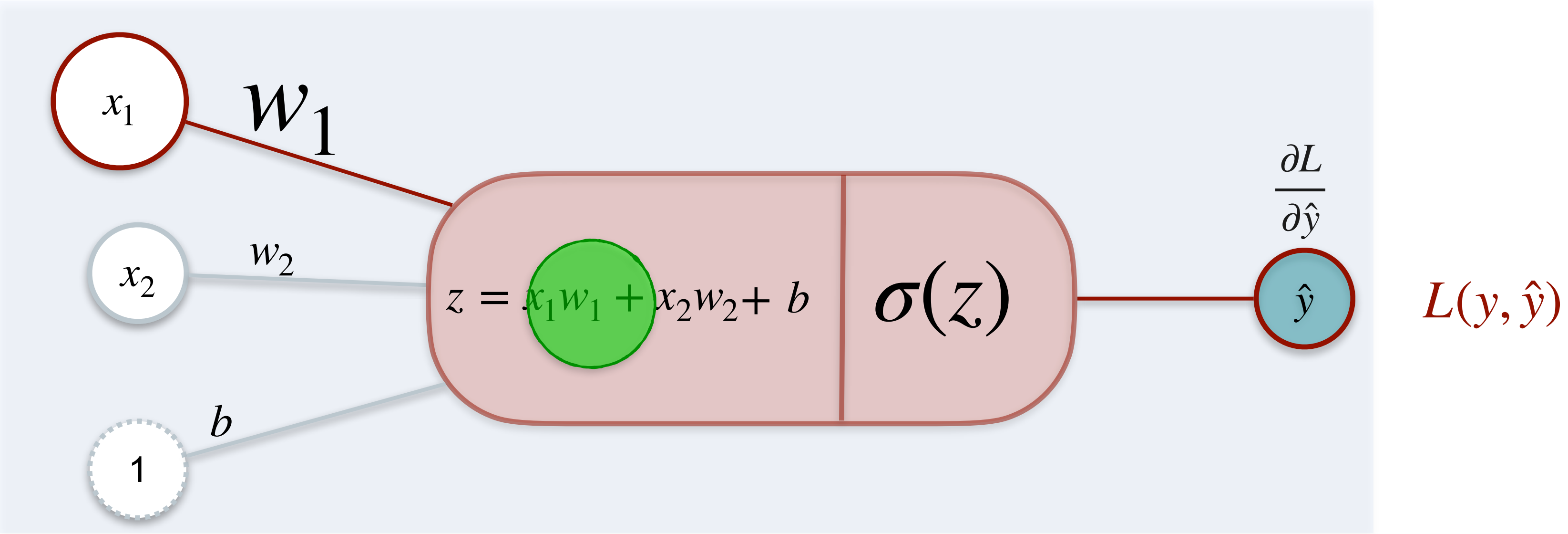
Classification With a Perceptron



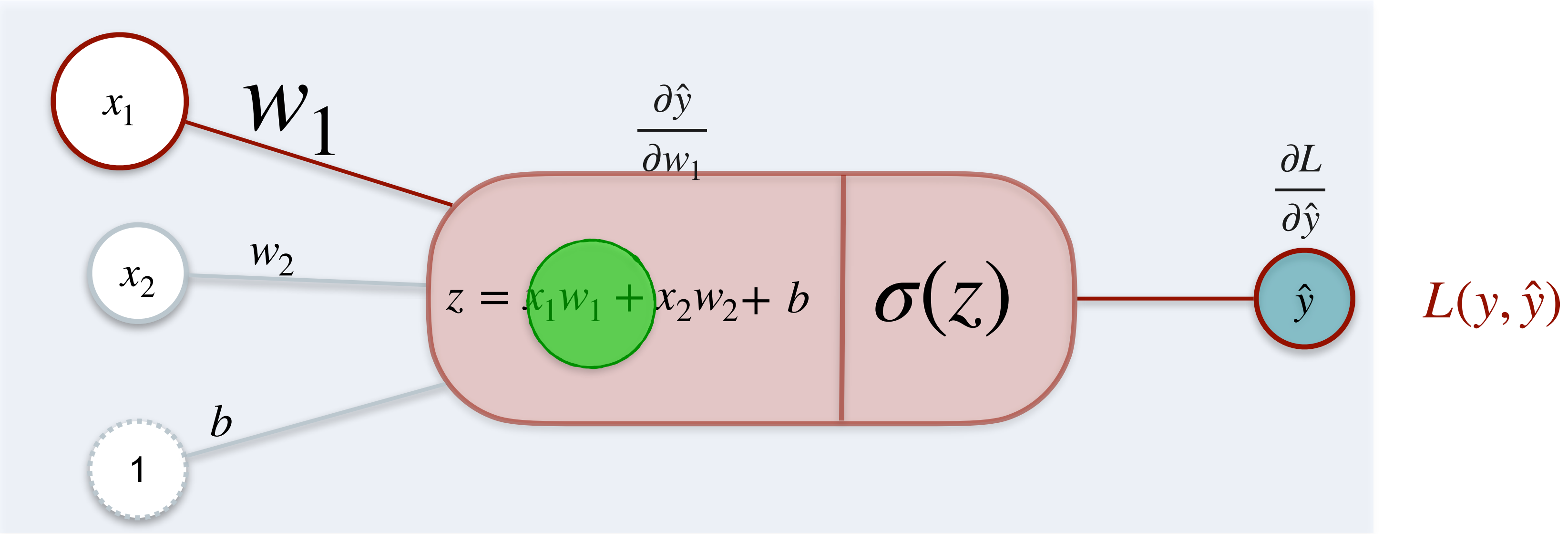
Classification With a Perceptron



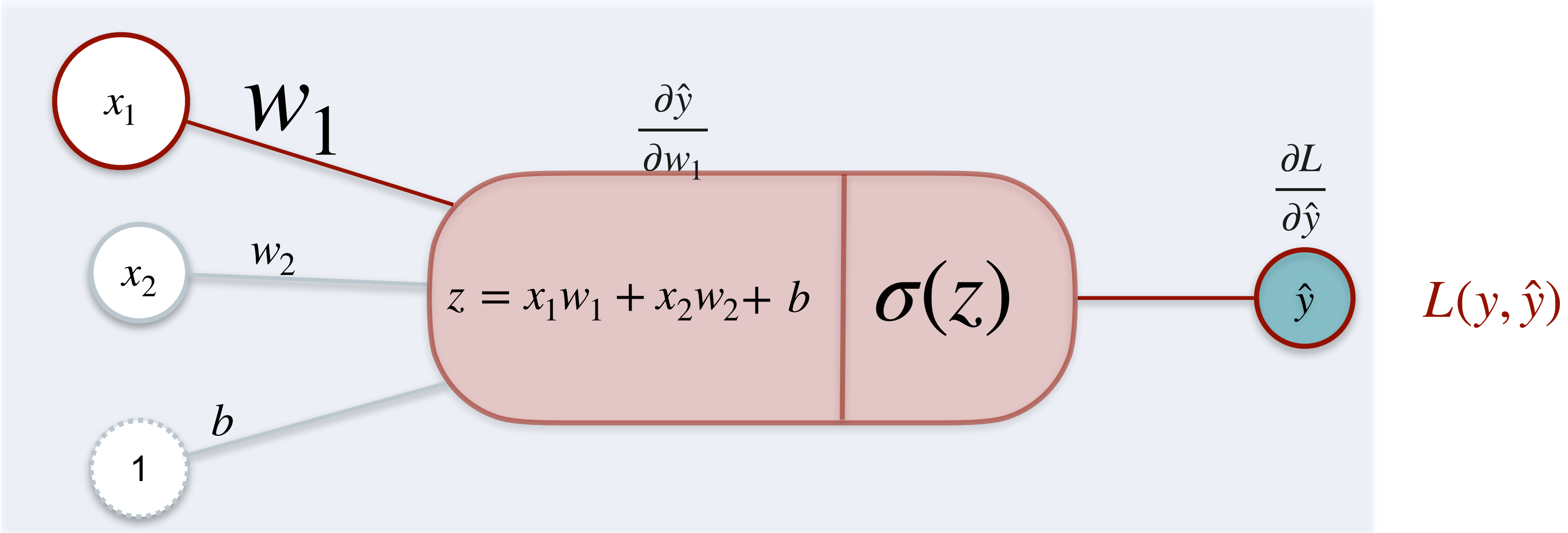
Classification With a Perceptron



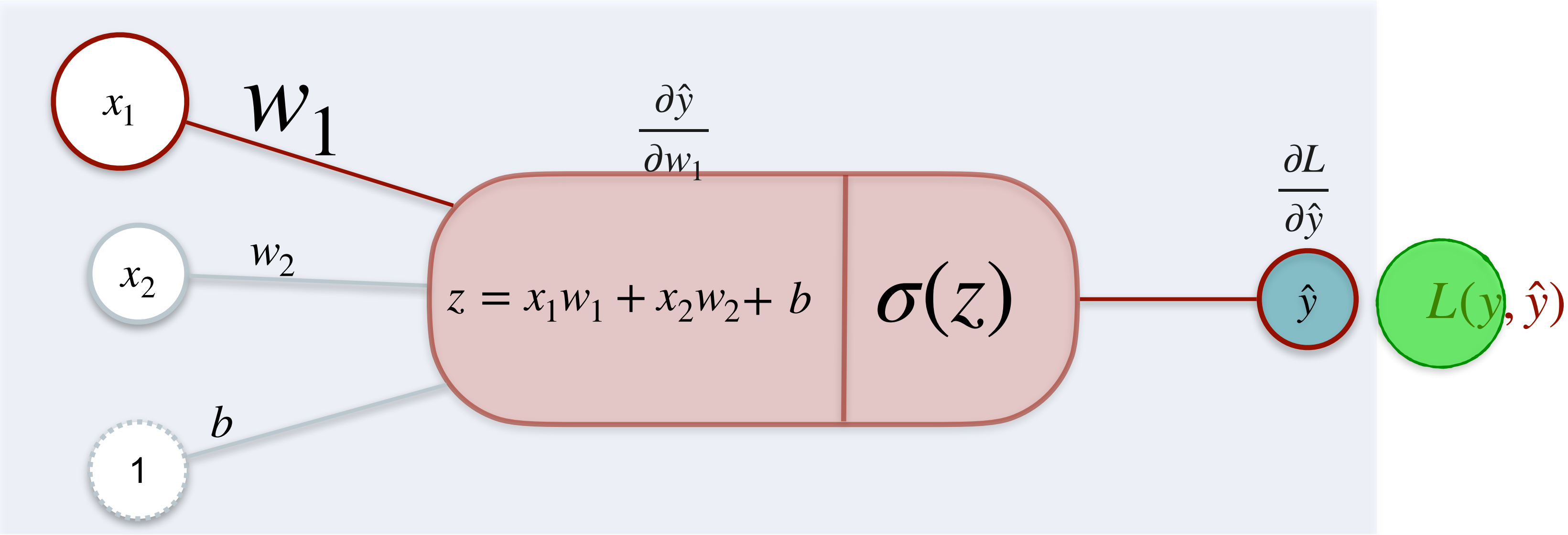
Classification With a Perceptron



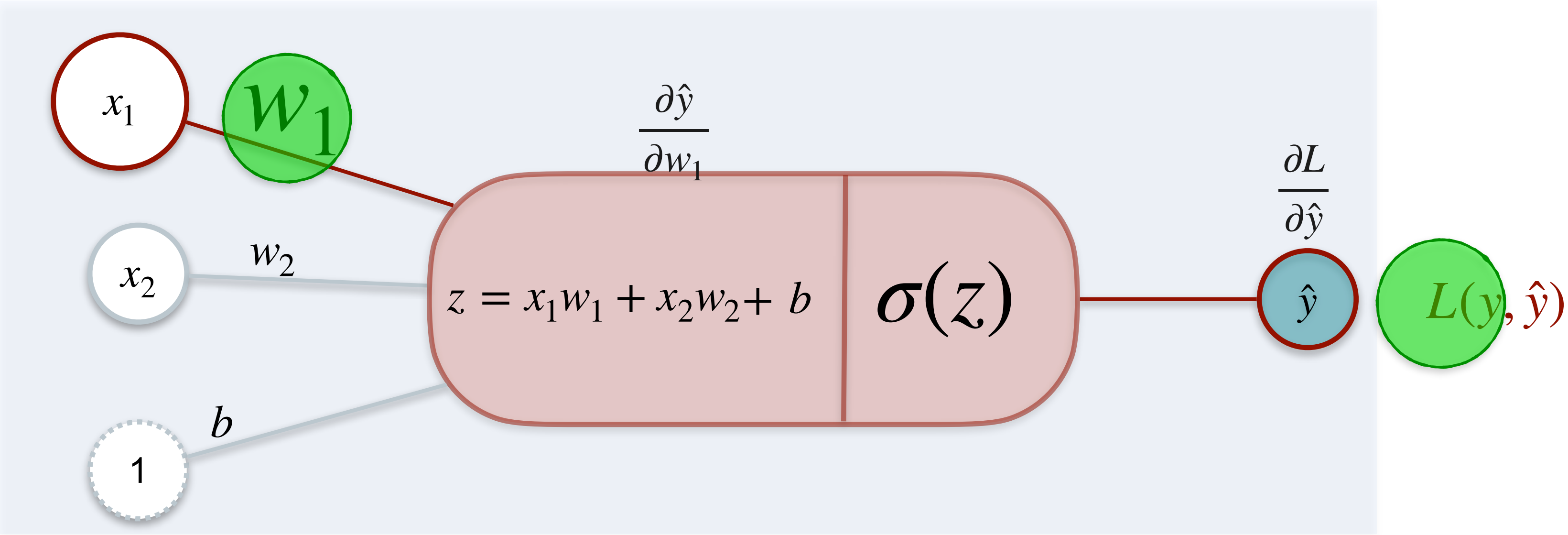
Classification With a Perceptron



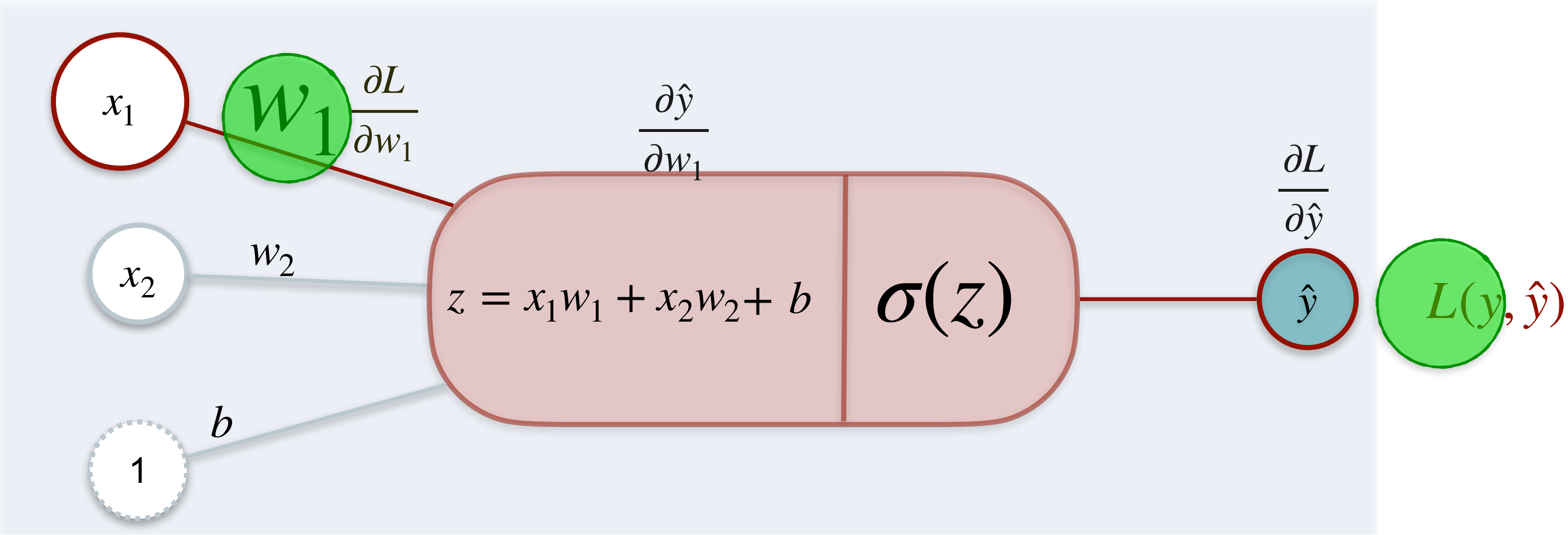
Classification With a Perceptron



Classification With a Perceptron

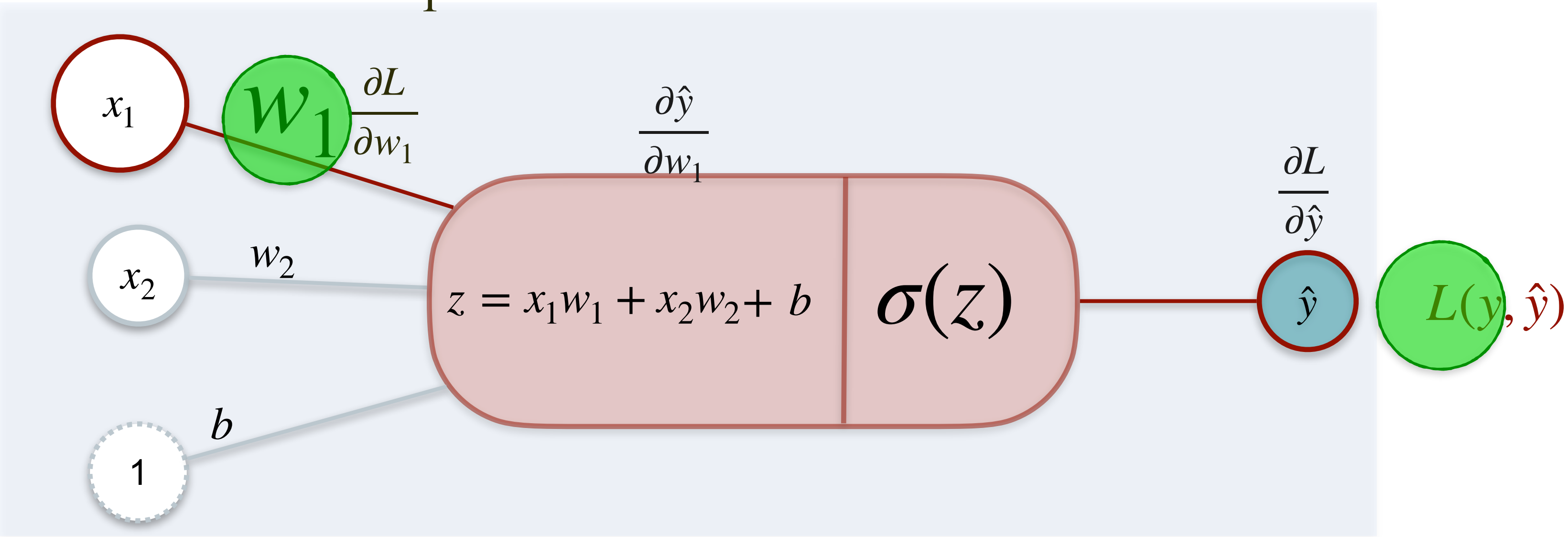


Classification With a Perceptron



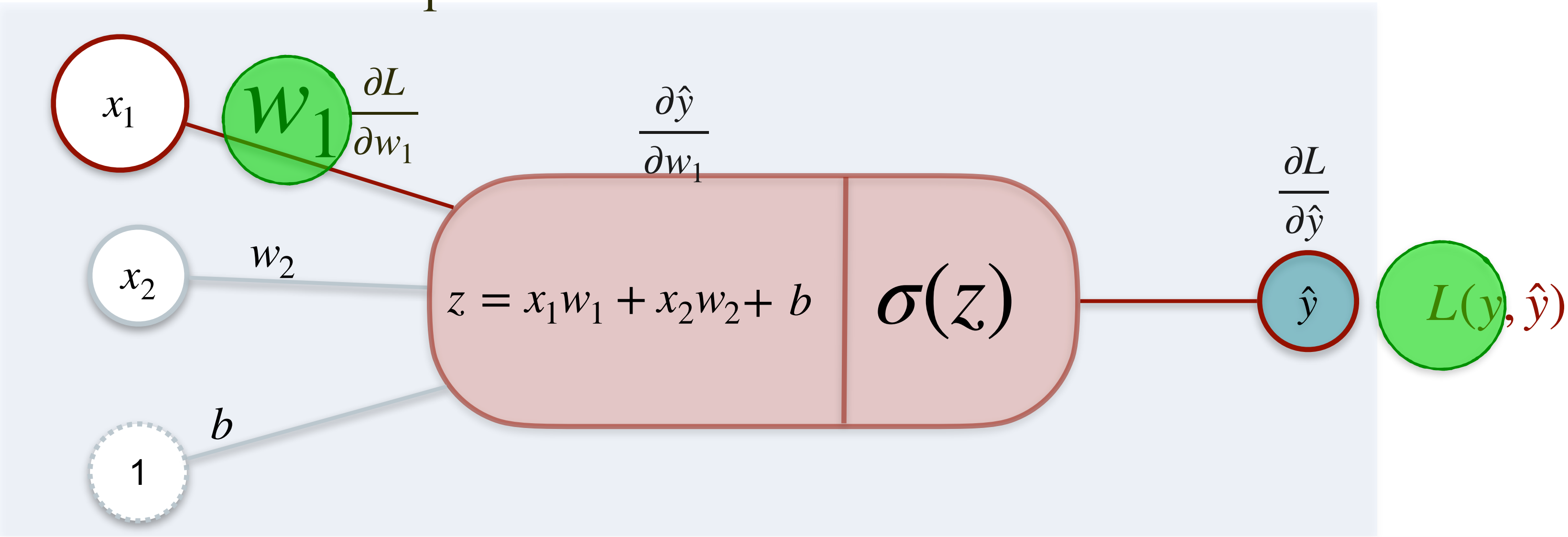
Classification With a Perceptron

$$\frac{\partial L}{\partial w_1}$$



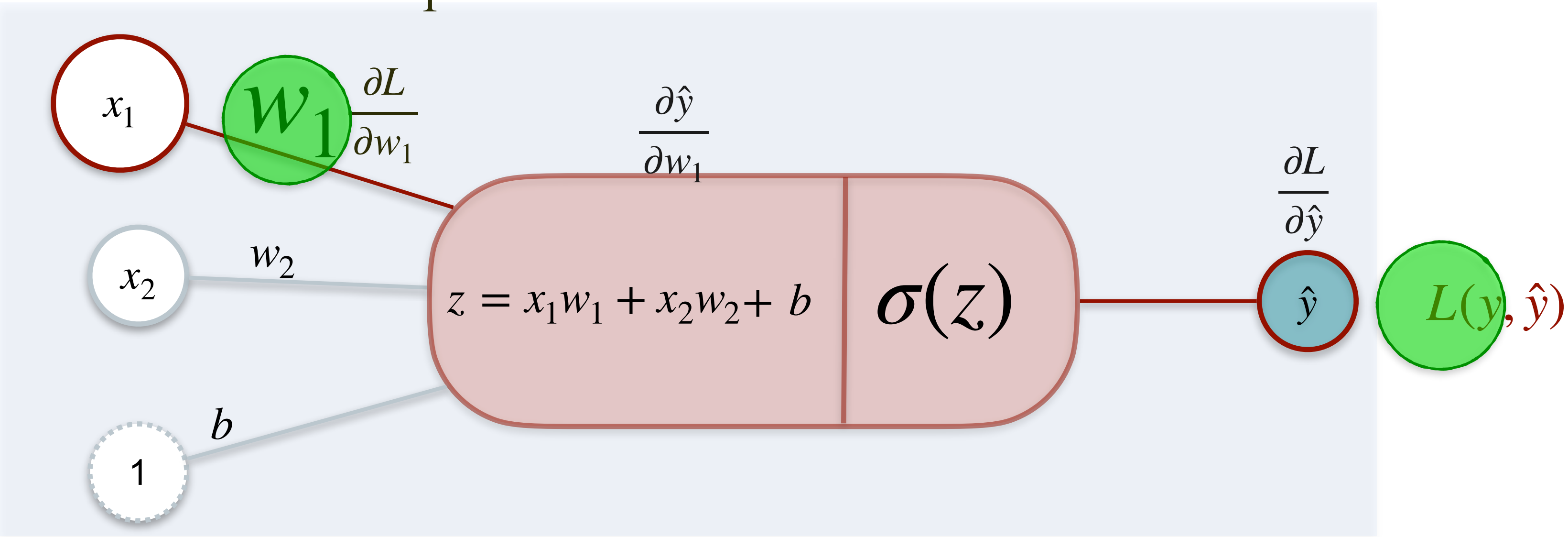
Classification With a Perceptron

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



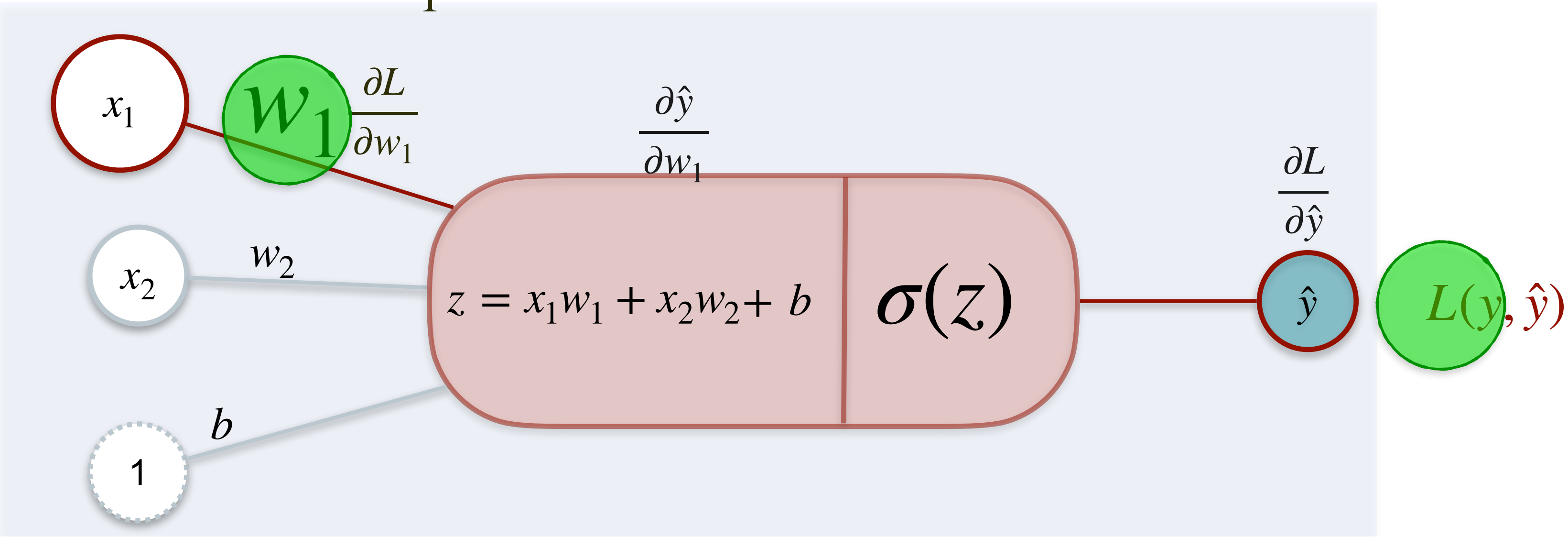
Classification With a Perceptron

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



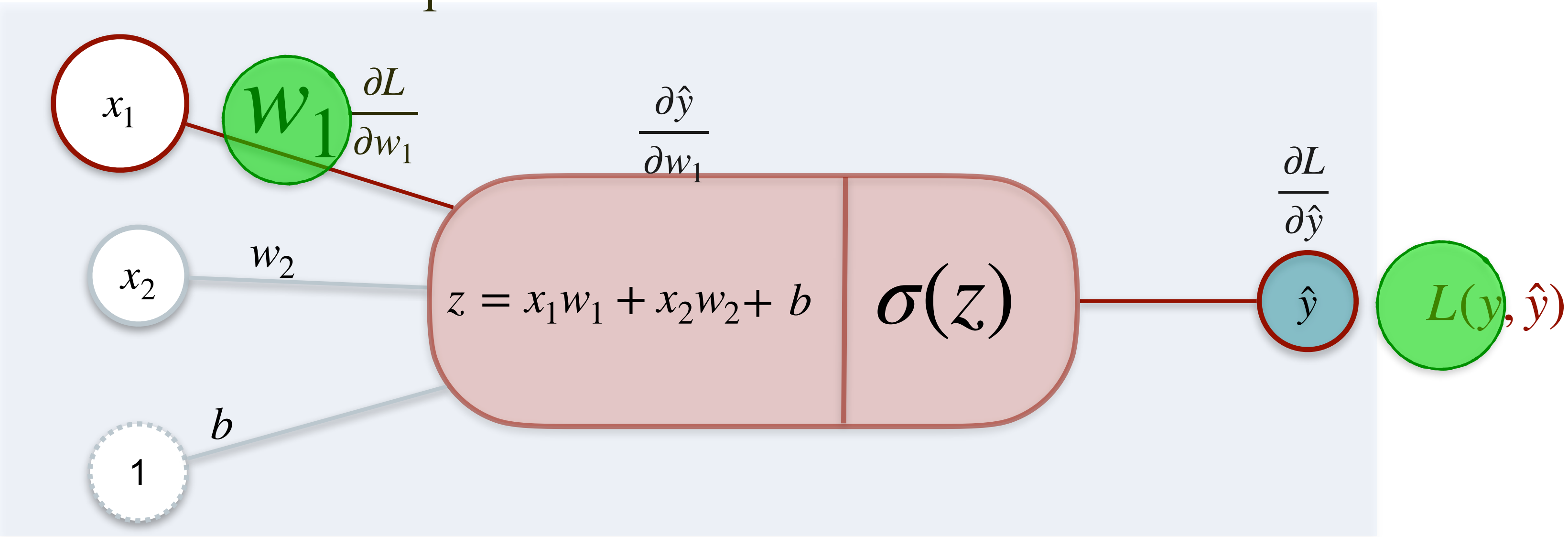
Classification With a Perceptron

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot$$

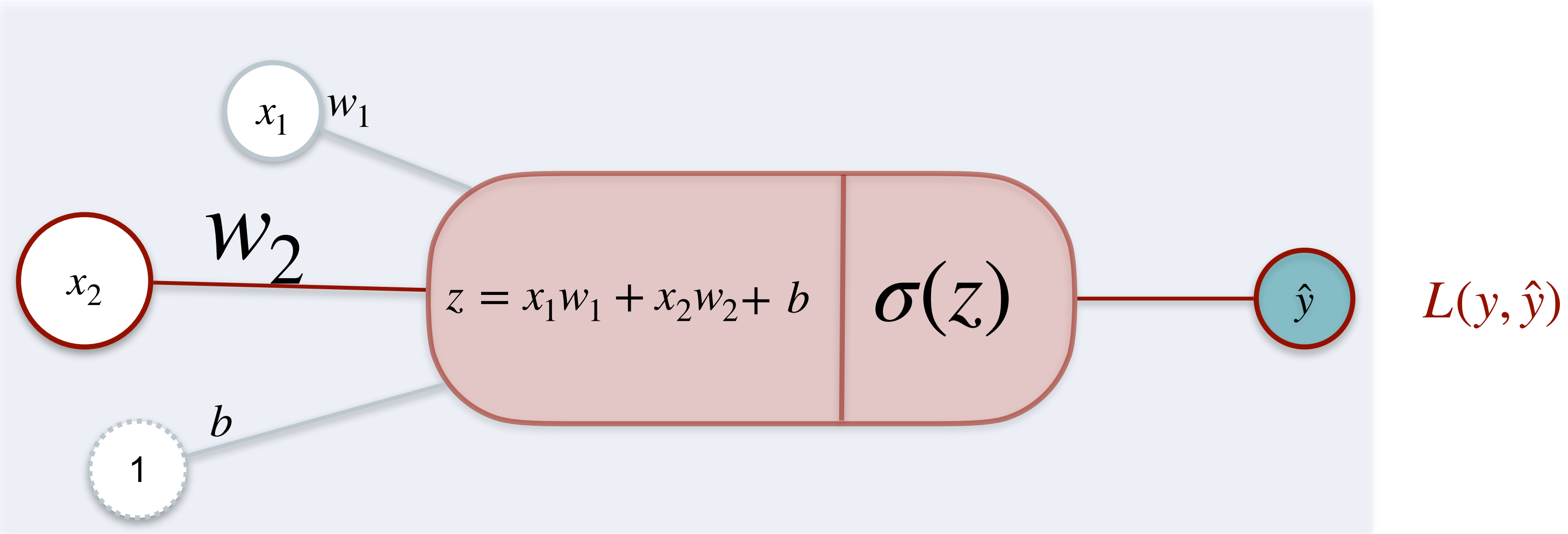


Classification With a Perceptron

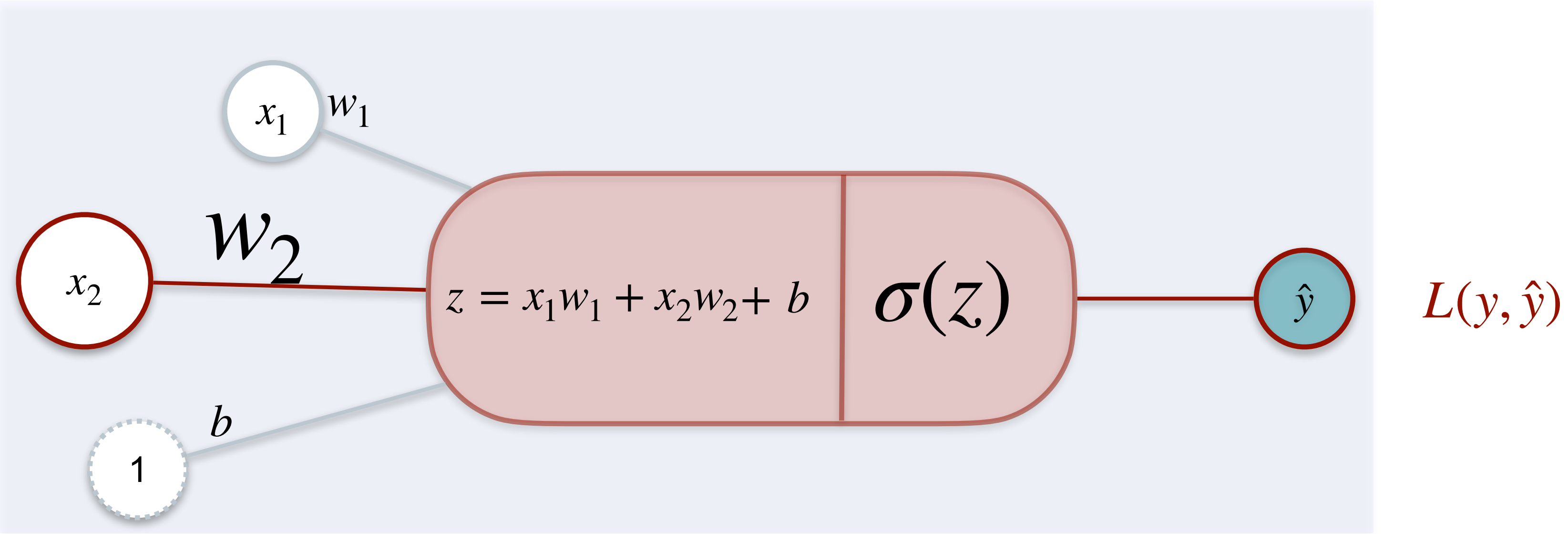
$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$



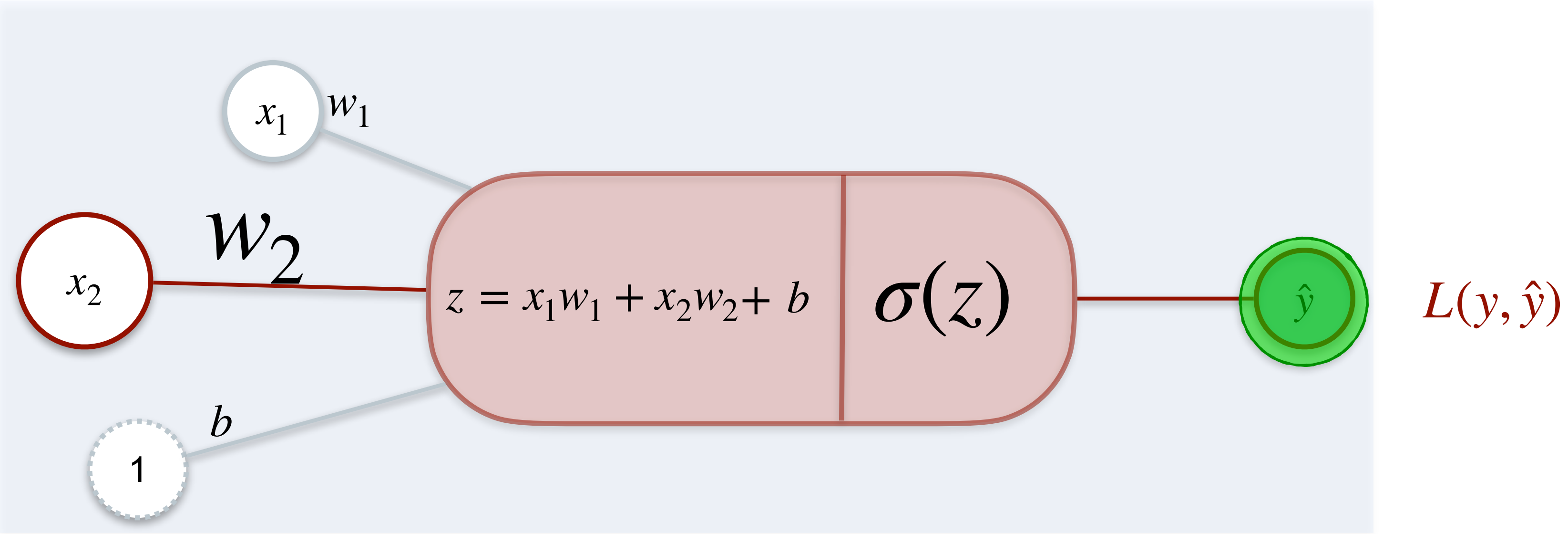
Classification With a Perceptron



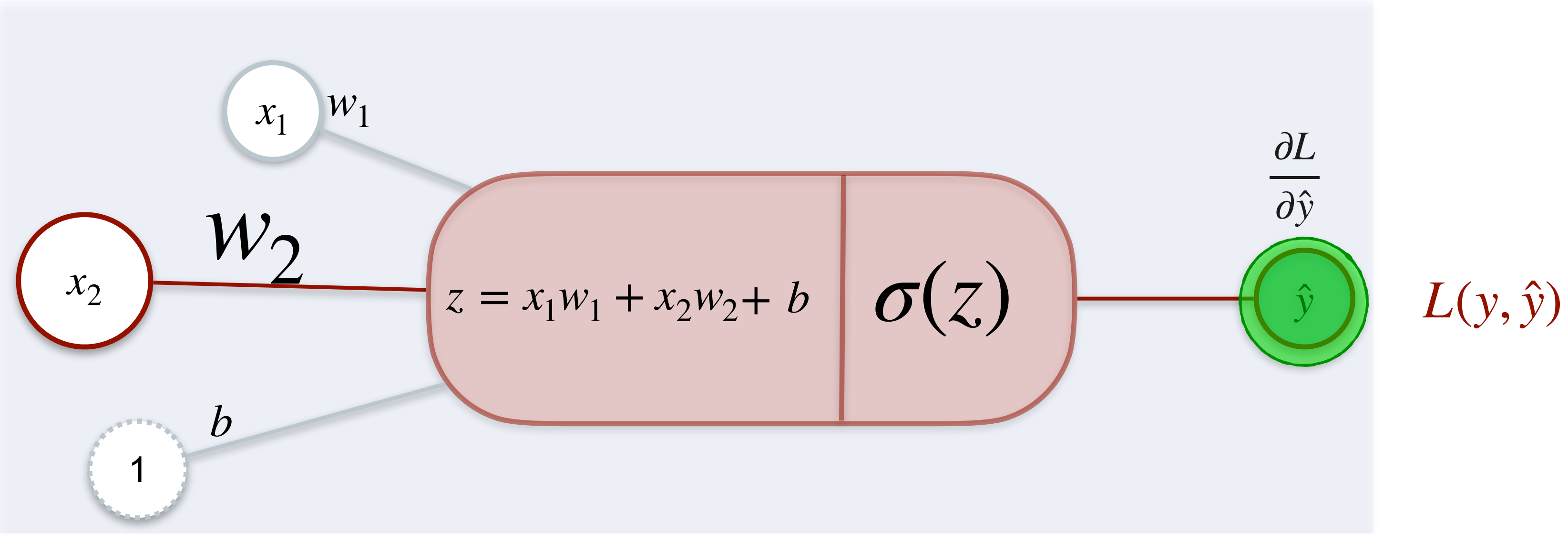
Classification With a Perceptron



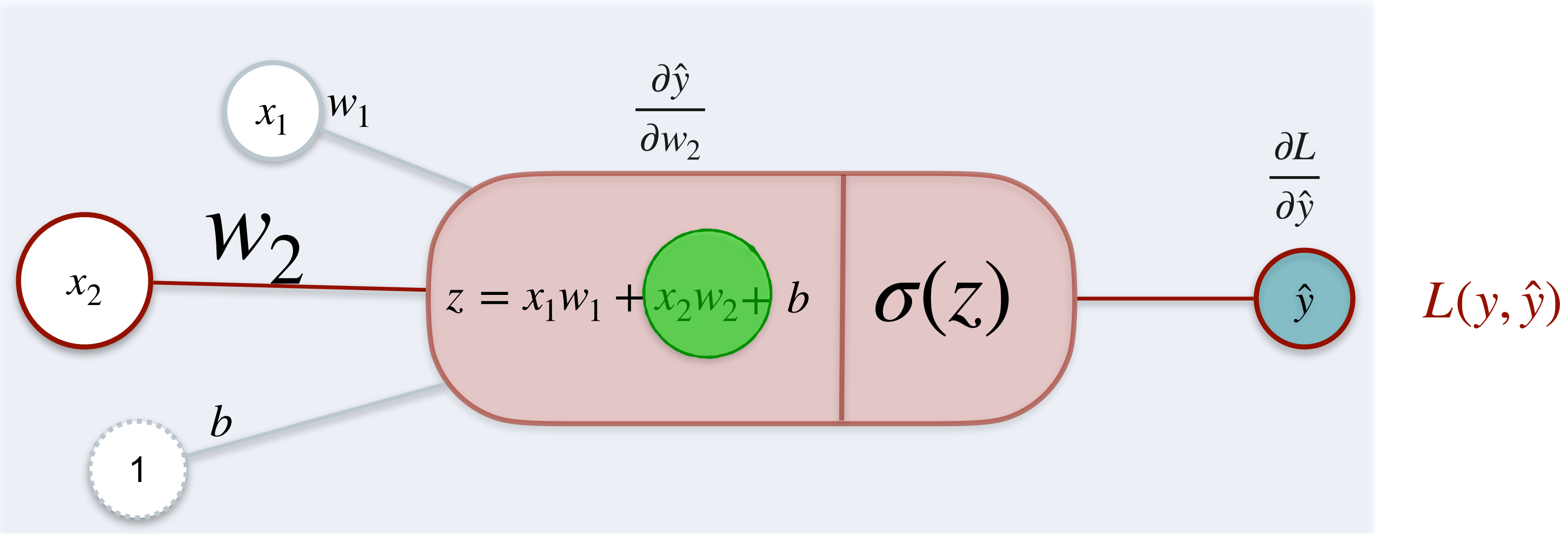
Classification With a Perceptron



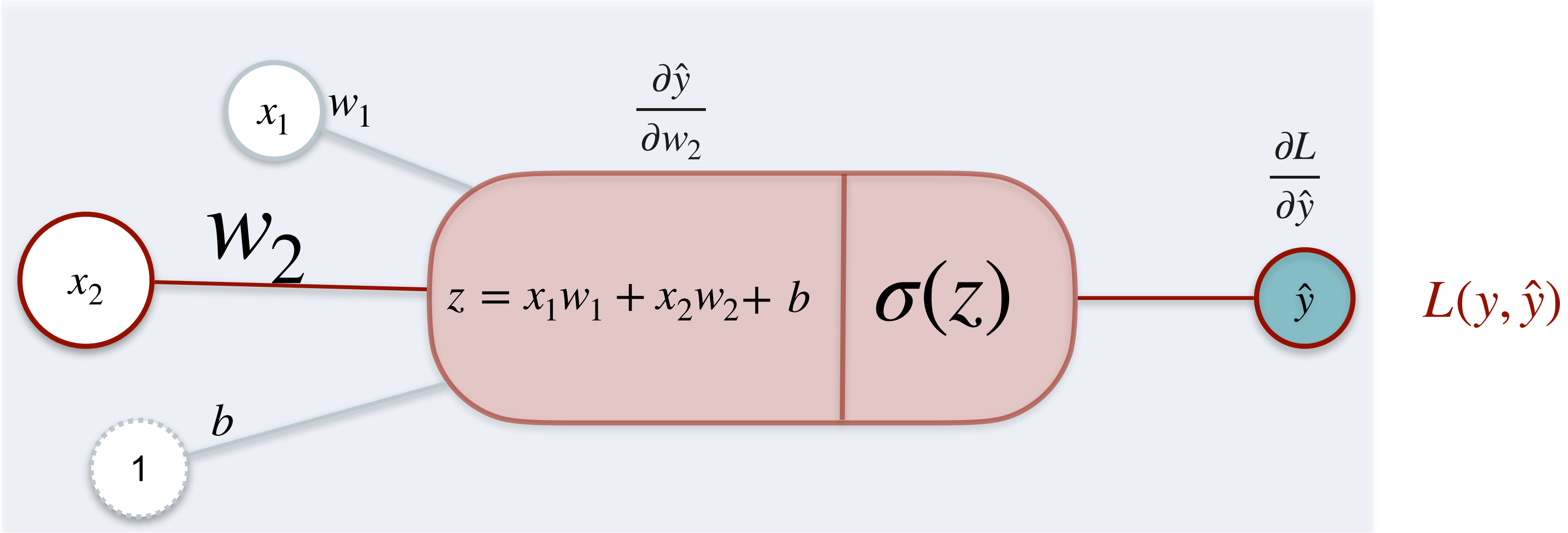
Classification With a Perceptron



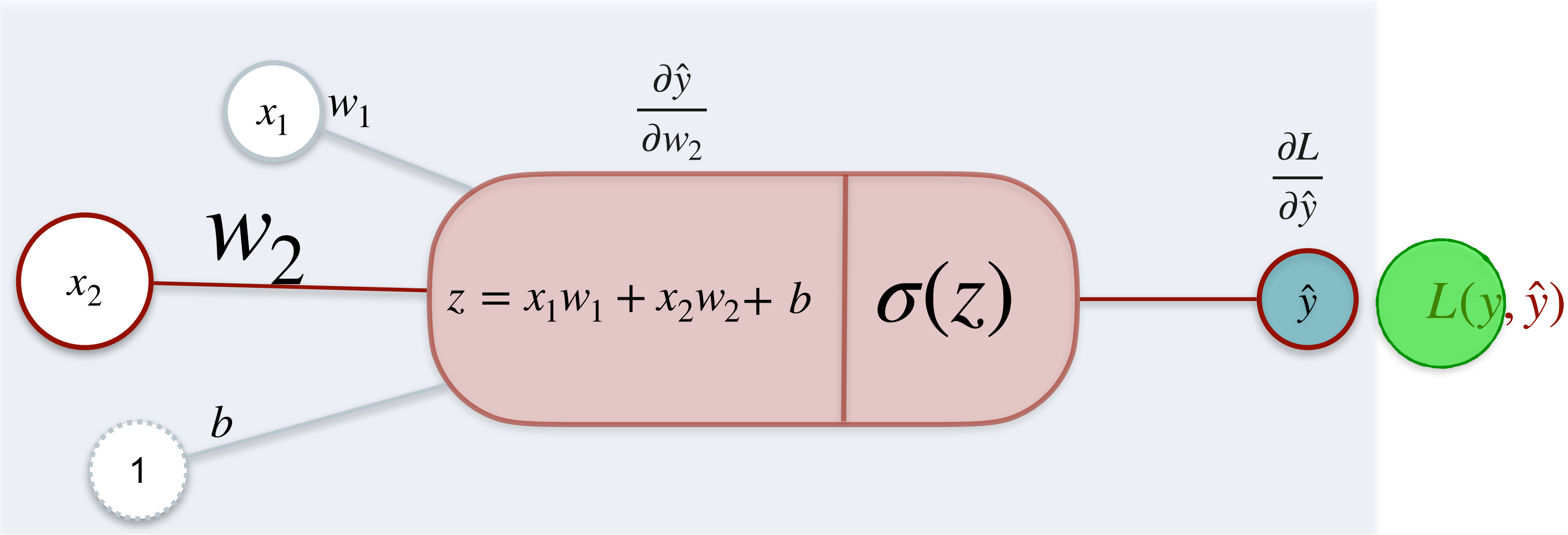
Classification With a Perceptron



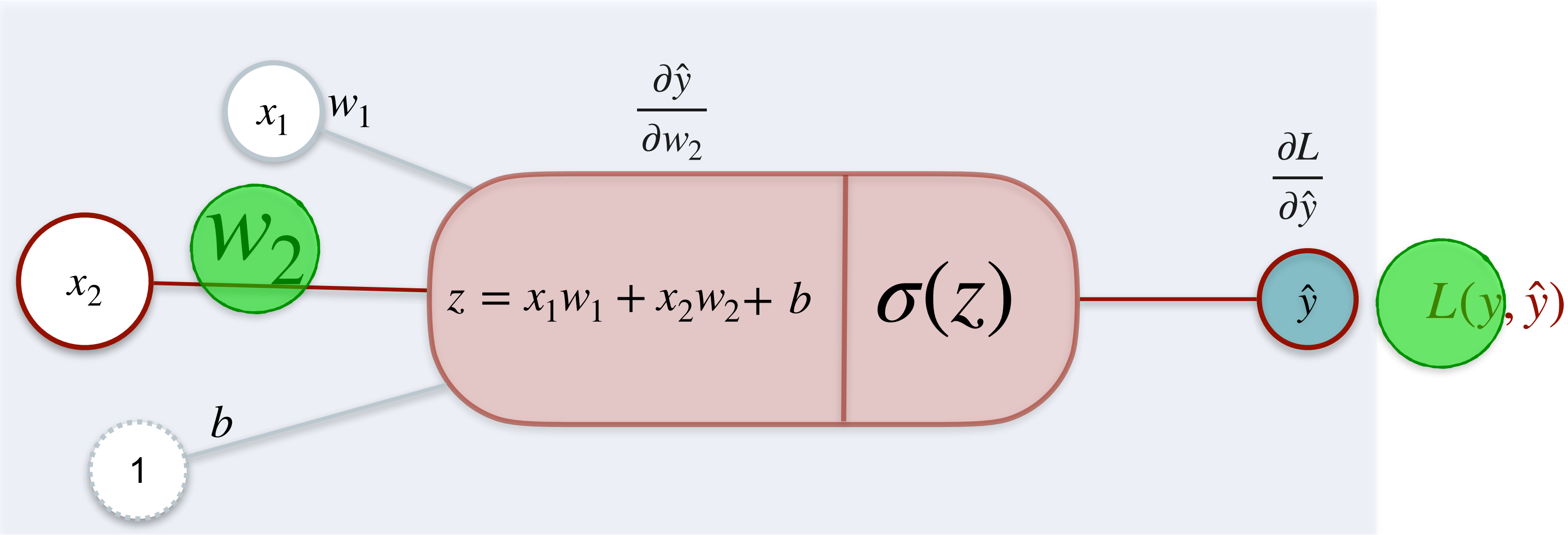
Classification With a Perceptron



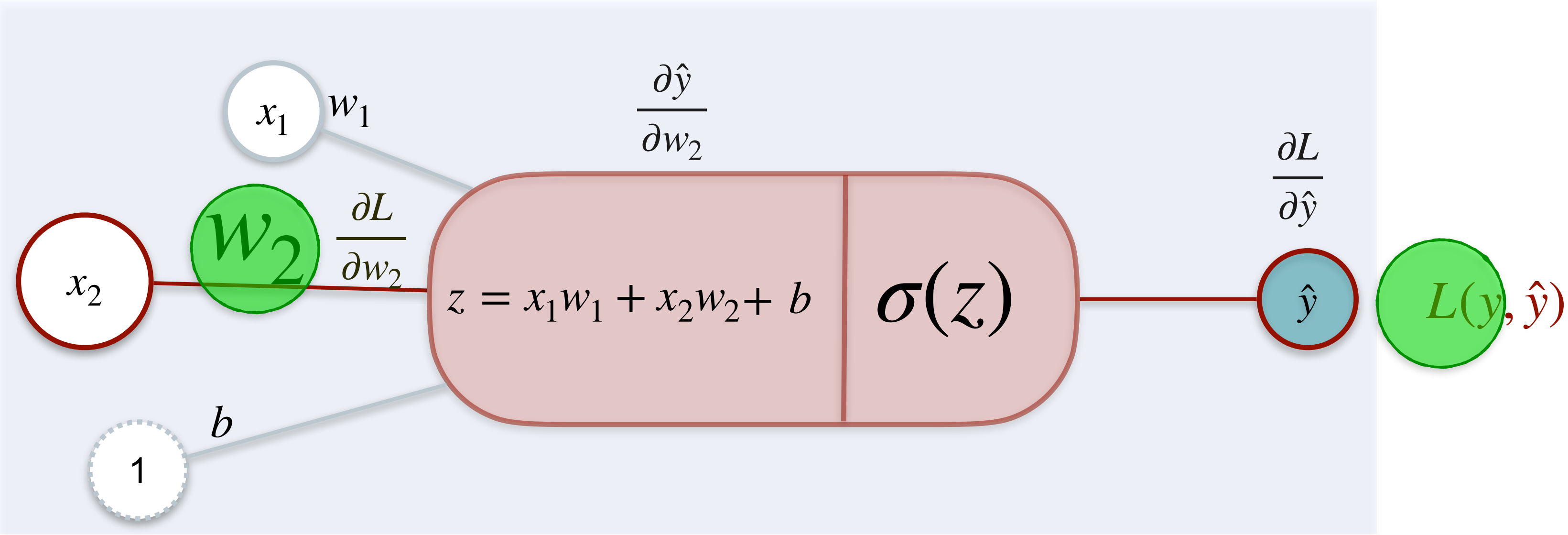
Classification With a Perceptron



Classification With a Perceptron

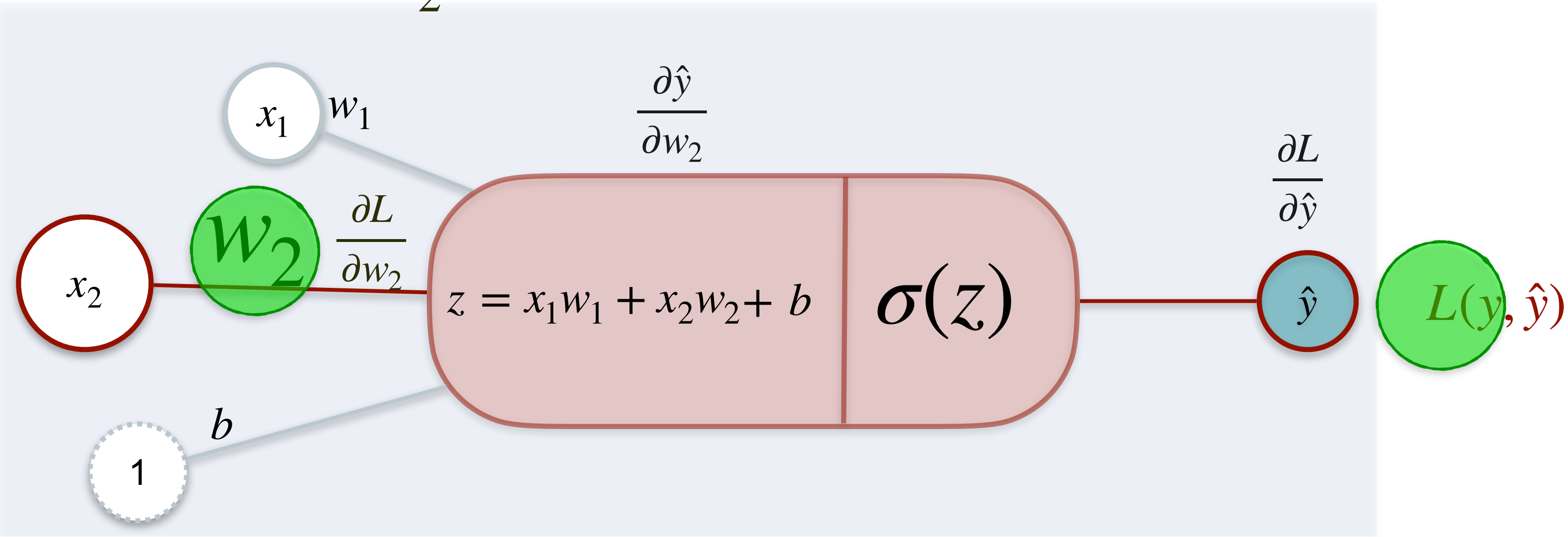


Classification With a Perceptron

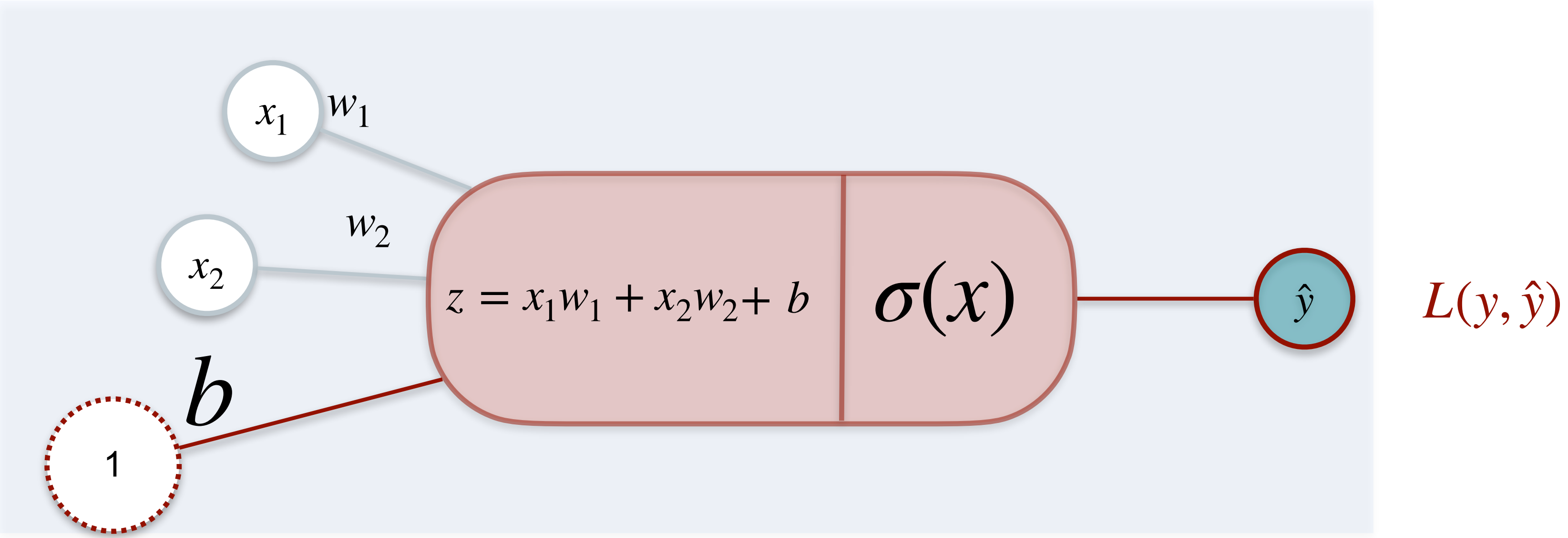


Classification With a Perceptron

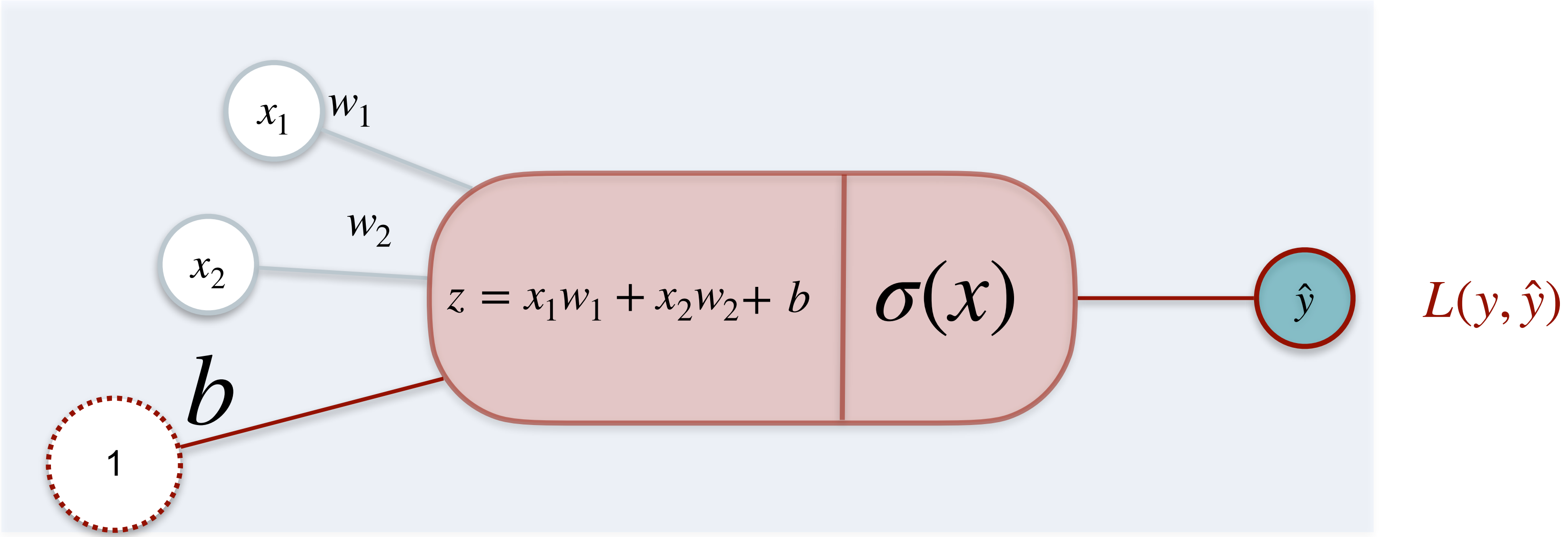
$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$



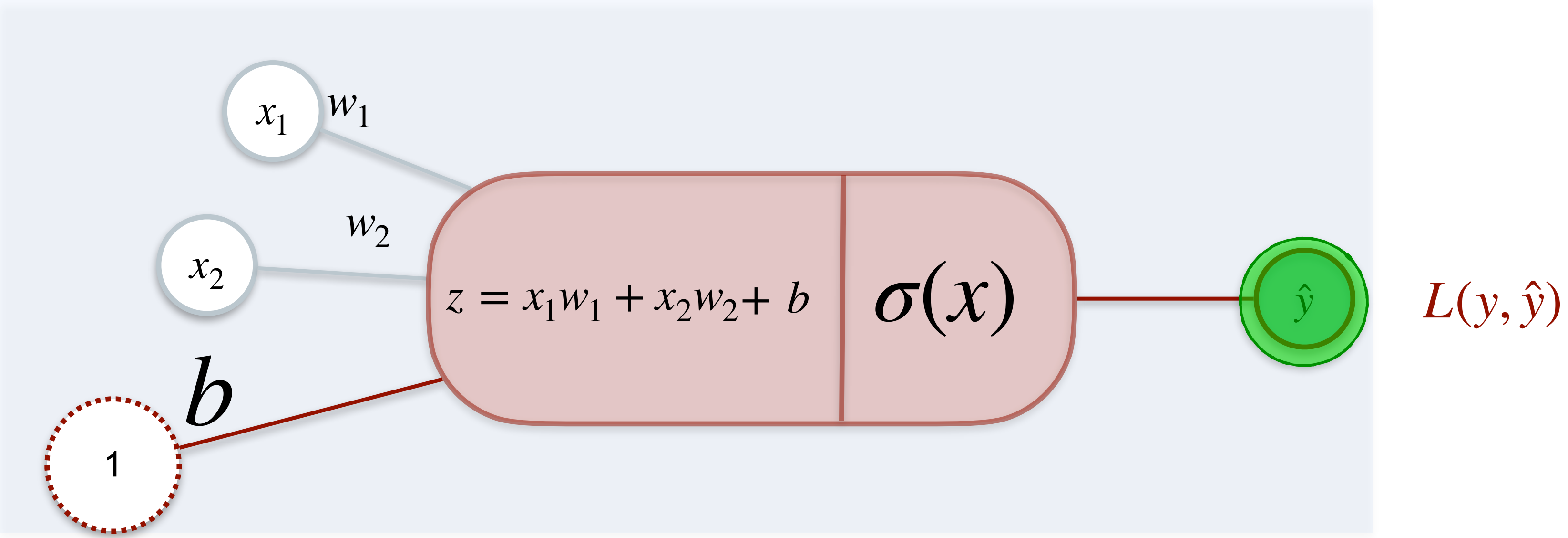
Classification With a Perceptron



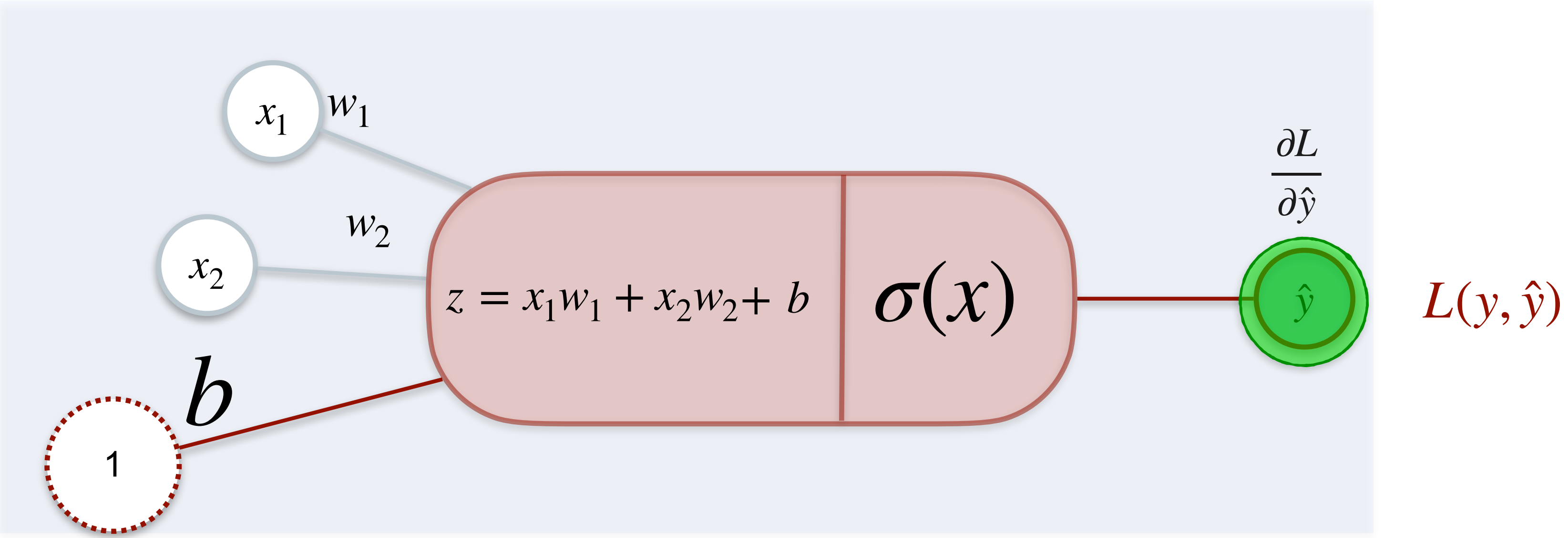
Classification With a Perceptron



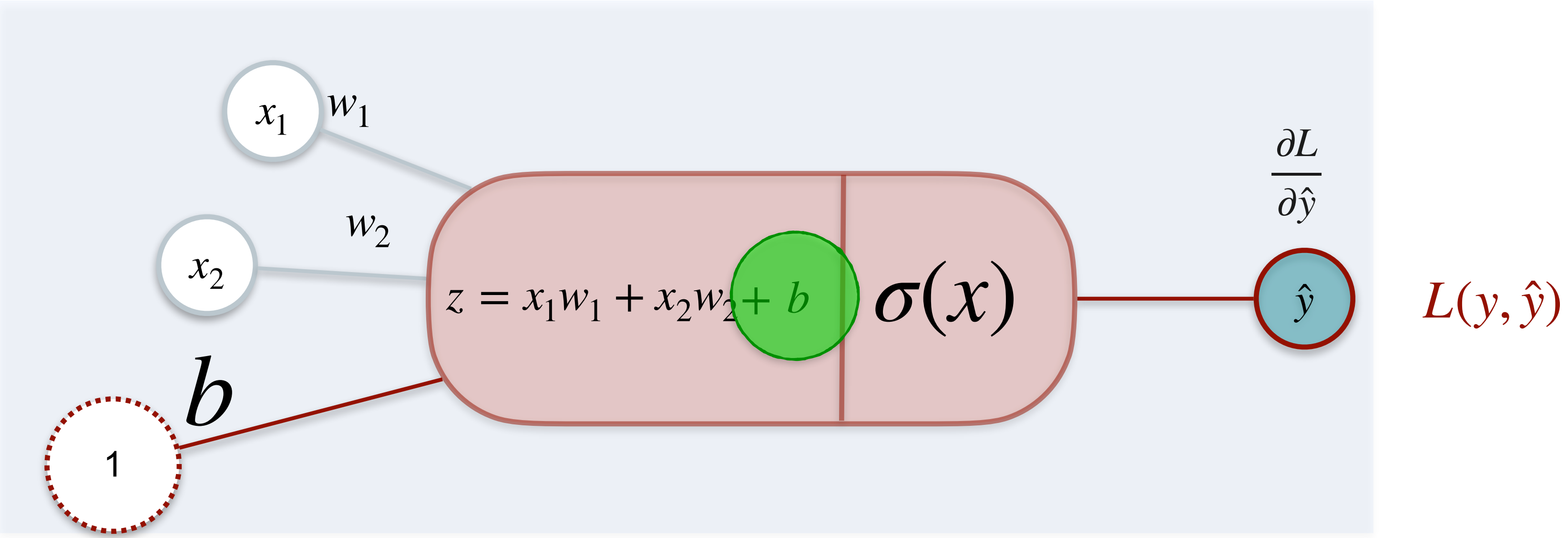
Classification With a Perceptron



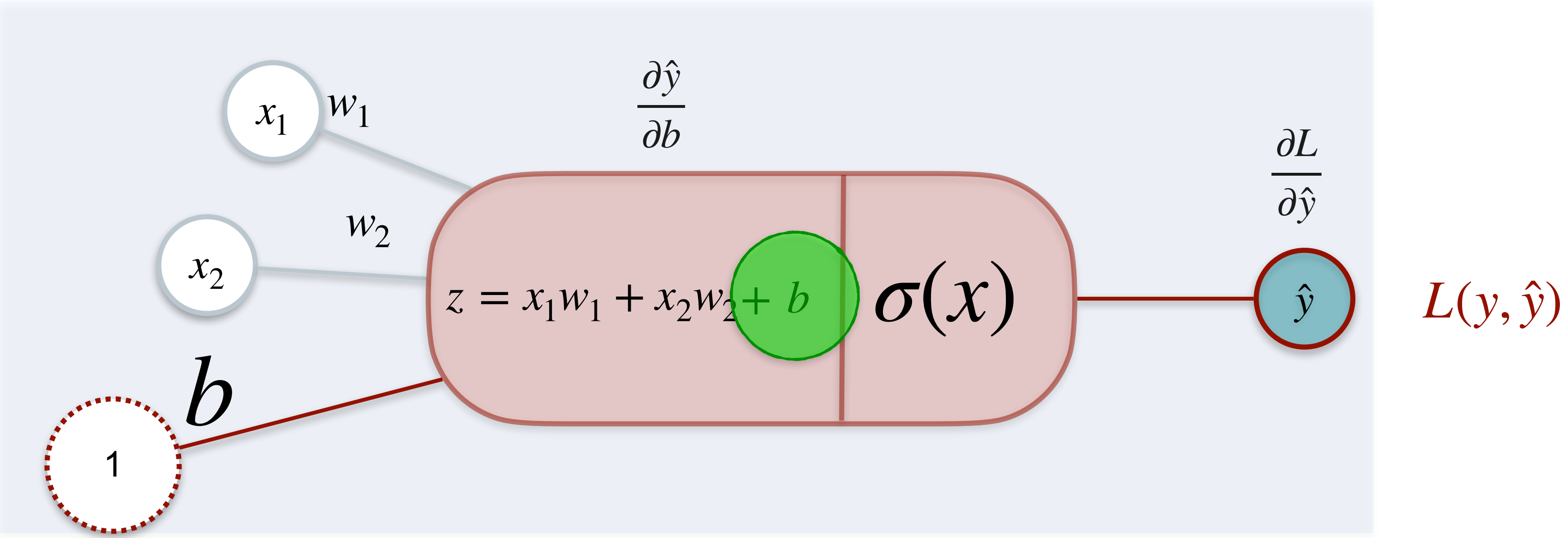
Classification With a Perceptron



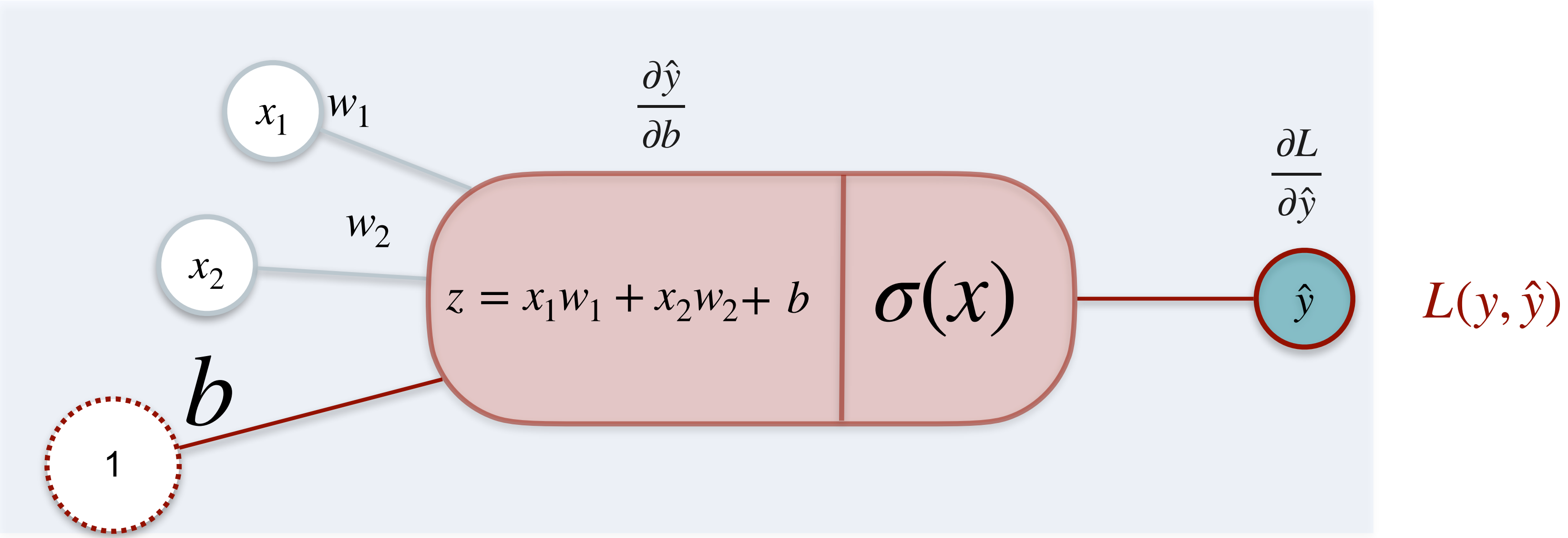
Classification With a Perceptron



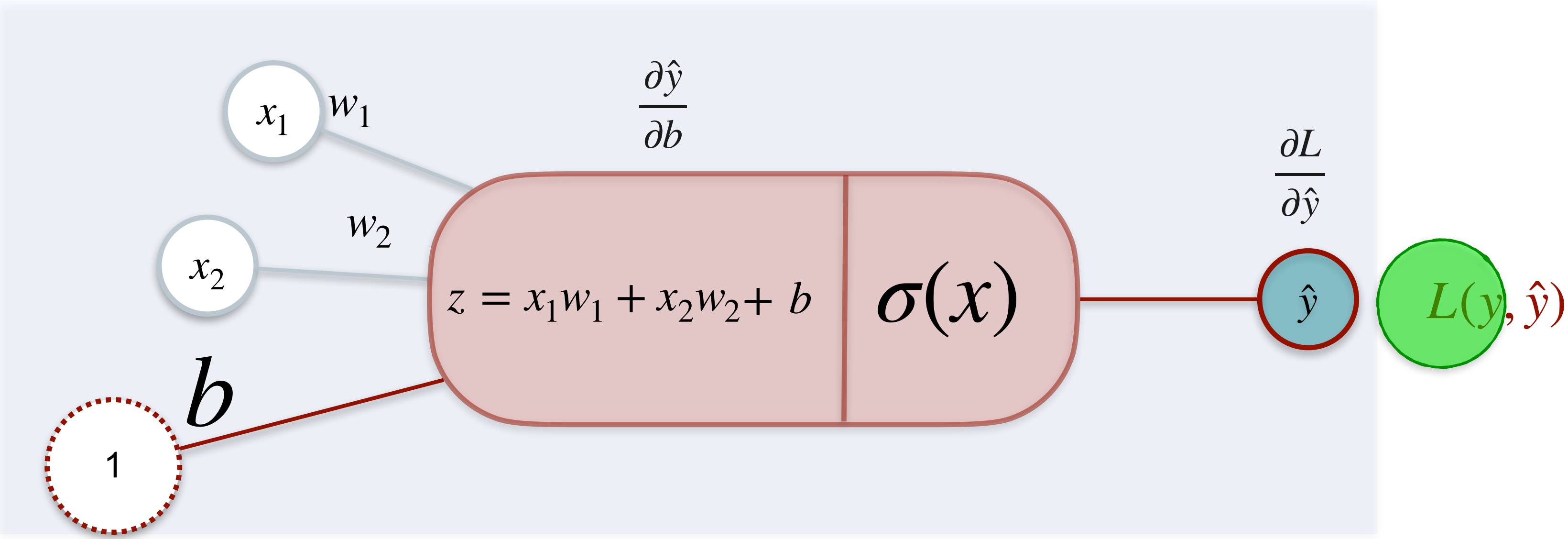
Classification With a Perceptron



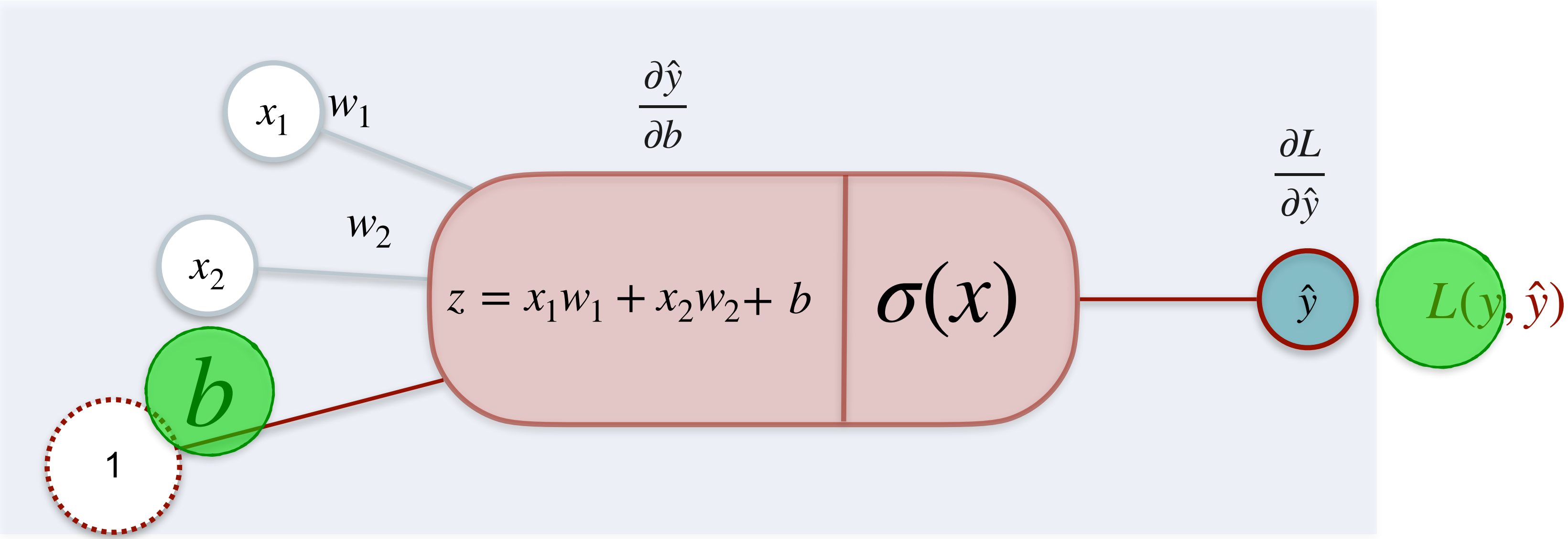
Classification With a Perceptron



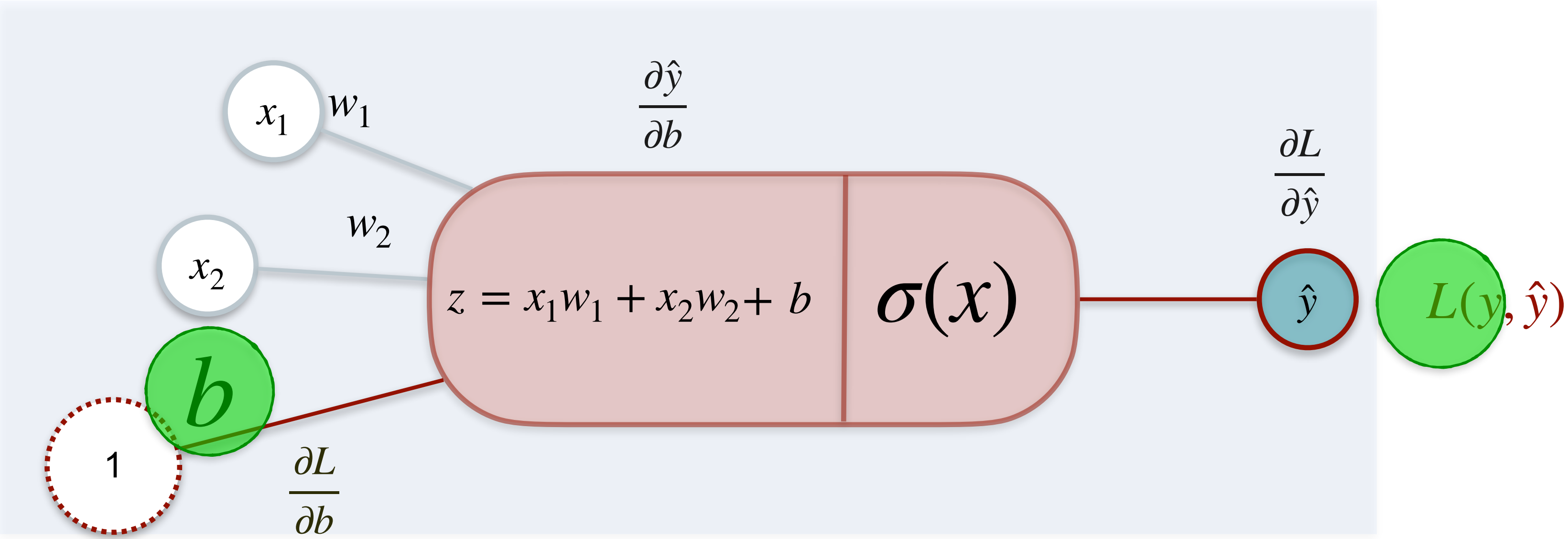
Classification With a Perceptron



Classification With a Perceptron

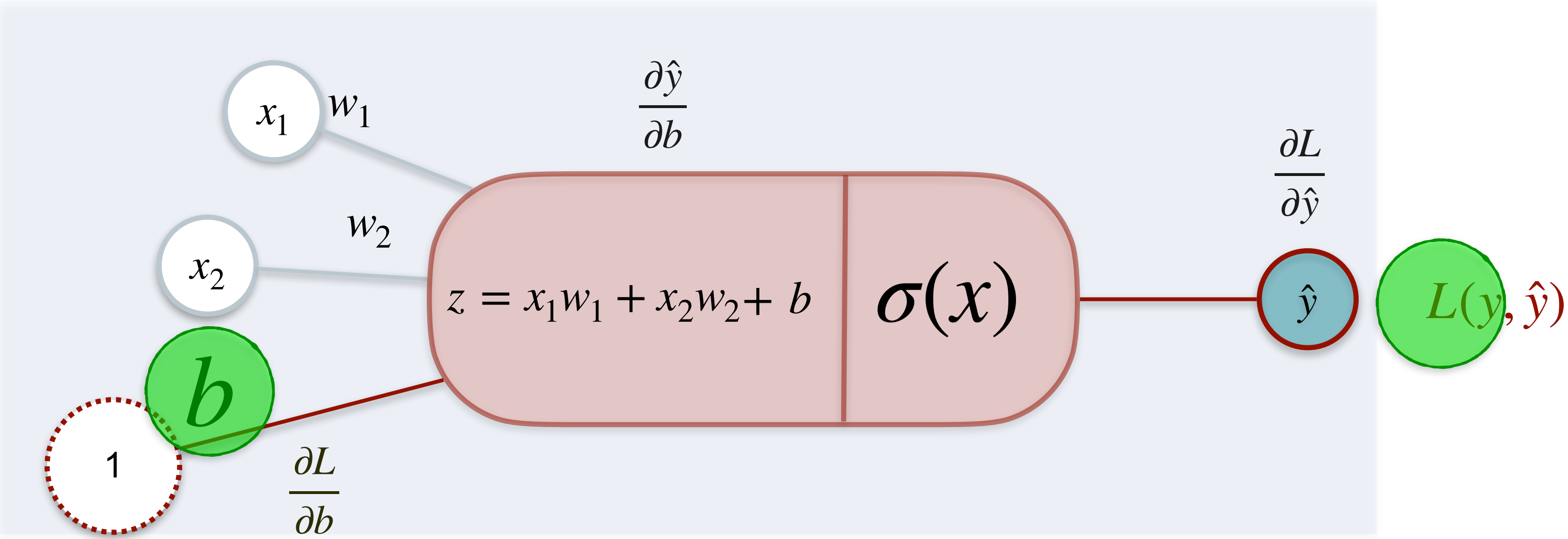


Classification With a Perceptron



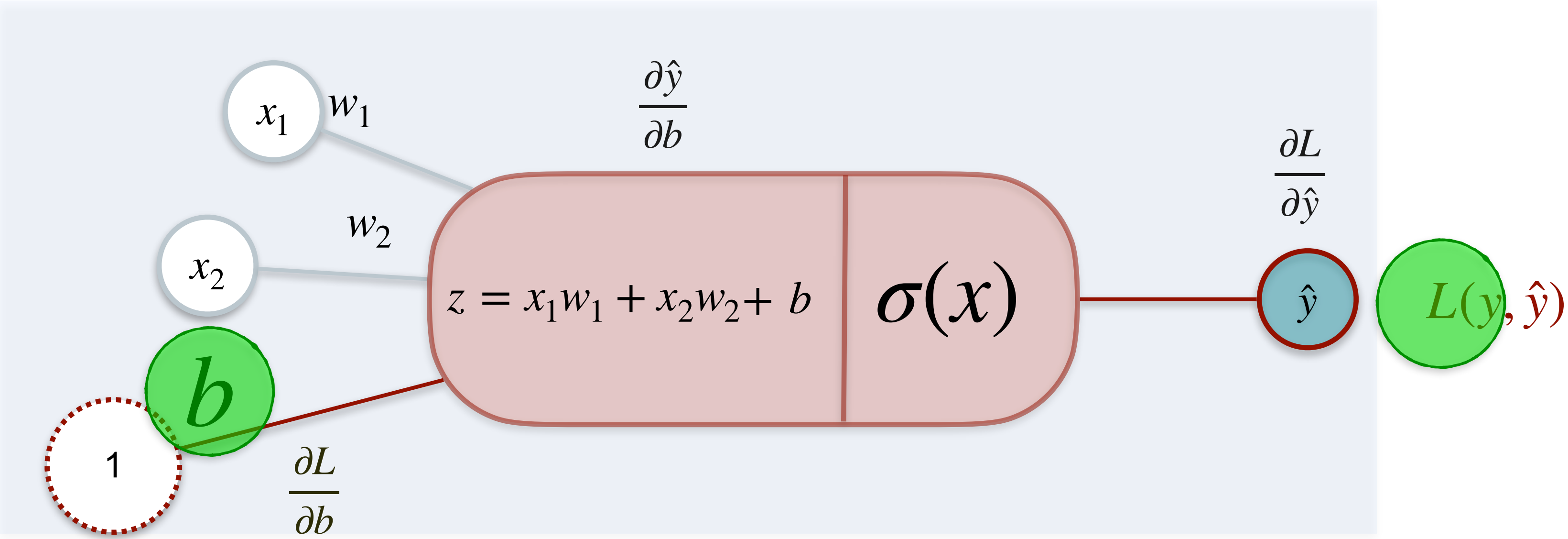
Classification With a Perceptron

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



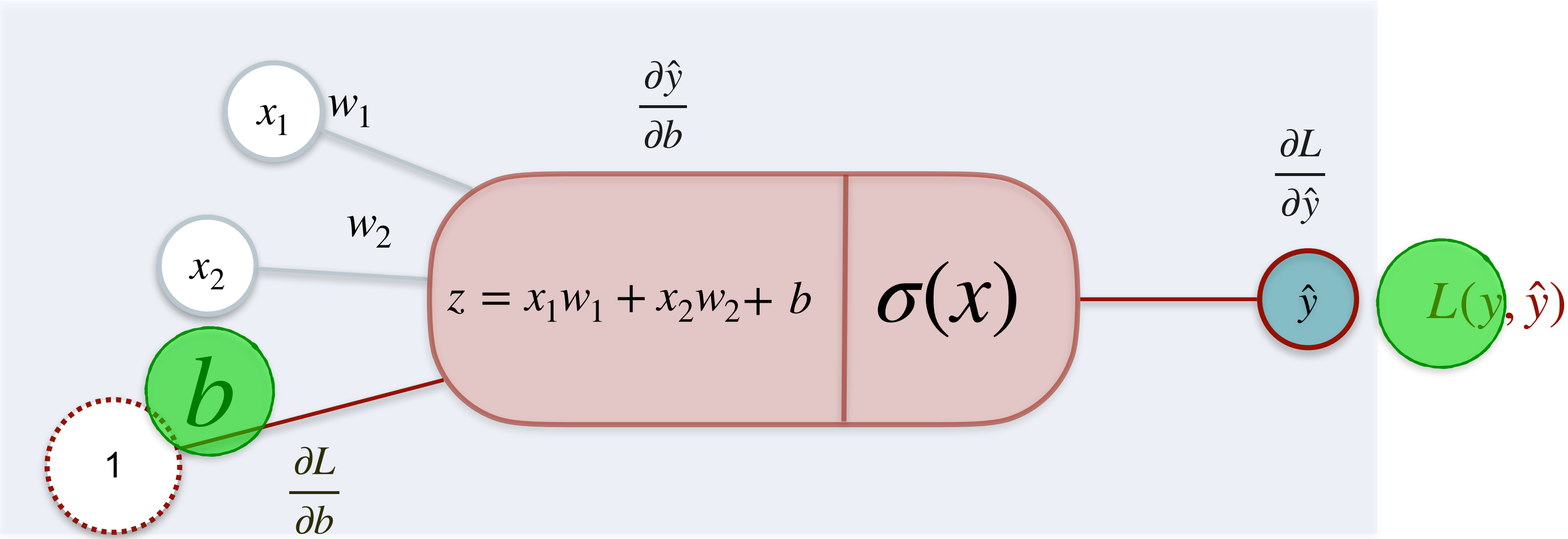
Classification With a Perceptron

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



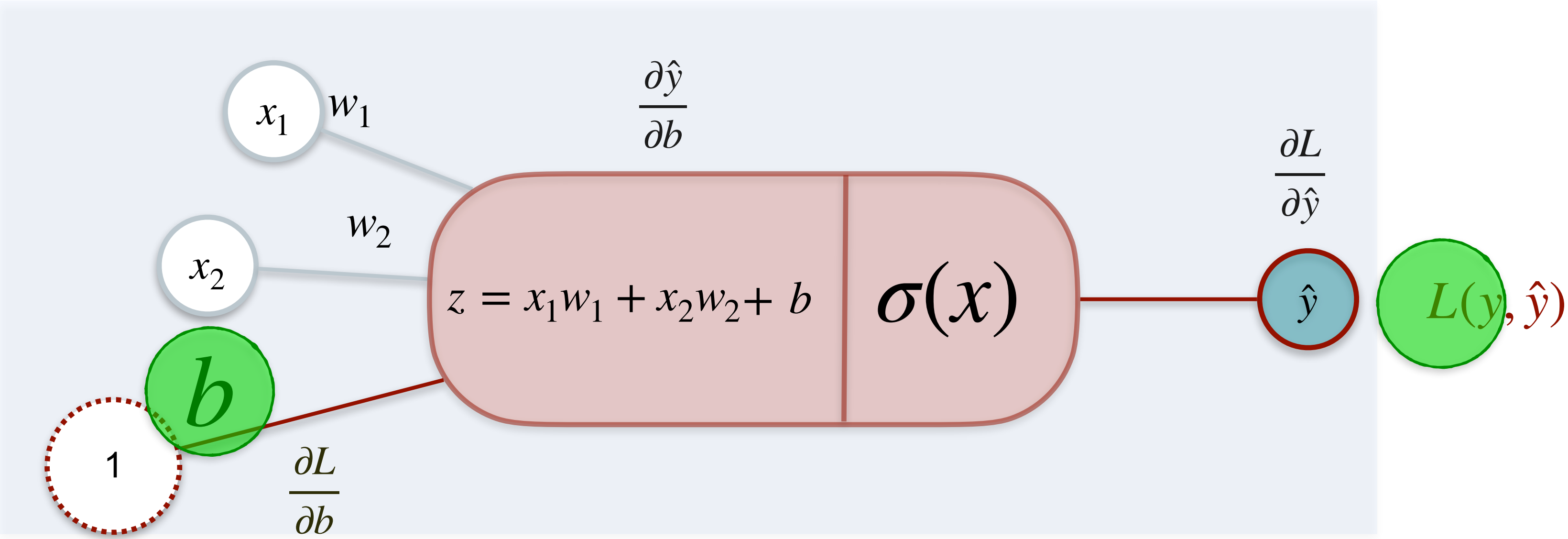
Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot$$



Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$



Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

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

$$\frac{\partial \hat{y}}{\partial b} =$$

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

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

Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

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

$$\frac{\partial \hat{y}}{\partial b} =$$

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

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

?

Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

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

$$\frac{\partial \hat{y}}{\partial b} =$$

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

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

?

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

Classification With a Perceptron

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

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

$$\frac{\partial \hat{y}}{\partial b} =$$

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

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

?

$$\hat{y} = \sigma(w_1x_1 + w_2x_2 + b)$$

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

Classification With a Perceptron

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

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

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

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

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

Classification With a Perceptron

$$\hat{y} = \sigma(w_1x_1 + w_2x_2 + b)$$

Classification With a Perceptron

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

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

Classification With a Perceptron

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

$\frac{\partial \hat{y}}{\partial w_1}$	$= \hat{y}(1 - \hat{y})x_1$

Classification With a Perceptron

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

$\frac{\partial \hat{y}}{\partial w_1}$	$= \hat{y}(1 - \hat{y})x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	

Classification With a Perceptron

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

$\frac{\partial \hat{y}}{\partial w_1}$	$= \hat{y}(1 - \hat{y})x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	$= \hat{y}(1 - \hat{y})x_2$

Classification With a Perceptron

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

$\frac{\partial \hat{y}}{\partial w_1}$	$= \hat{y}(1 - \hat{y})x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	$= \hat{y}(1 - \hat{y})x_2$
$\frac{\partial \hat{y}}{\partial b}$	

Classification With a Perceptron

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

$\frac{\partial \hat{y}}{\partial w_1}$	$= \hat{y}(1 - \hat{y})x_1$
$\frac{\partial \hat{y}}{\partial w_2}$	$= \hat{y}(1 - \hat{y})x_2$
$\frac{\partial \hat{y}}{\partial b}$	$= \hat{y}(1 - \hat{y})$

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

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

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

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

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

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})} \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2}$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_1$$

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

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

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

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

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})} \hat{y}(1-\hat{y})x_1$$

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

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})} \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})} \hat{y}(1-\hat{y})x_2$$

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

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

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

Classification With a Perceptron

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

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

$$\frac{\partial \hat{y}}{\partial w_1} = \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = \hat{y}(1-\hat{y})x_2$$

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

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

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

Classification With a Perceptron

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})} \hat{y}(1-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})} \hat{y}(1-\hat{y})x_2$$

Classification With a Perceptron

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

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}} \cancel{\hat{y}(1-\hat{y})}x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}} \cancel{\hat{y}(1-\hat{y})}x_2$$

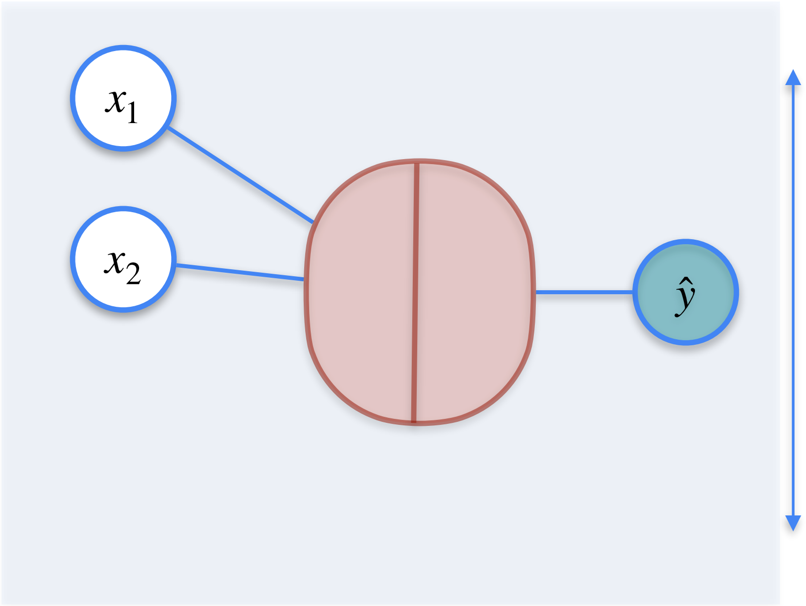
Classification With a Perceptron

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

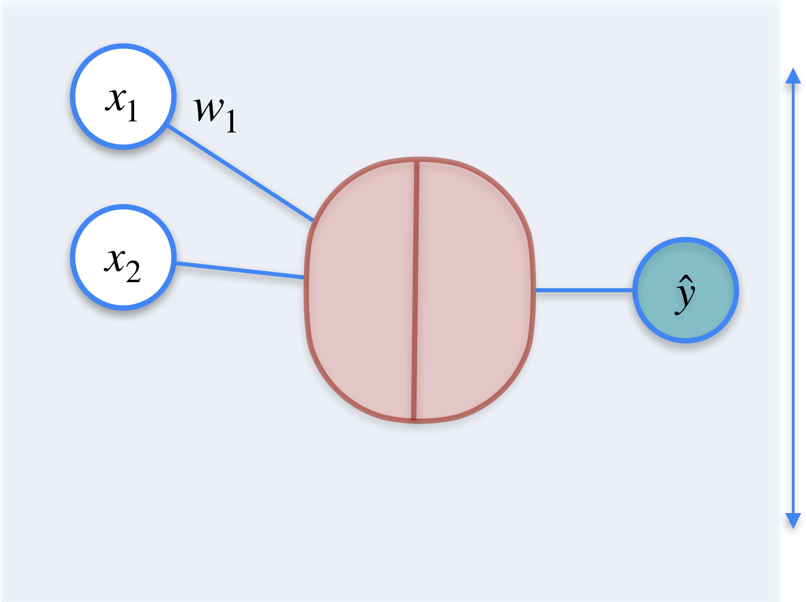
$$\frac{\partial L}{\partial w_1} = -(y - \hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = -(y - \hat{y})x_2$$

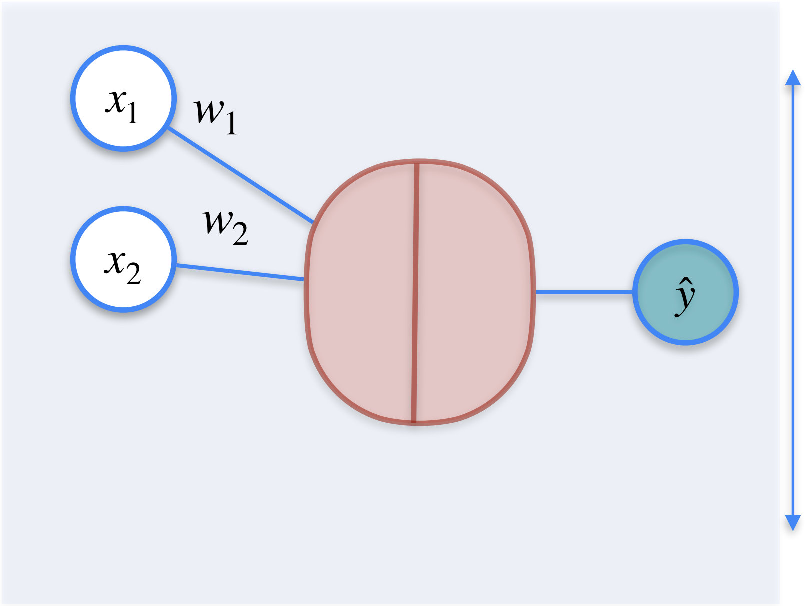
Classification With a Perceptron



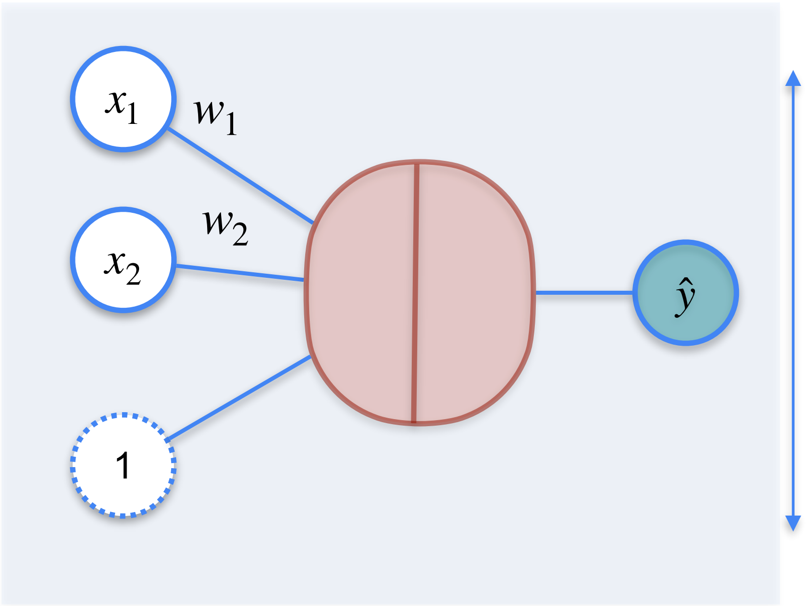
Classification With a Perceptron



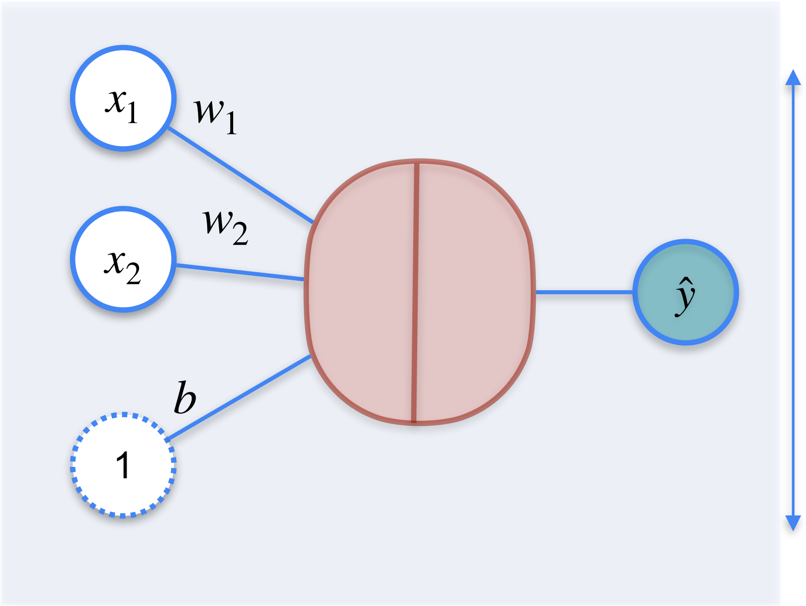
Classification With a Perceptron



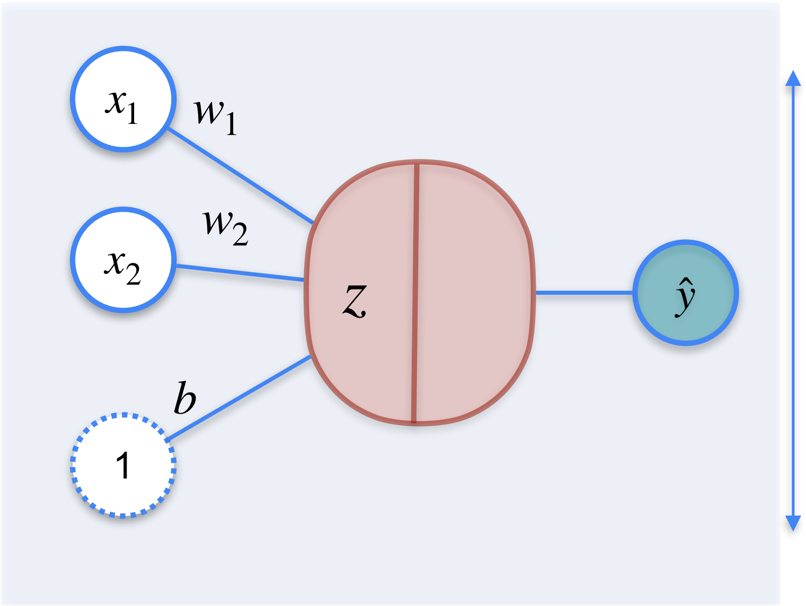
Classification With a Perceptron



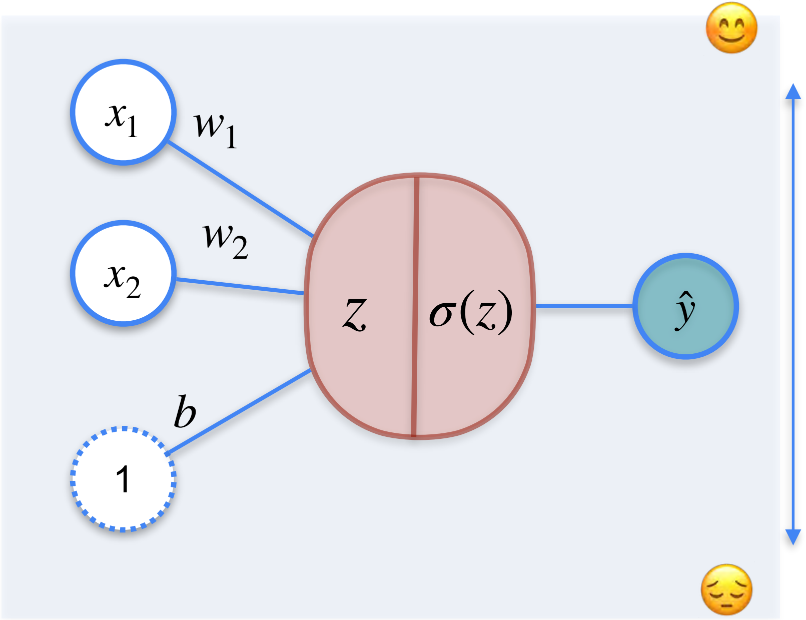
Classification With a Perceptron



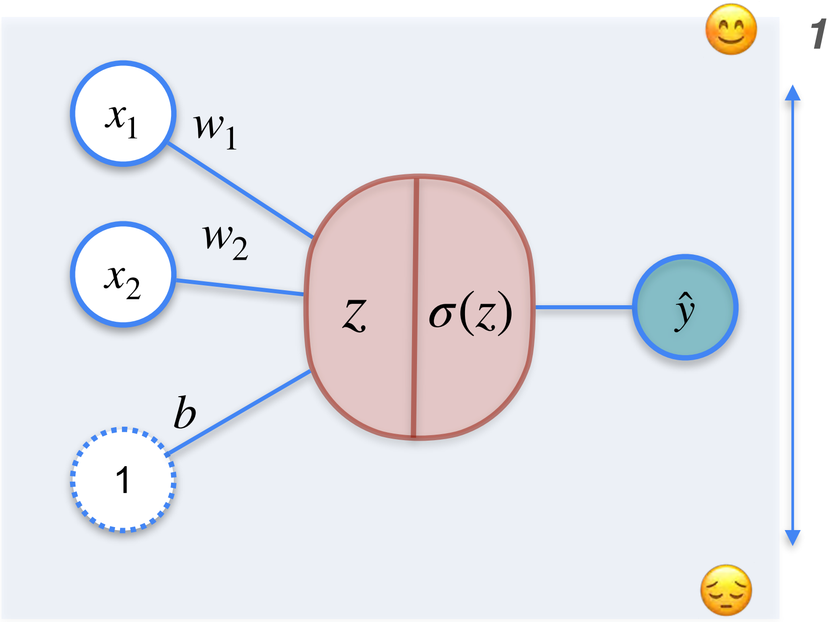
Classification With a Perceptron



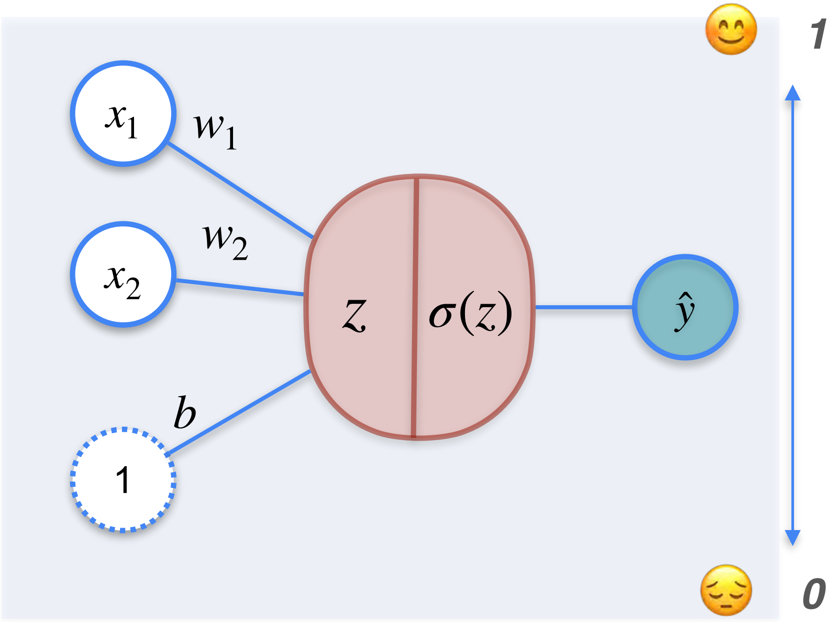
Classification With a Perceptron



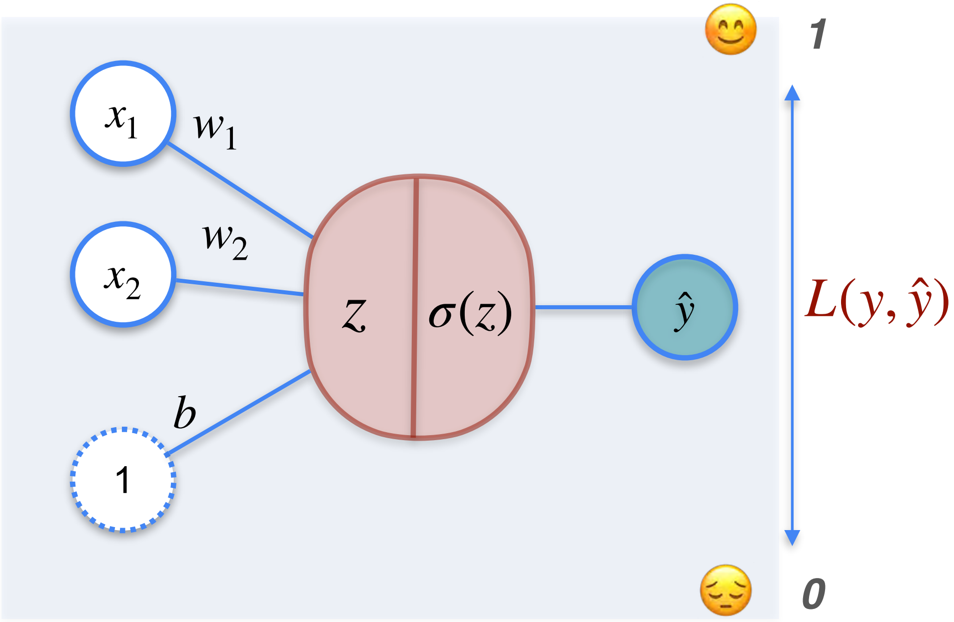
Classification With a Perceptron



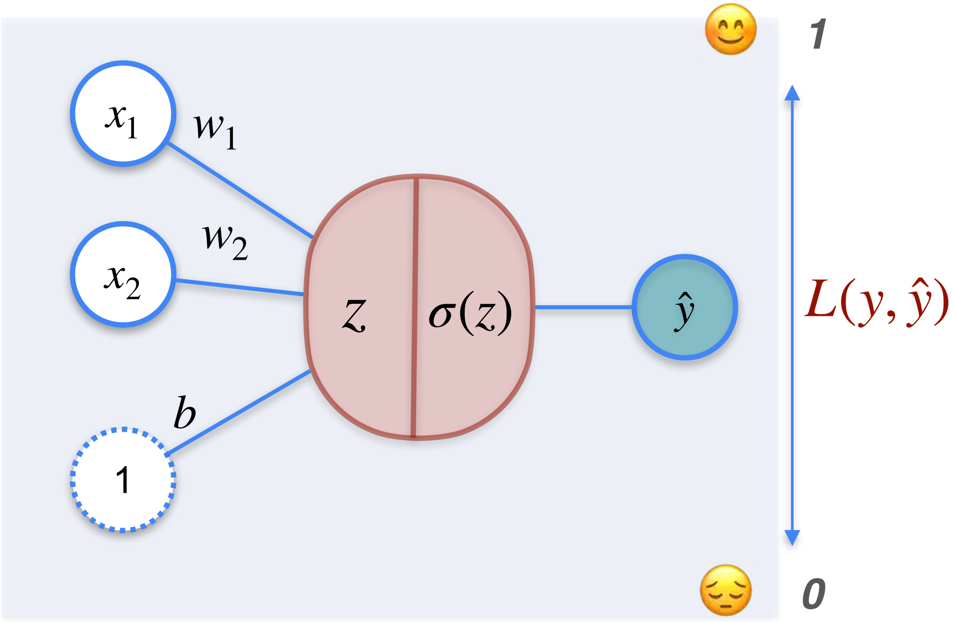
Classification With a Perceptron



Classification With a Perceptron

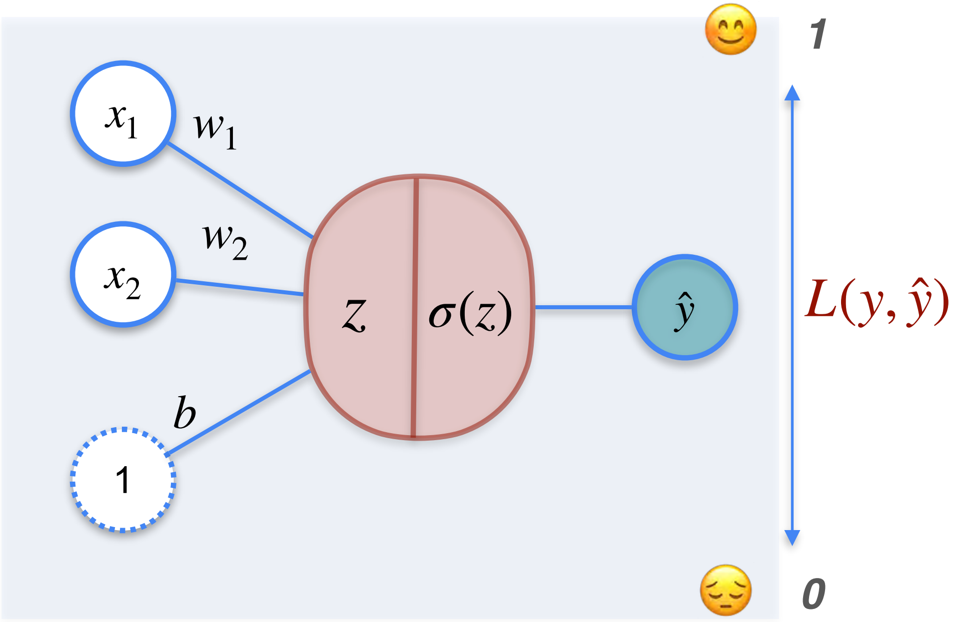


Classification With a Perceptron



To find optimal values for:

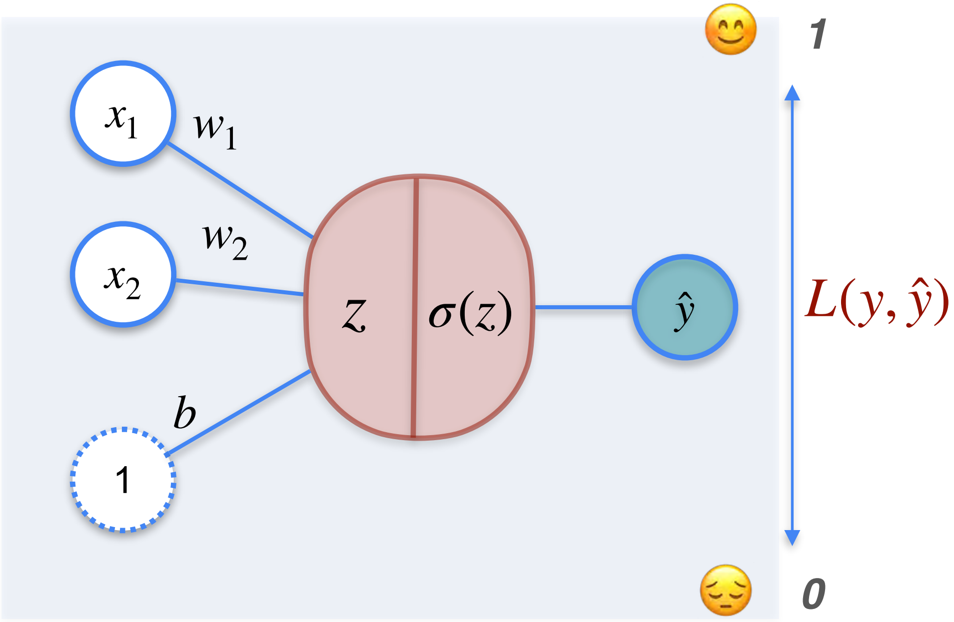
Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

Classification With a Perceptron

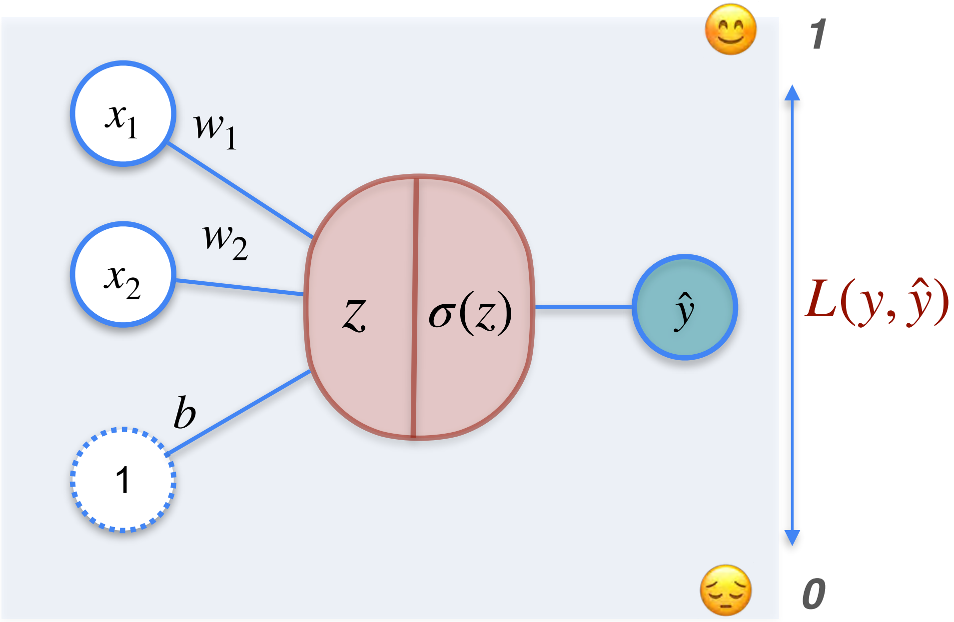


To find optimal values for:

w_1 , w_2 , b

You need gradient descent

Classification With a Perceptron



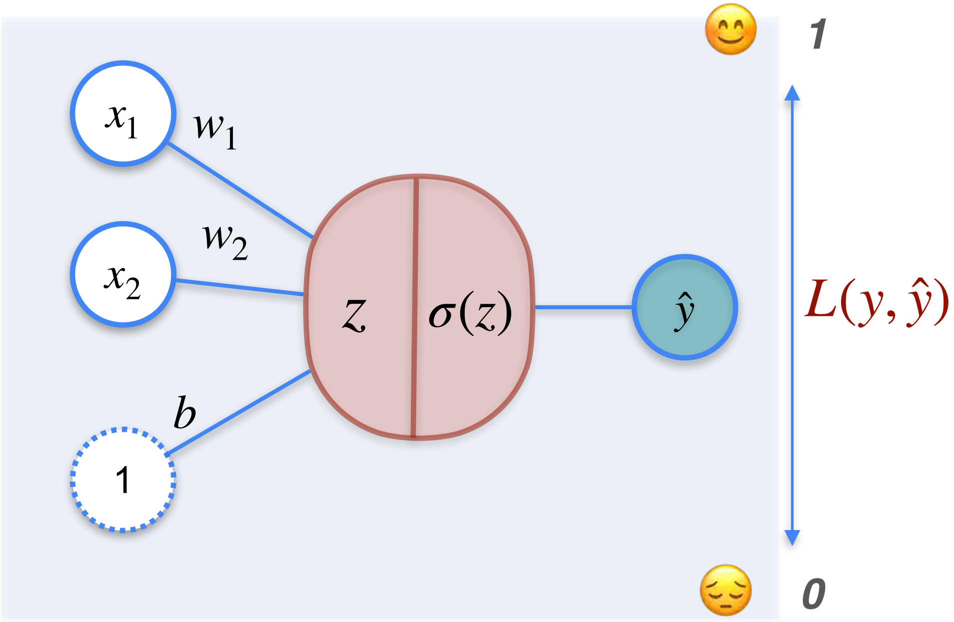
To find optimal values for:

w_1 , w_2 , b

You need gradient descent

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

Classification With a Perceptron



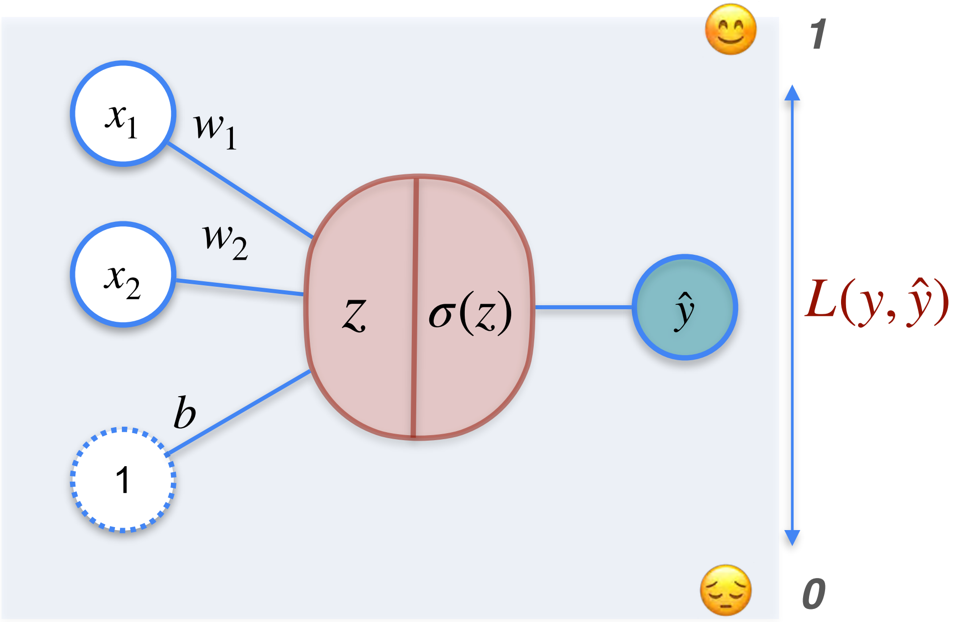
To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$$w_1 \rightarrow w_1 - \alpha$$

Classification With a Perceptron



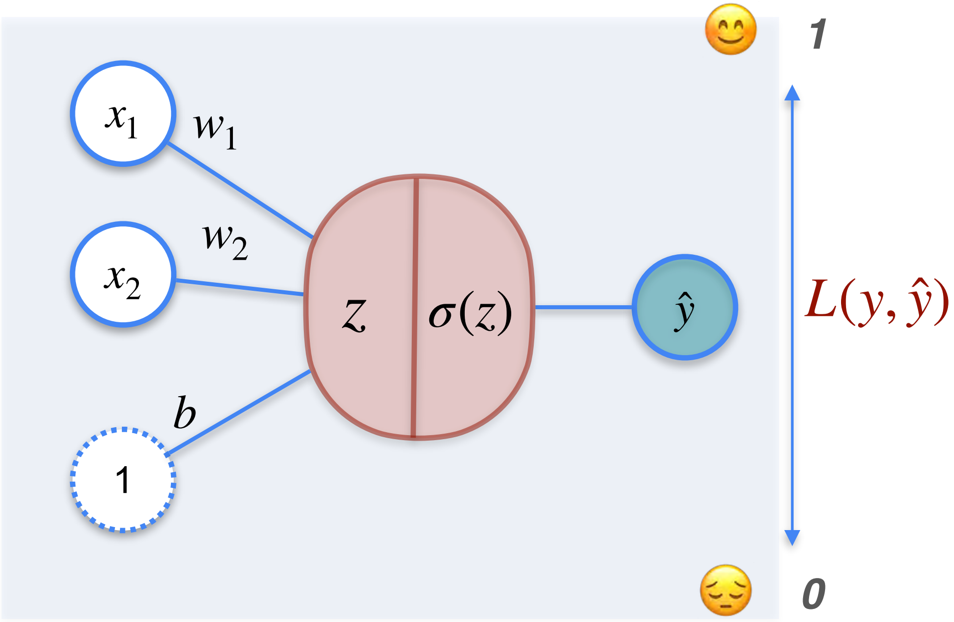
To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$$w_1 \rightarrow w_1 - \alpha(-x_1(y - \hat{y}))$$

Classification With a Perceptron



To find optimal values for:

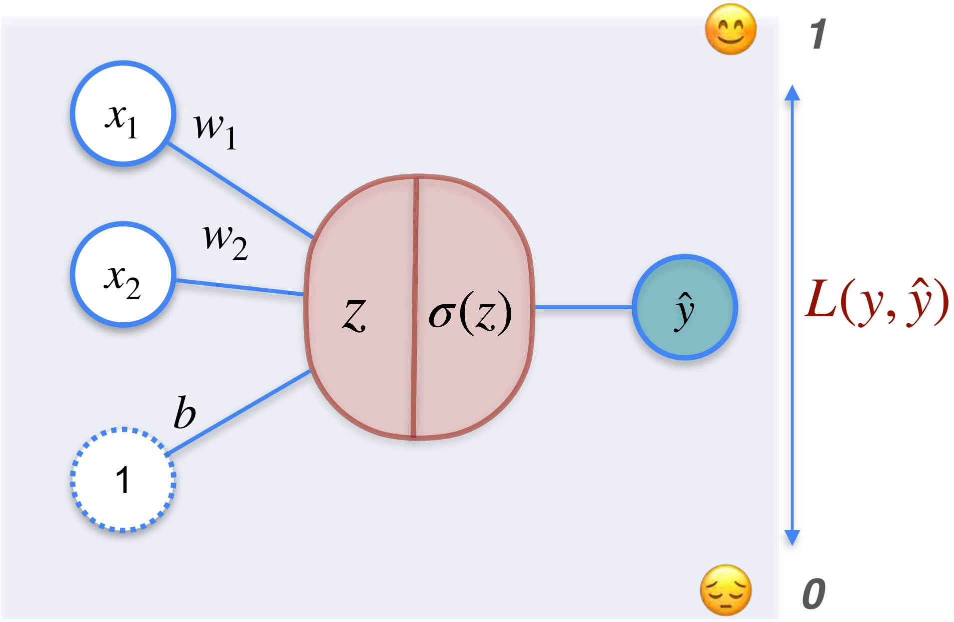
w_1 , w_2 , b

You need gradient descent

$$w_1 \rightarrow w_1 - \alpha(-x_1(y - \hat{y}))$$

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

Classification With a Perceptron



To find optimal values for:

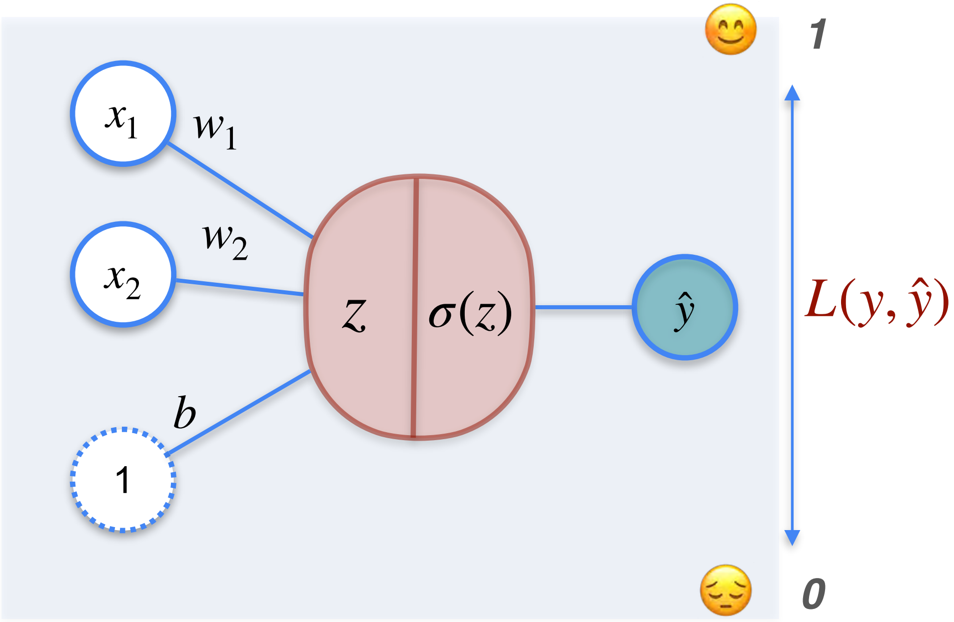
w_1 , w_2 , b

You need gradient descent

$$w_1 \rightarrow w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 \rightarrow w_2 - \alpha$$

Classification With a Perceptron



To find optimal values for:

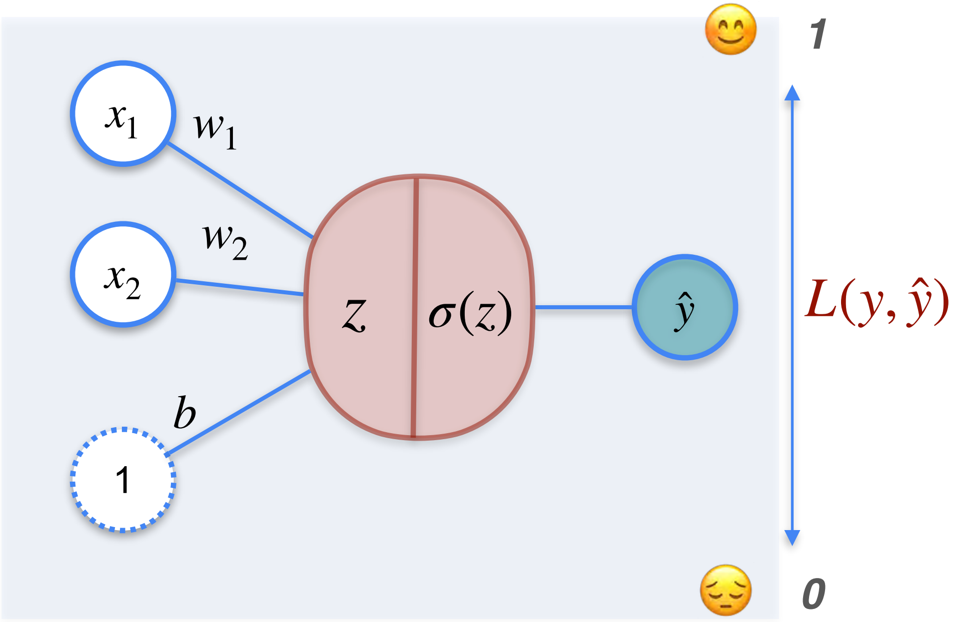
w_1 , w_2 , b

You need gradient descent

$$w_1 \rightarrow w_1 - \alpha(\text{-}x_1(y - \hat{y}))$$

$$w_2 \rightarrow w_2 - \alpha(\text{-}x_2(y - \hat{y}))$$

Classification With a Perceptron



To find optimal values for:

$w_1\ ,\ w_2\ ,\ b$

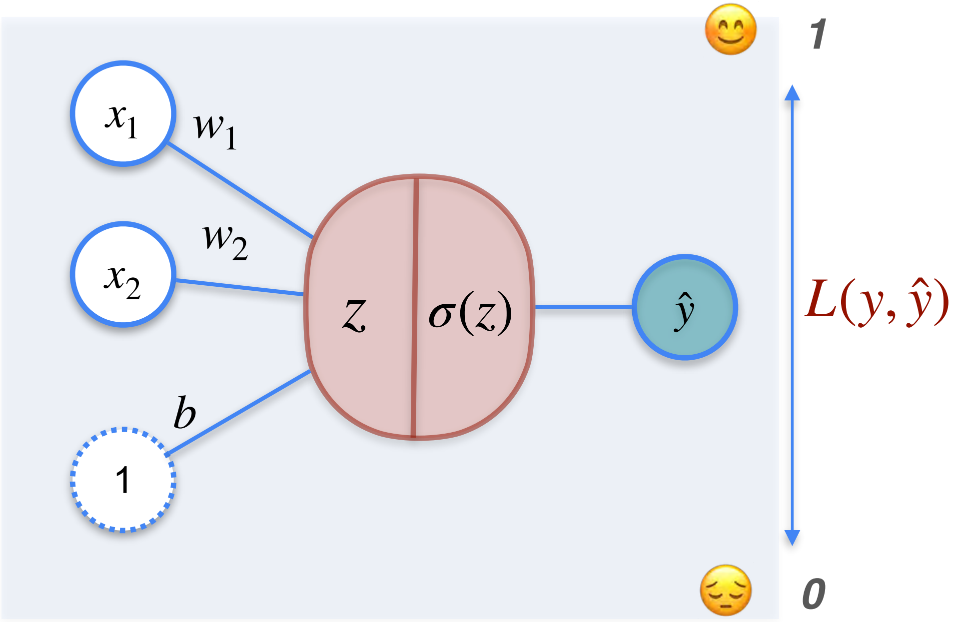
You need gradient descent

$$w_1 \rightarrow w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 \rightarrow w_2 - \alpha(-x_2(y - \hat{y}))$$

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

Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

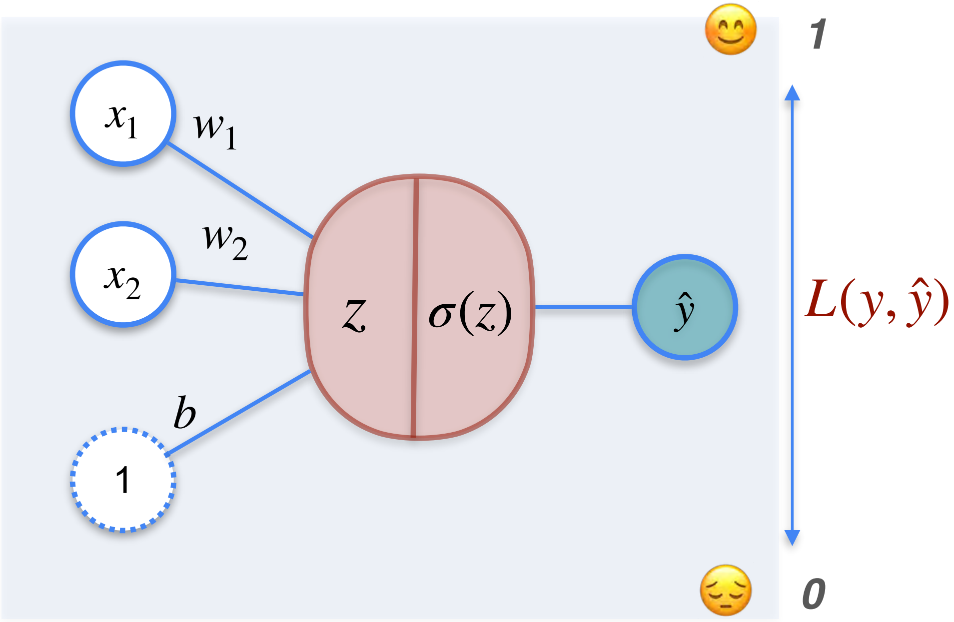
You need gradient descent

$$w_1 \to w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 \to w_2 - \alpha(-x_2(y - \hat{y}))$$

$$b \to b - \alpha$$

Classification With a Perceptron



To find optimal values for:

w_1 , w_2 , b

You need gradient descent

$$w_1 \rightarrow w_1 - \alpha(-x_1(y - \hat{y}))$$

$$w_2 \rightarrow w_2 - \alpha(-x_2(y - \hat{y}))$$

$$b \rightarrow b - \alpha(-(y - \hat{y}))$$

DeepLearning.AI

Optimization in Neural Networks and Newton's Method

Classification with a Neural Network

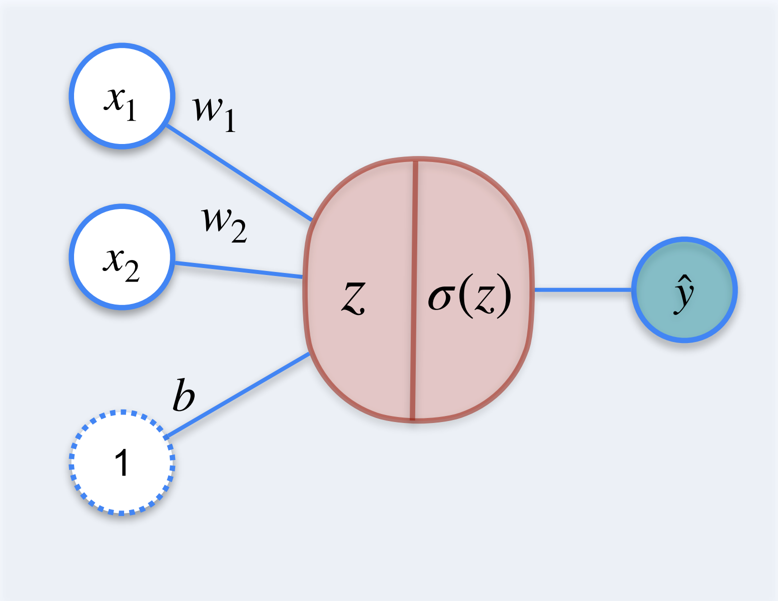
Classification Problem Motivation

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* 😊

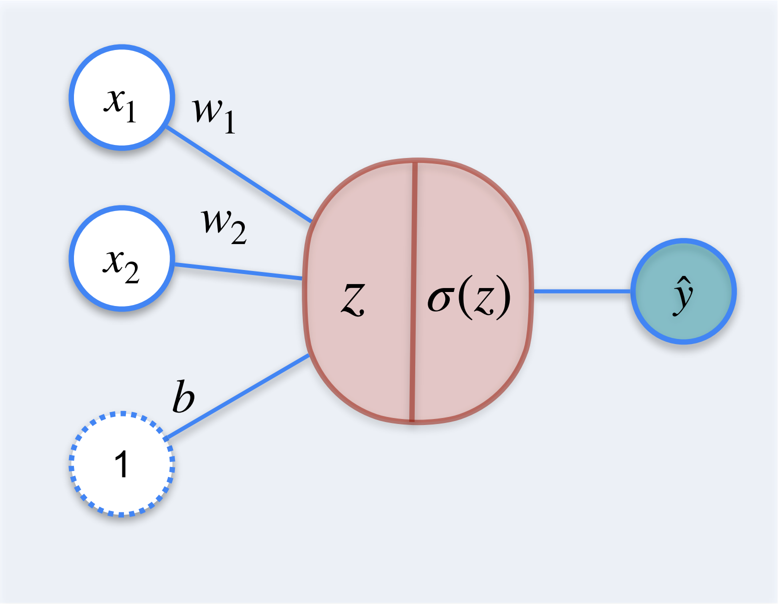
Classification Problem Motivation

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* 😊



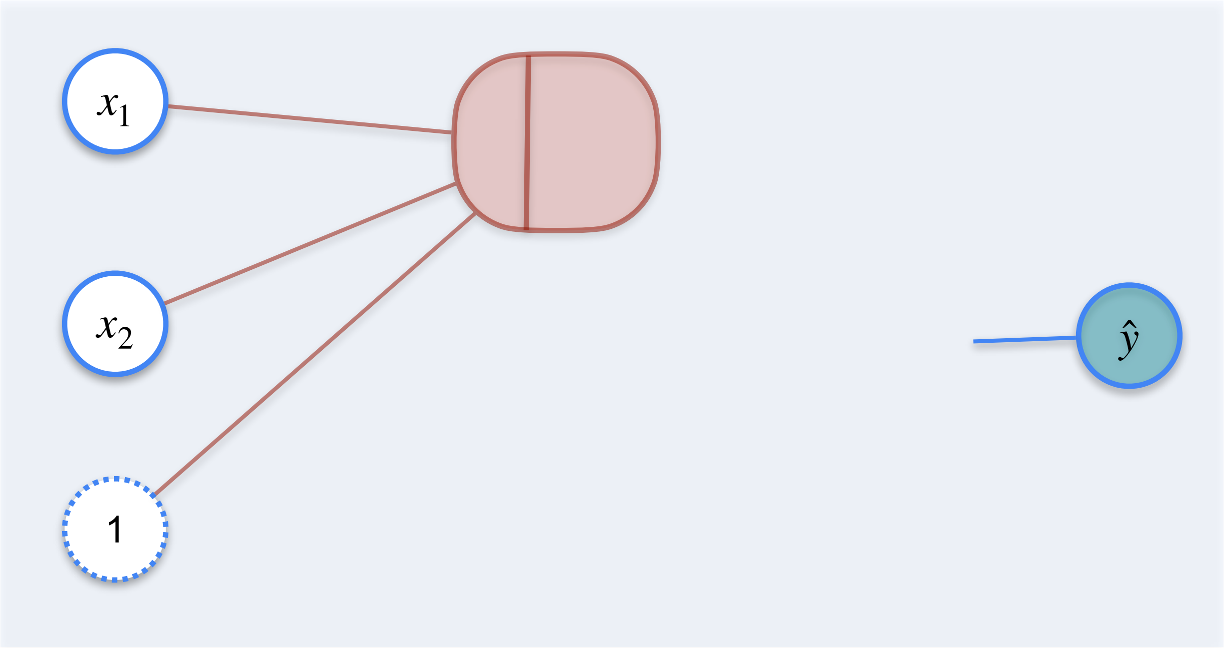
Classification Problem Motivation

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* 😊

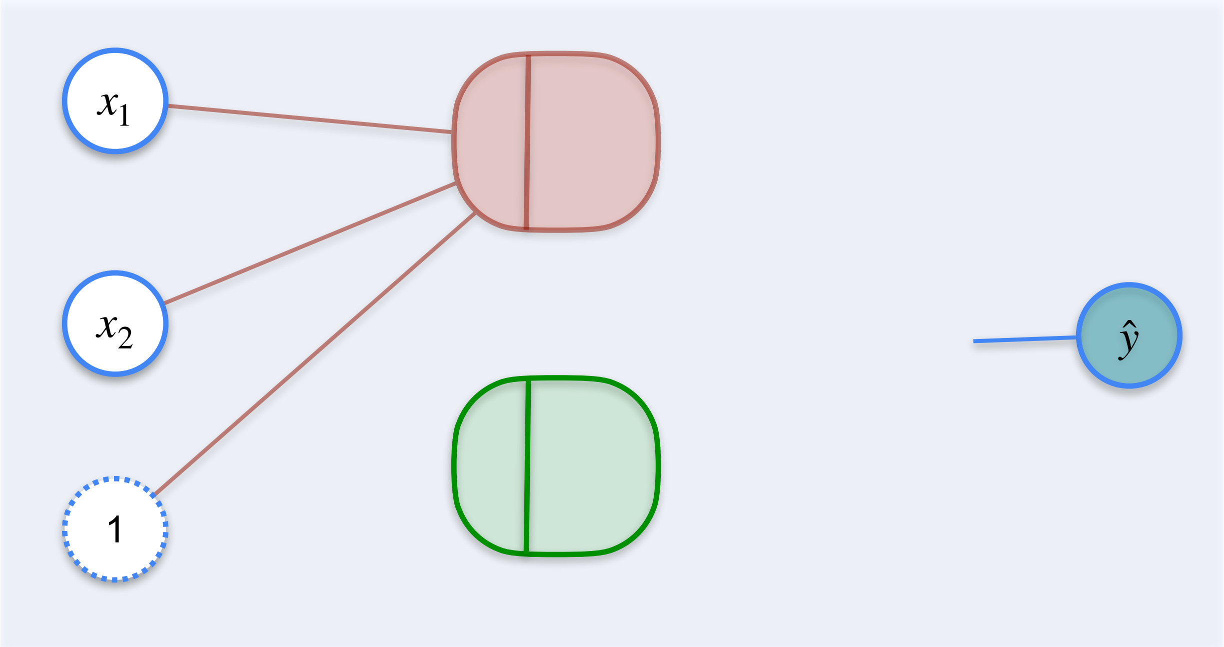


$$z = x_1 w_1 + x_2 w_2 + b$$

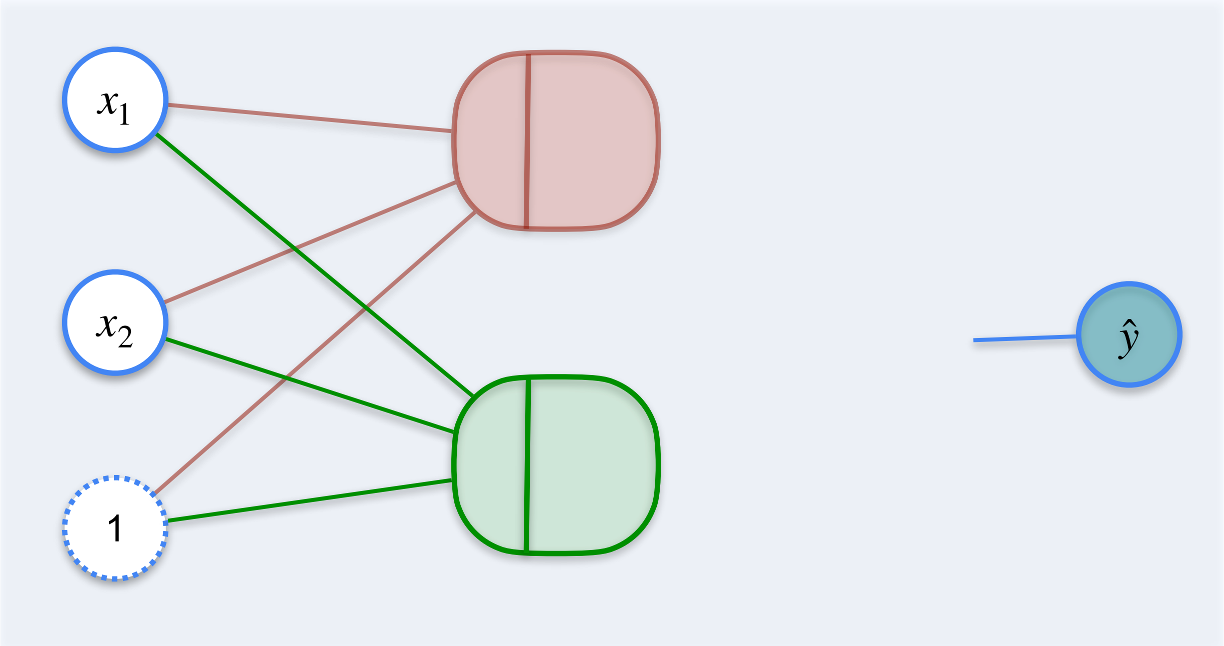
2,2,1 Neural Network



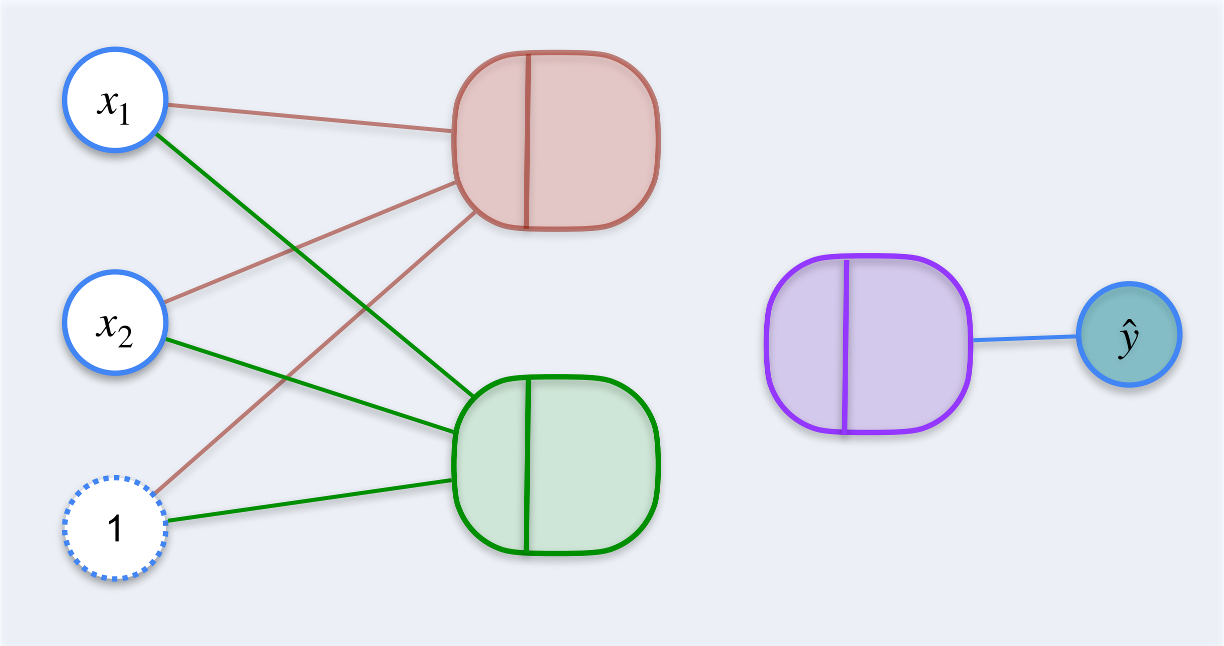
2,2,1 Neural Network



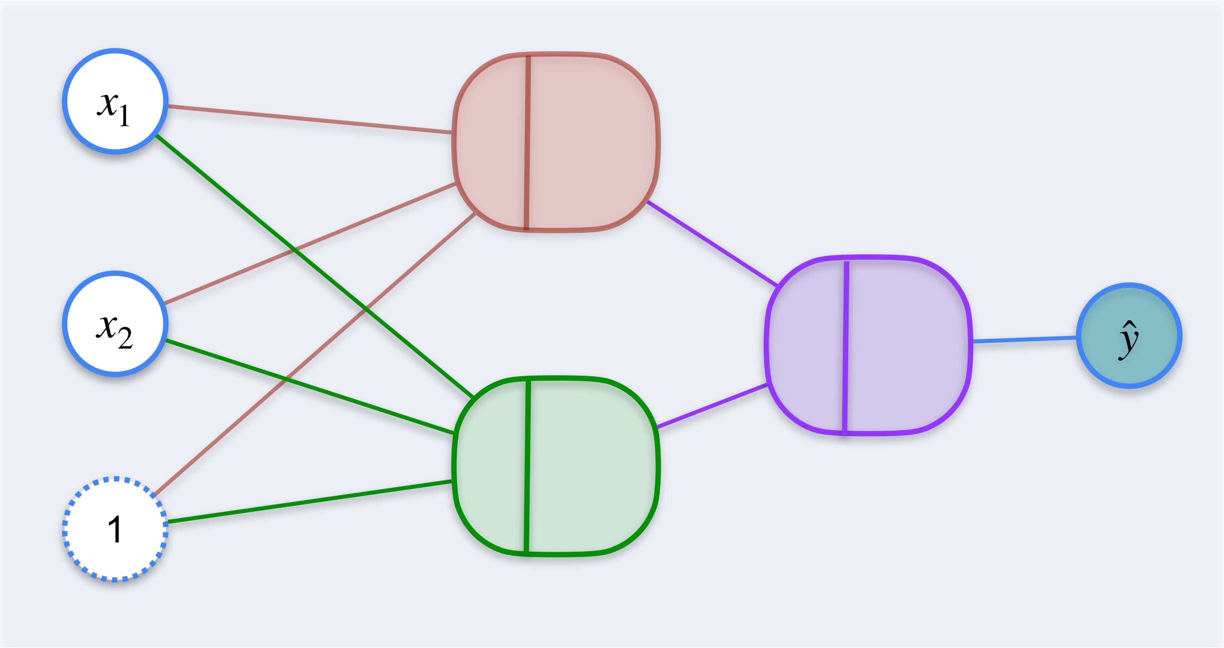
2,2,1 Neural Network



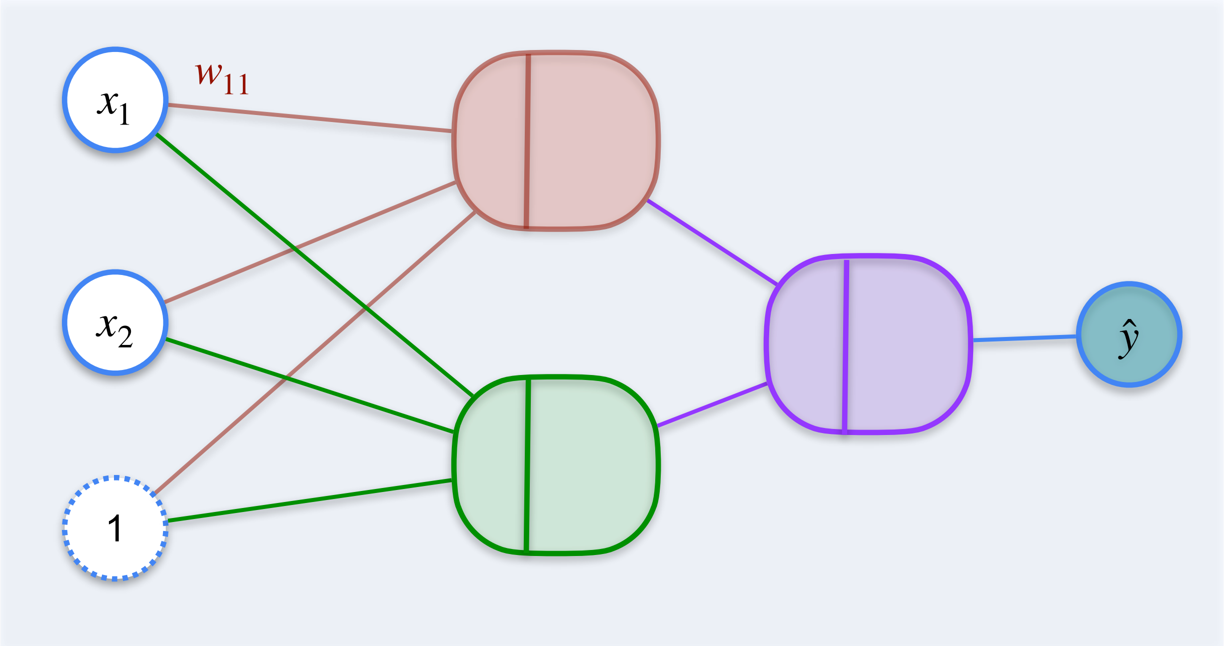
2,2,1 Neural Network



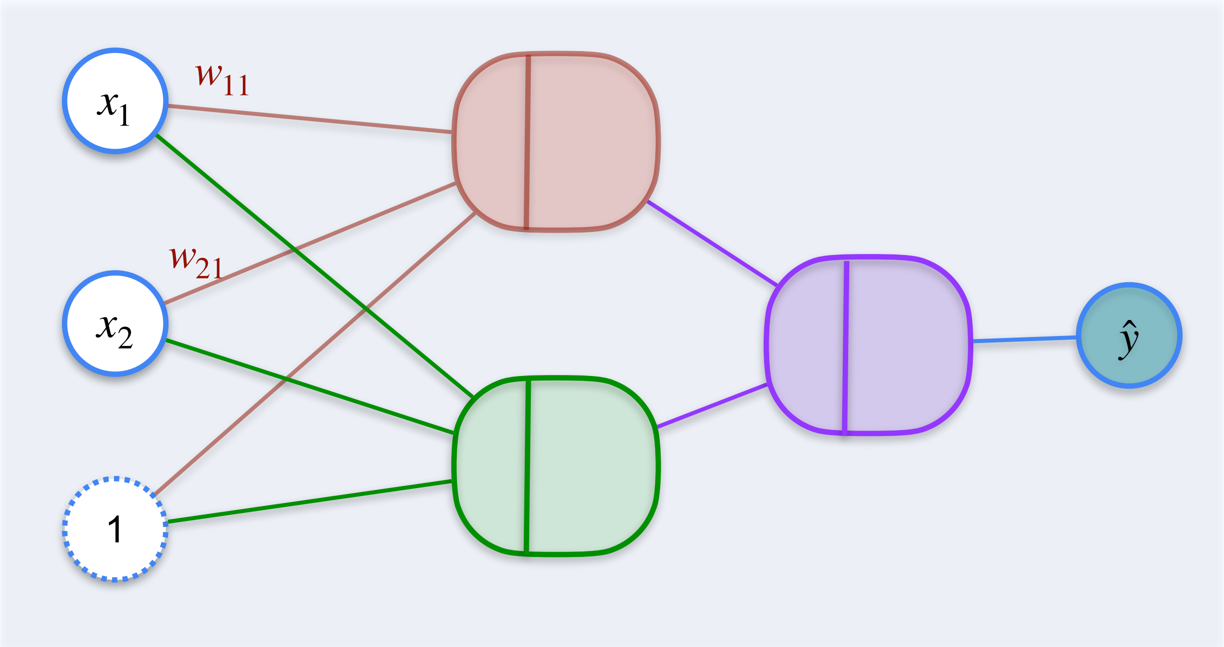
2,2,1 Neural Network



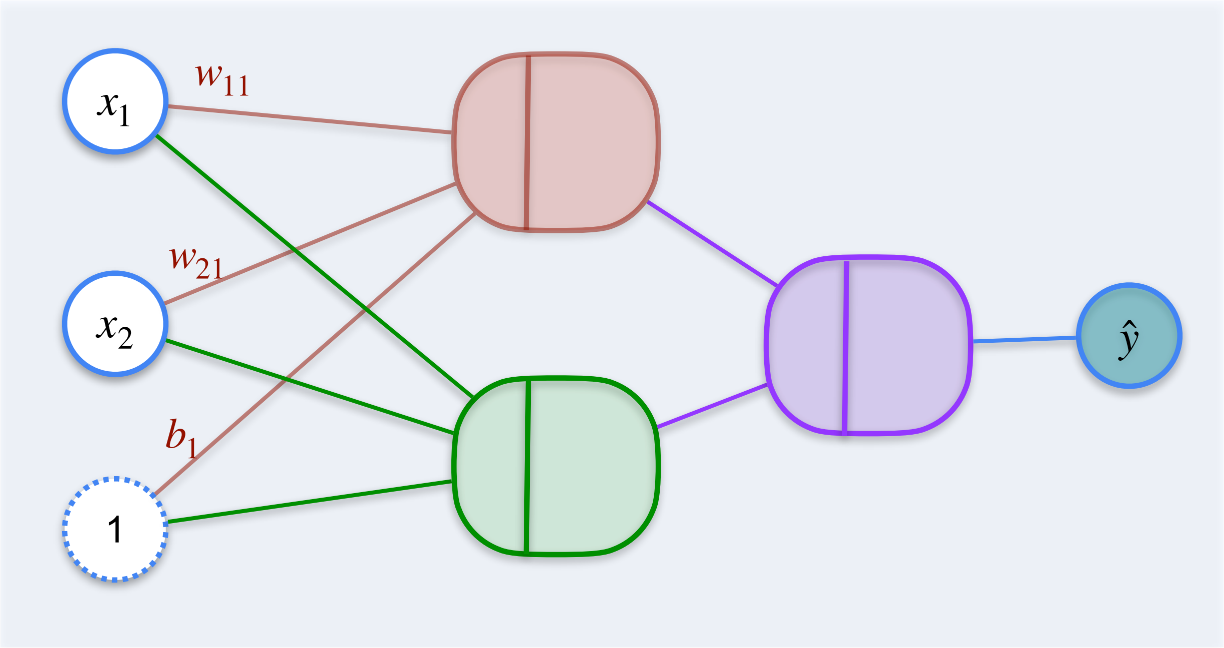
2,2,1 Neural Network



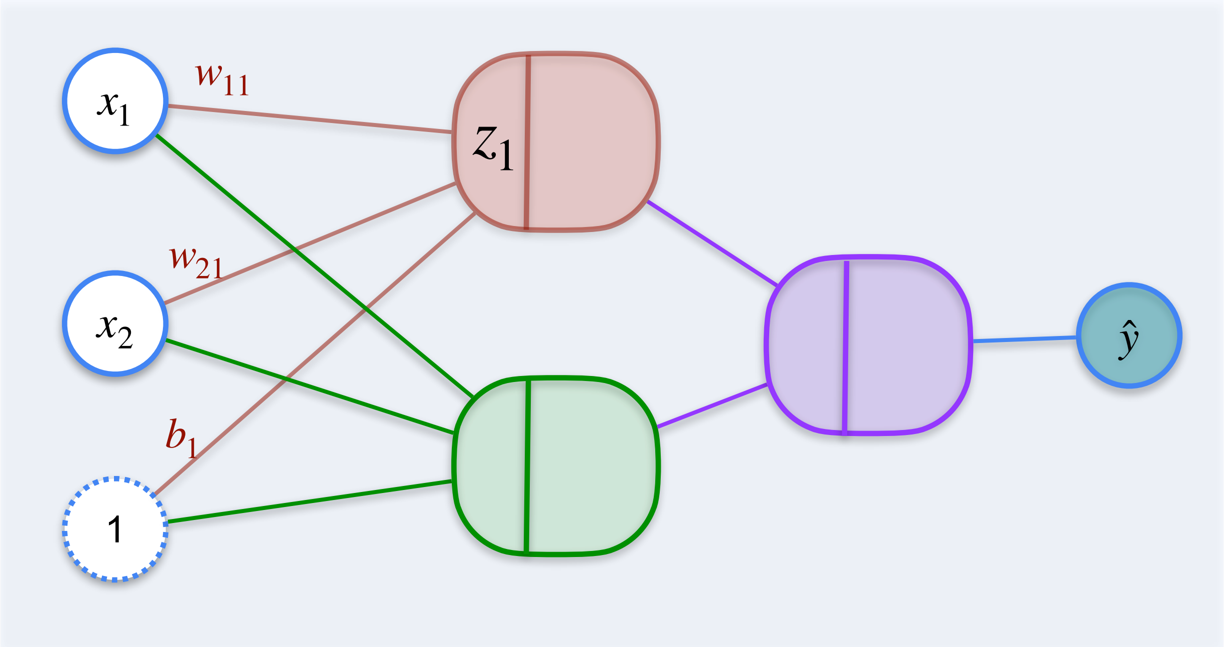
2,2,1 Neural Network



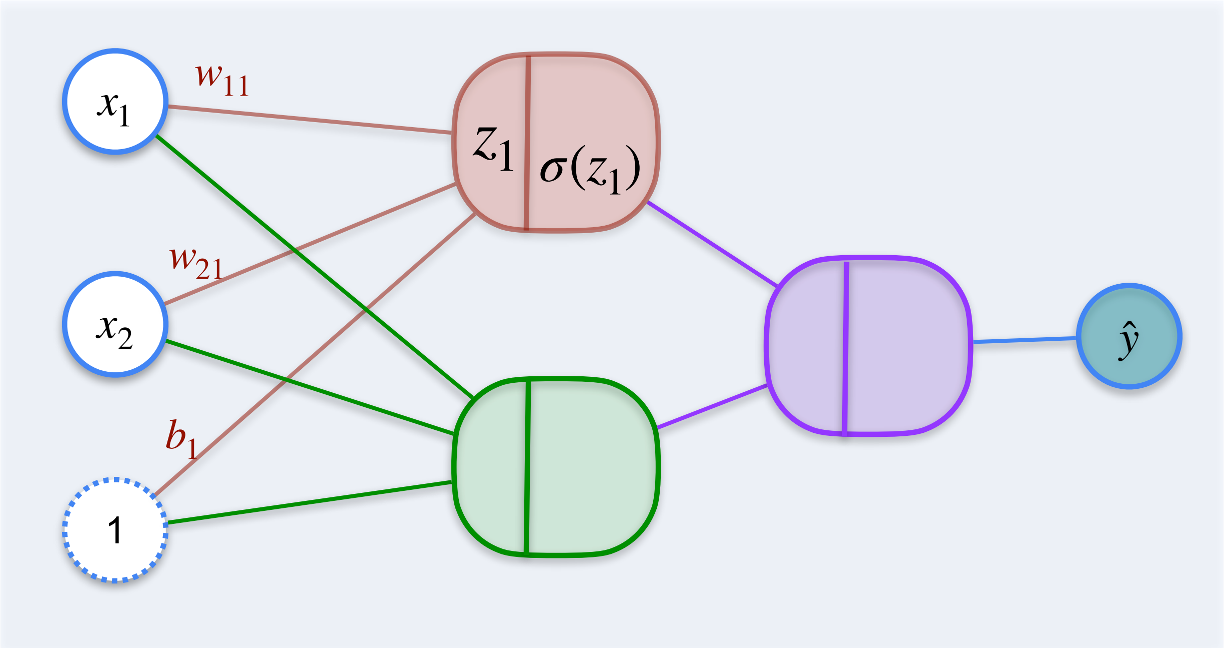
2,2,1 Neural Network



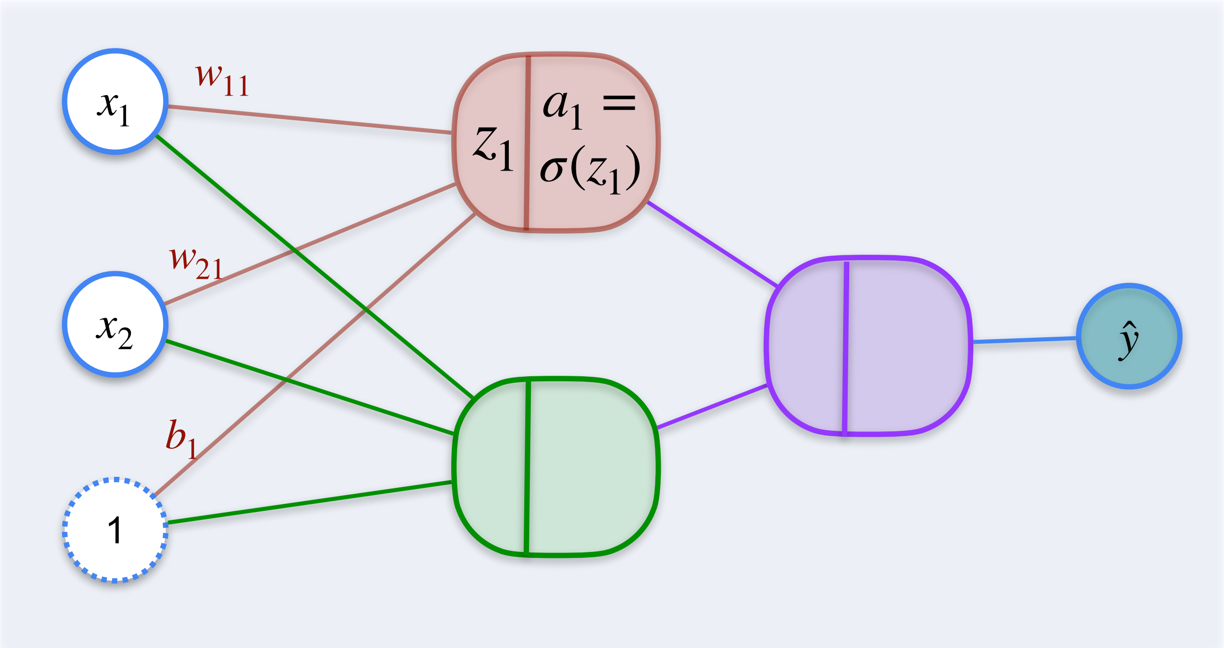
2,2,1 Neural Network



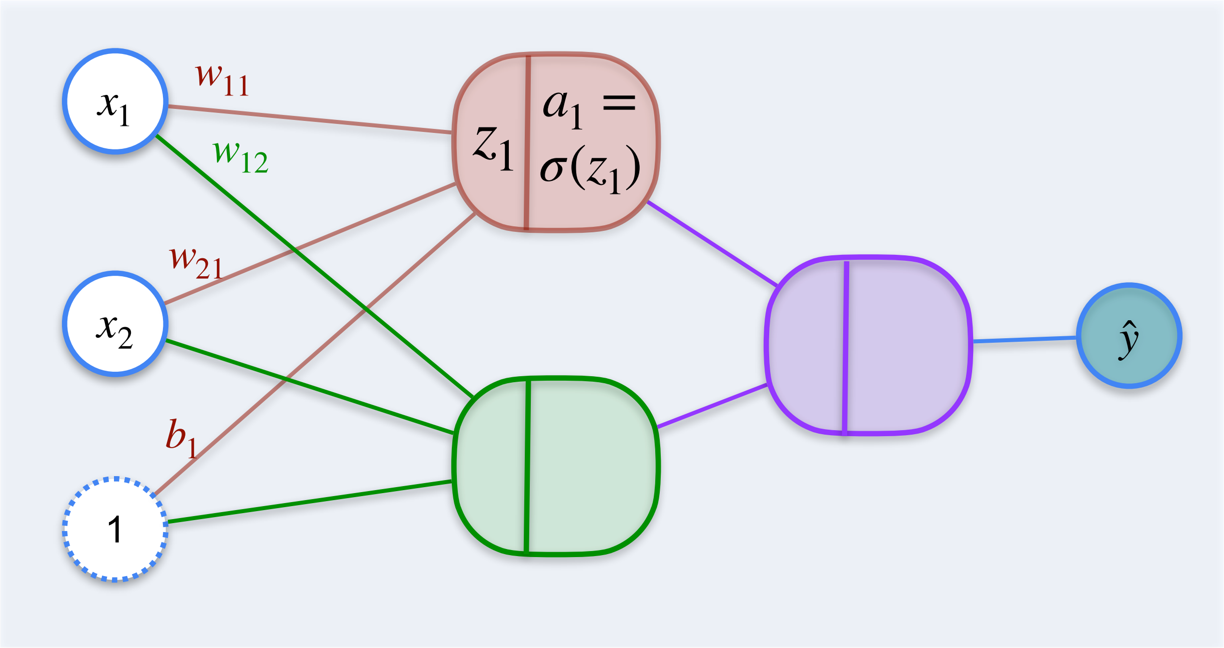
2,2,1 Neural Network



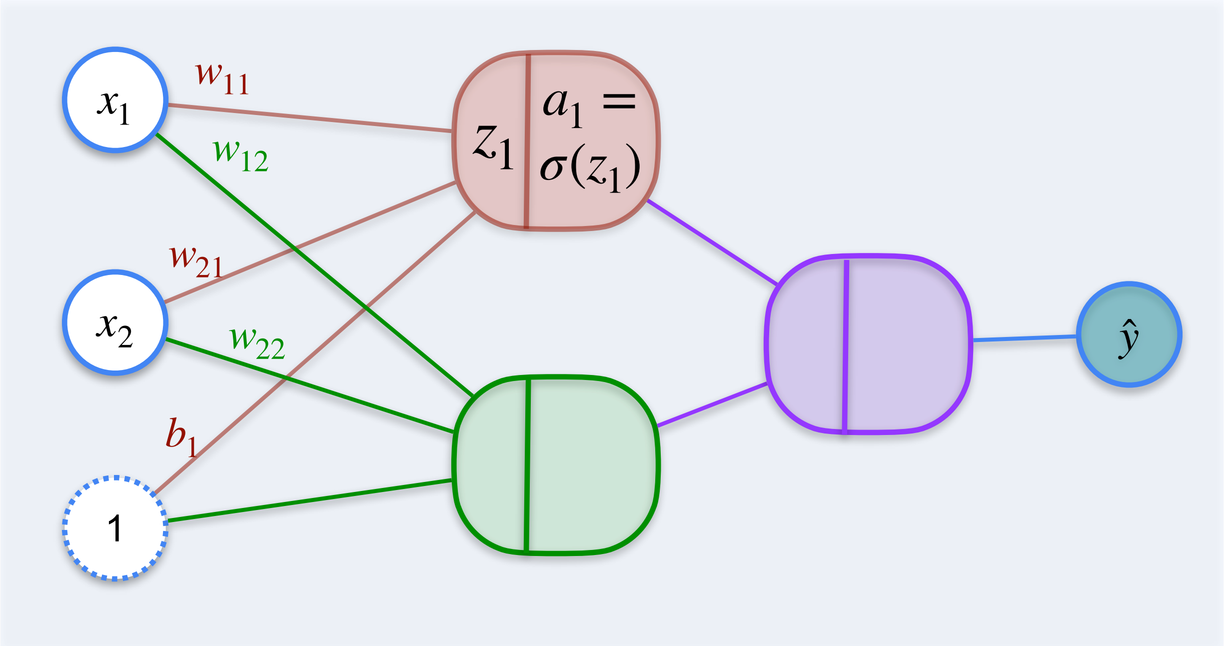
2,2,1 Neural Network



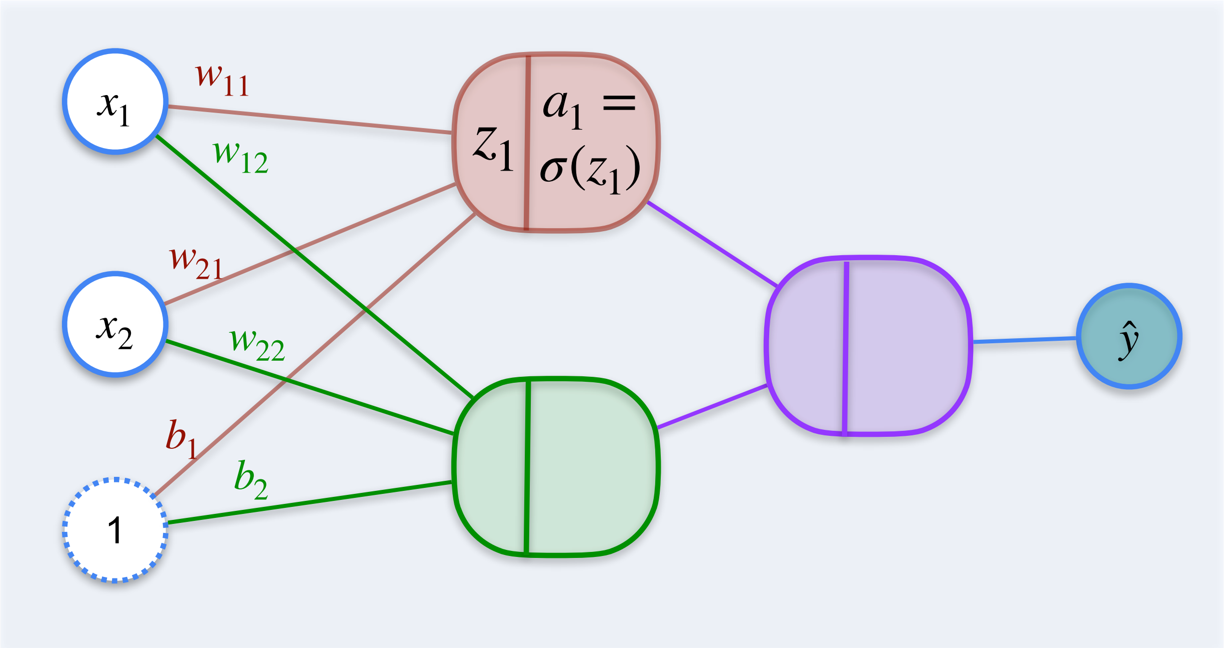
2,2,1 Neural Network



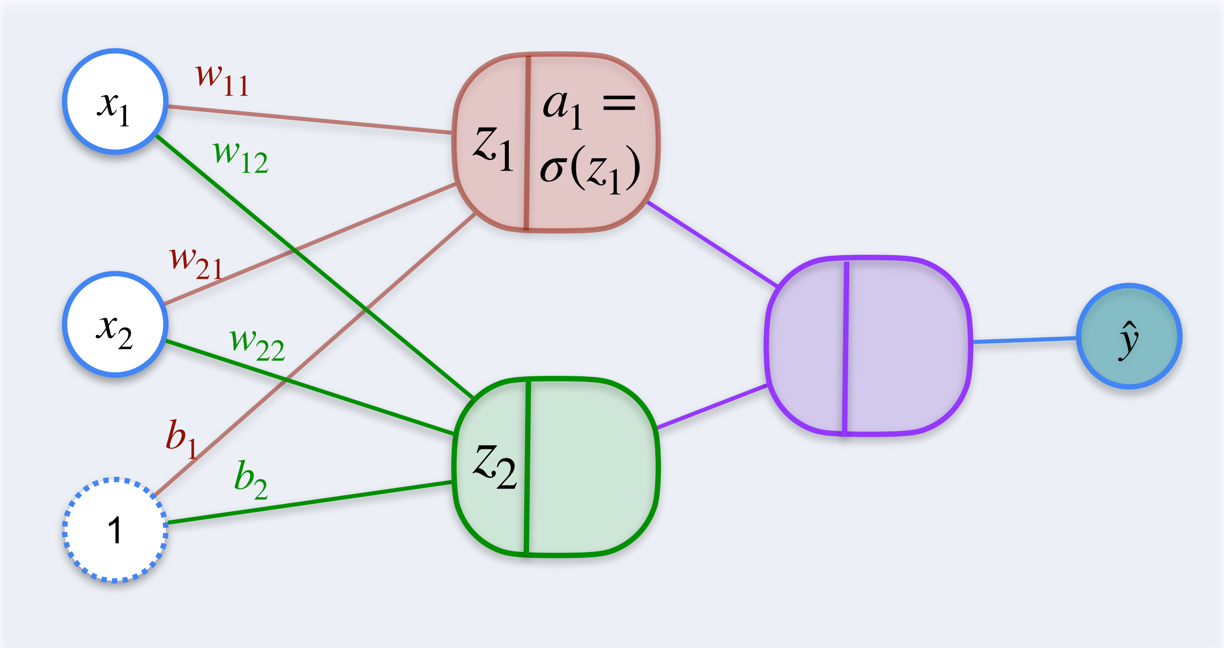
2,2,1 Neural Network



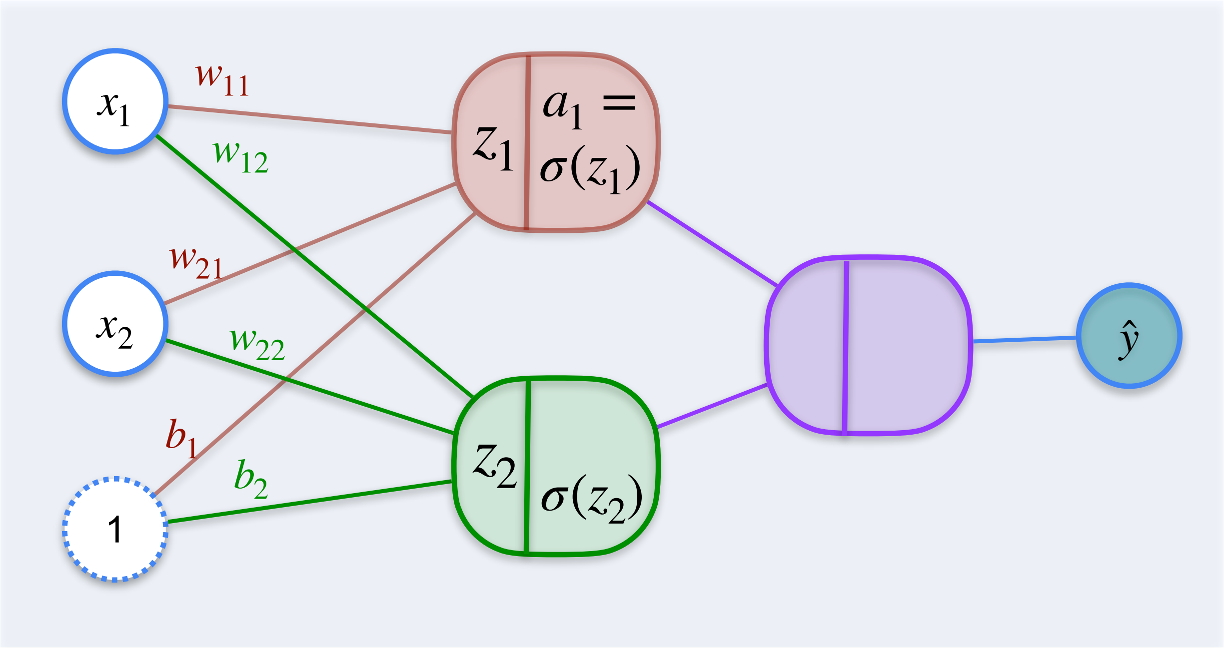
2,2,1 Neural Network



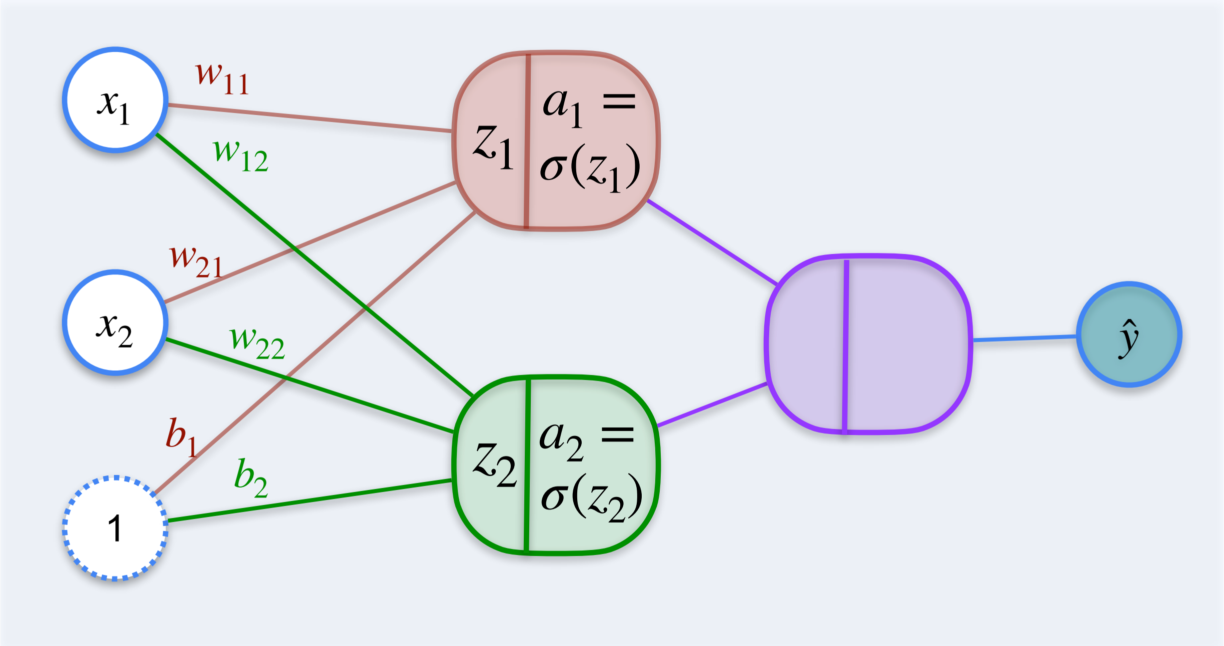
2,2,1 Neural Network



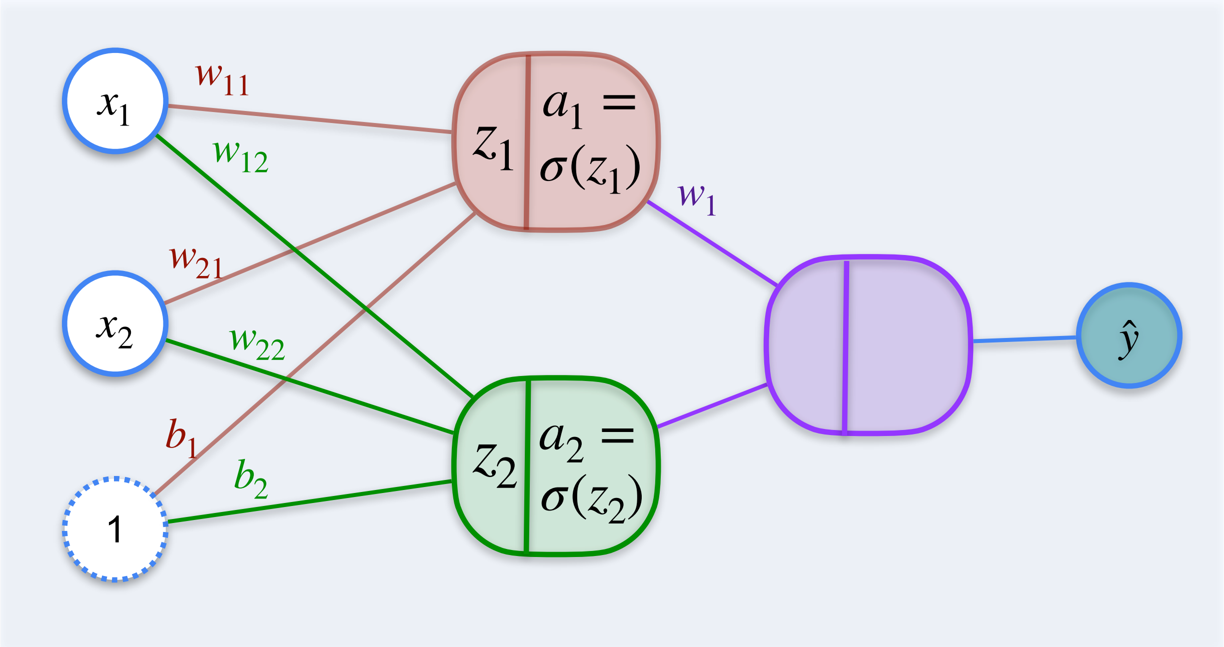
2,2,1 Neural Network



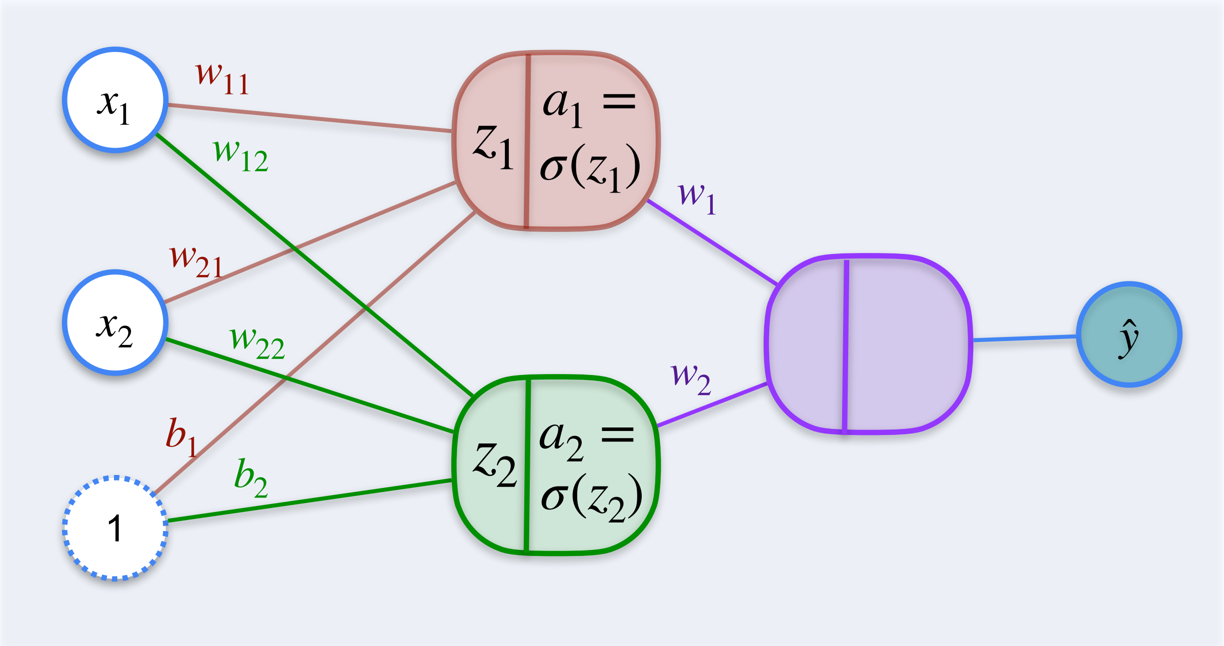
2,2,1 Neural Network



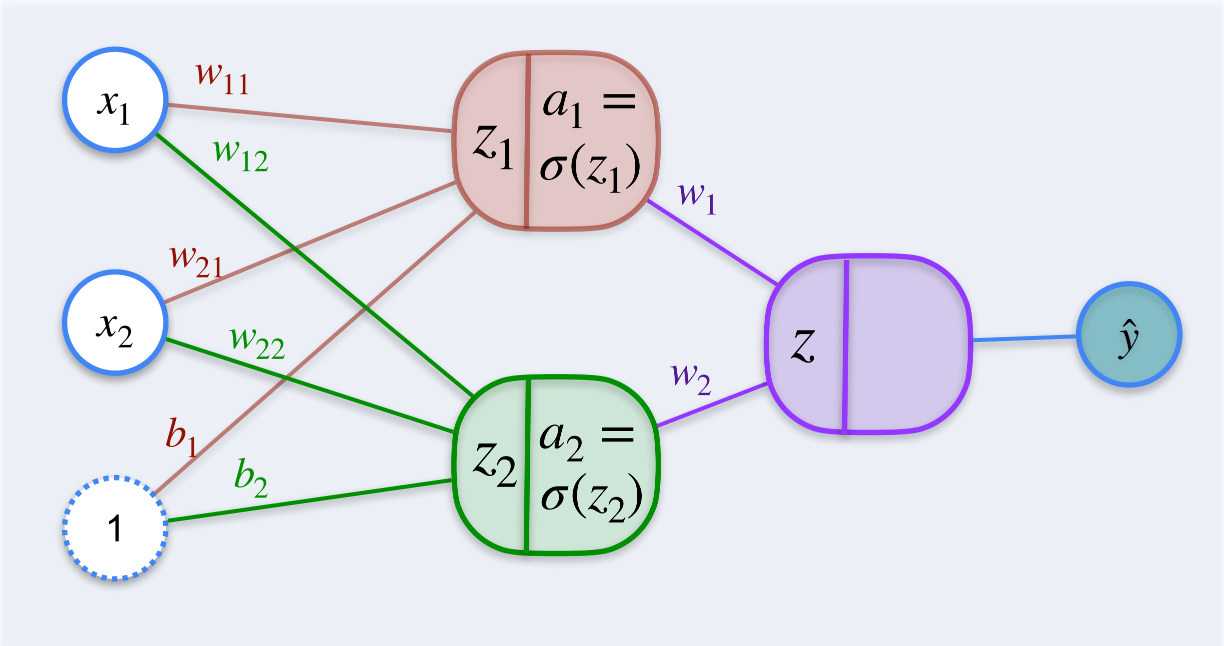
2,2,1 Neural Network



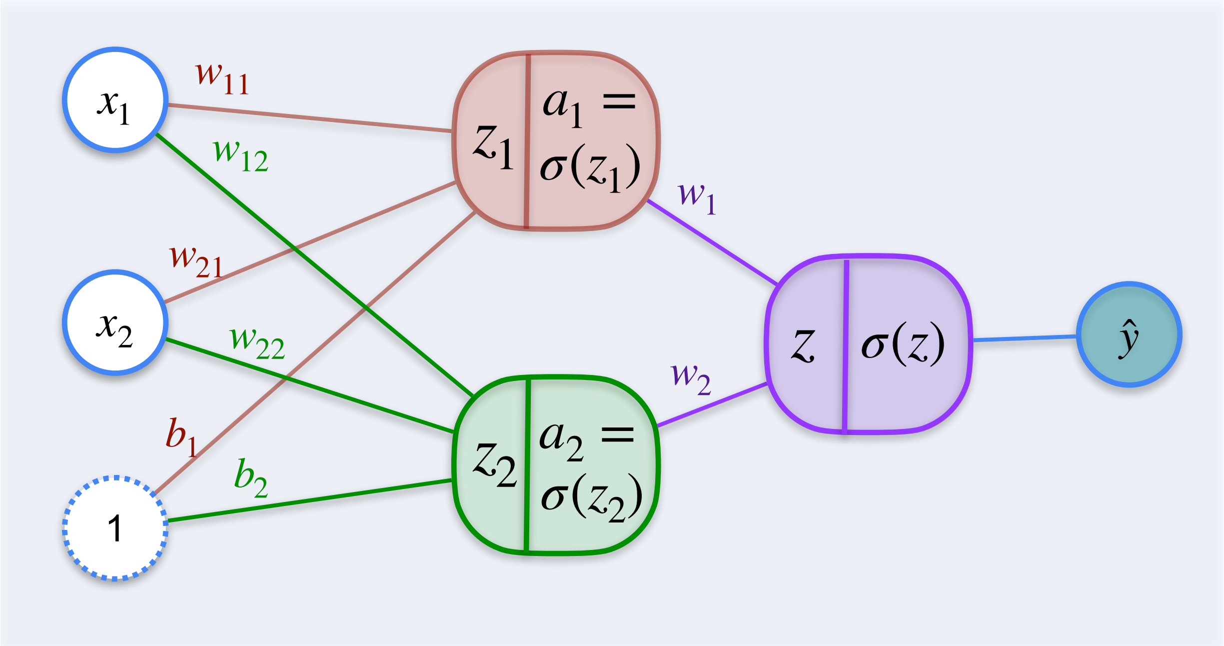
2,2,1 Neural Network



2,2,1 Neural Network



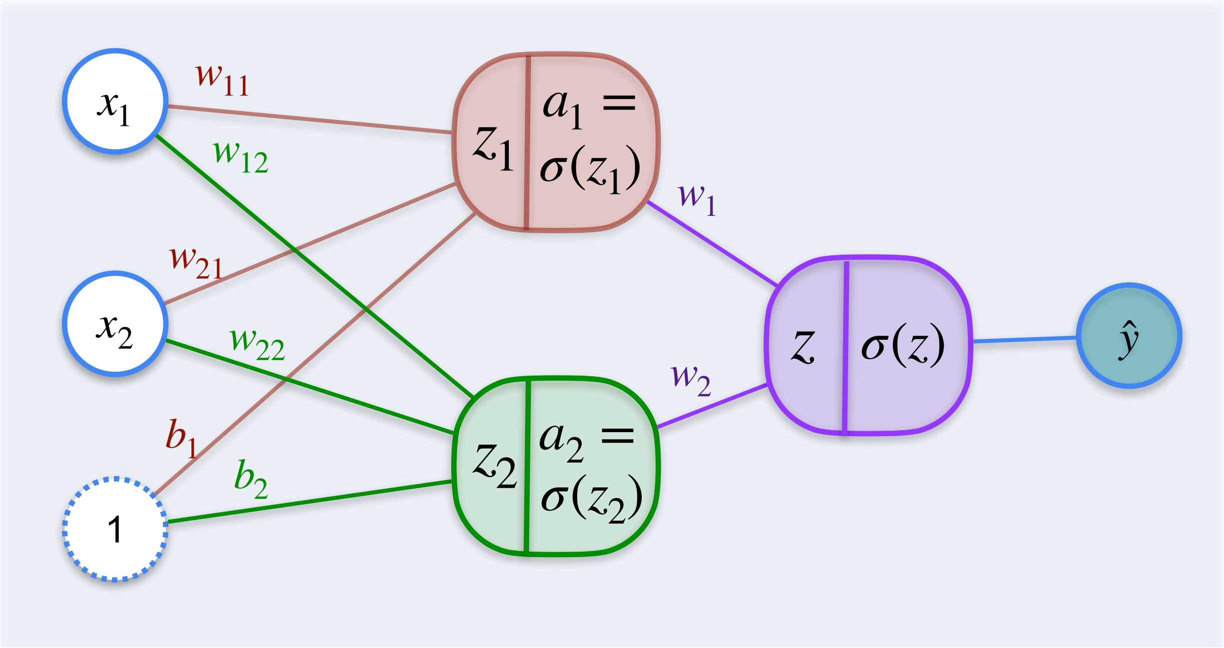
2,2,1 Neural Network



2,2,1 Neural Network

Neural network of depth 2

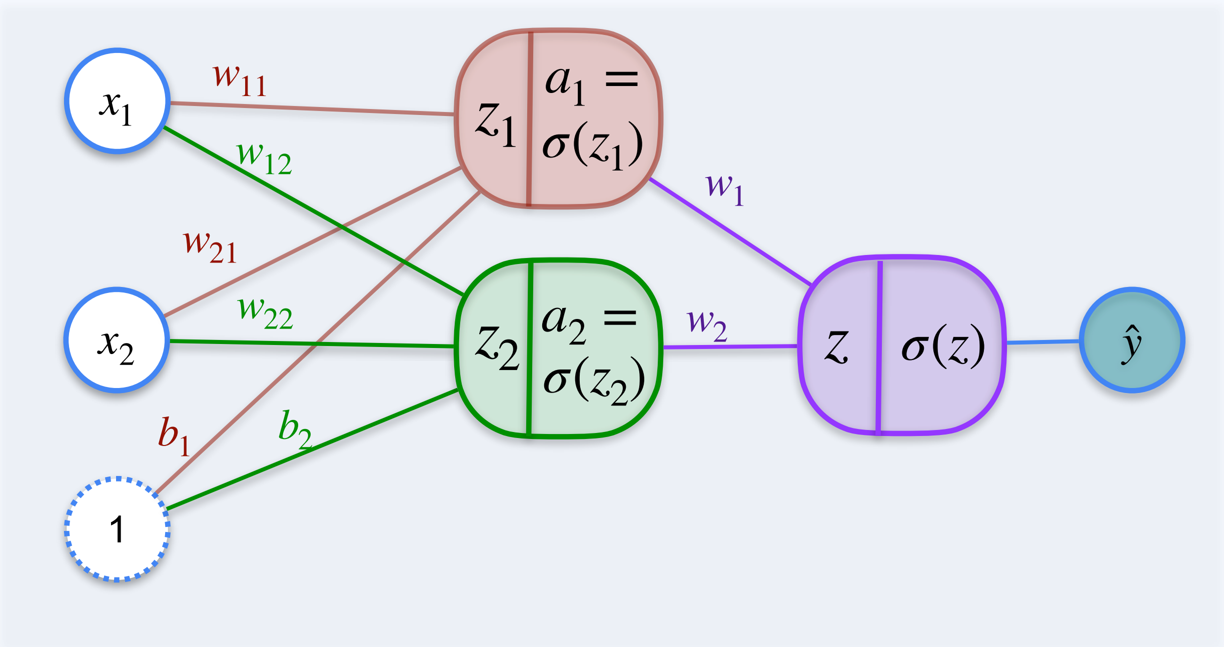
- one input layer
- one hidden layer
- one output layer



2,2,1 Neural Network

Neural network of depth 2

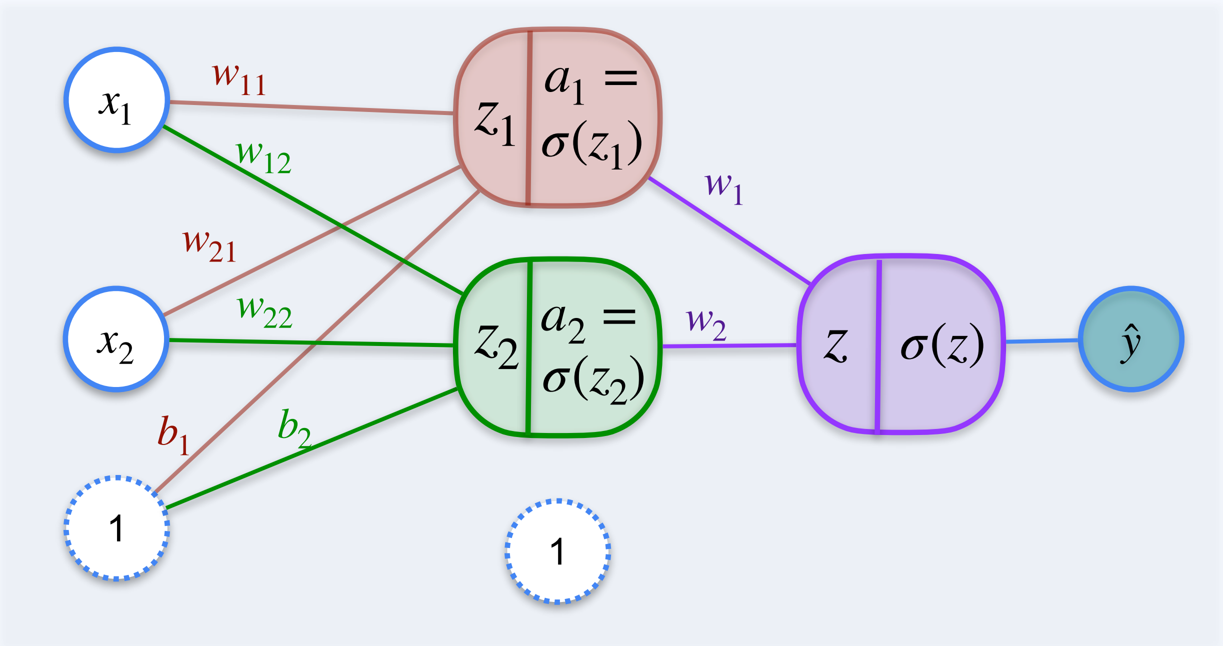
- one input layer
- one hidden layer
- one output layer



2,2,1 Neural Network

Neural network of depth 2

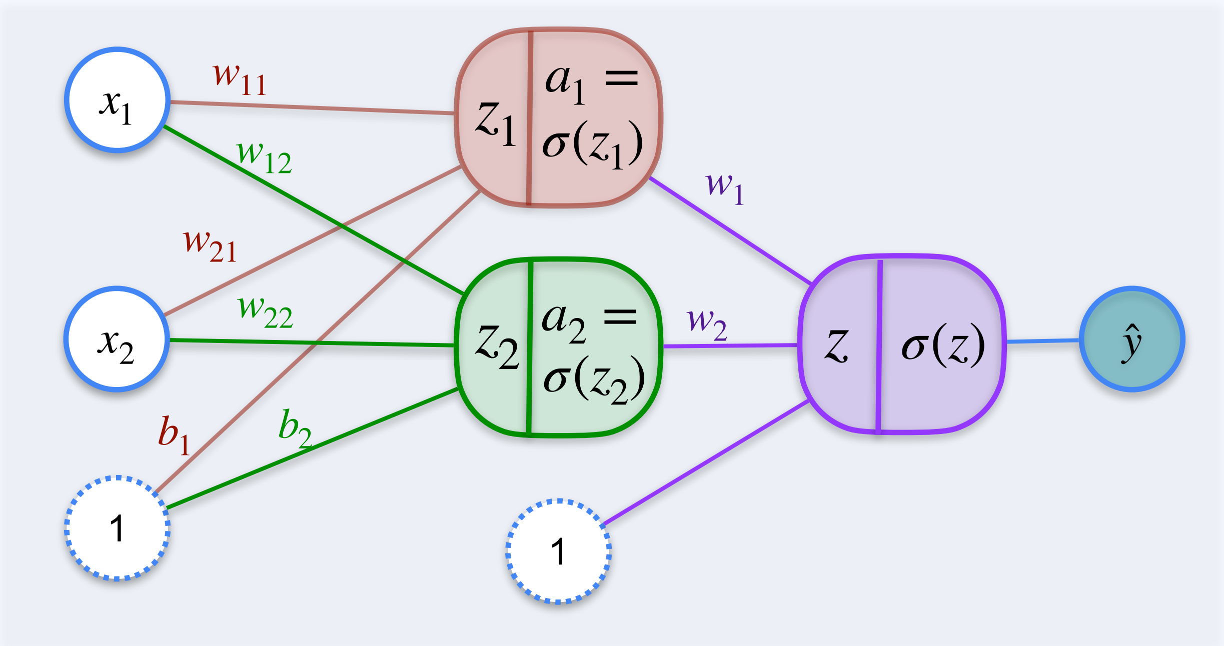
- one input layer
- one hidden layer
- one output layer



2,2,1 Neural Network

Neural network of depth 2

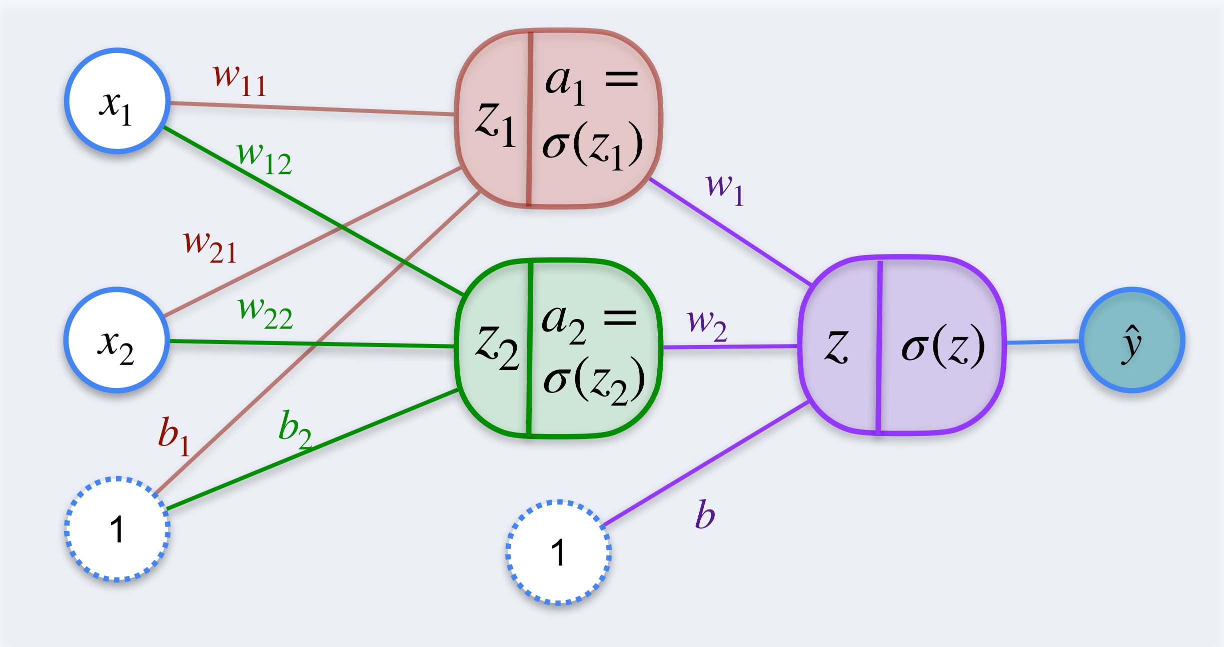
- one input layer
- one hidden layer
- one output layer



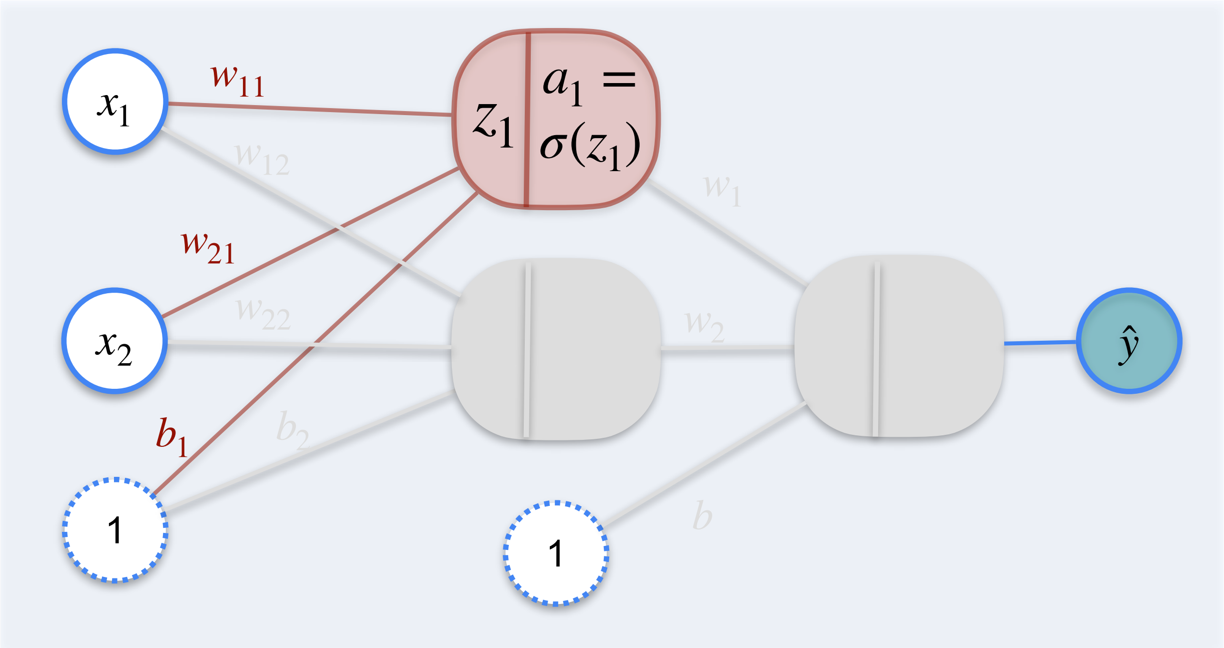
2,2,1 Neural Network

Neural network of depth 2

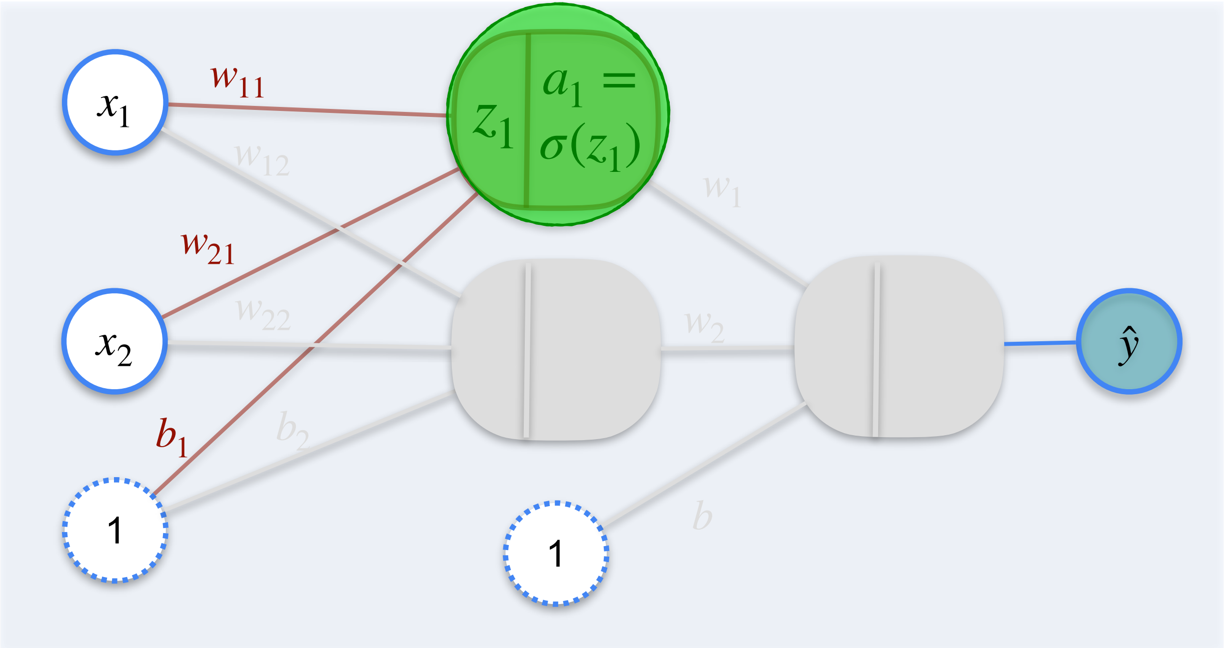
- one input layer
- one hidden layer
- one output layer



2,2,1 Neural Network

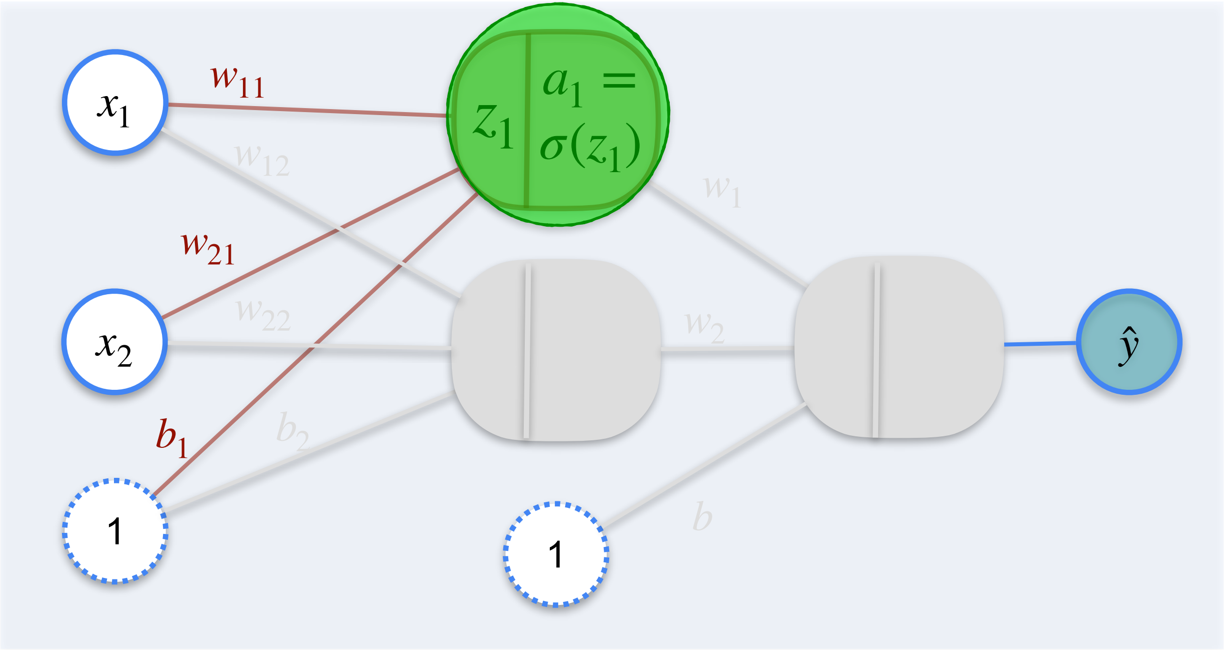


2,2,1 Neural Network



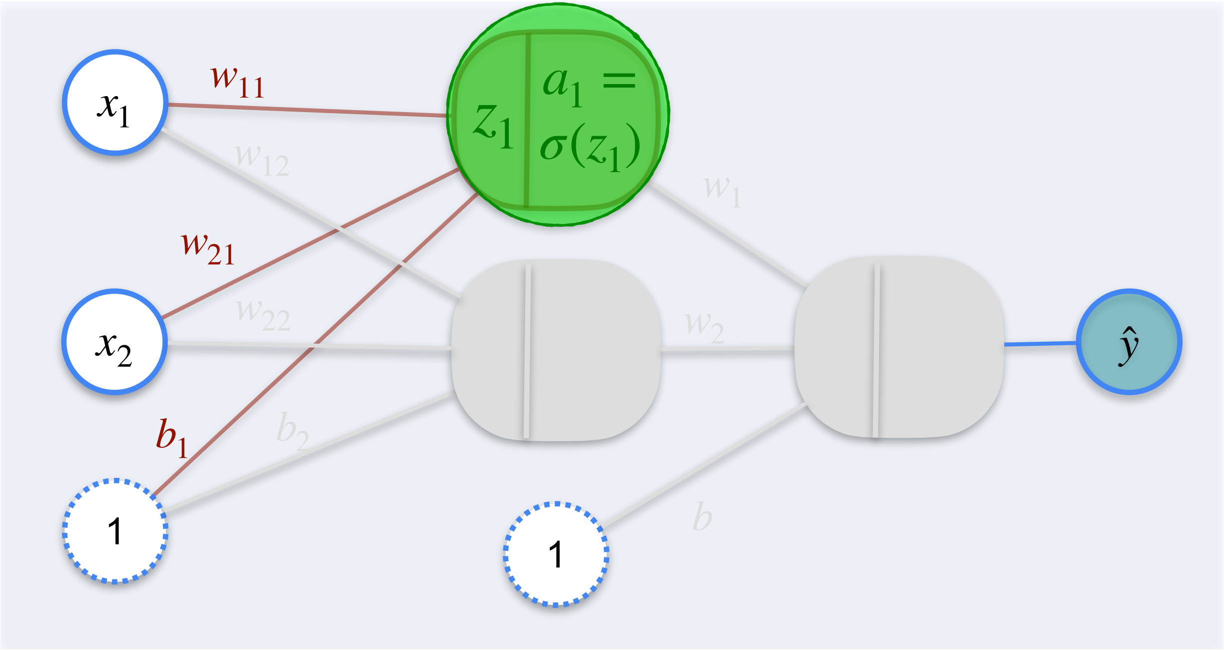
2,2,1 Neural Network

a_1



2,2,1 Neural Network

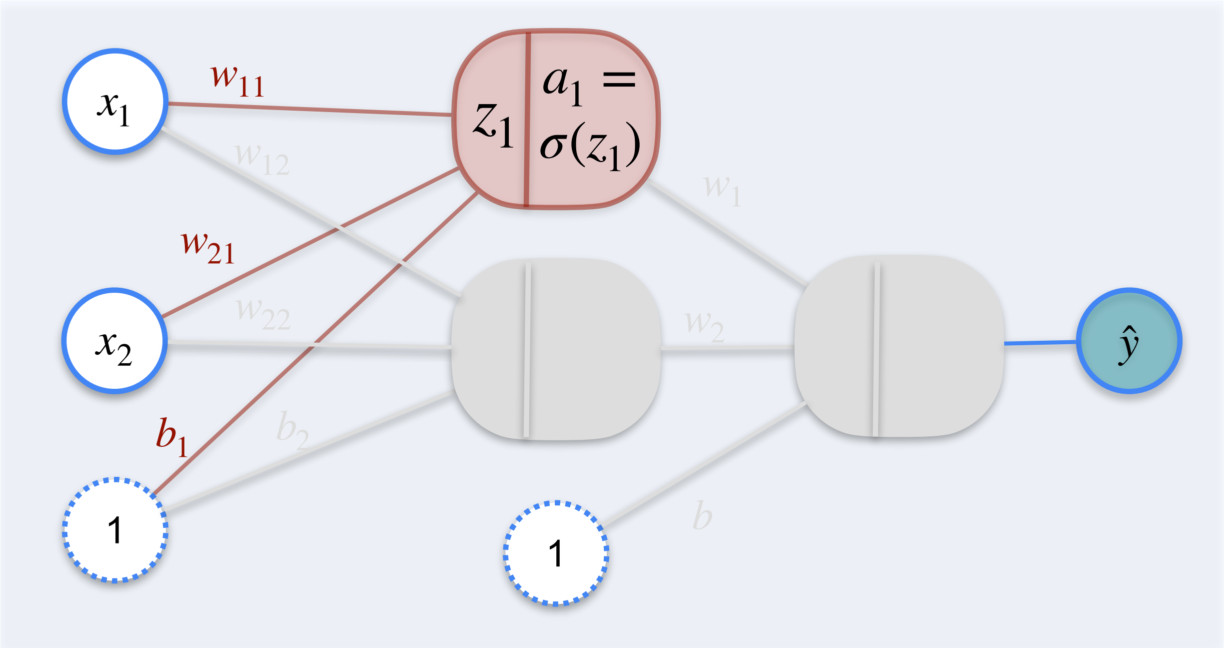
$$a_1 = \sigma(z_1)$$



2,2,1 Neural Network

$a_1 = \sigma(z_1)$

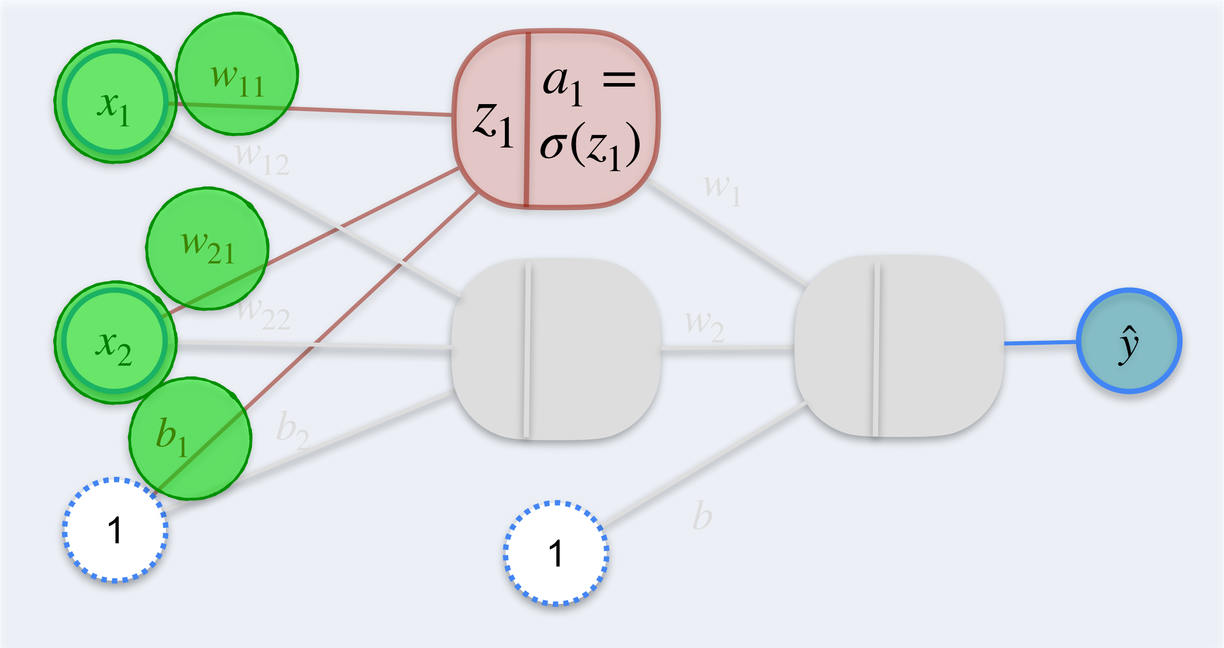
z_1



2,2,1 Neural Network

$a_1 = \sigma(z_1)$

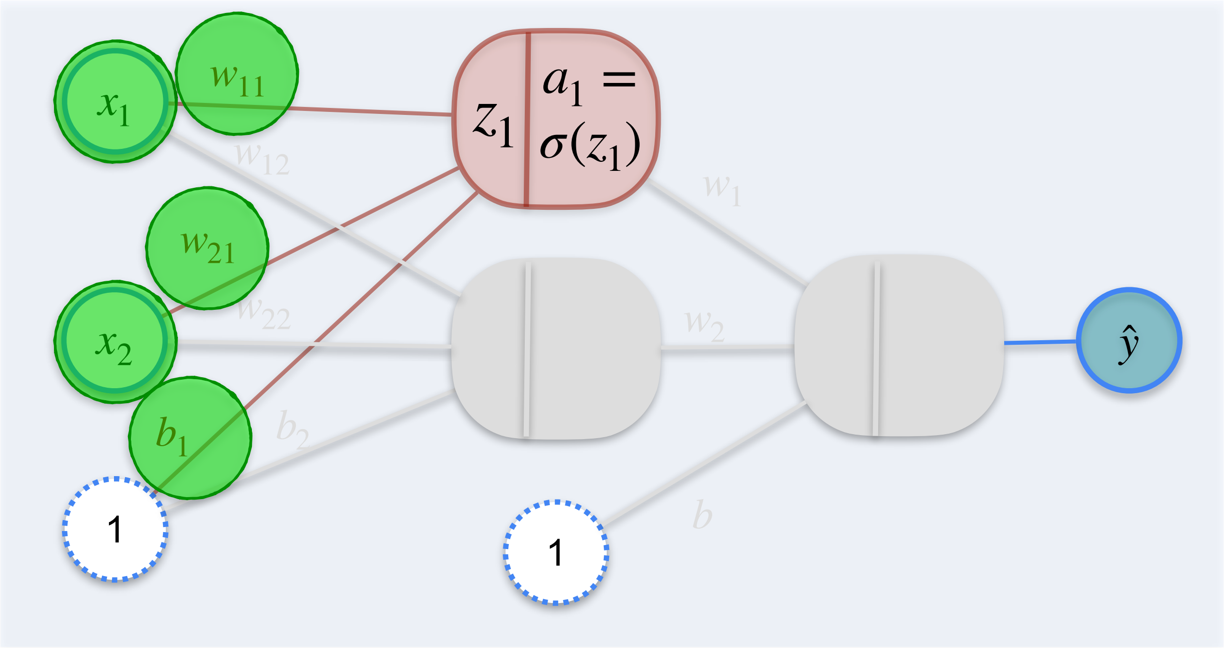
z_1



2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

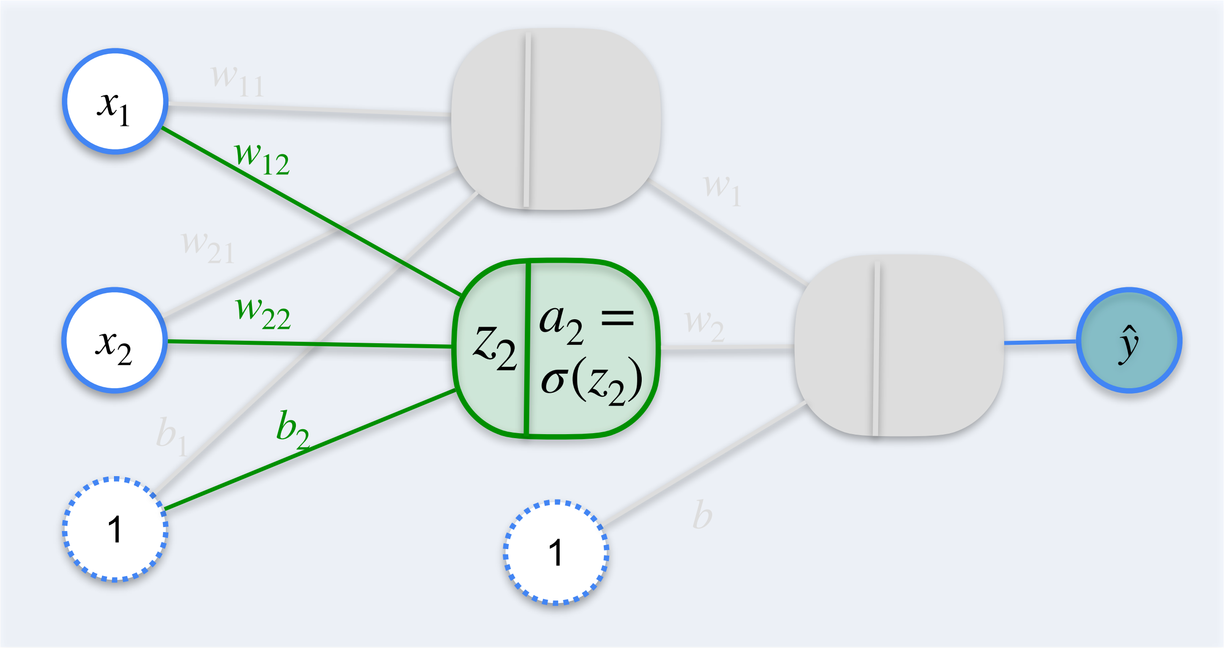
$$z_1 = x_1 w_{11} \; + \; x_2 w_{21} \; + \; b_1$$



2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

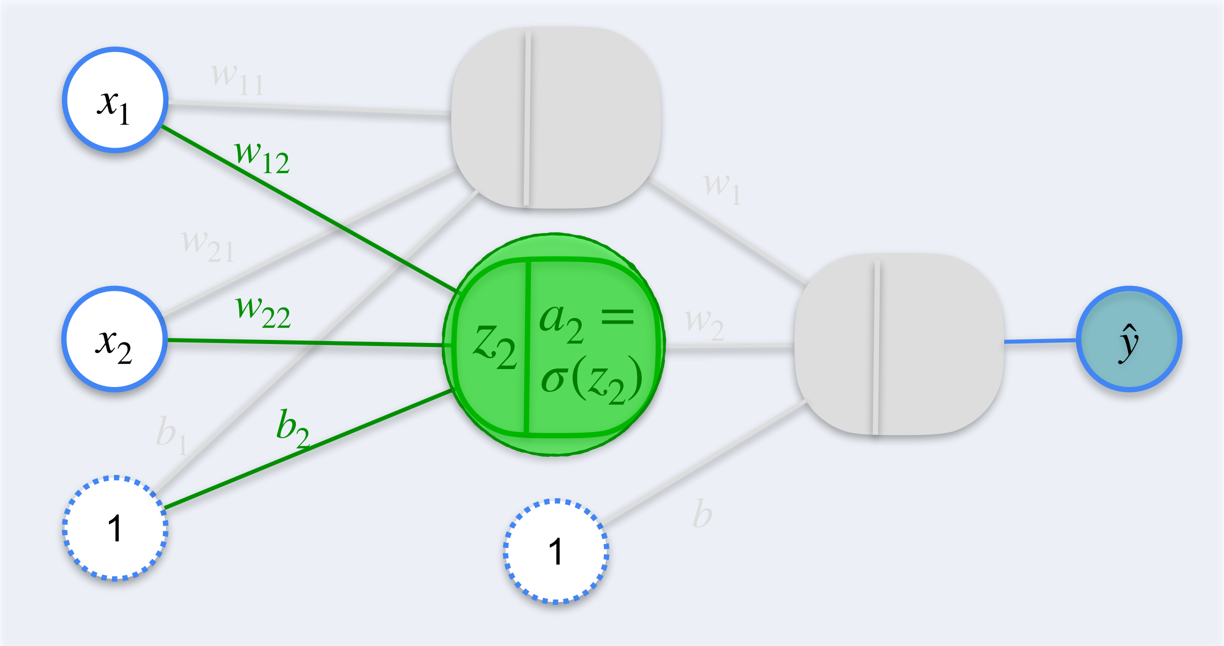
$$z_1 = x_1 w_{11} \ + \ x_2 w_{21} \ + \ b_1$$



2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

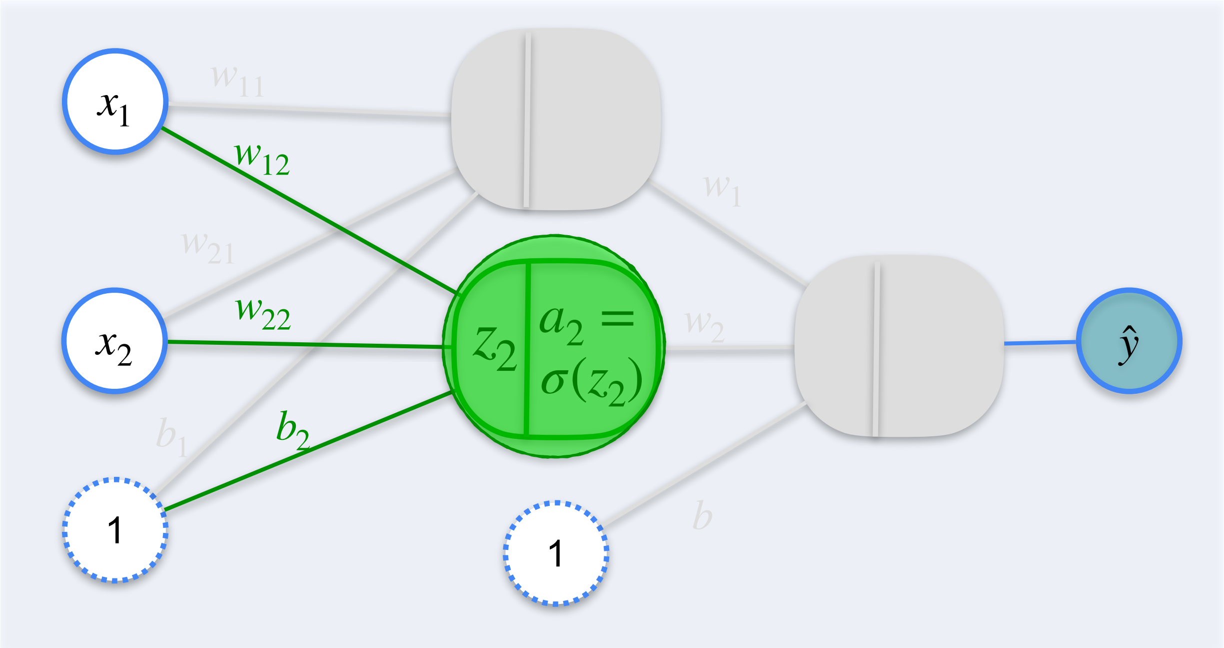


2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} \ + \ x_2 w_{21} \ + \ b_1$$

$$a_2$$

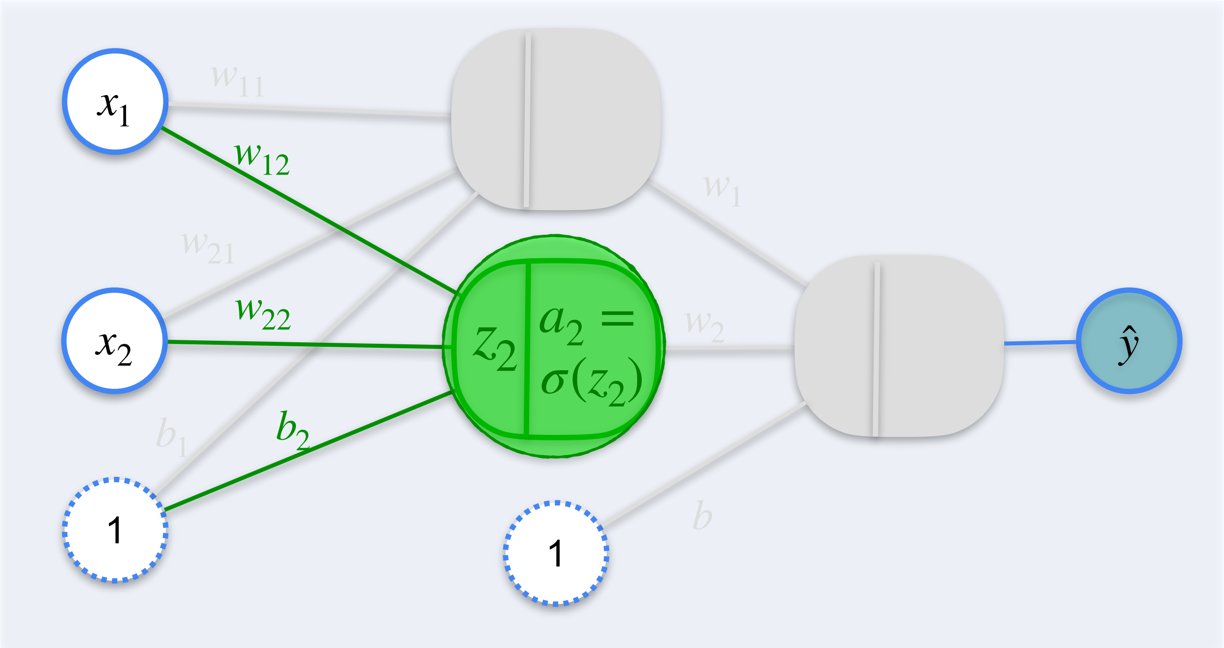


2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} \; + \; x_2 w_{21} \; + \; b_1$$

$$a_2 = \sigma(z_2)$$



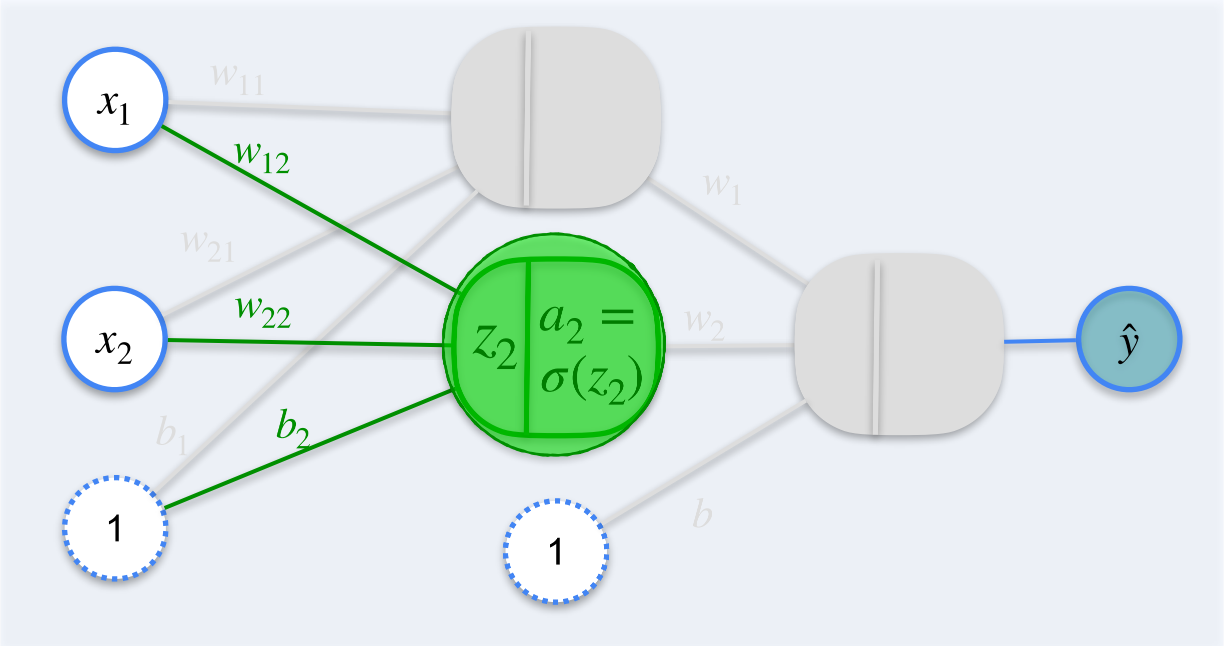
2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2$$



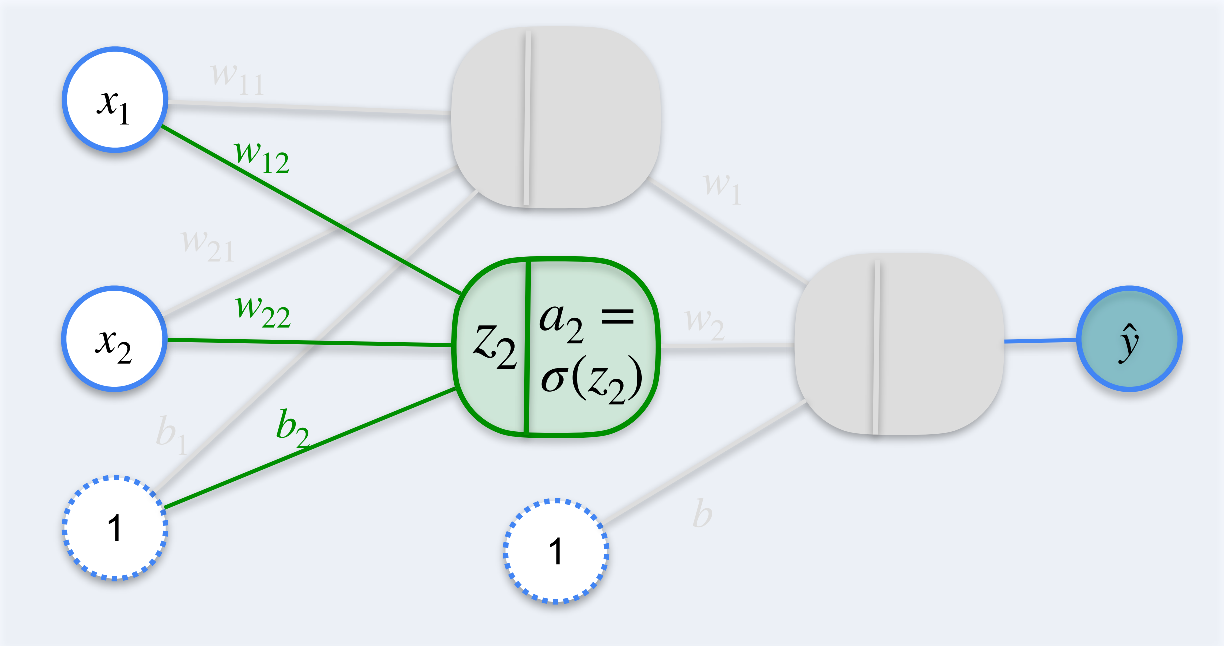
2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} \;+\; x_2 w_{21} \;+\; b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2$$



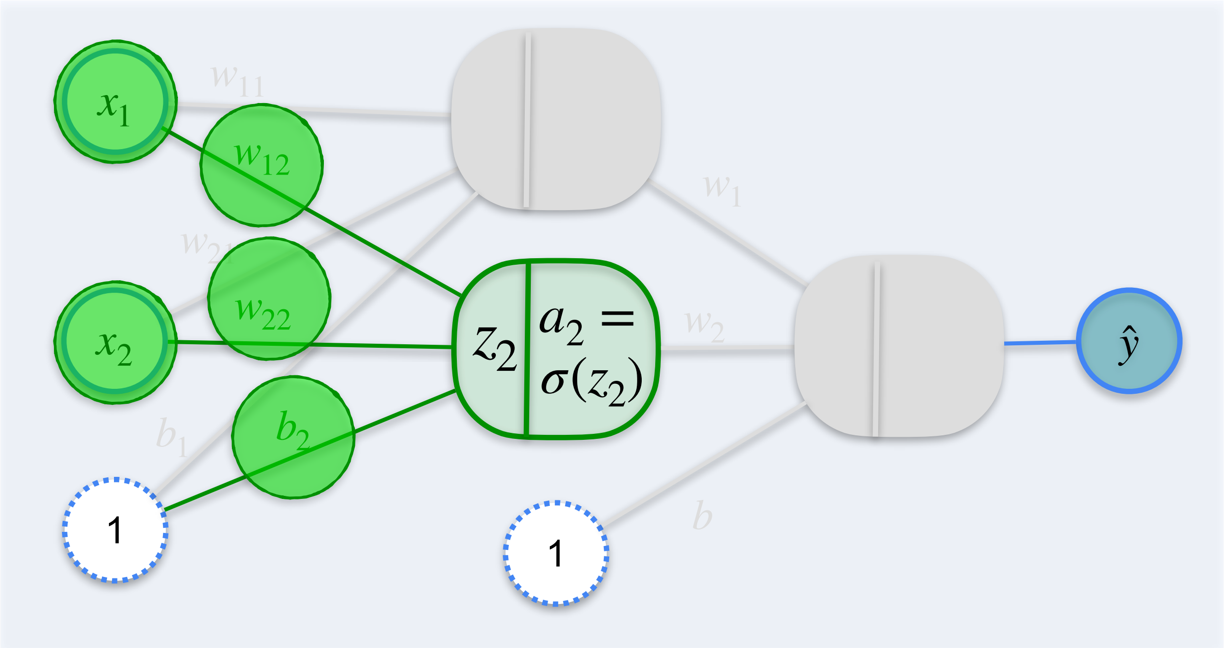
2,2,1 Neural Network

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} \;+\; x_2 w_{21} \;+\; b_1$

$a_2 = \sigma(z_2)$

z_2



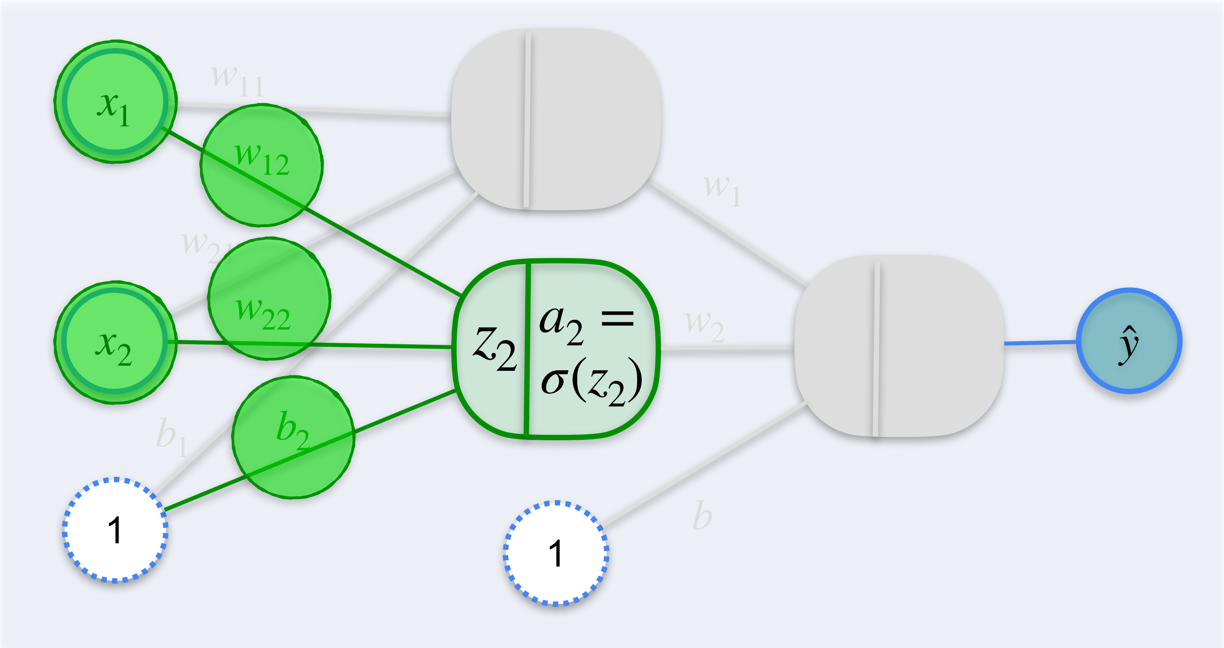
2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$$



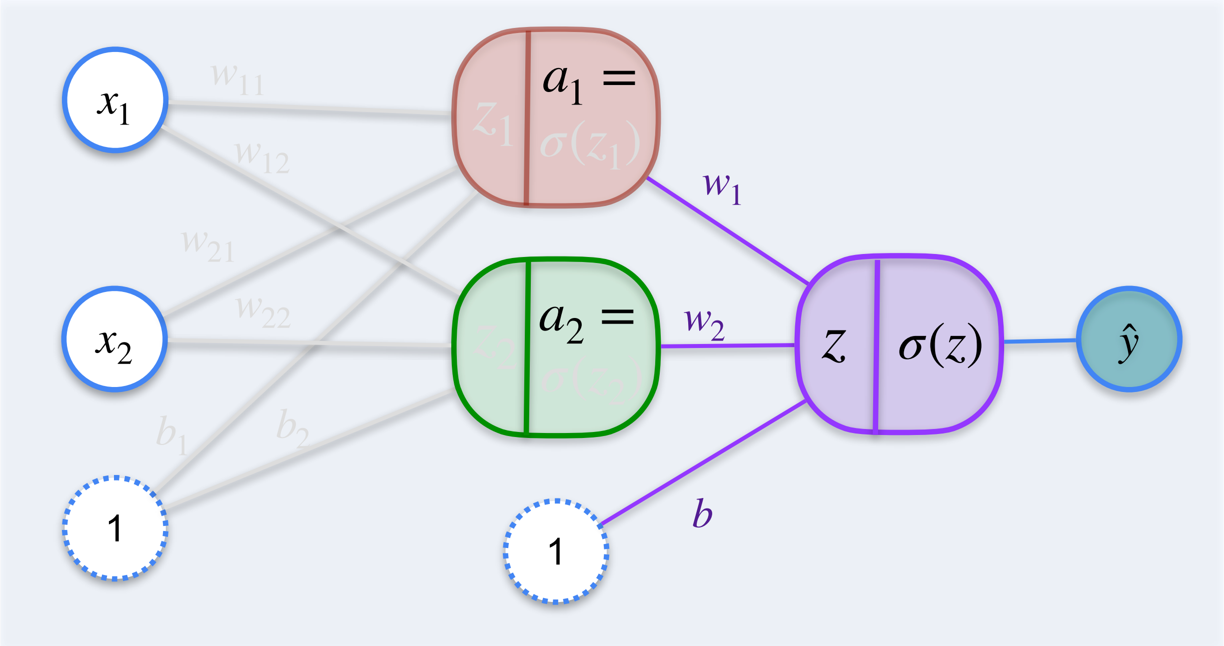
2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} \; + \; x_2 w_{21} \; + \; b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} \; + \; x_2 w_{22} \; + \; b_2$$



2,2,1 Neural Network

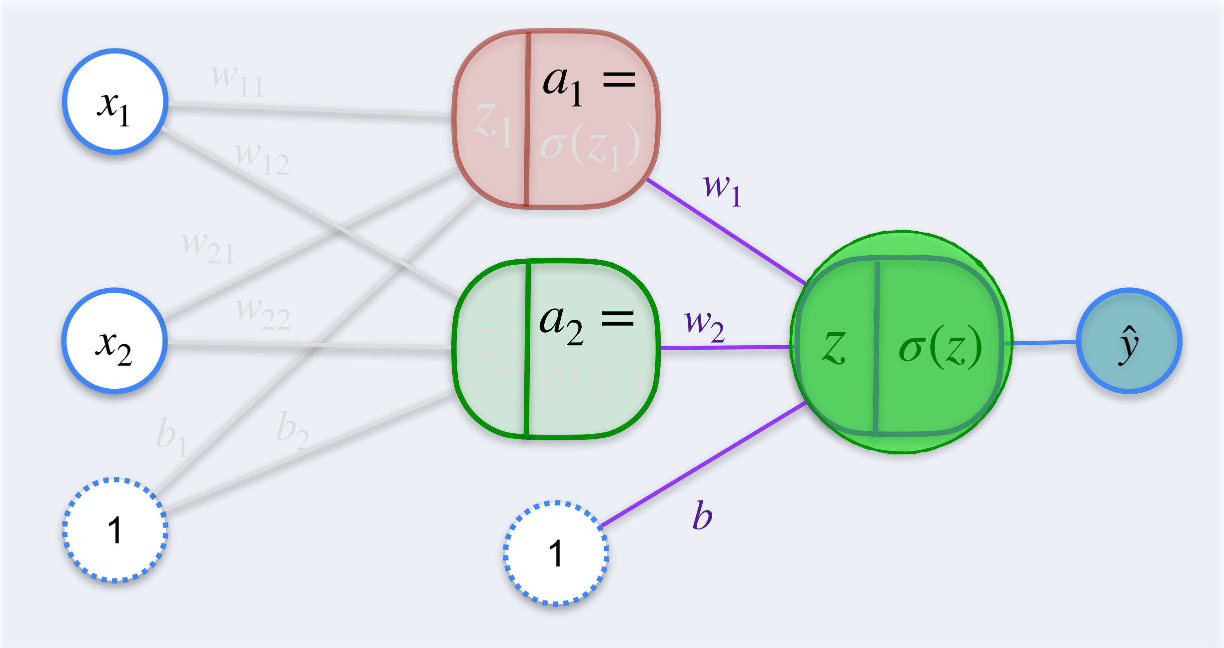
$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} \ + \ x_2 w_{21} \ + \ b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} \ + \ x_2 w_{22} \ + \ b_2$$

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



2,2,1 Neural Network

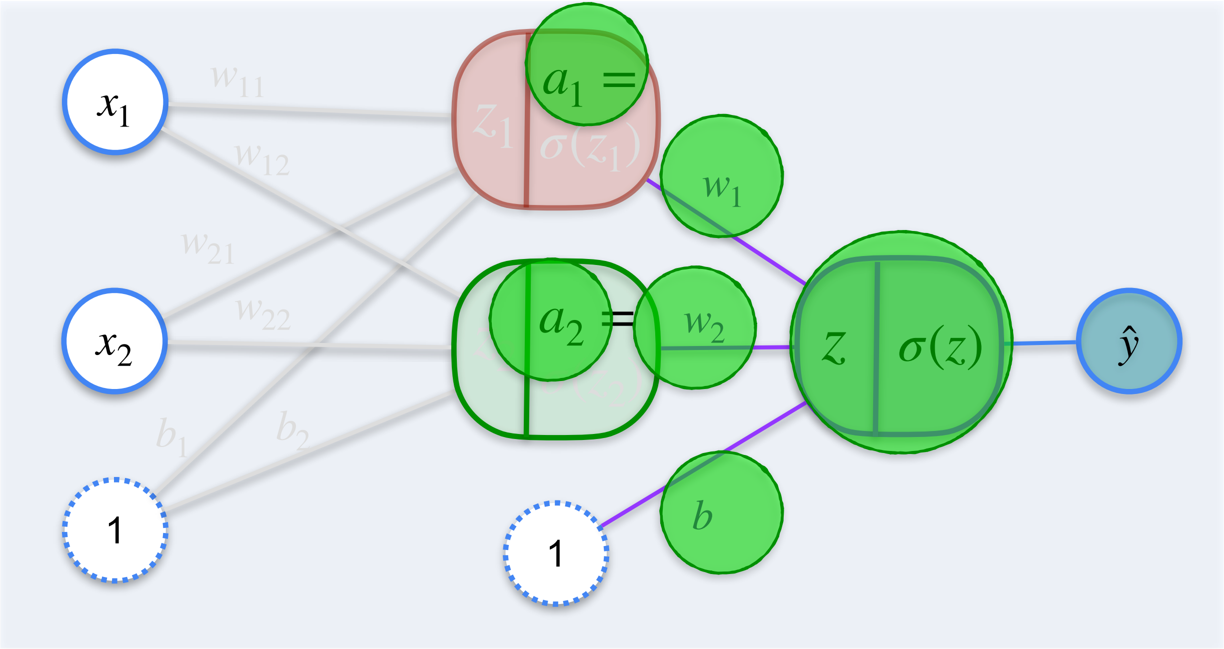
$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} \ + \ x_2 w_{21} \ + \ b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} \ + \ x_2 w_{22} \ + \ b_2$$

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



2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

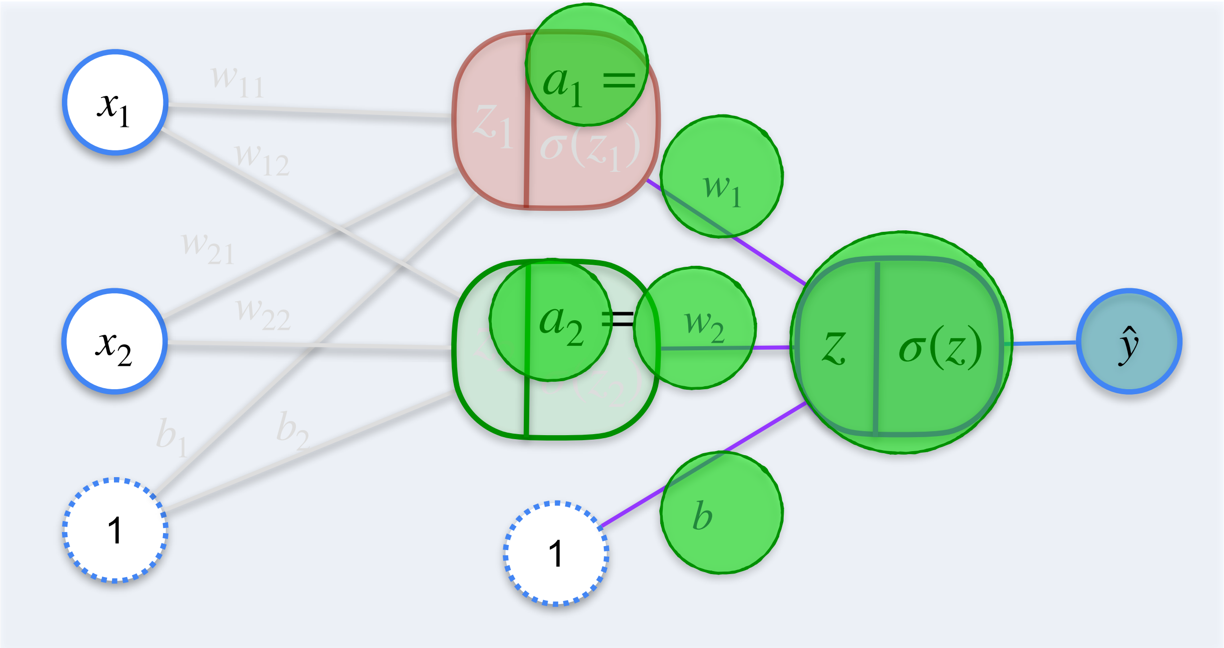
$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$$

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

$$z = a_1 w_1 + a_2 w_2 + b$$



2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

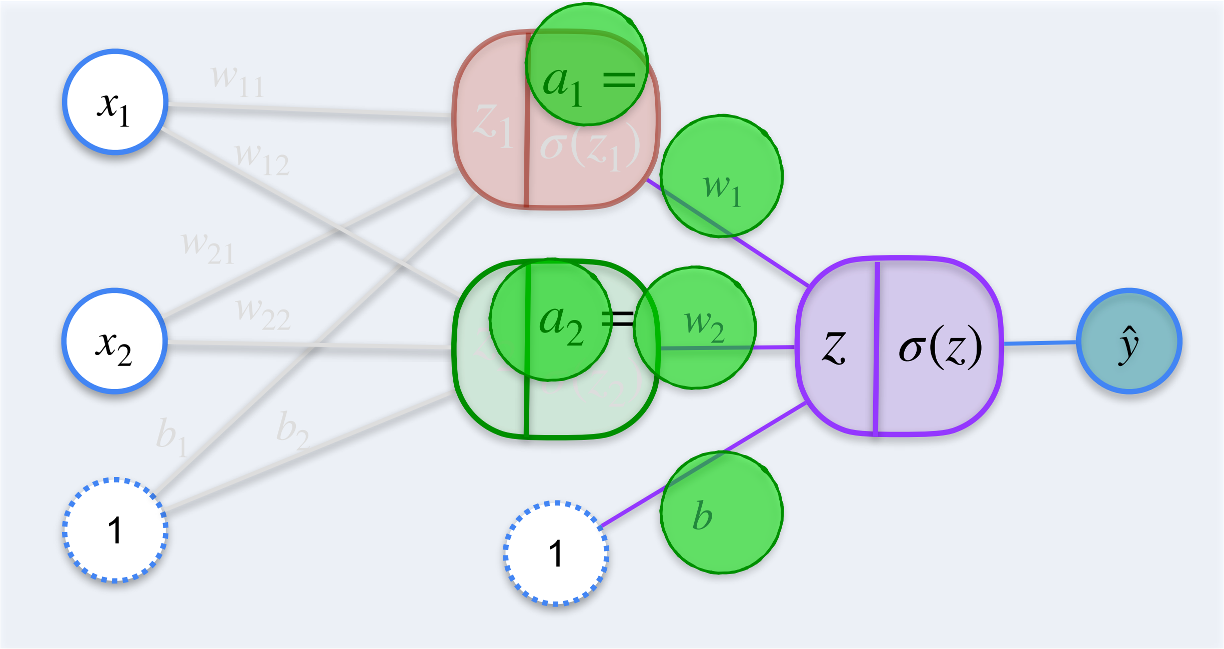
$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$$

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

$$z = a_1 w_1 + a_2 w_2 + b$$



2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

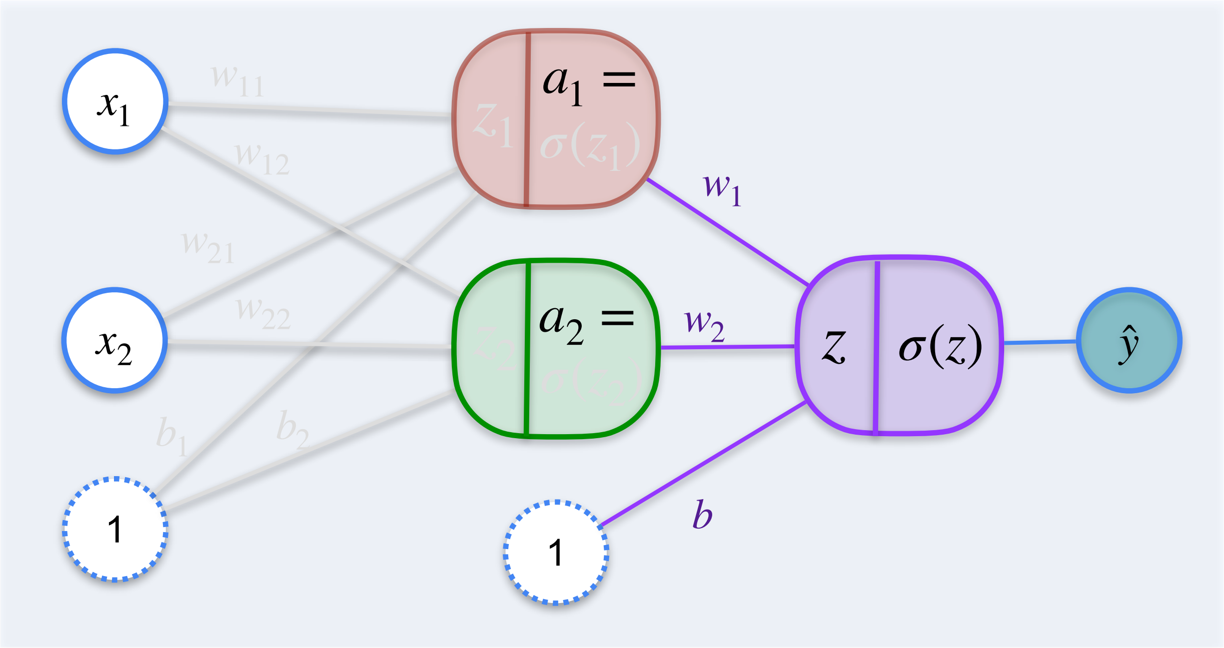
$$z_1 = x_1 w_{11} \ + \ x_2 w_{21} \ + \ b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} \ + \ x_2 w_{22} \ + \ b_2$$

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

$$z = a_1 w_1 \ + \ a_2 w_2 \ + \ b$$



2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

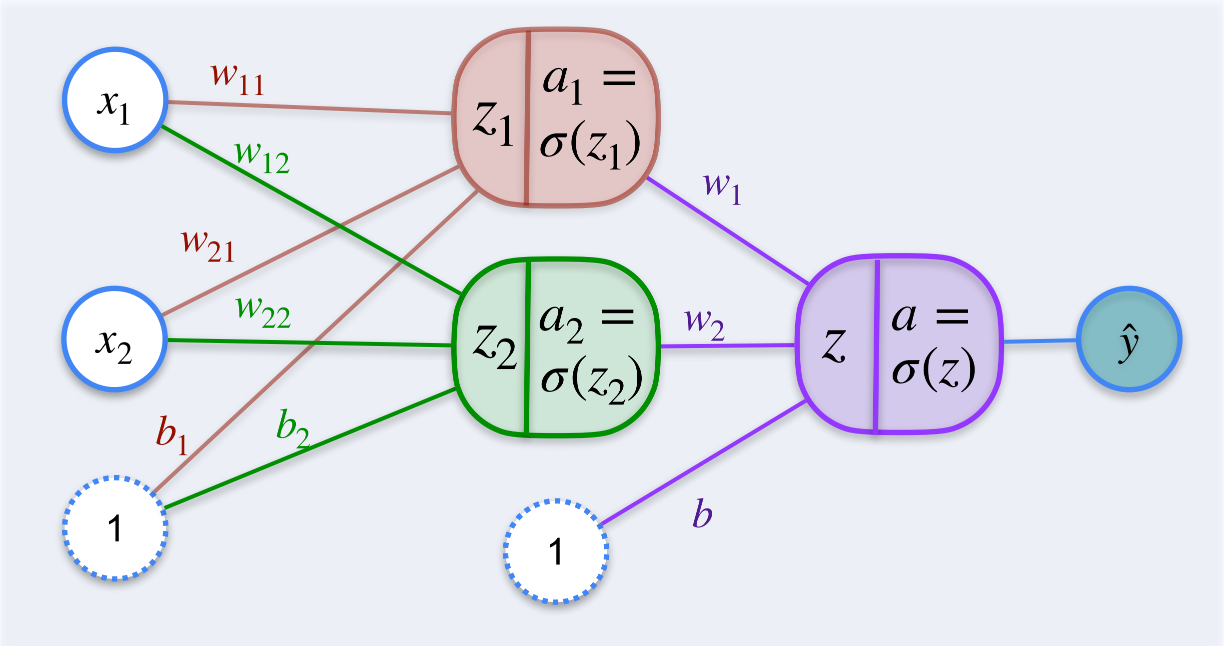
$$z_1 = x_1 w_{11} \ + \ x_2 w_{21} \ + \ b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} \ + \ x_2 w_{22} \ + \ b_2$$

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

$$z = a_1 w_1 \ + \ a_2 w_2 \ + \ b$$



2,2,1 Neural Network

$$a_1 = \sigma(z_1)$$

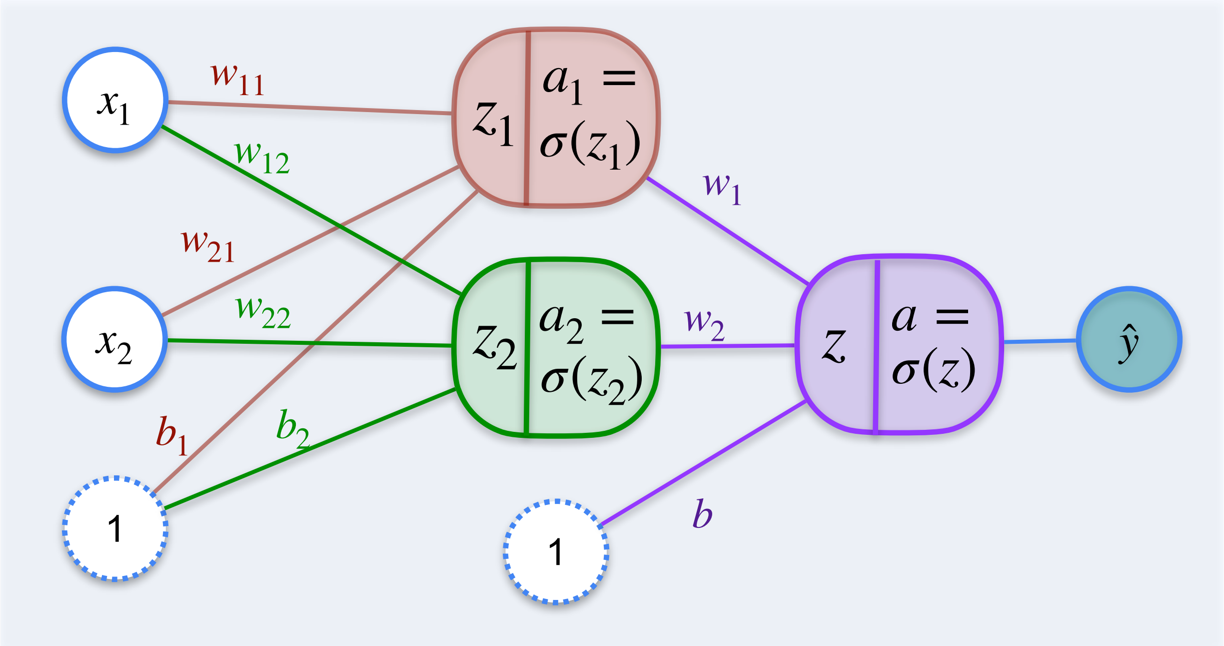
$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$a_2 = \sigma(z_2)$$

$$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$$

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

$$z = a_1 w_1 + a_2 w_2 + b$$



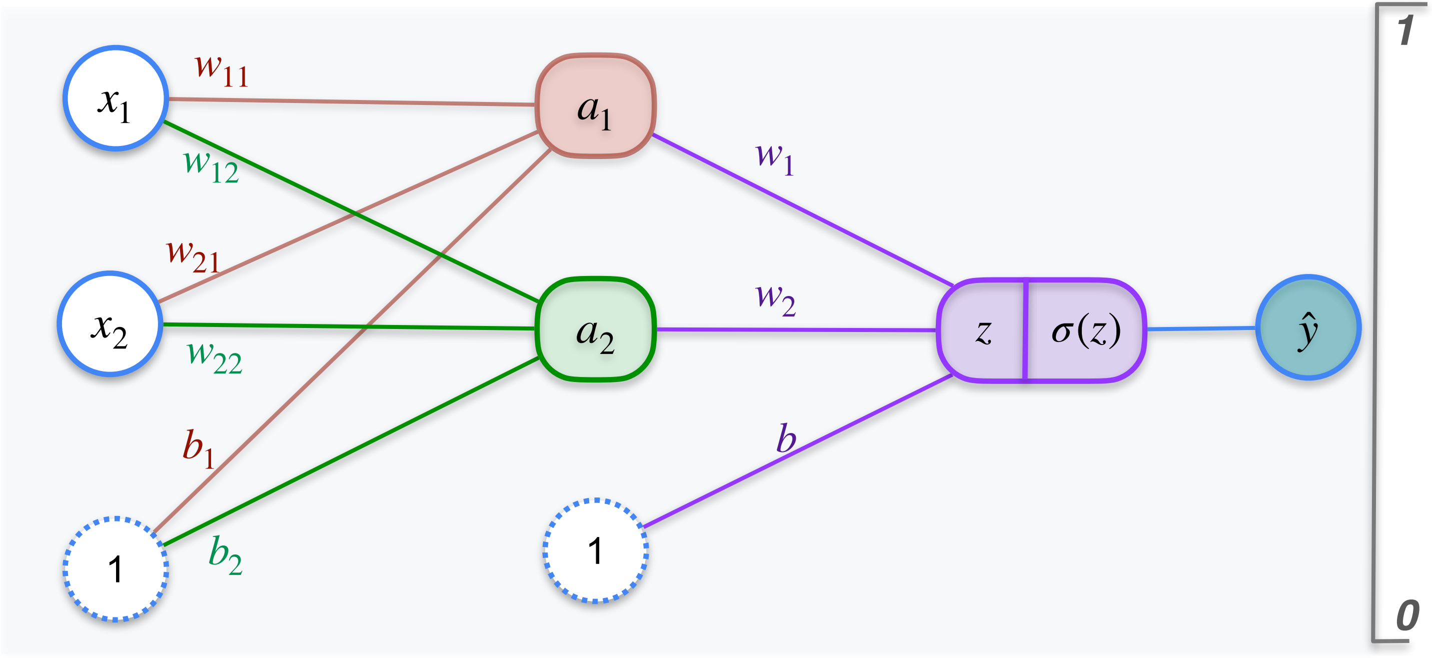
$$L(y, \hat{y}) = -y log(\hat{y}) - (1 - y) log(1 - \hat{y})$$

DeepLearning.AI

Optimization in Neural Networks and Newton's Method

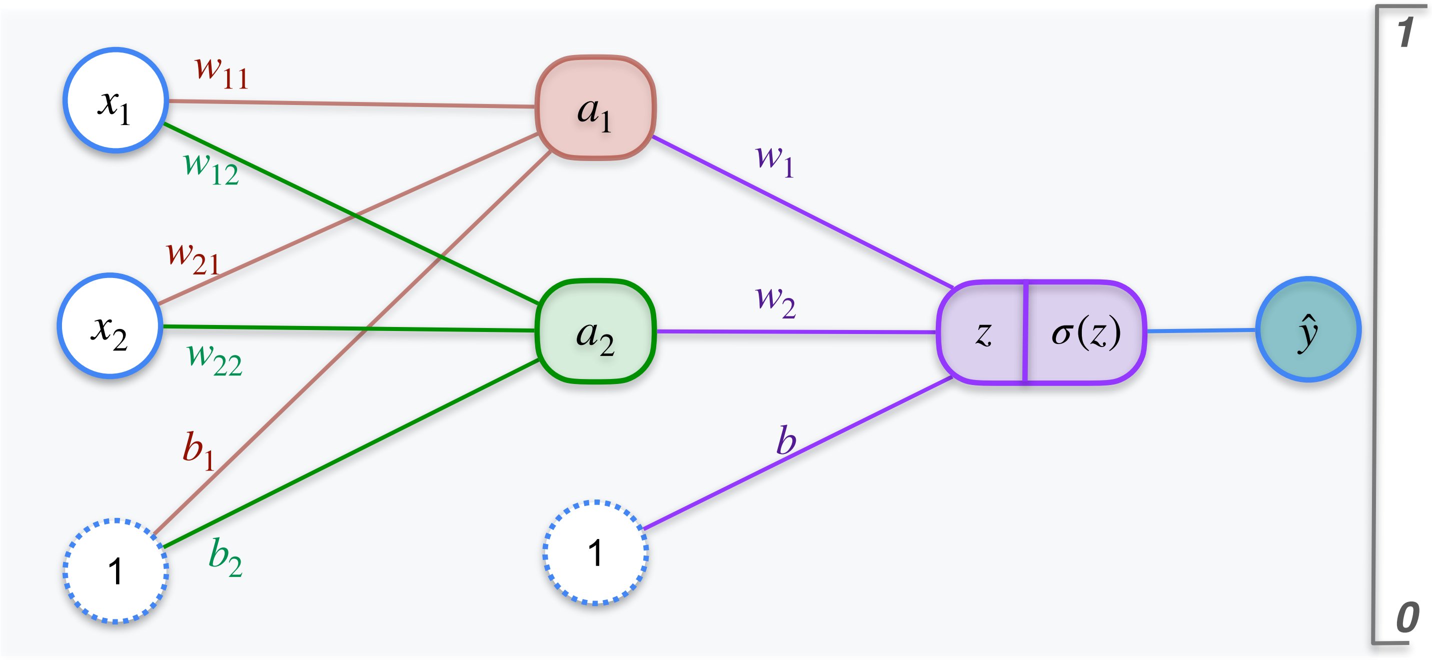
Classification with a Neural Network: Minimizing log-loss

2,2,1 Neural Network



2,2,1 Neural Network

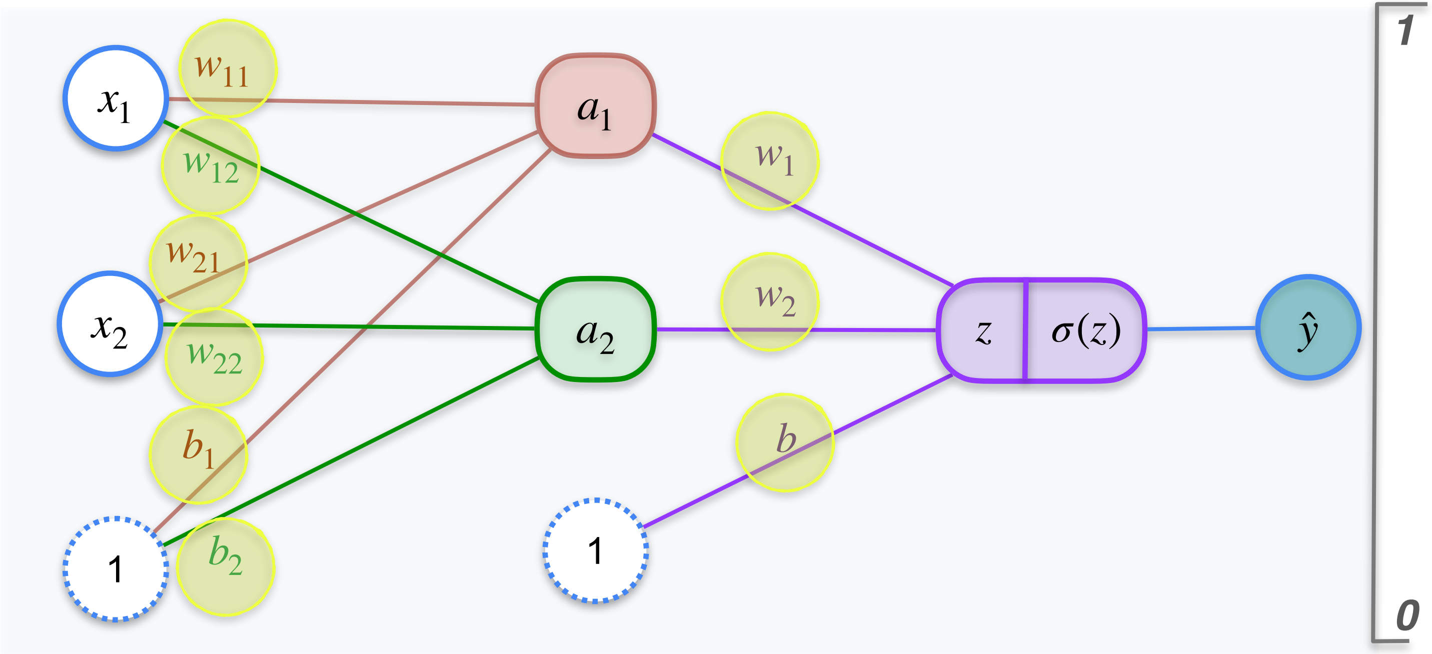
Goal



2,2,1 Neural Network

Goal

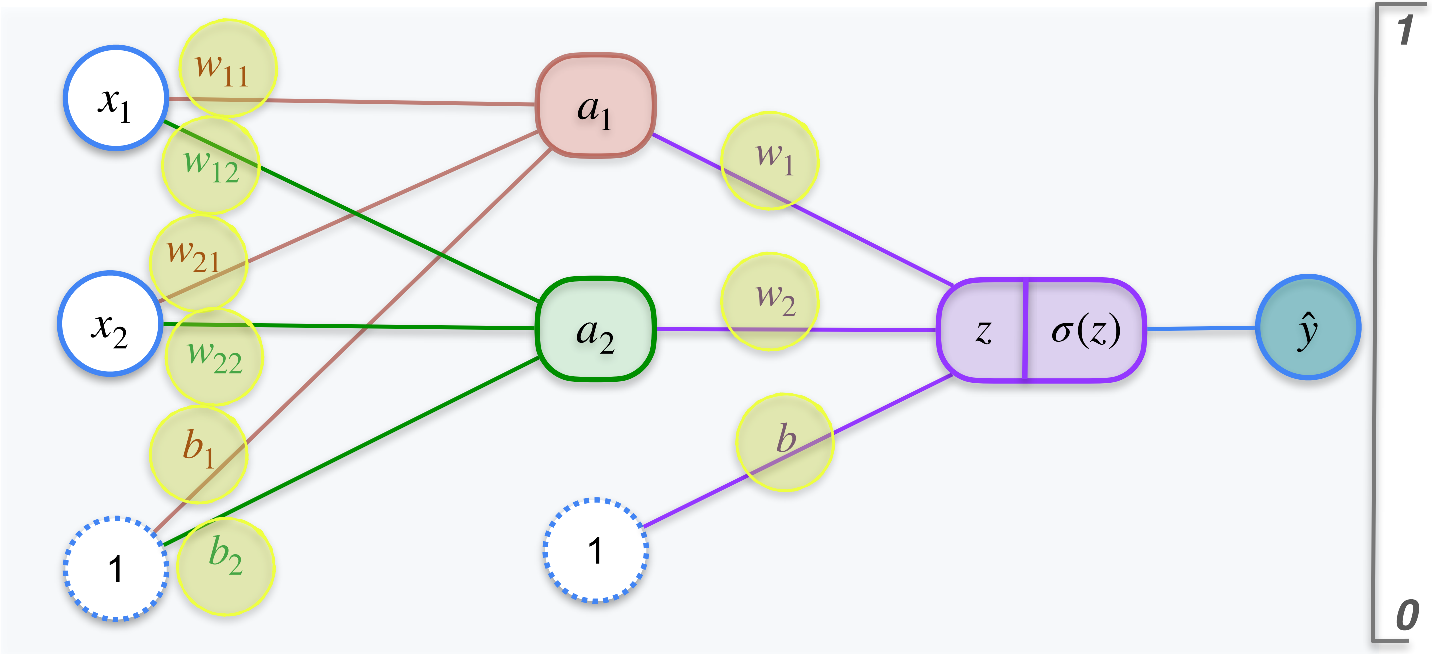
Adjust each of the highlighted weights and biases



2,2,1 Neural Network

Goal

Adjust each of the highlighted weights and biases

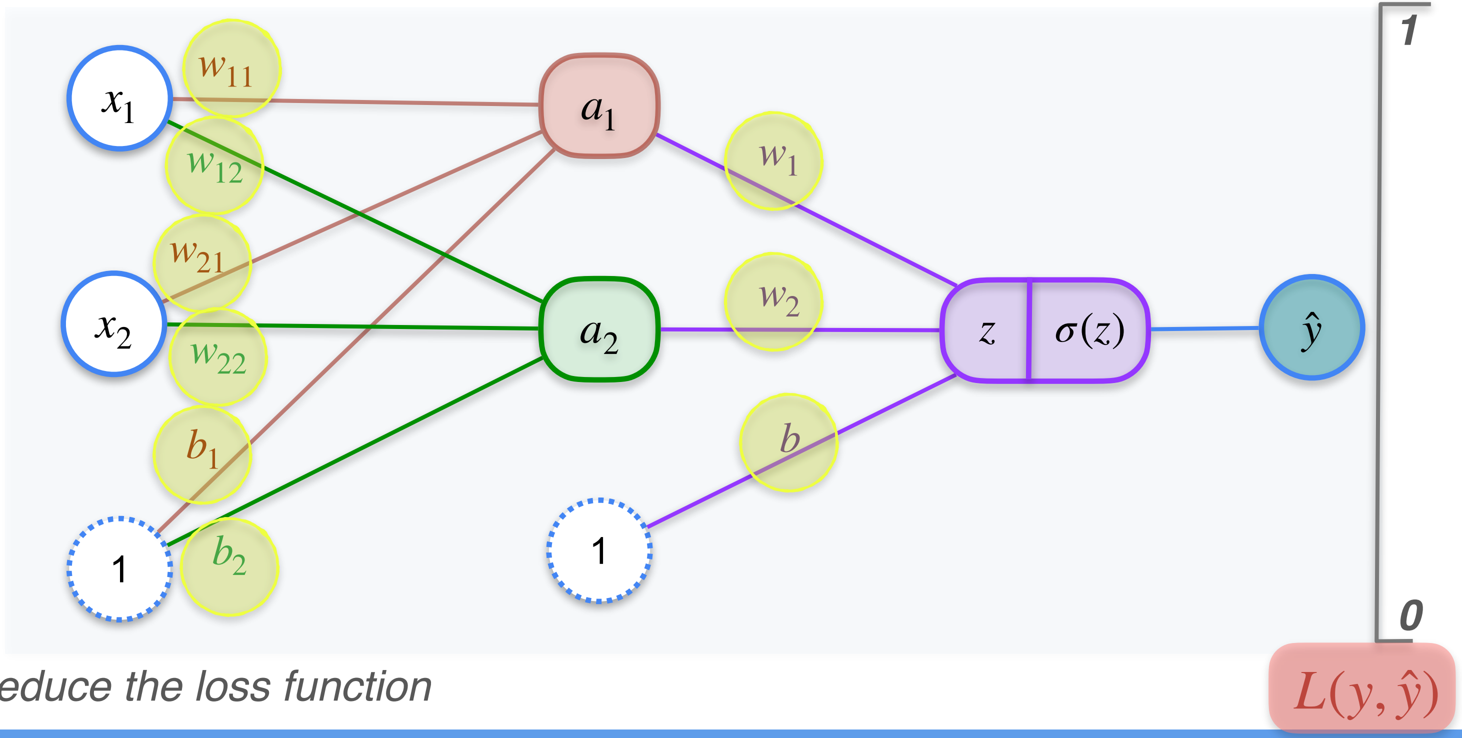


To reduce the loss function

2,2,1 Neural Network

Goal

Adjust each of the highlighted weights and biases

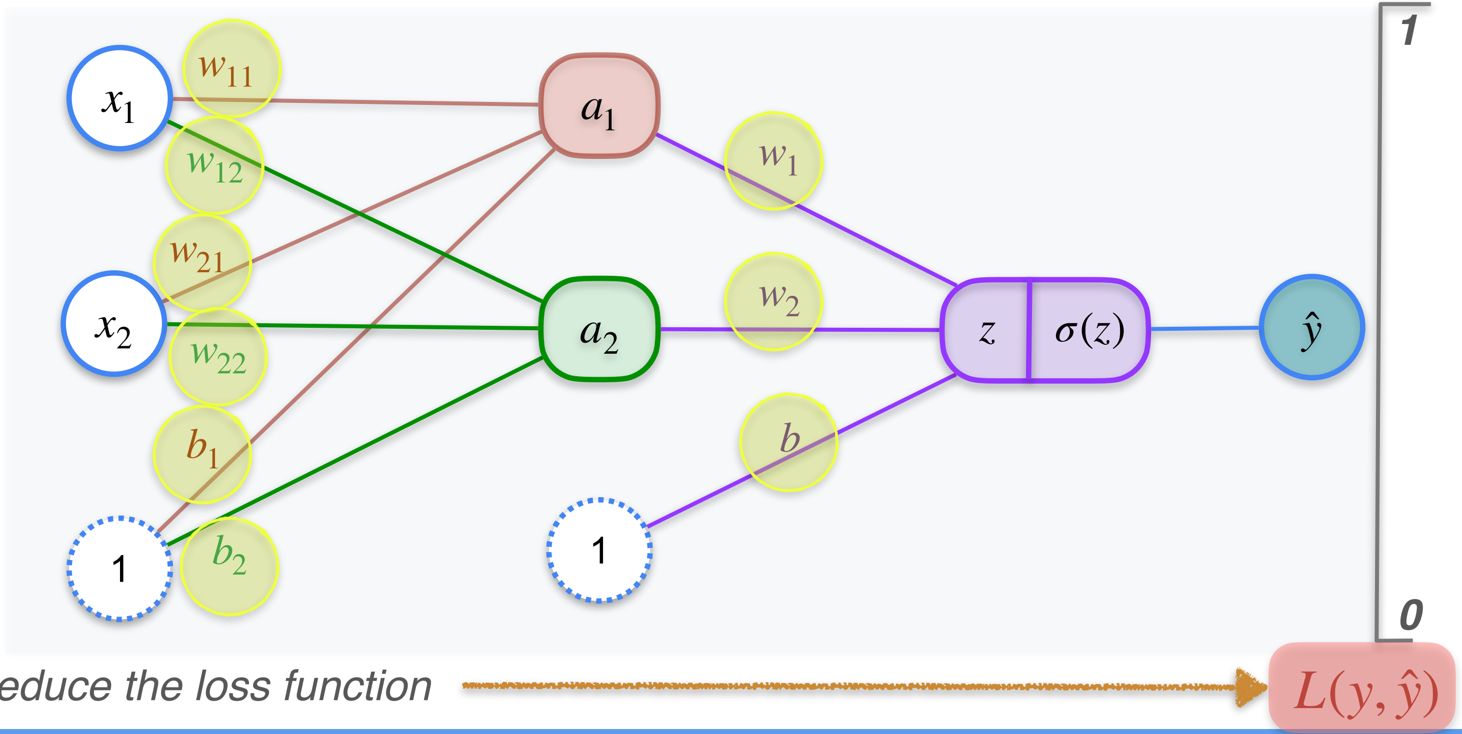


To reduce the loss function

2,2,1 Neural Network

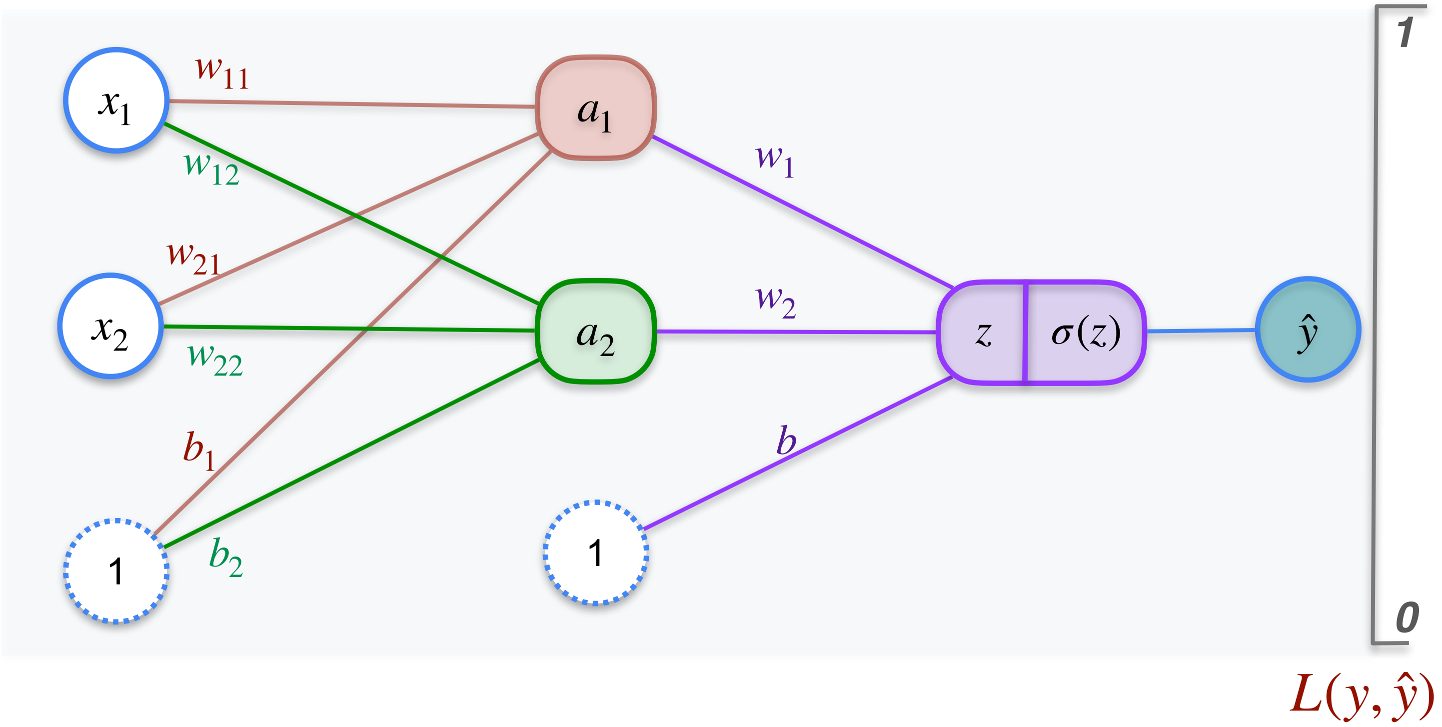
Goal

Adjust each of the highlighted weights and biases

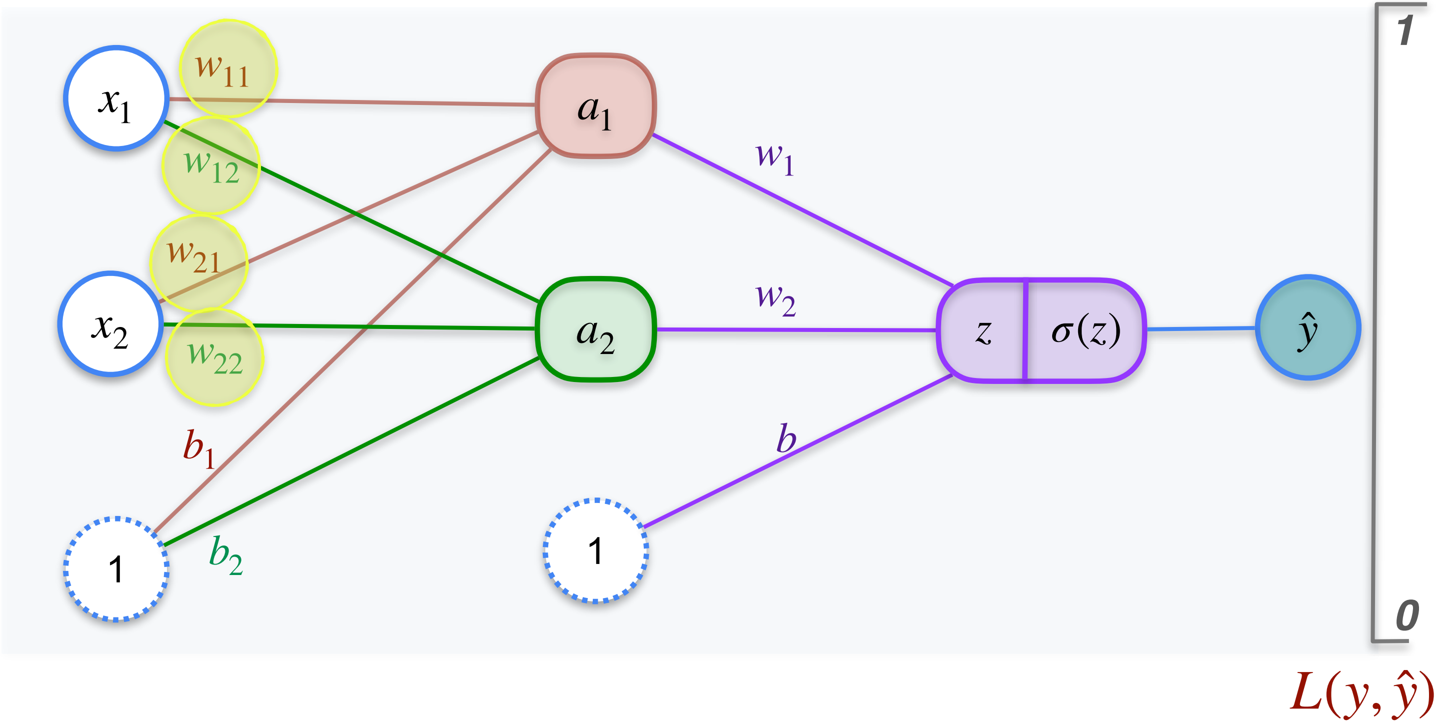


To reduce the loss function

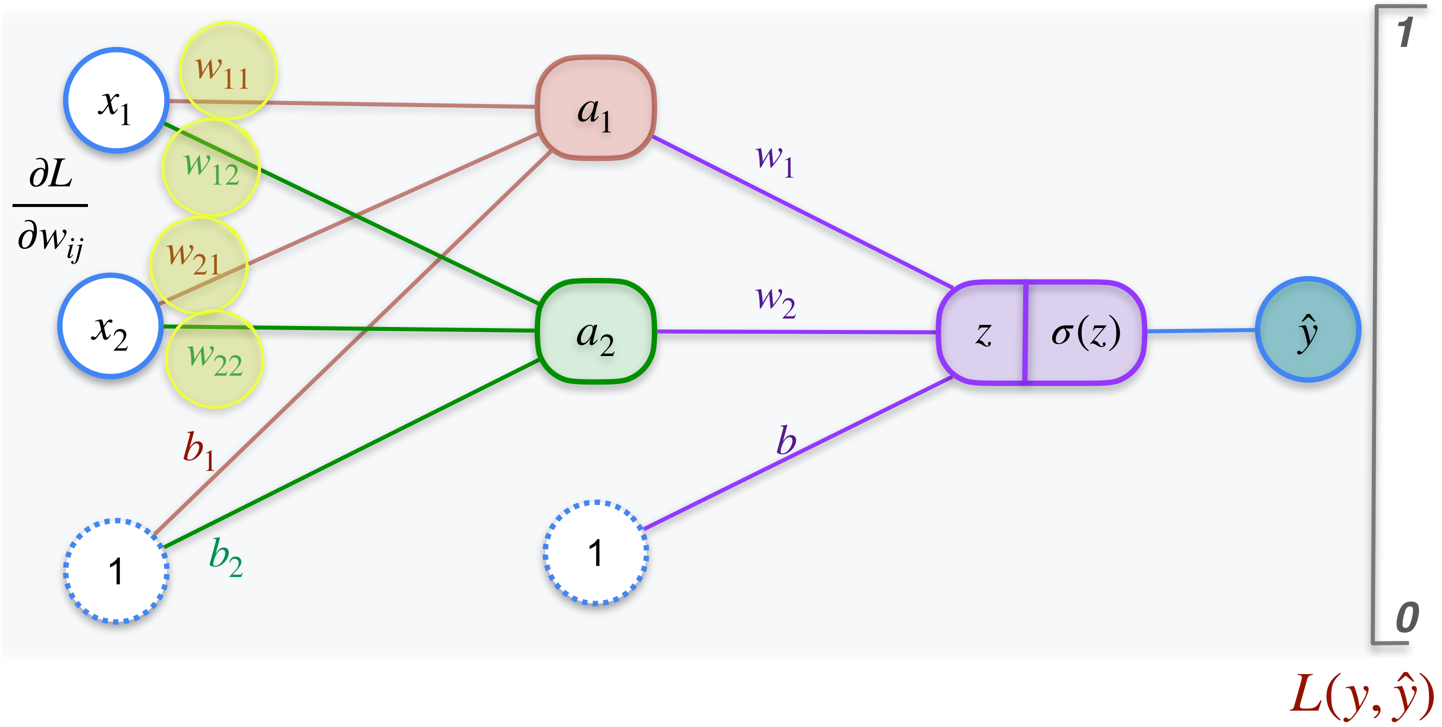
2,2,1 Neural Network



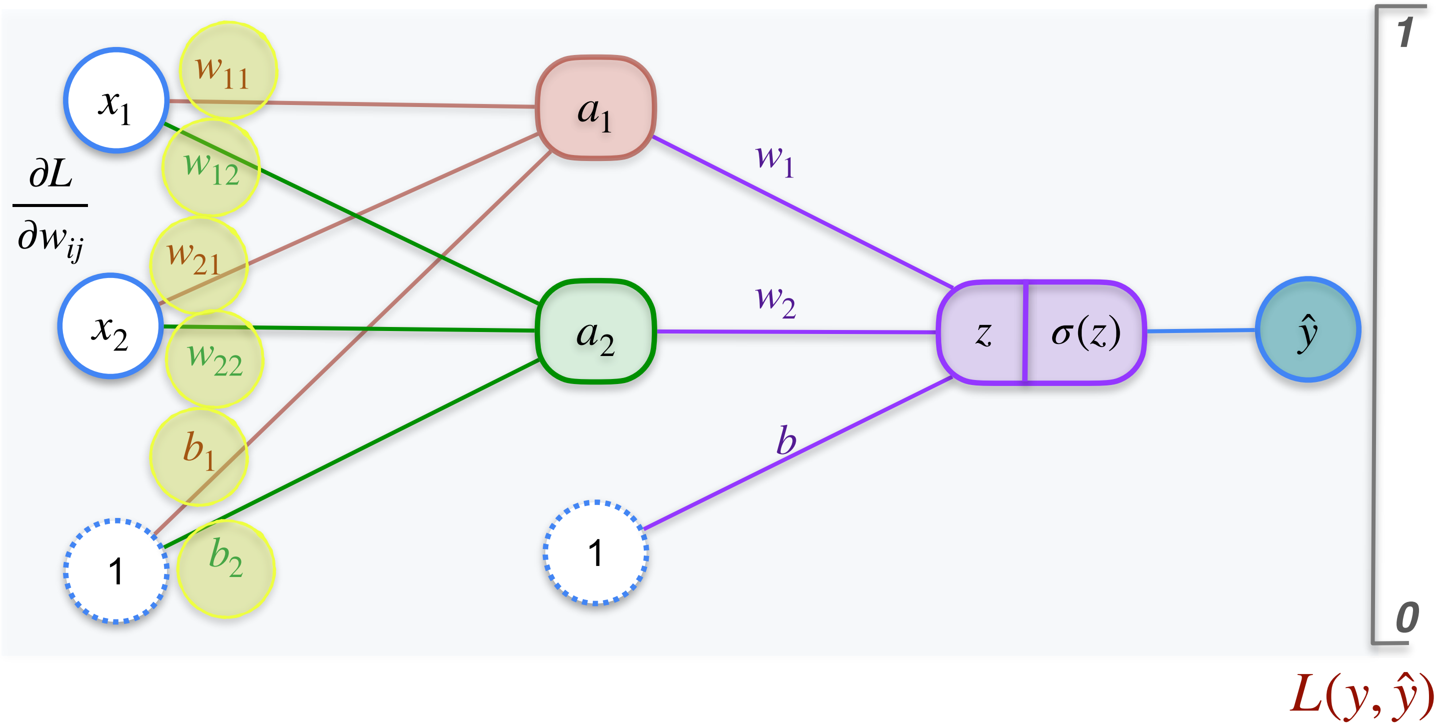
2,2,1 Neural Network



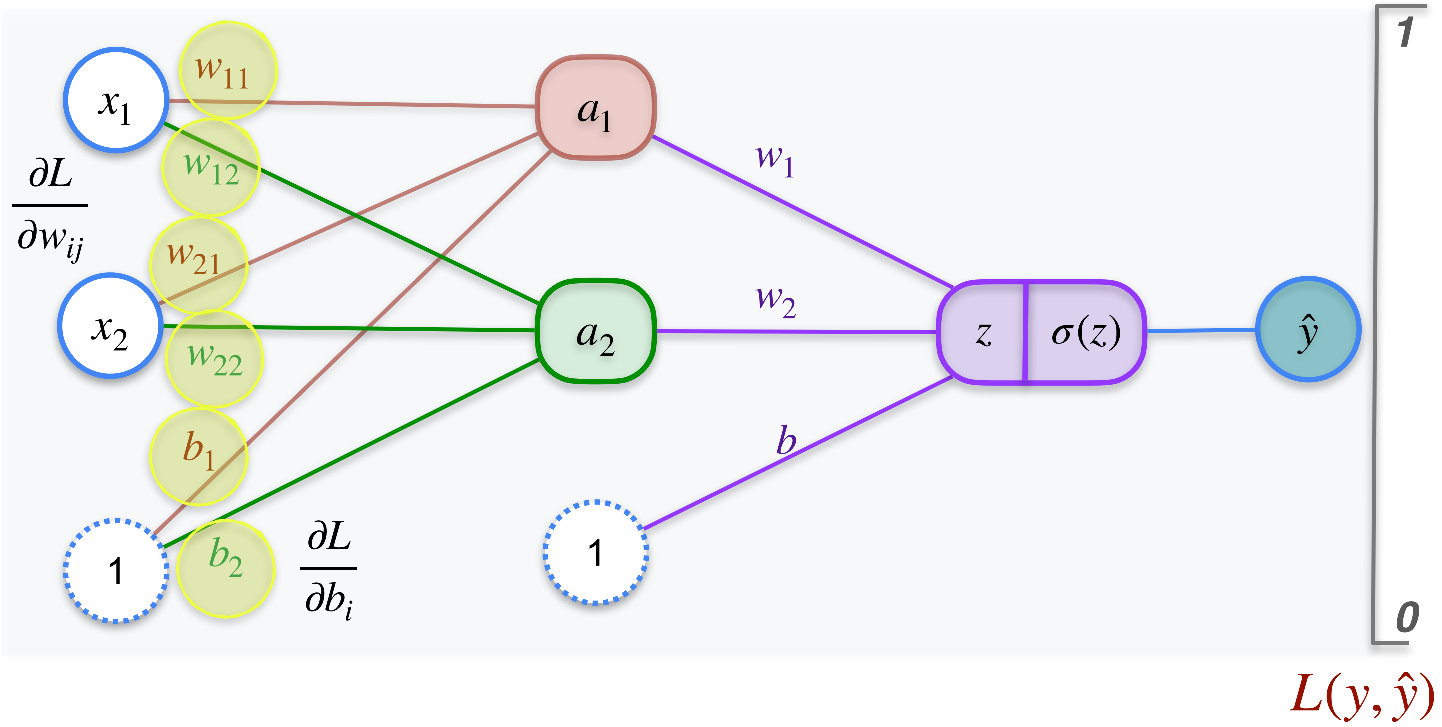
2,2,1 Neural Network



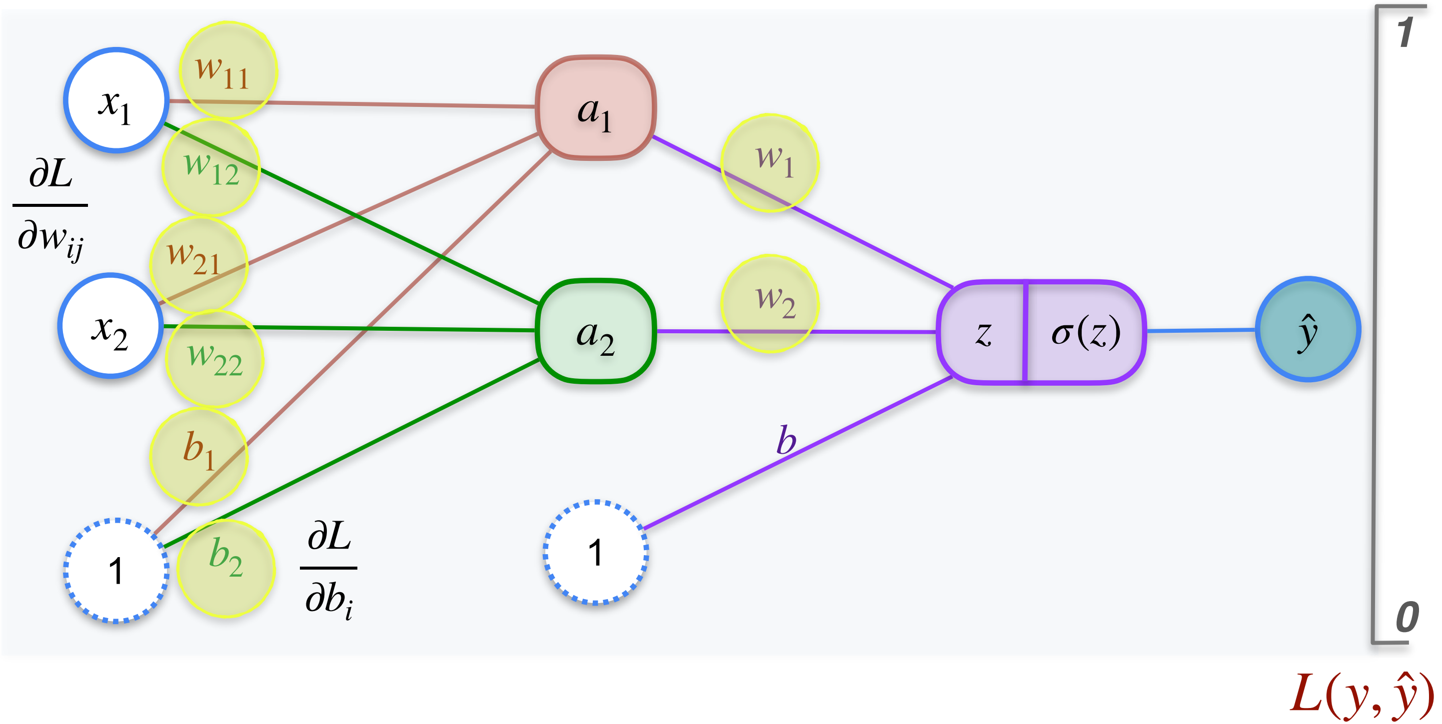
2,2,1 Neural Network



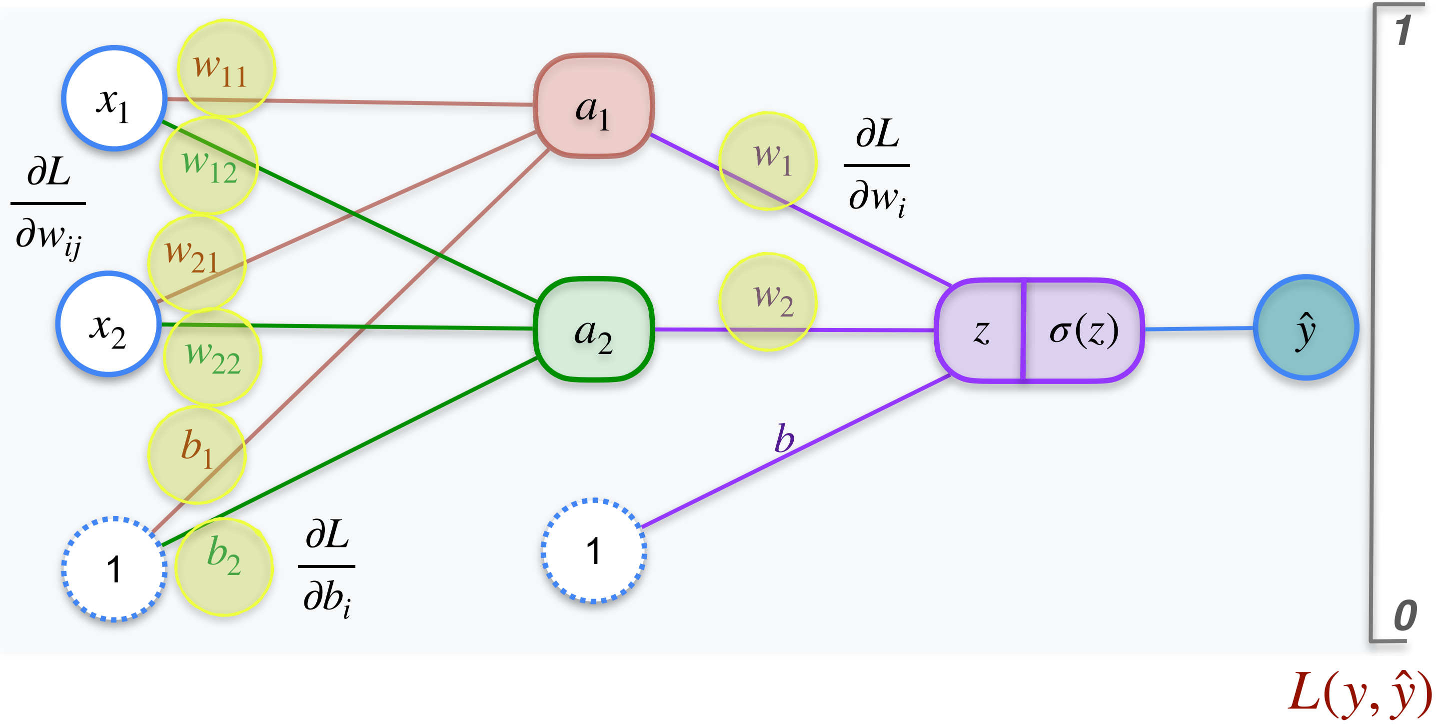
2,2,1 Neural Network



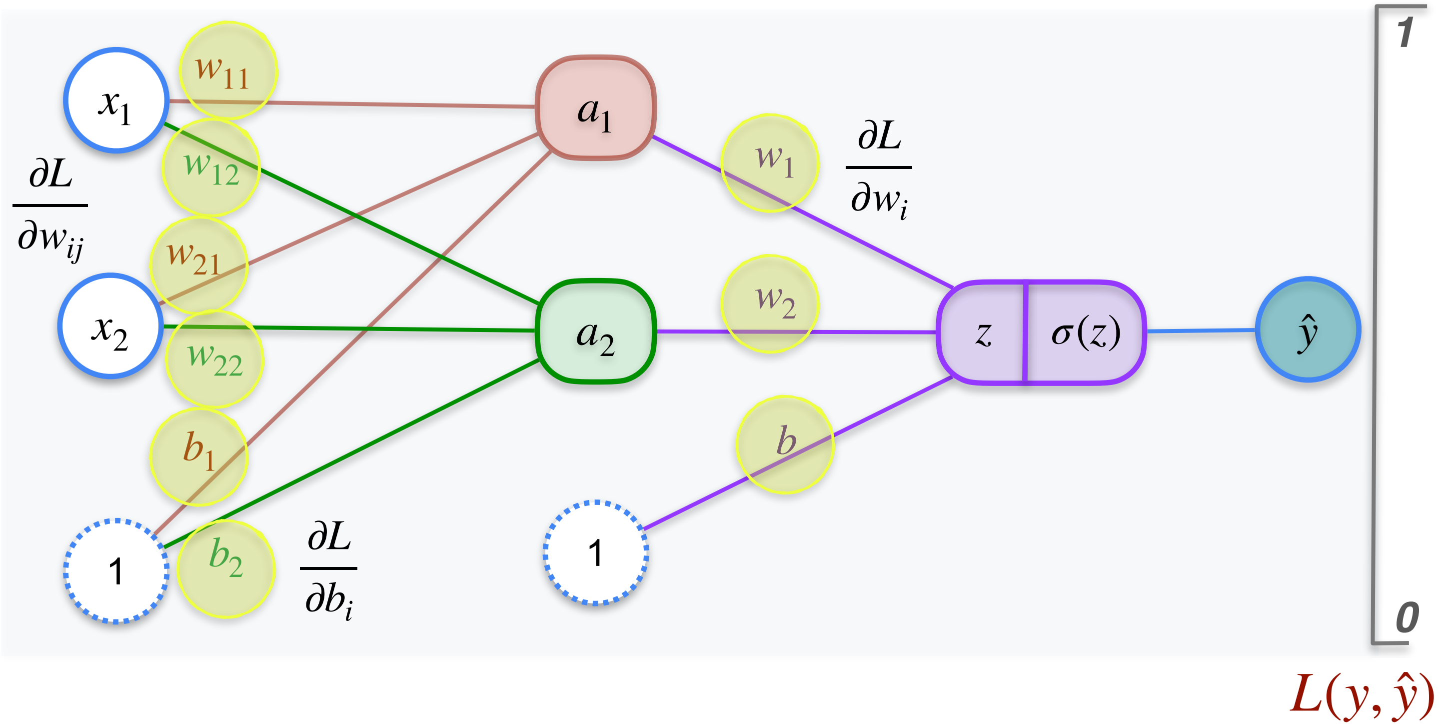
2,2,1 Neural Network



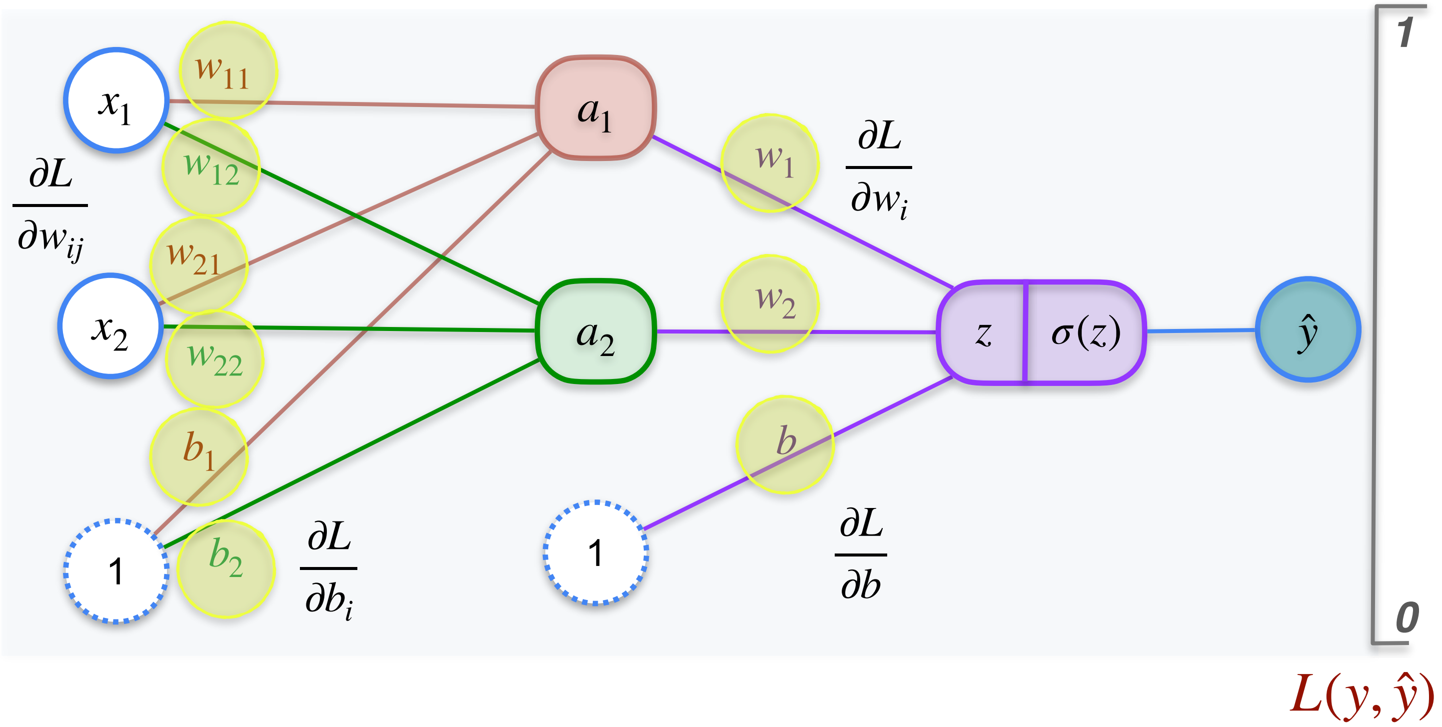
2,2,1 Neural Network



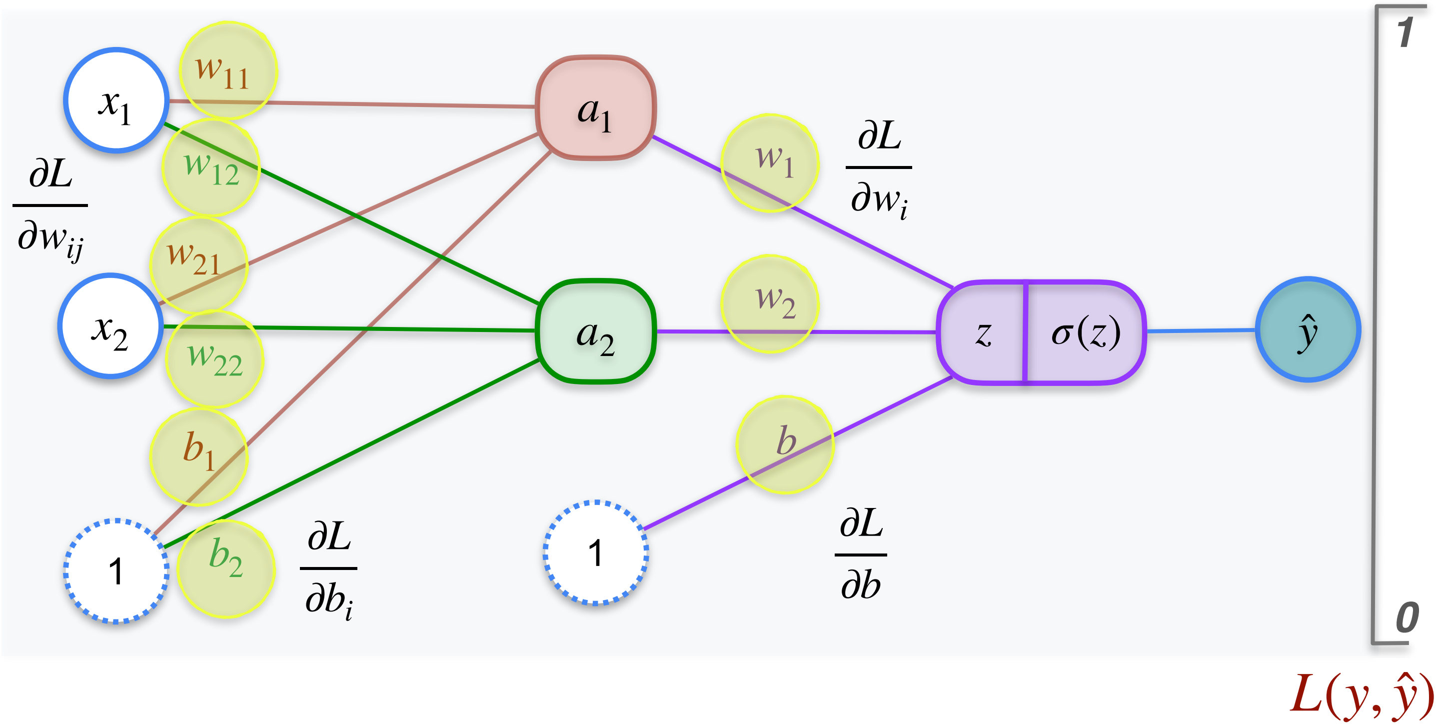
2,2,1 Neural Network



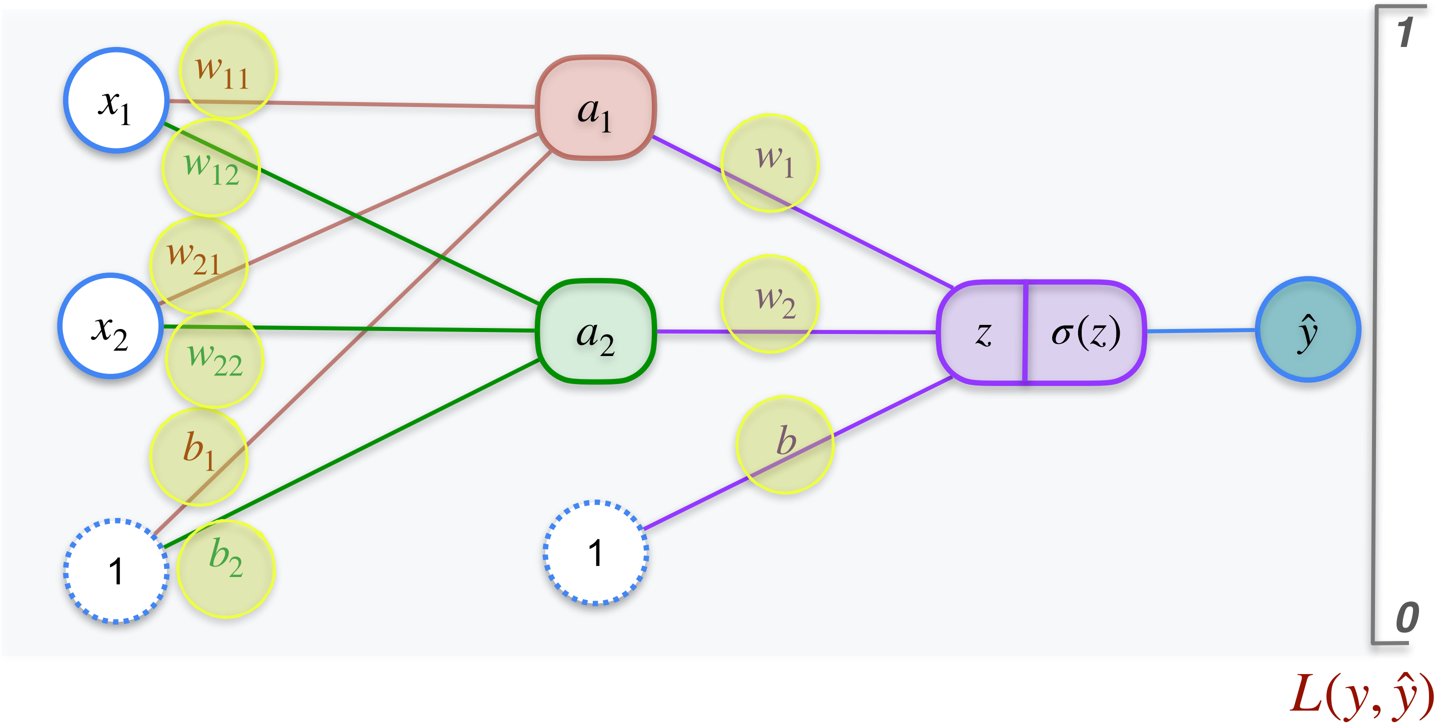
2,2,1 Neural Network



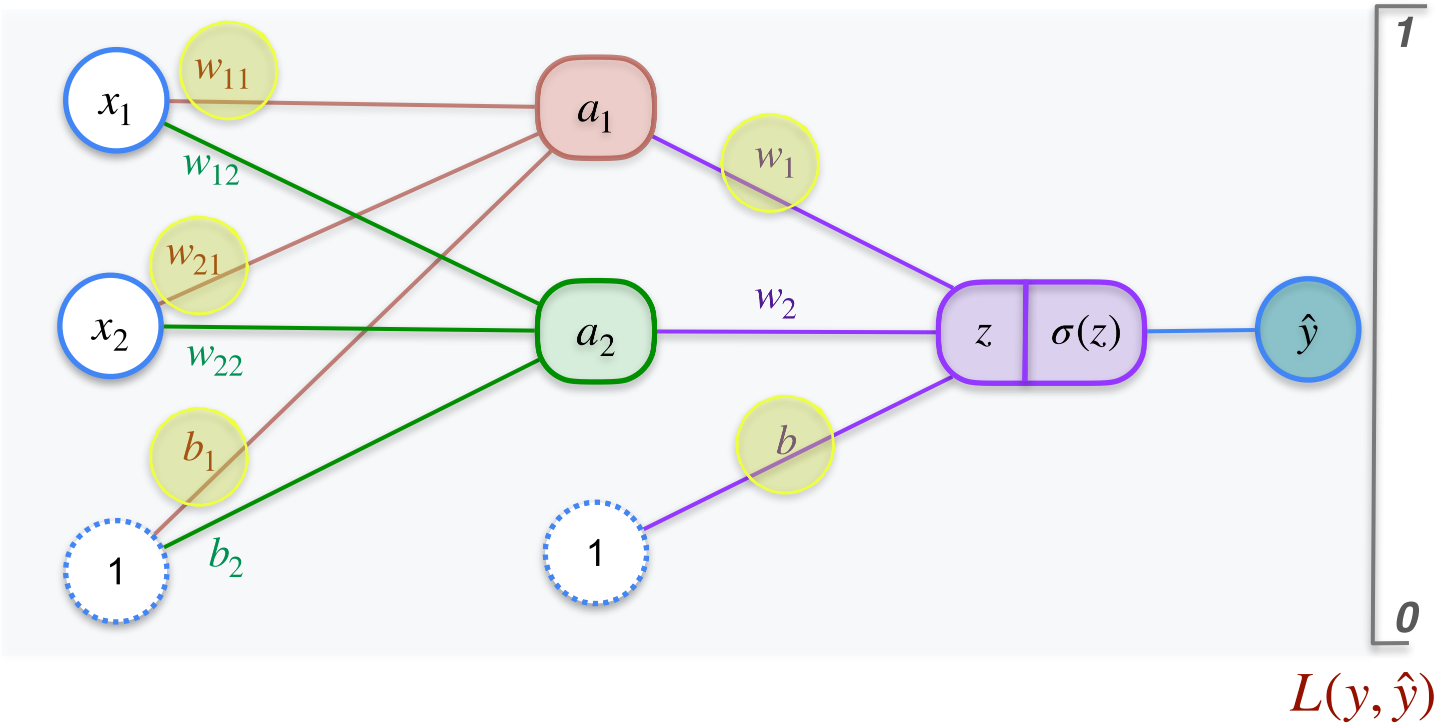
2,2,1 Neural Network



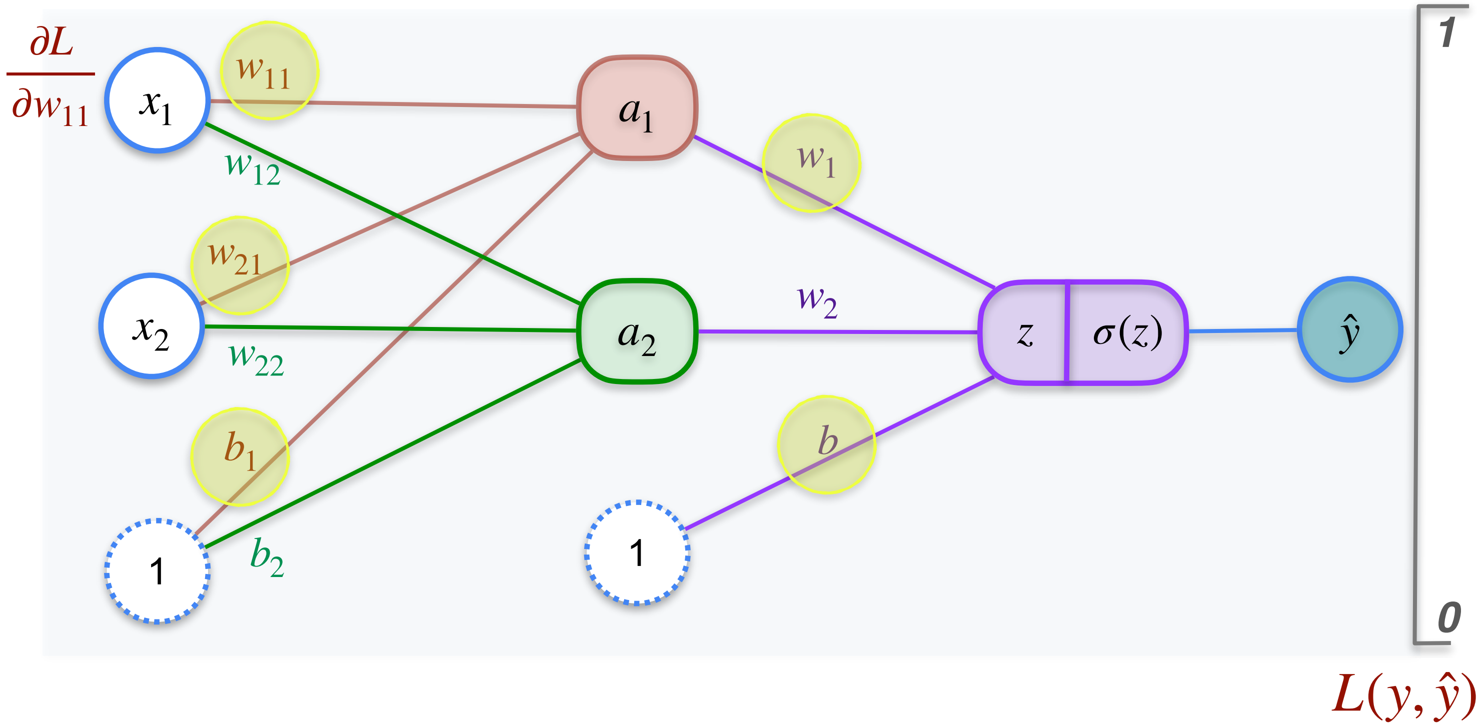
2,2,1 Neural Network



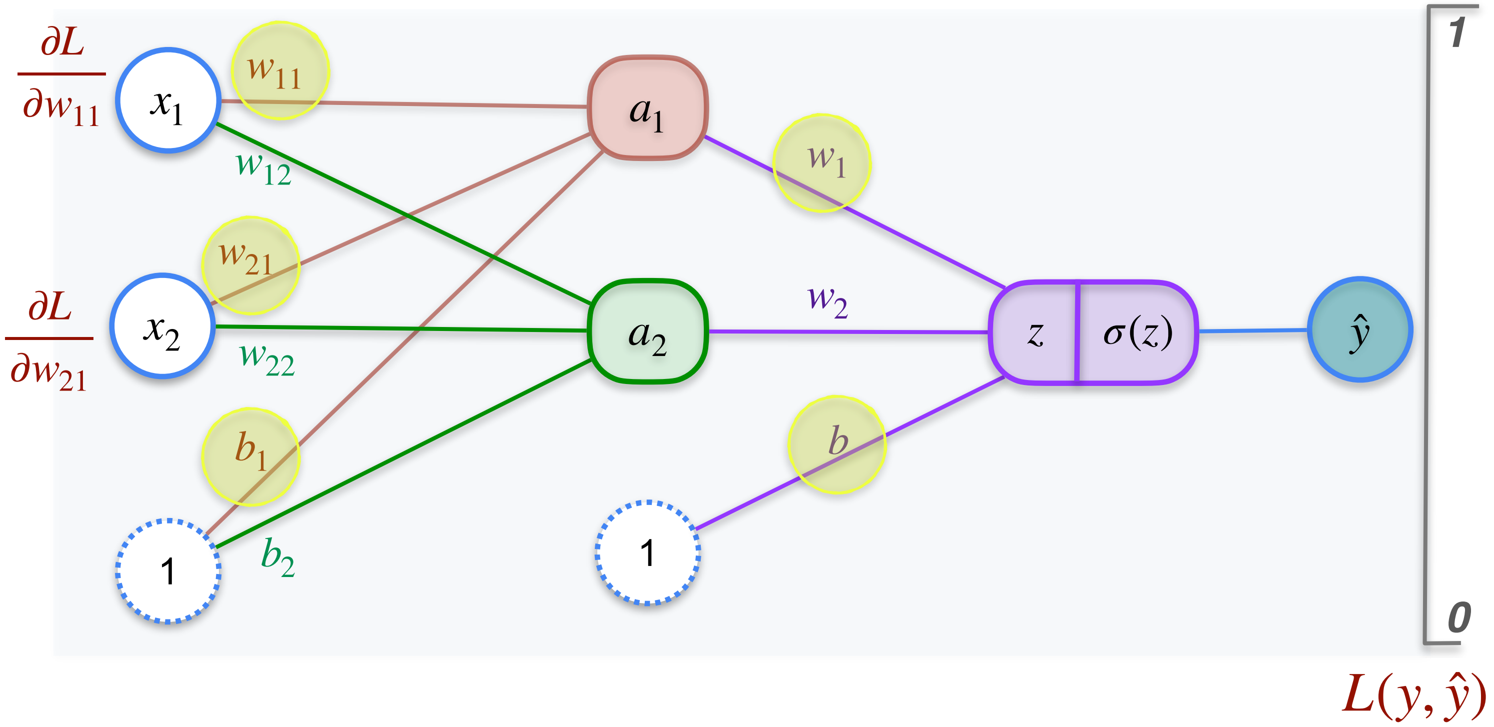
2,2,1 Neural Network



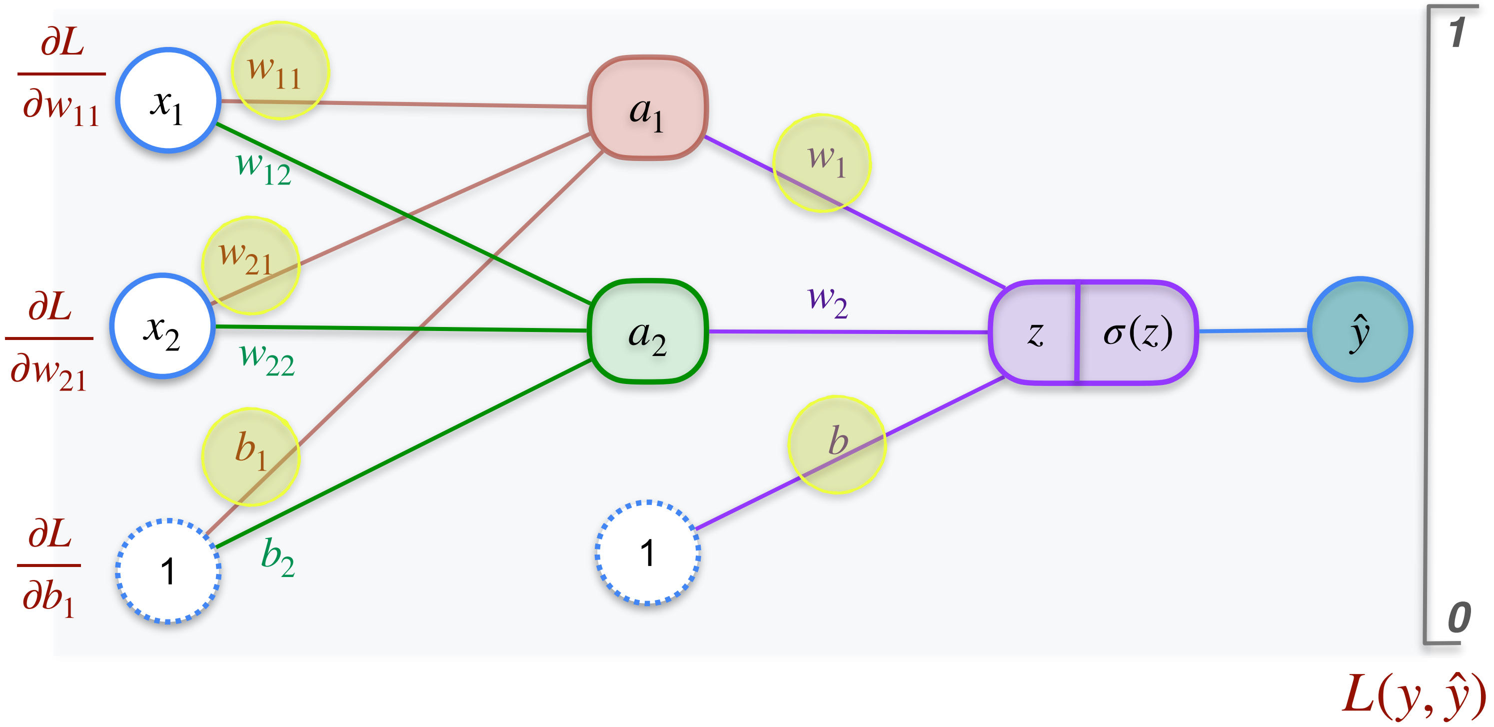
2,2,1 Neural Network



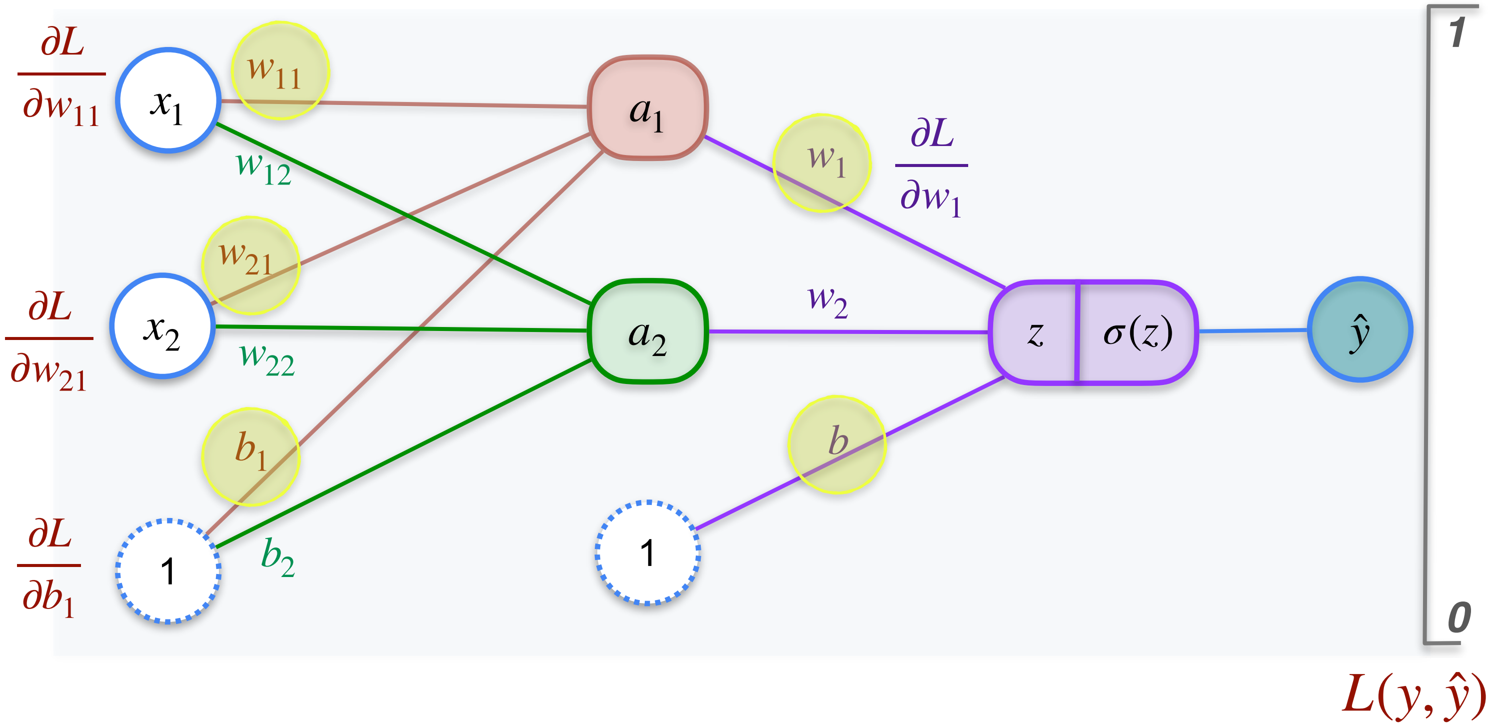
2,2,1 Neural Network



2,2,1 Neural Network

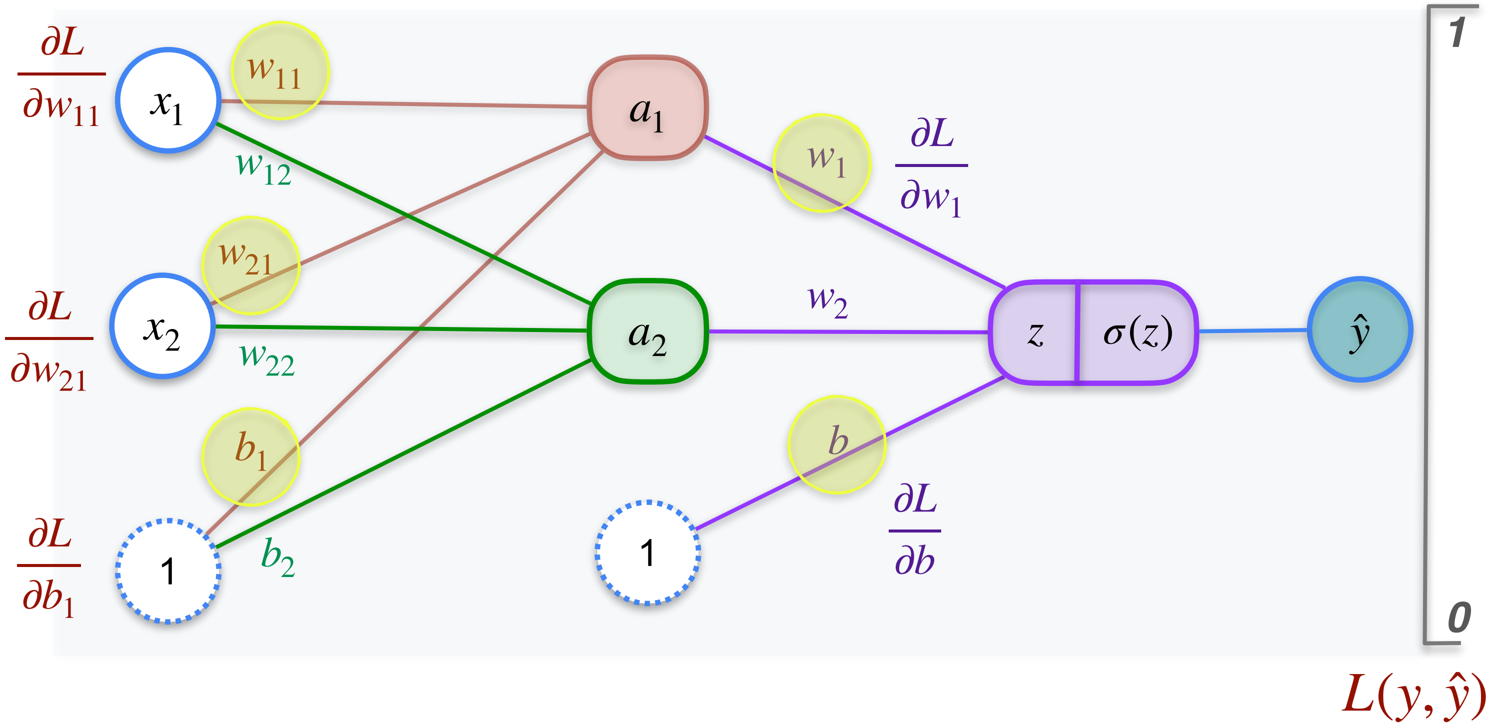


2,2,1 Neural Network

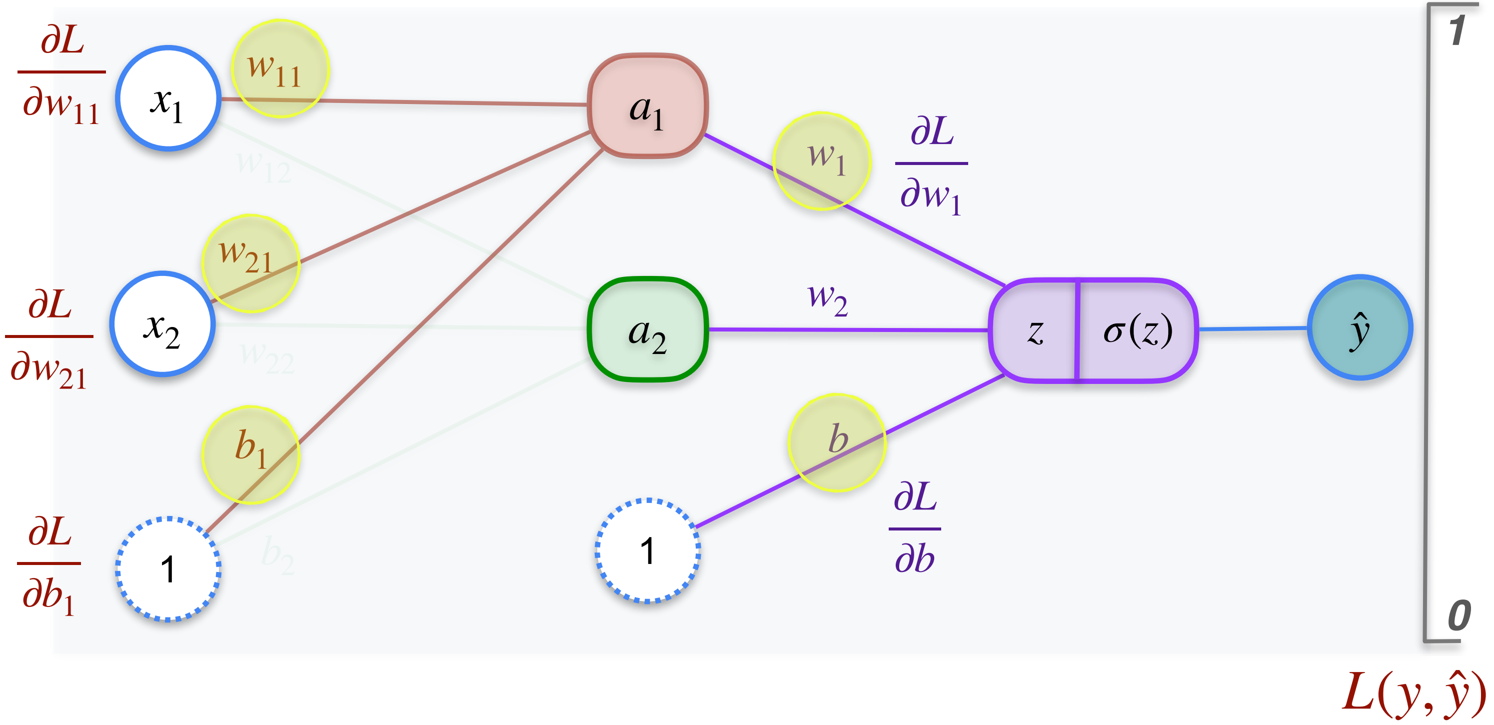


DeepLearning.AI

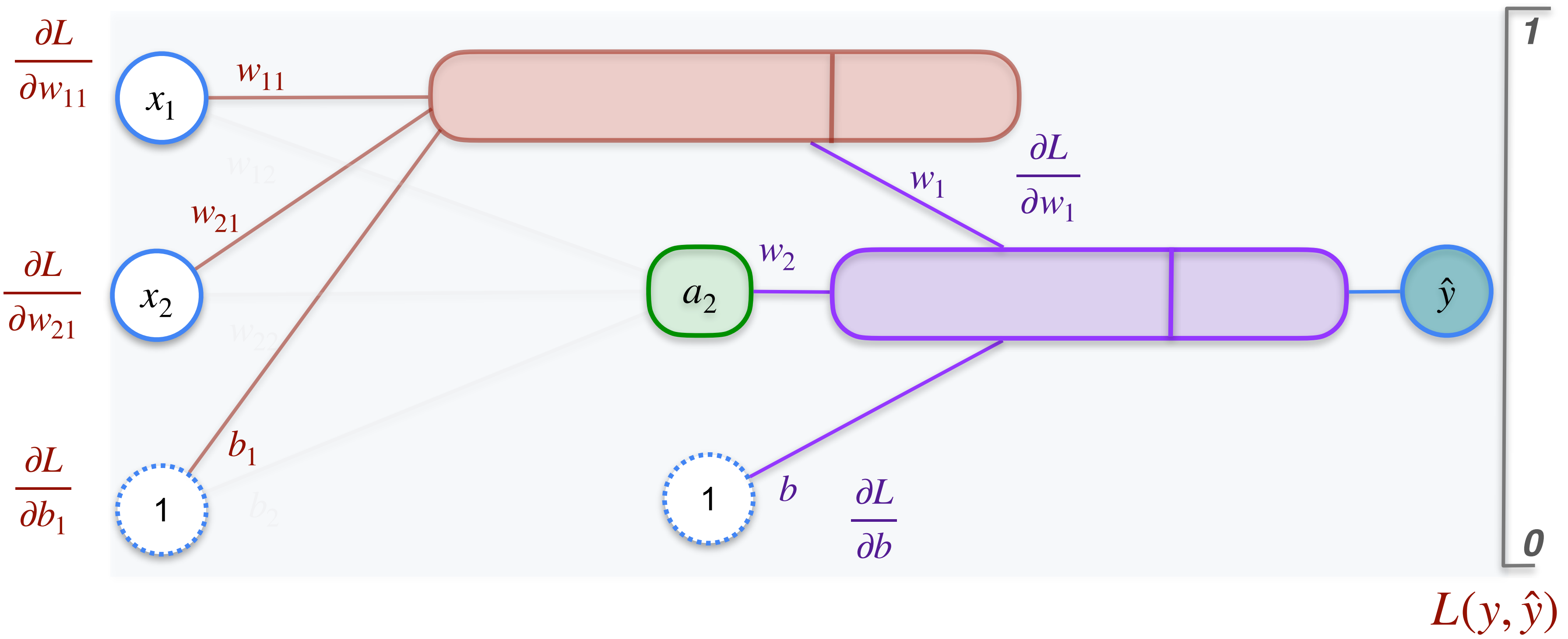
2,2,1 Neural Network



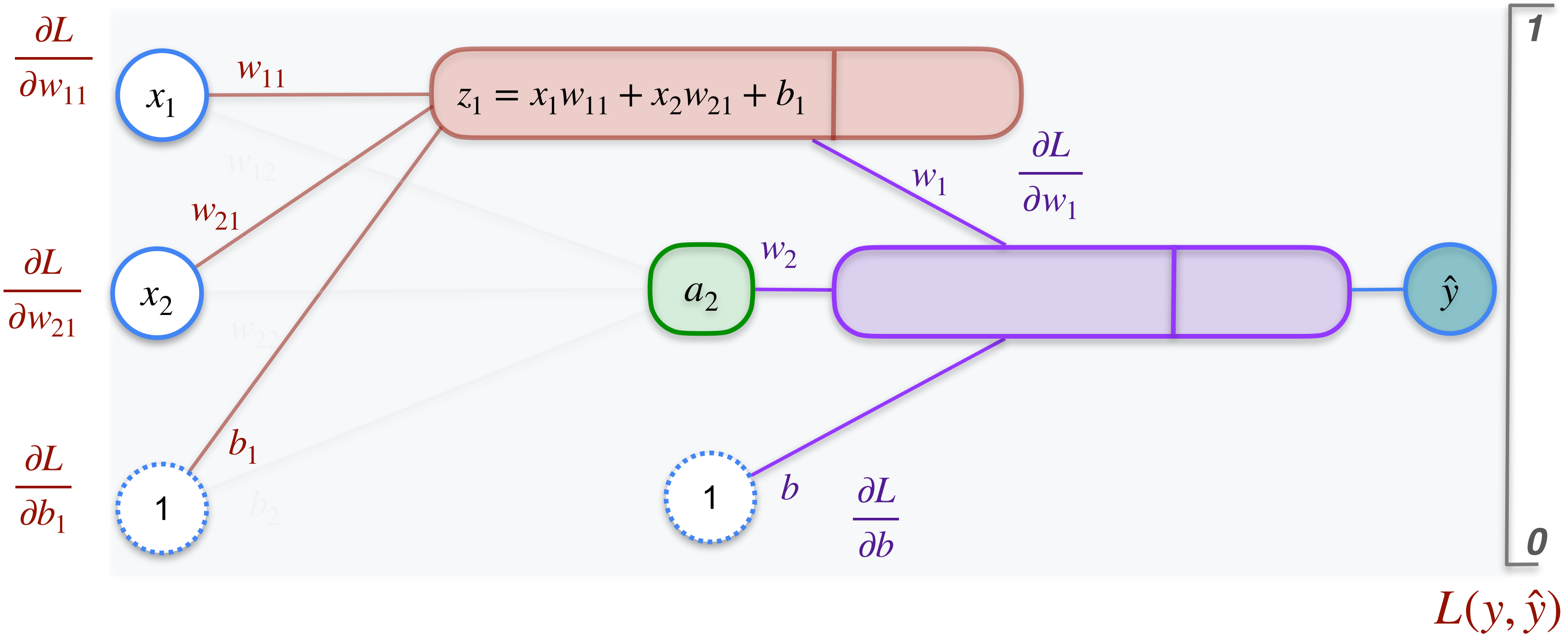
2,2,1 Neural Network



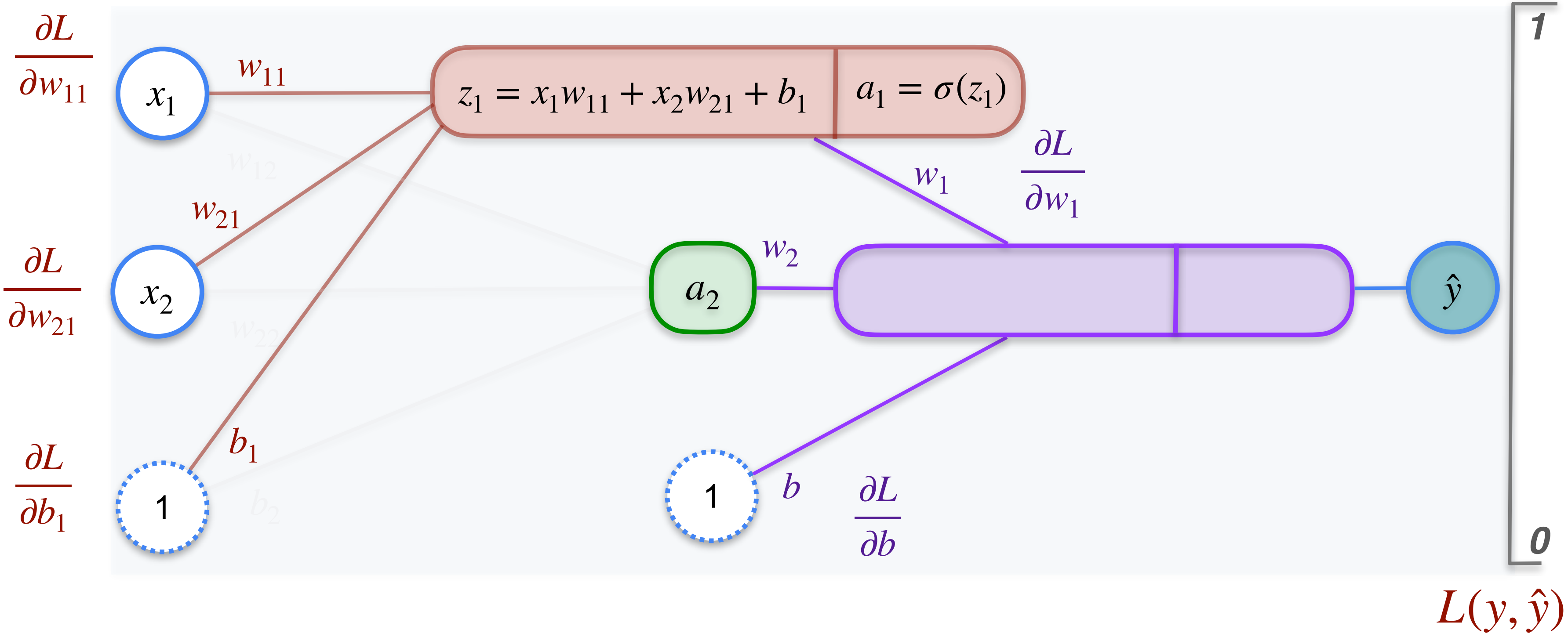
2,2,1 Neural Network



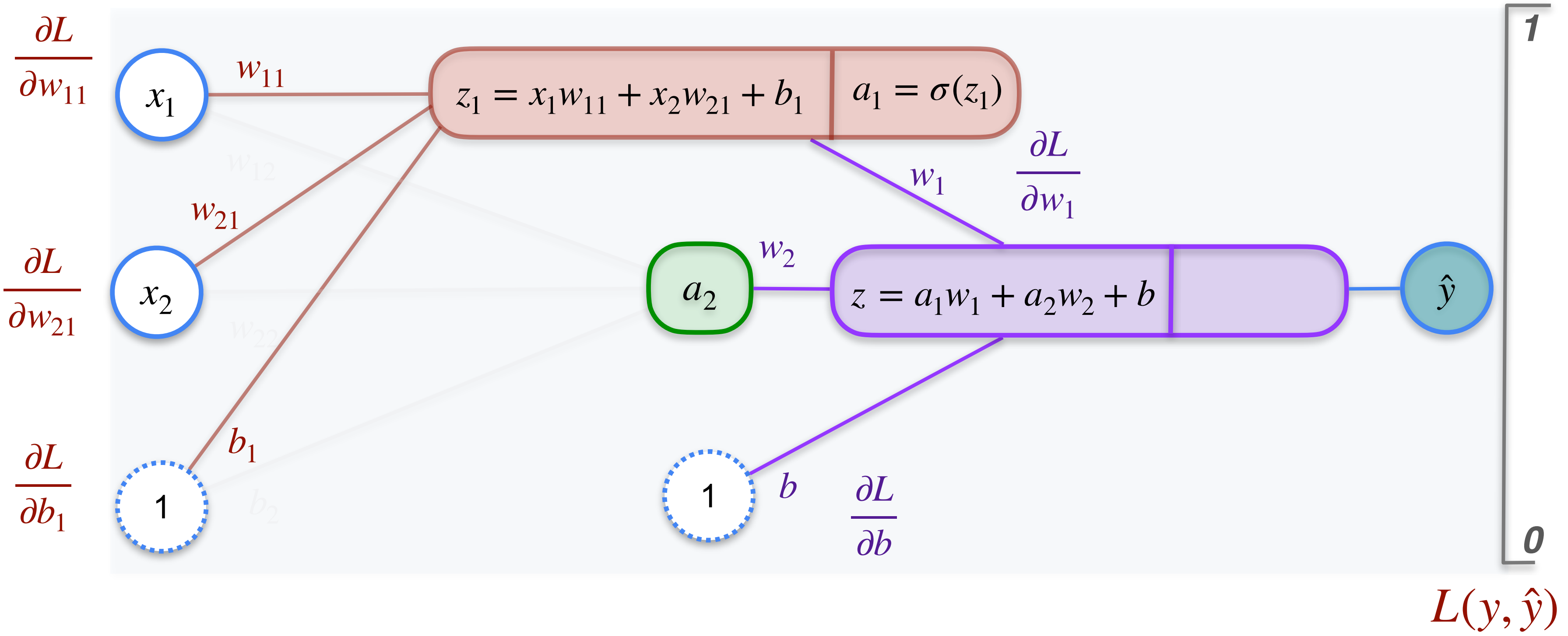
2,2,1 Neural Network



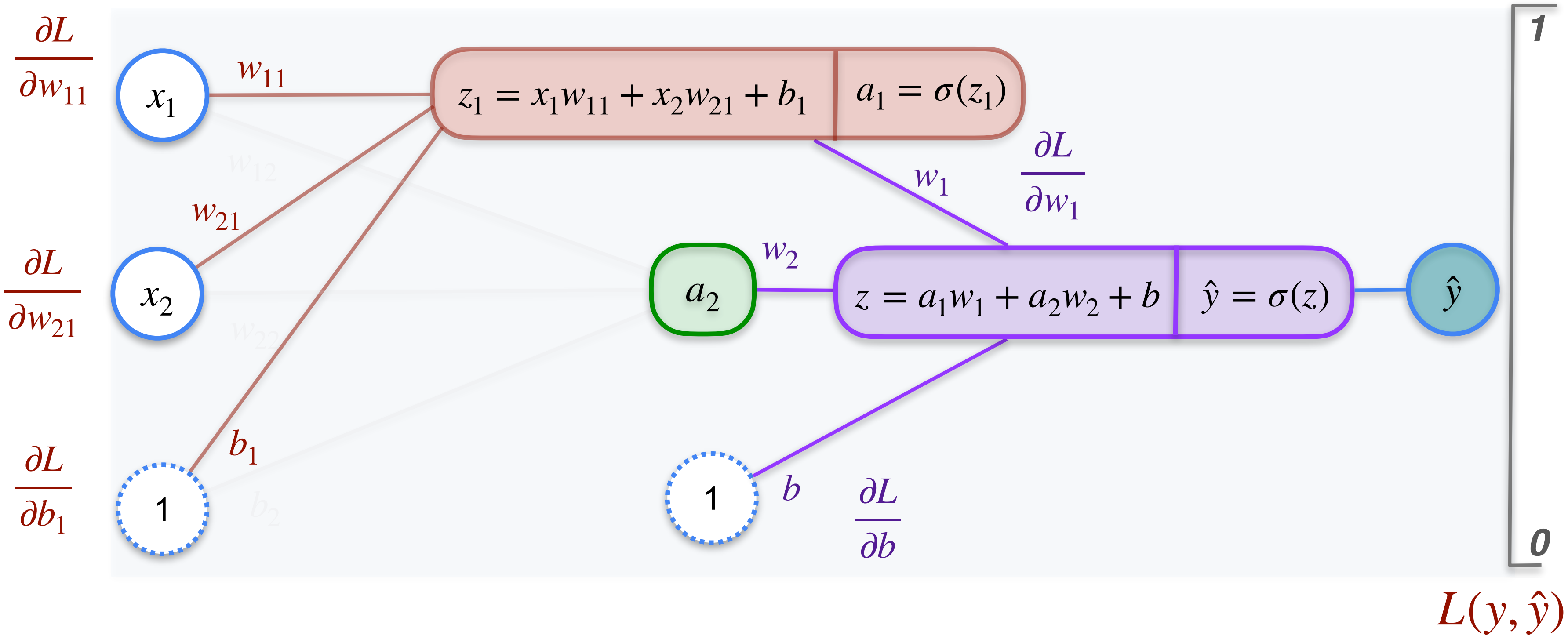
2,2,1 Neural Network



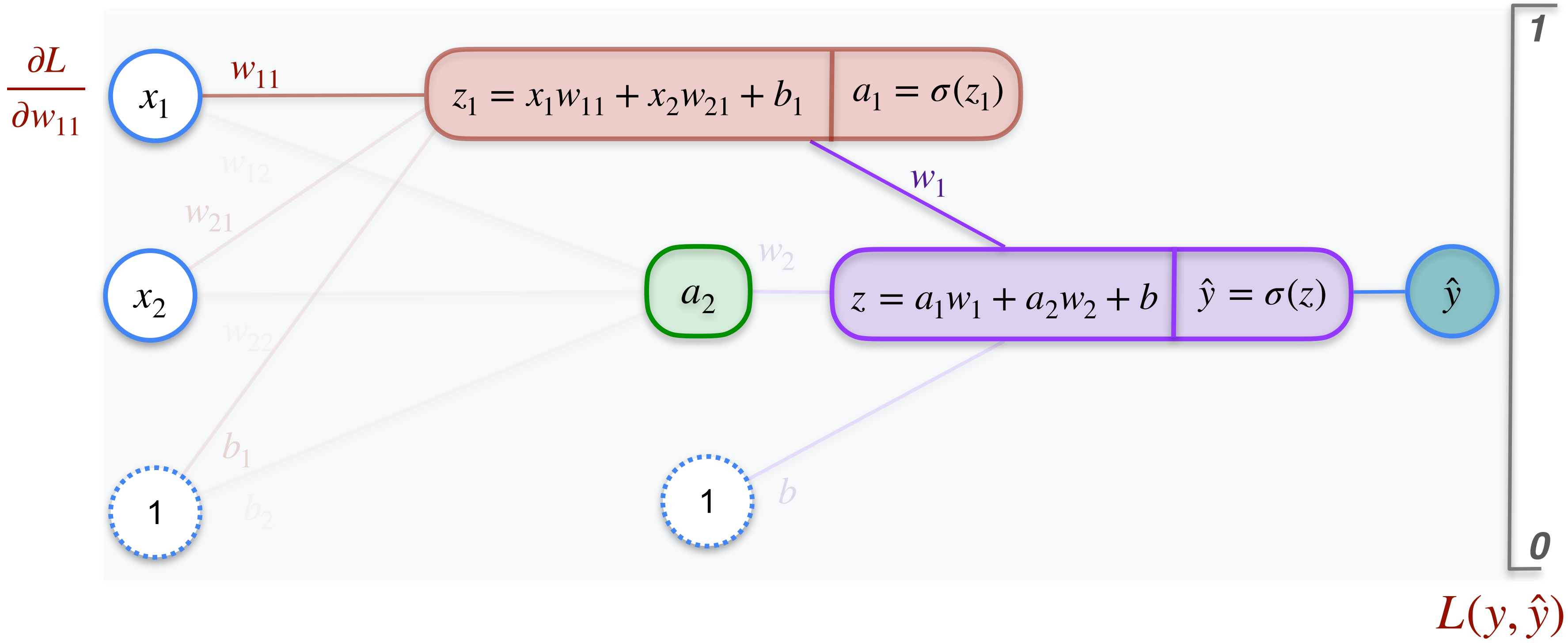
2,2,1 Neural Network



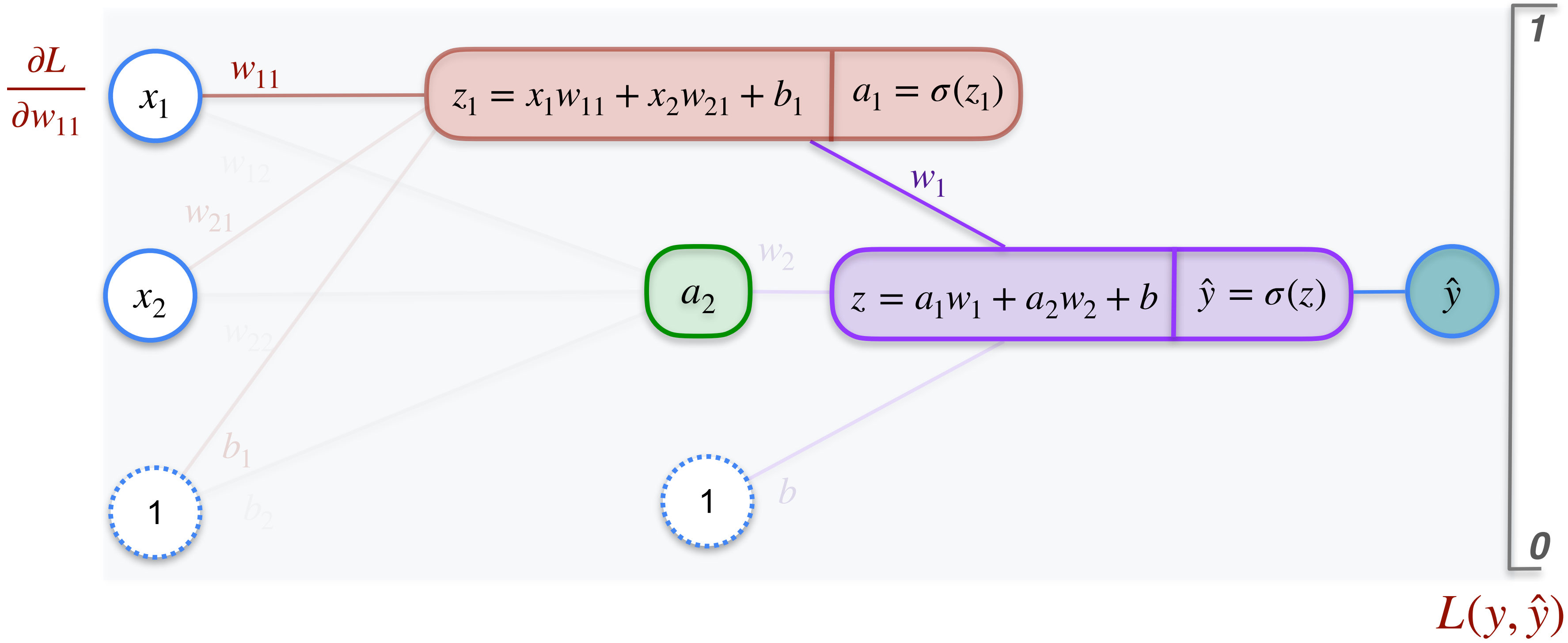
2,2,1 Neural Network



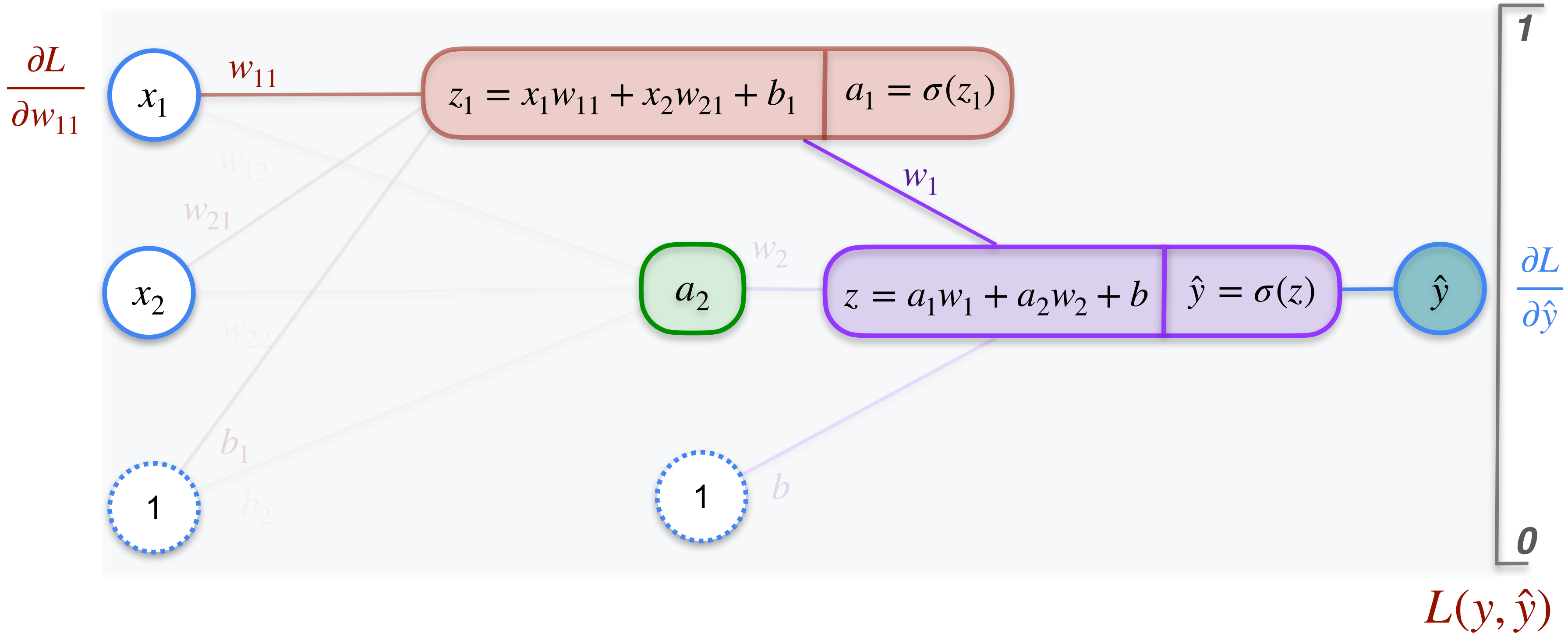
2,2,1 Neural Network



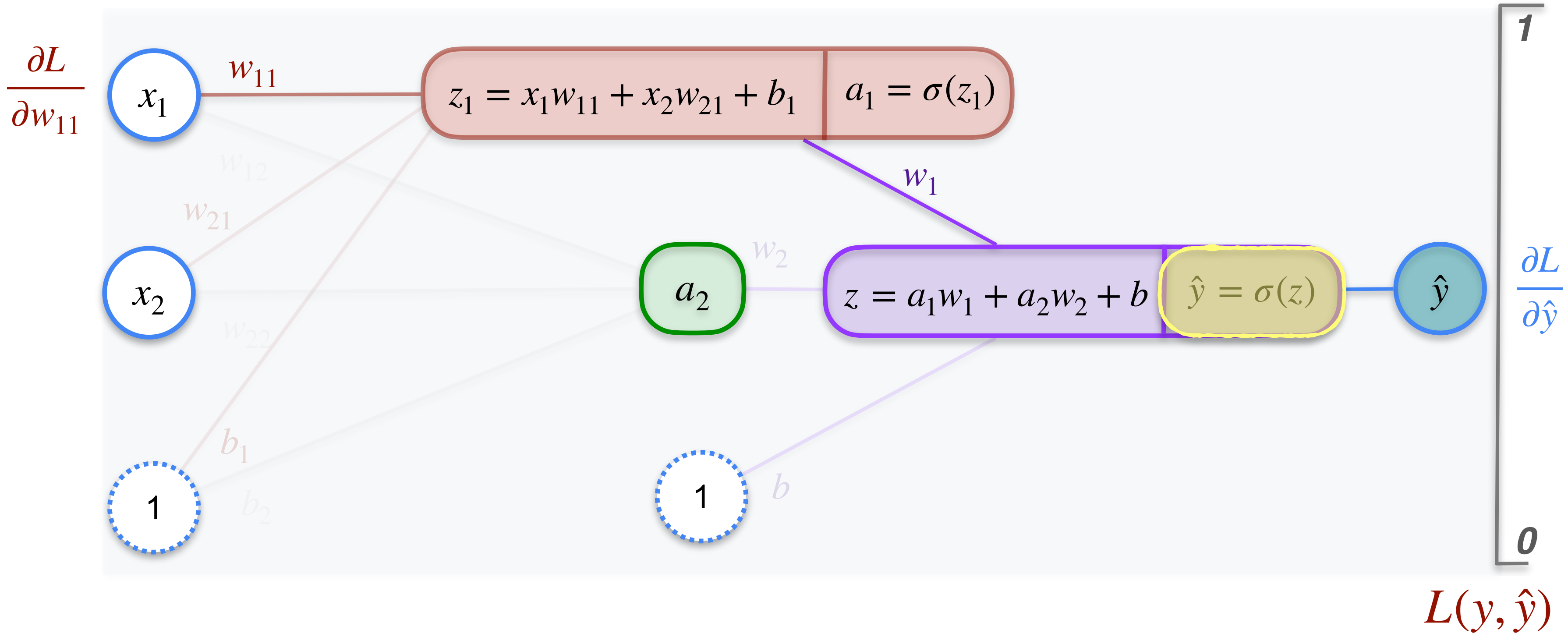
2,2,1 Neural Network



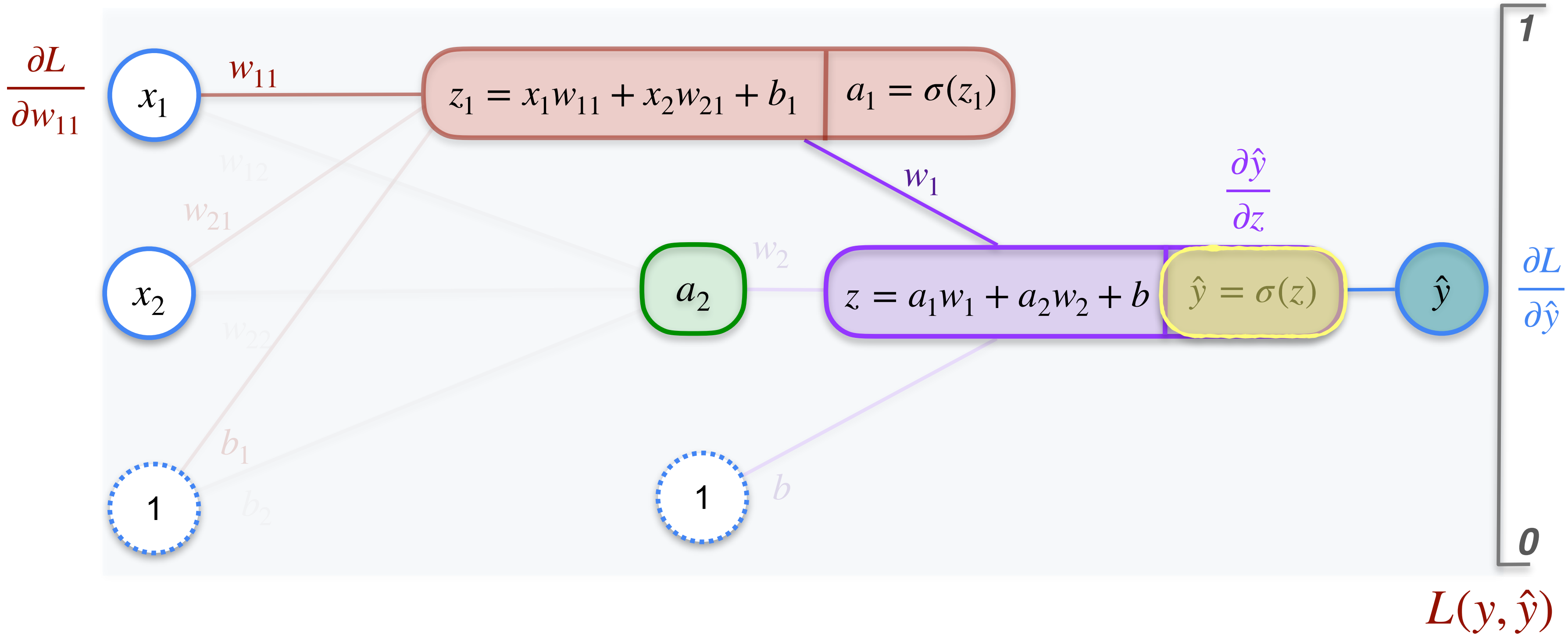
2,2,1 Neural Network



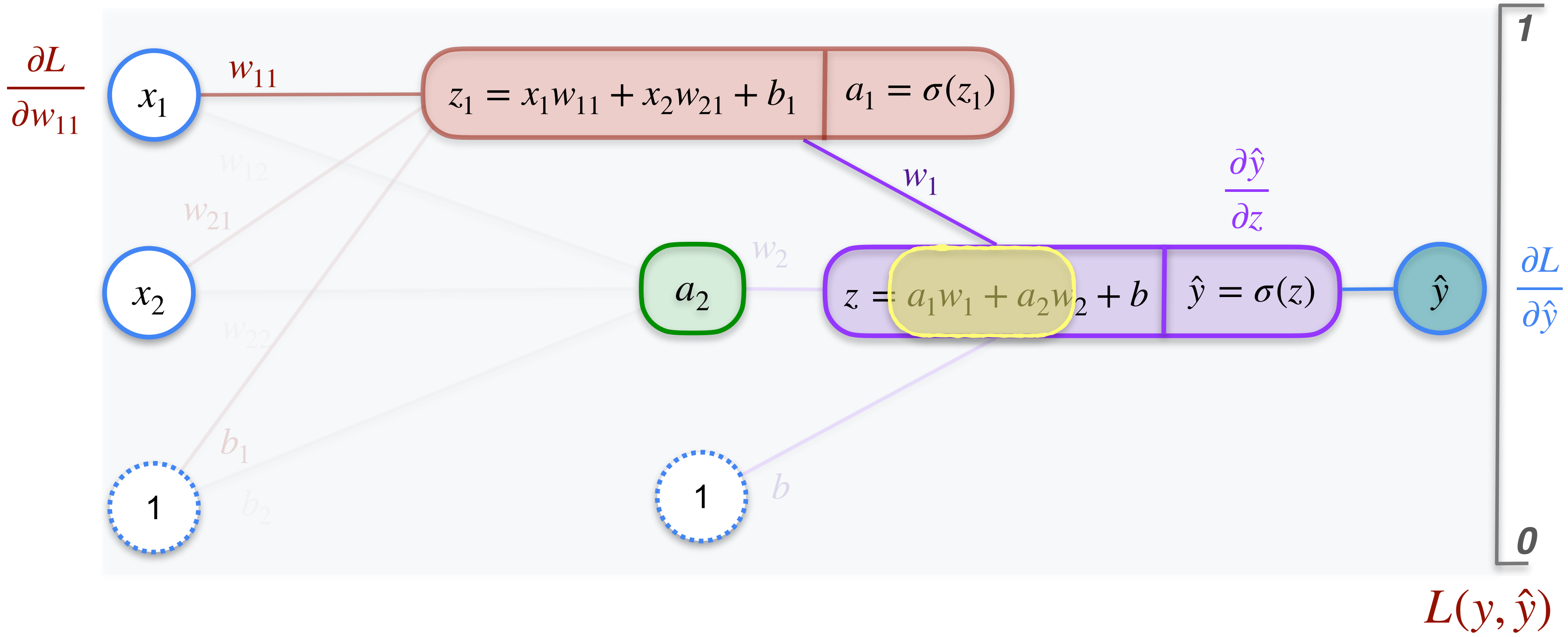
2,2,1 Neural Network



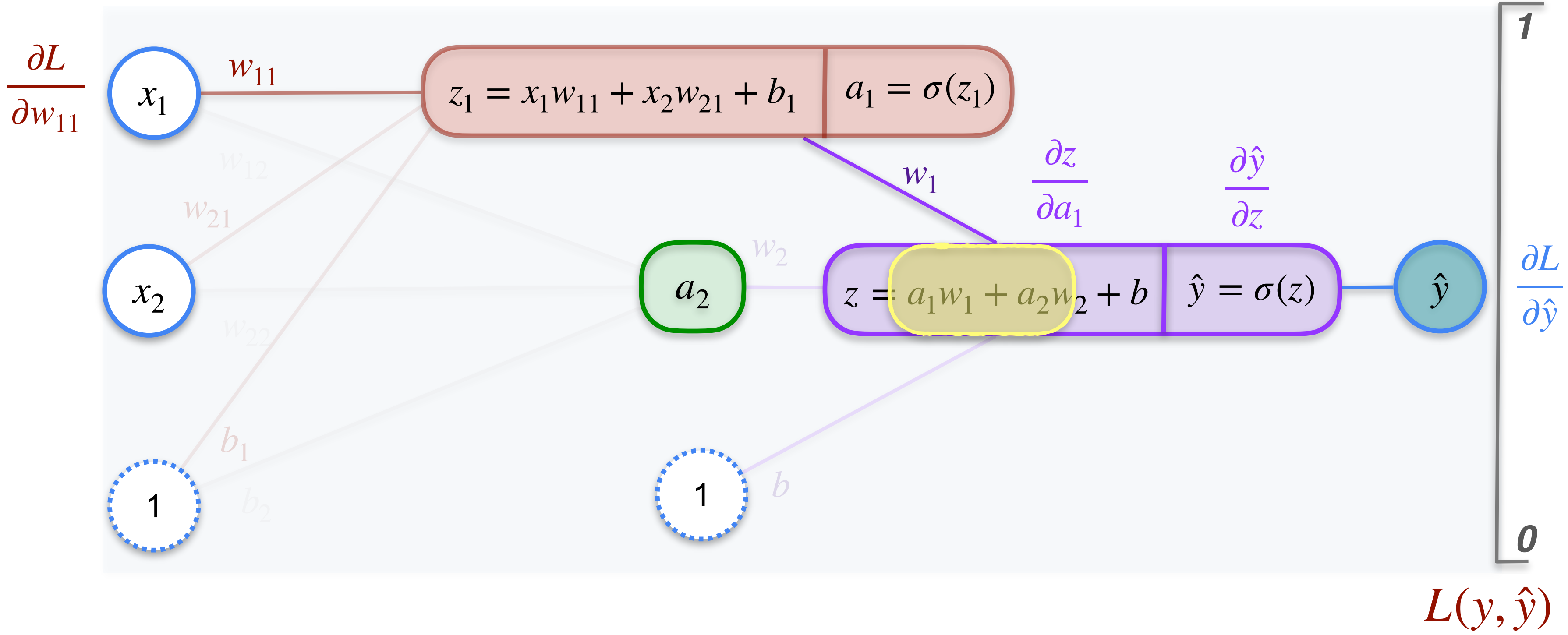
2,2,1 Neural Network



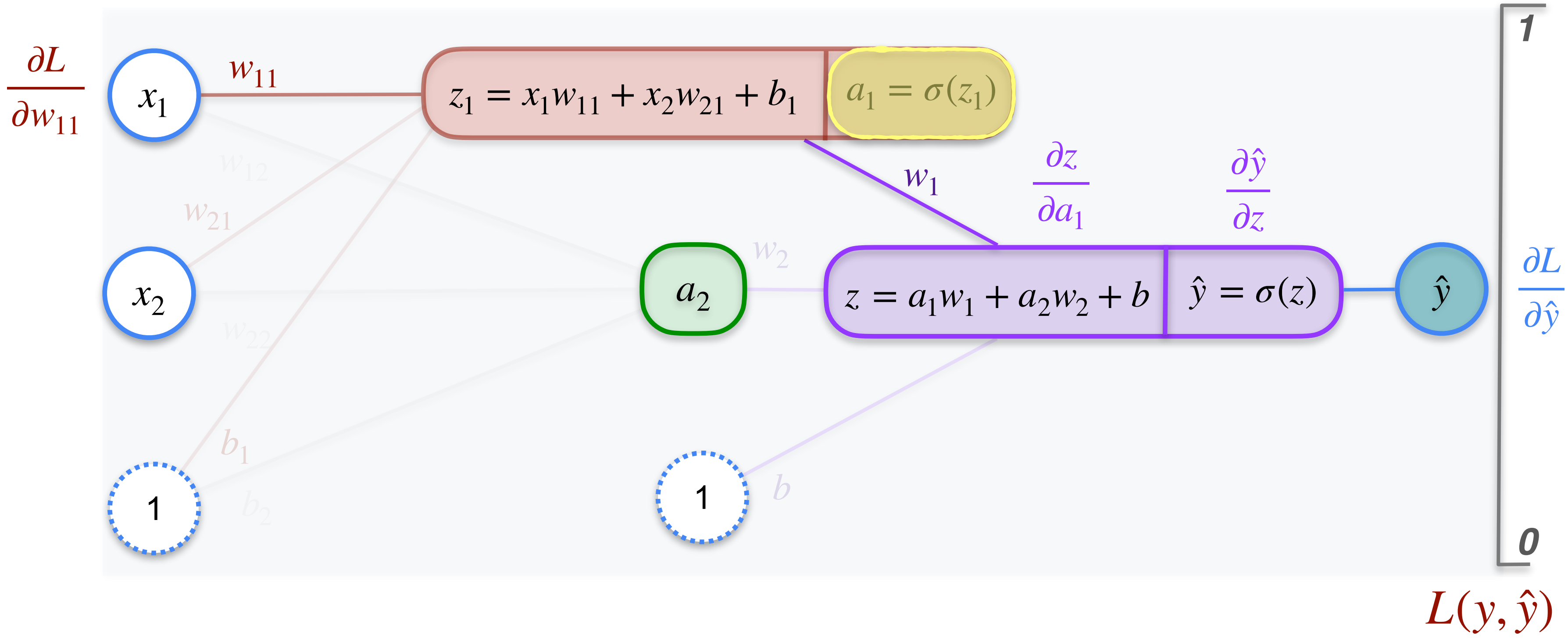
2,2,1 Neural Network



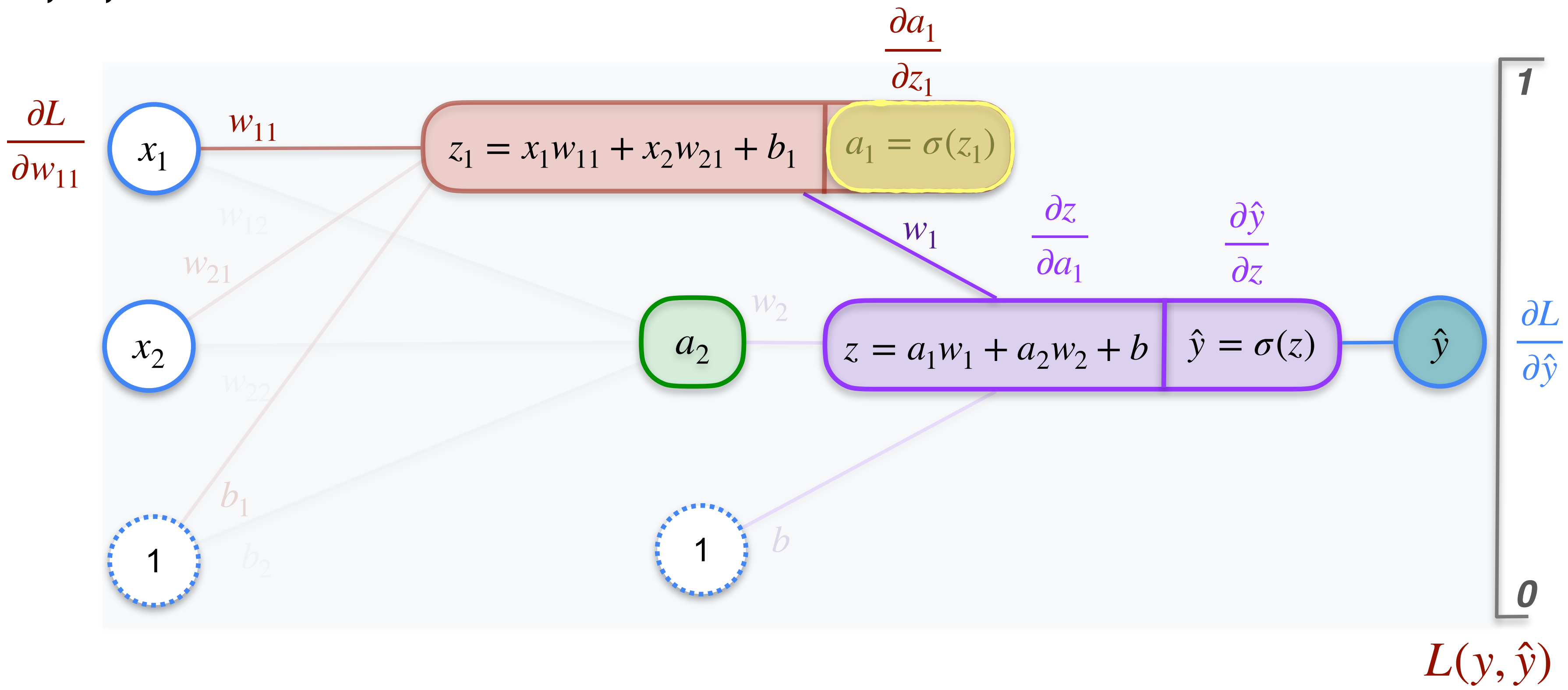
2,2,1 Neural Network



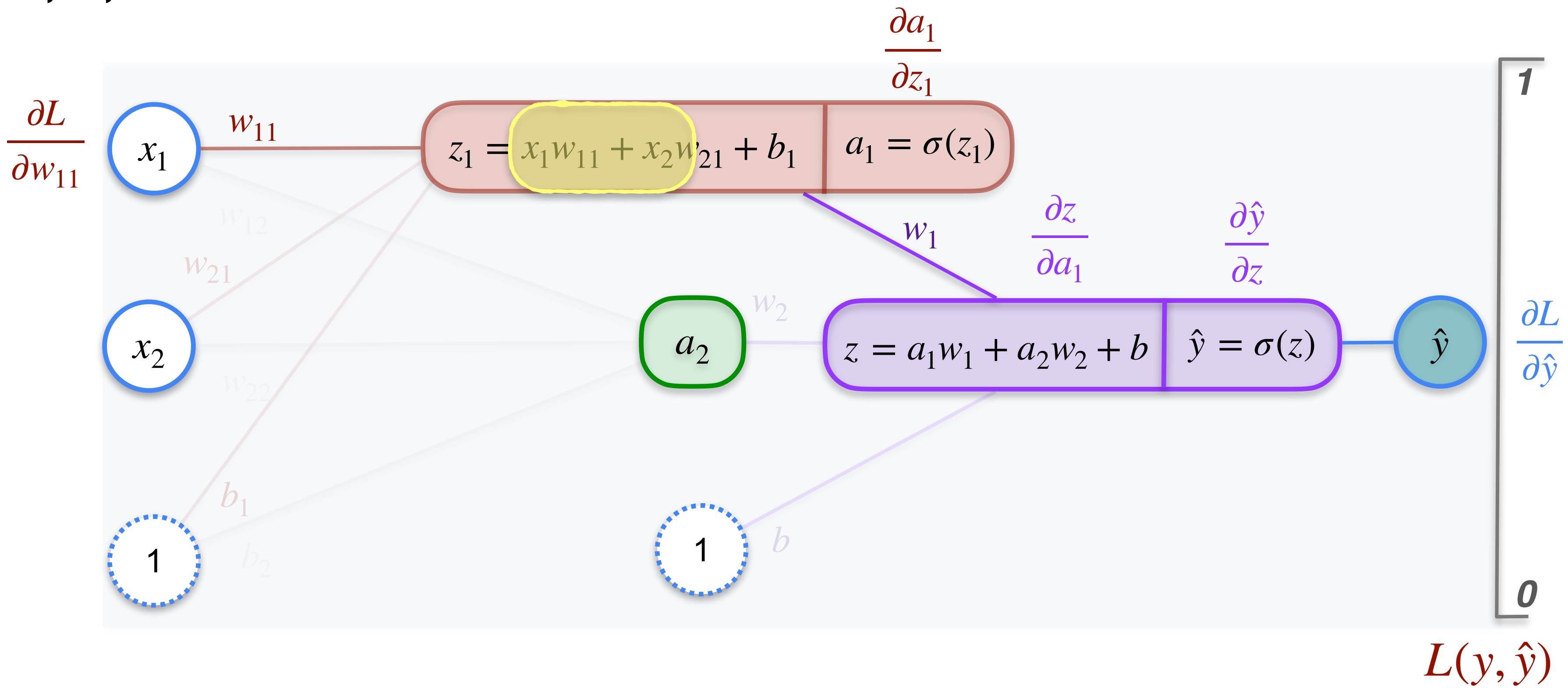
2,2,1 Neural Network



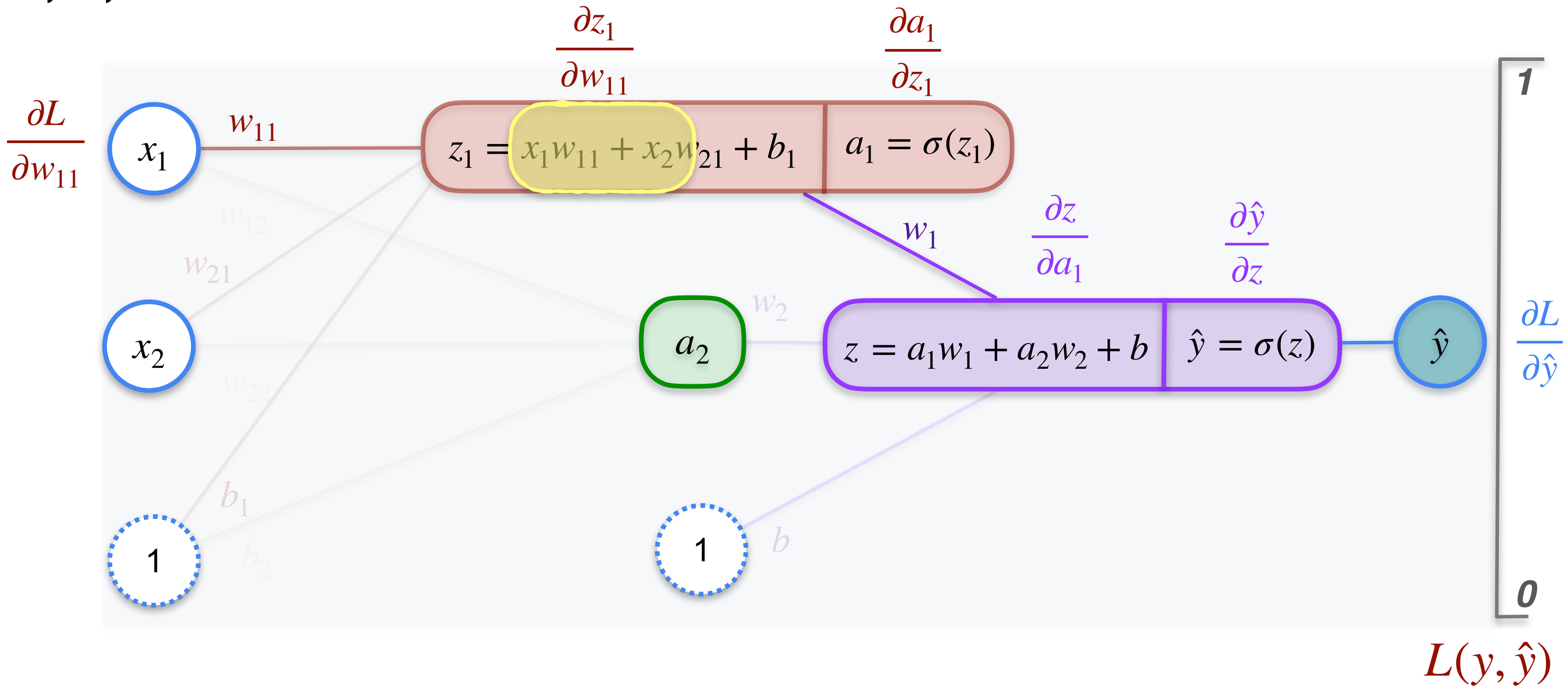
2,2,1 Neural Network



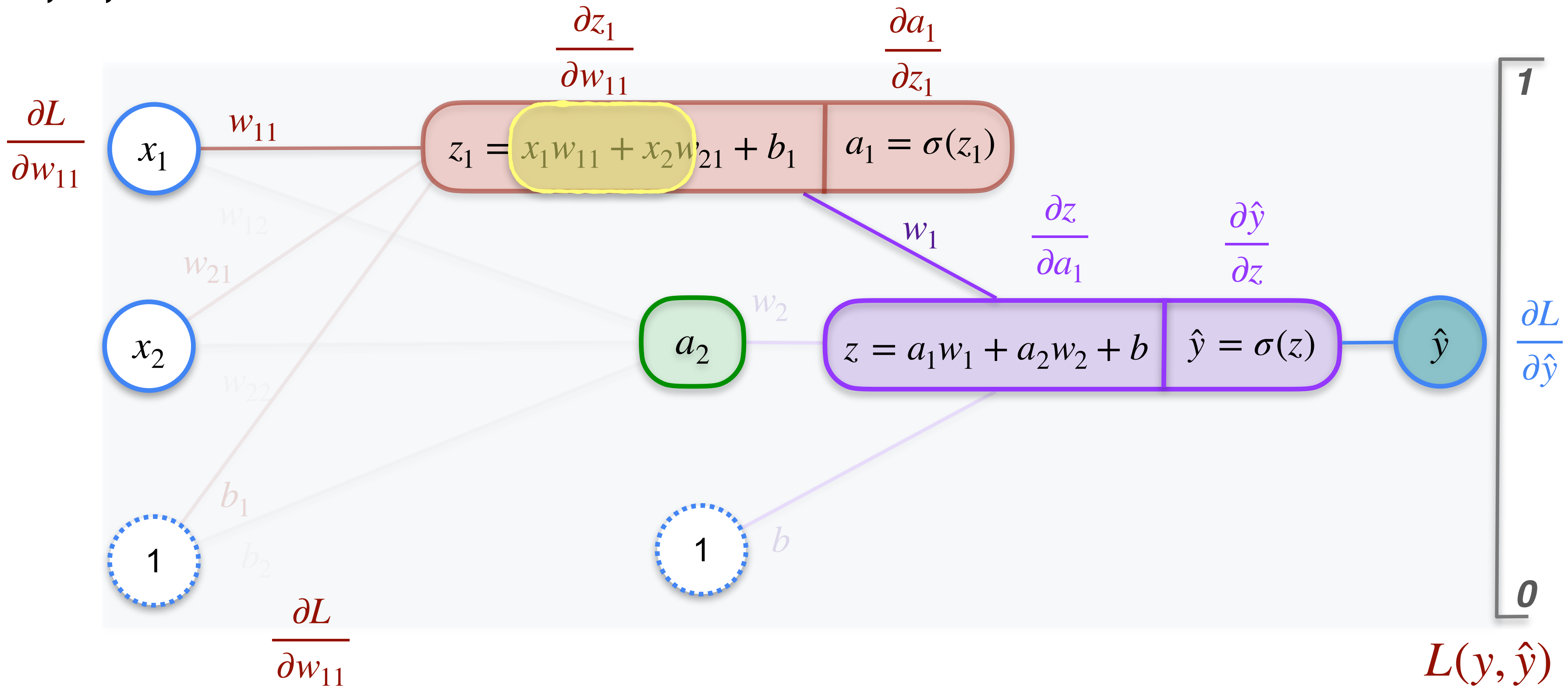
2,2,1 Neural Network



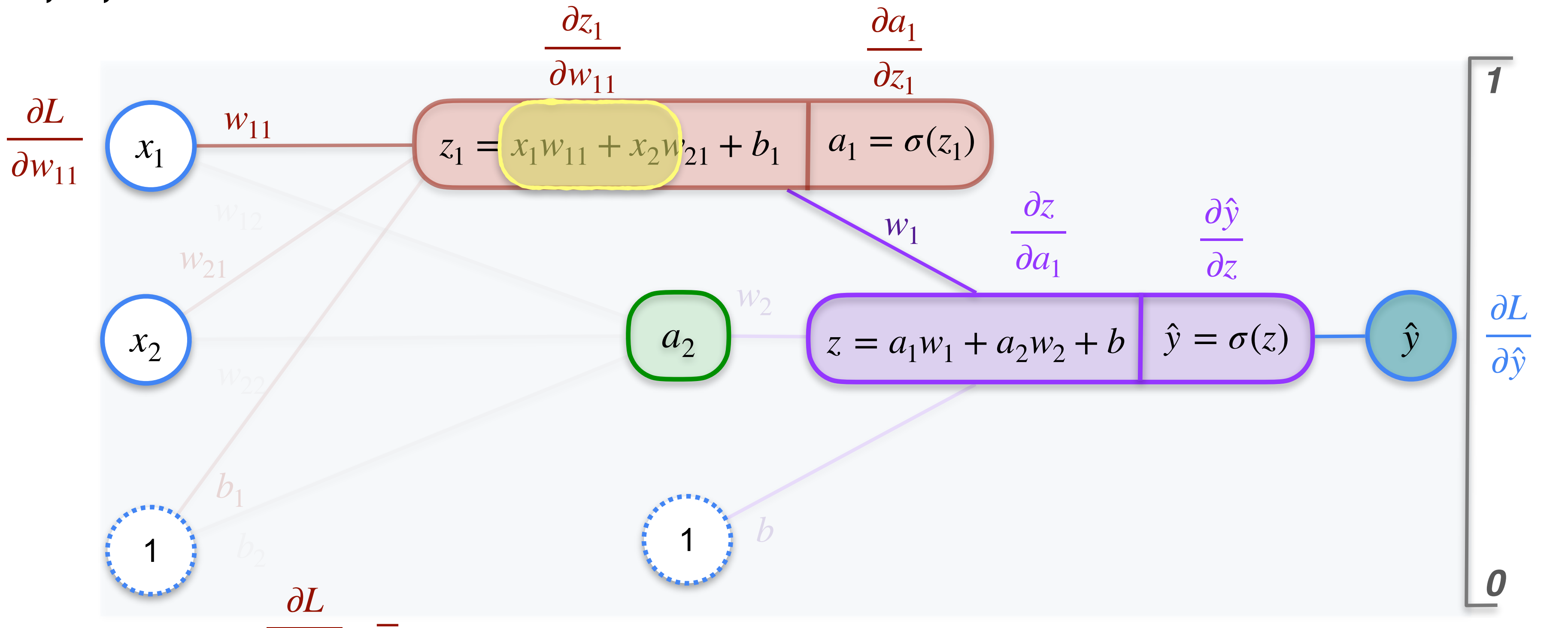
2,2,1 Neural Network



2,2,1 Neural Network

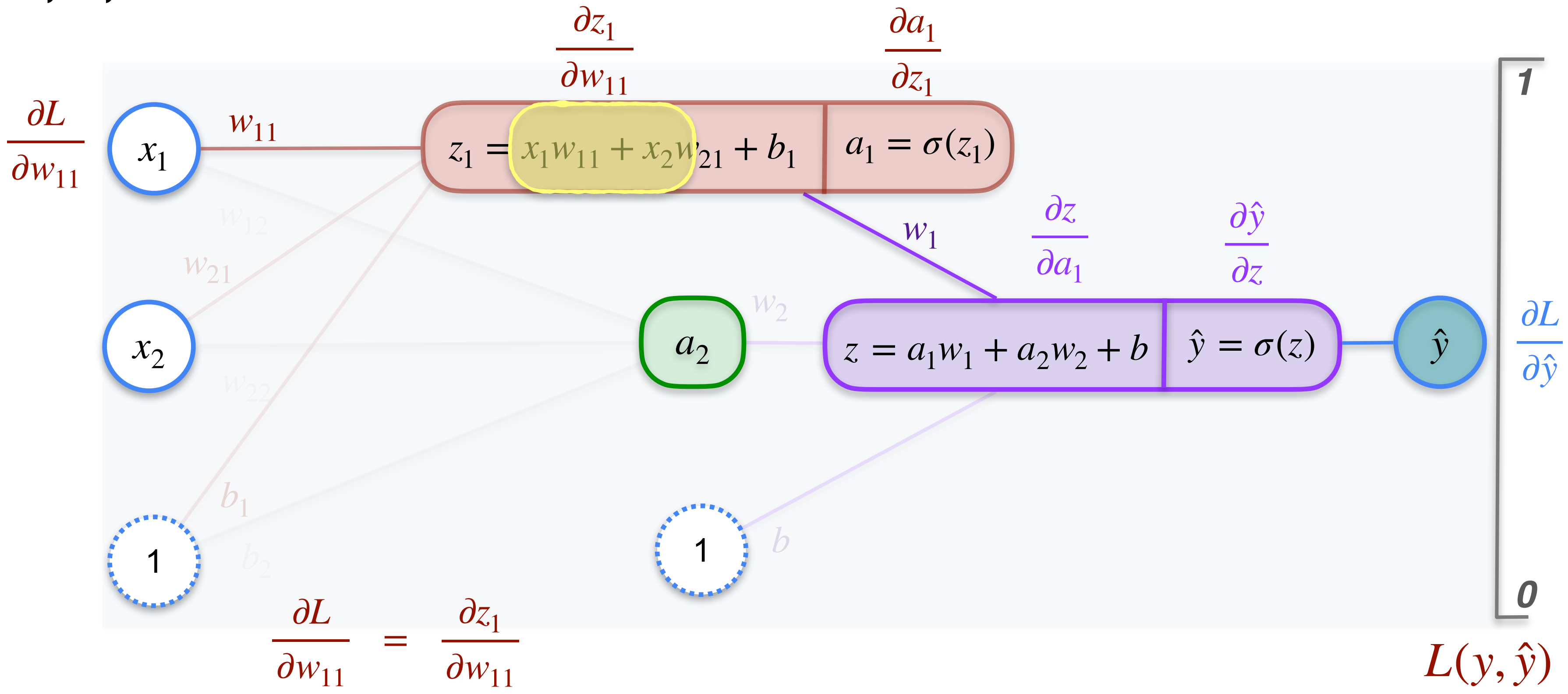


2,2,1 Neural Network



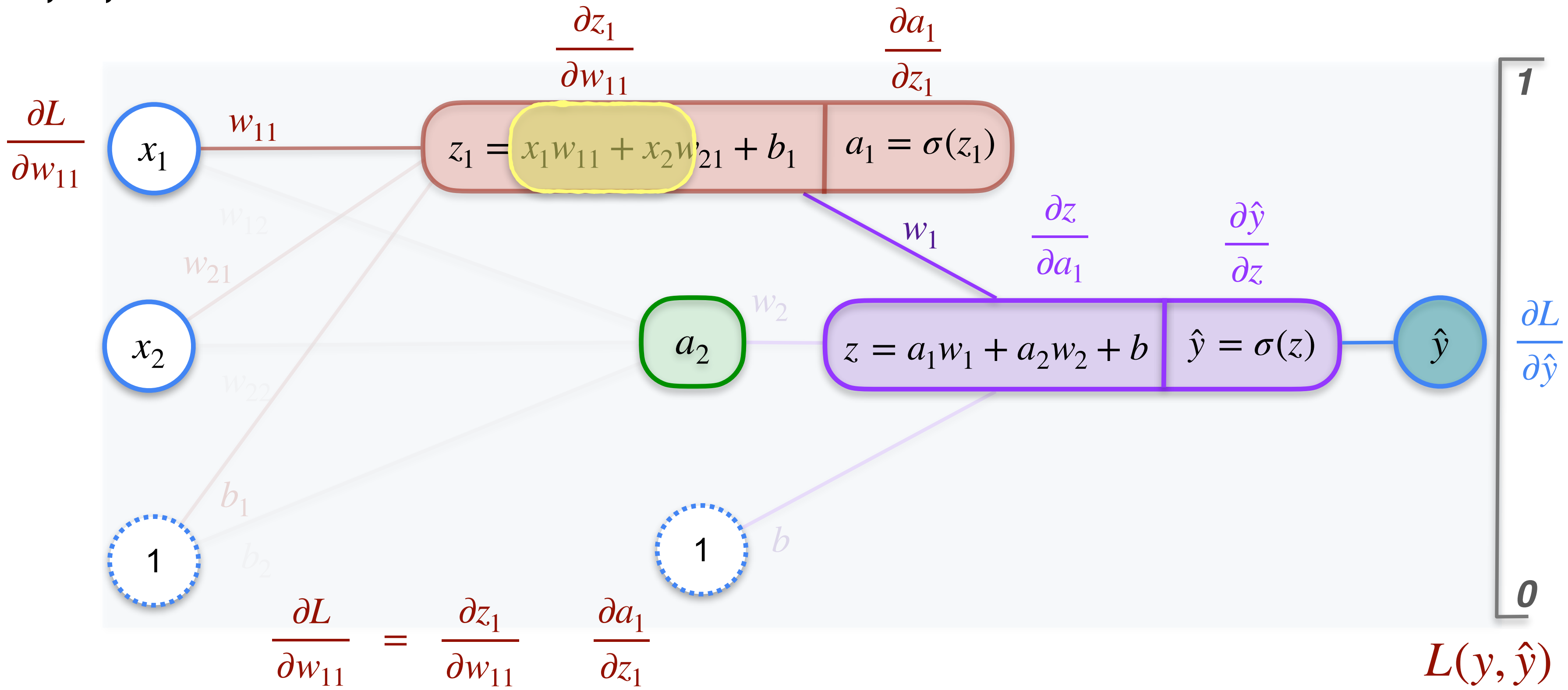
$$\frac{\partial L}{\partial w_{11}} =$$

2,2,1 Neural Network



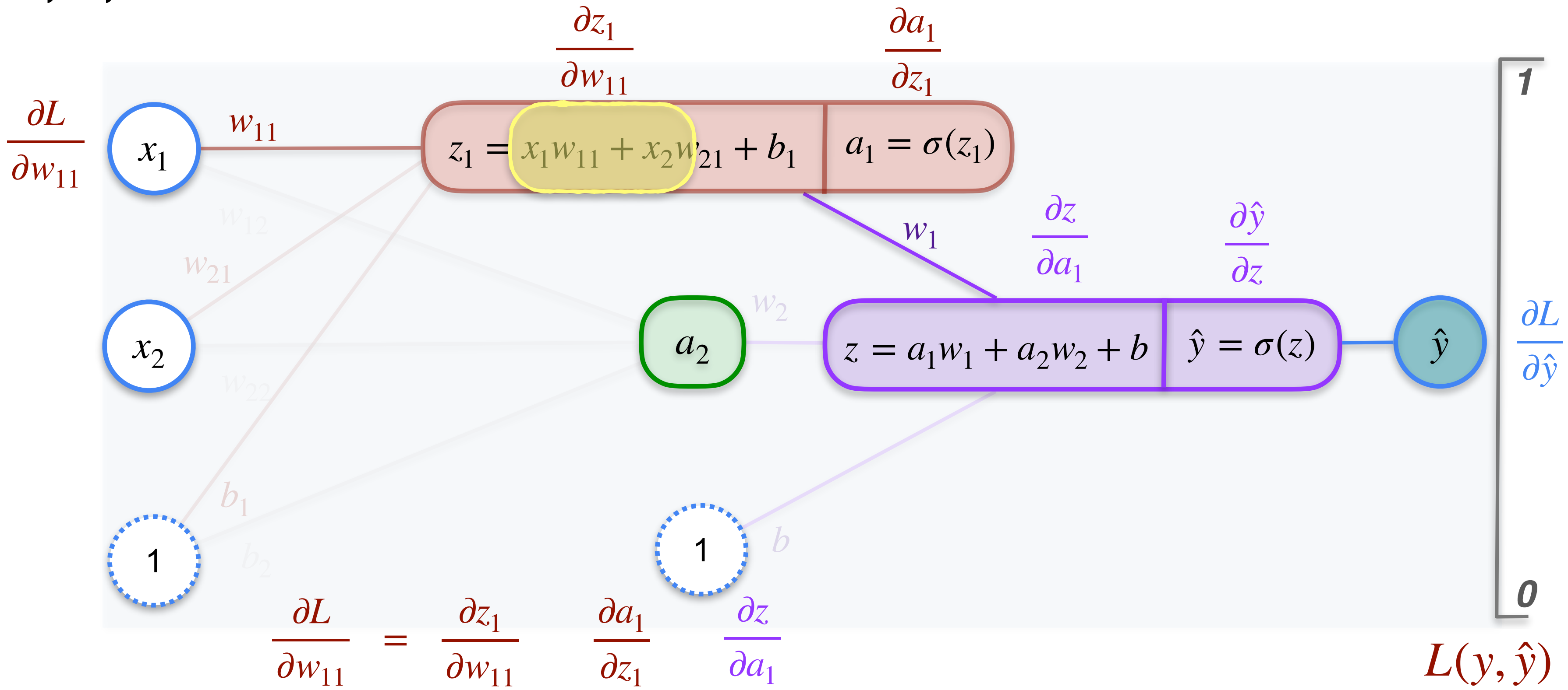
$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}}$$

2,2,1 Neural Network



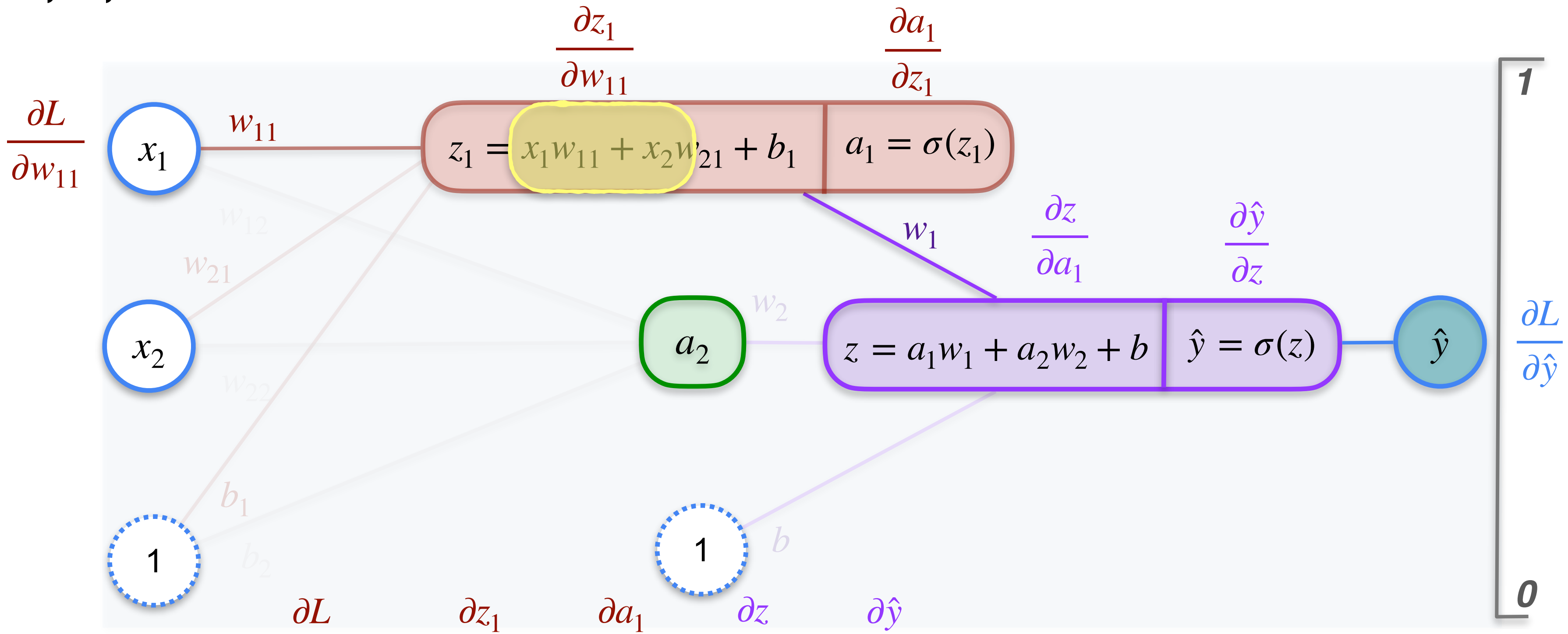
$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \quad \frac{\partial a_1}{\partial z_1}$$

2,2,1 Neural Network



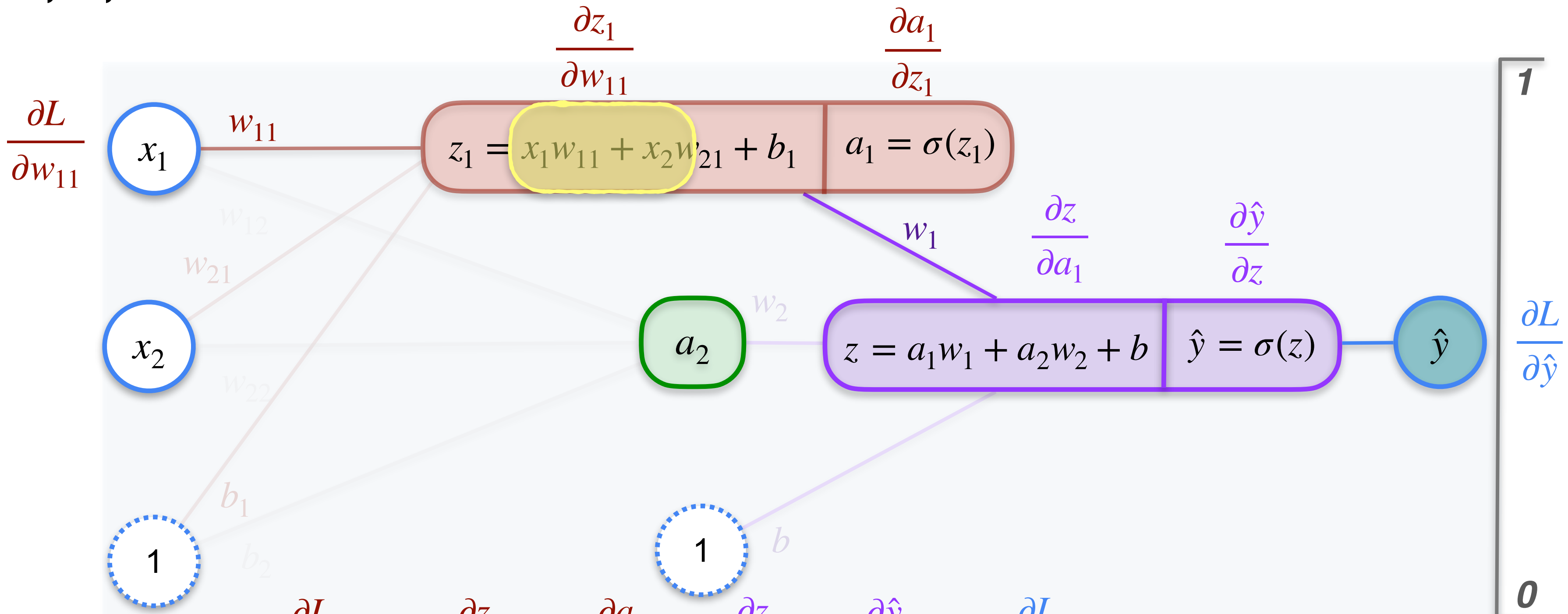
$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1}$$

2,2,1 Neural Network



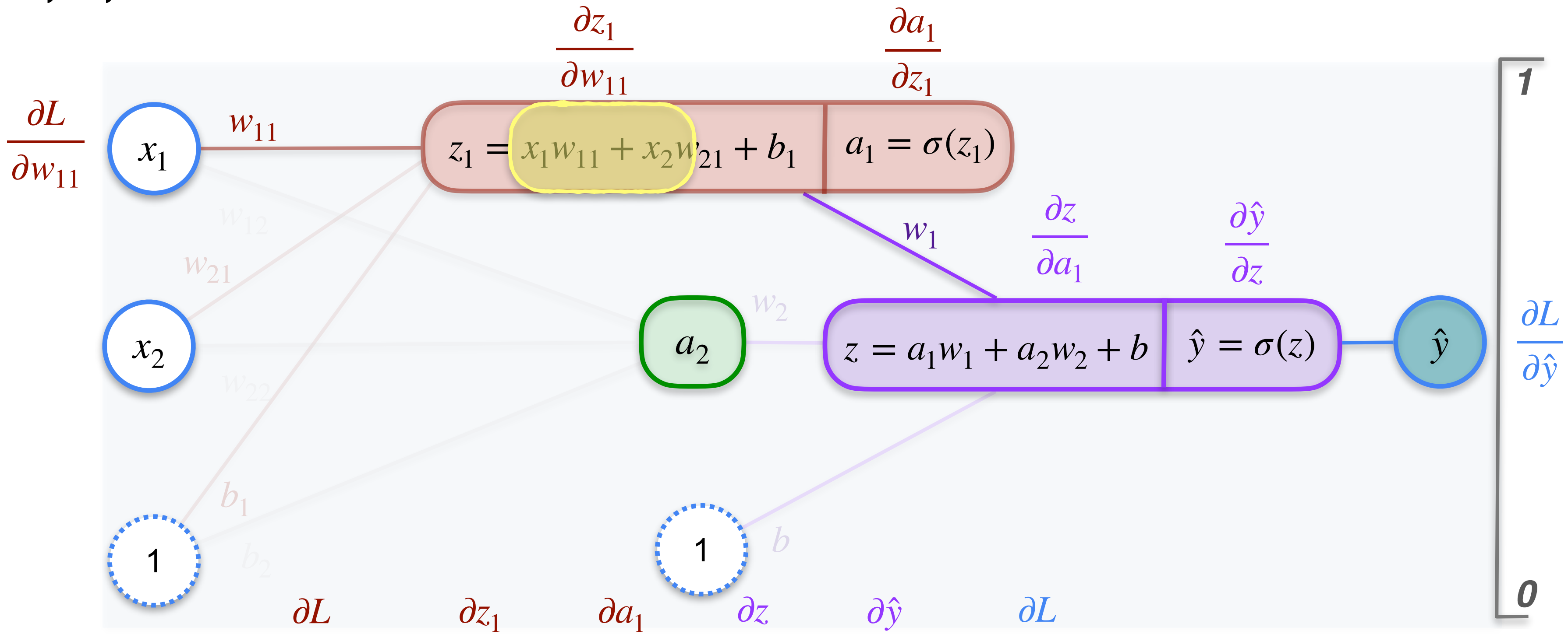
$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1} \quad \frac{\partial \hat{y}}{\partial z}$$

2,2,1 Neural Network



$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1} \quad \frac{\partial \hat{y}}{\partial z} \quad \frac{\partial L}{\partial \hat{y}}$$

2,2,1 Neural Network



$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

2,2,1 Neural Network

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}}$$

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} =$$

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} =$$

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} =$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1$$

DeepLearning.AI

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1$$

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1-a_1)$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1 - a_1)$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1-a_1)$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1-a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1-a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1-a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1-a_1) \quad w_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1 - a_1) \quad w_1 \quad \hat{y}(1 - \hat{y})$$

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1-a_1) \quad w_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \quad a_1(1 - a_1) \quad w_1 \quad \hat{y}(1 - \hat{y}) \quad \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1 - a_1) \cdot w_1 \cdot \hat{y}(1 - \hat{y}) \cdot \frac{\text{-}(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y \ - \ \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -x_1 w_1 a_1(1-a_1)(y - \hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -x_1 w_1 a_1 (1-a_1)(y - \hat{y})$$

Perform gradient descent with

to find optimal value of w_{11} that gives the least error

DeepLearning.AI

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -x_1 w_1 a_1 (1-a_1)(y - \hat{y})$$

Perform gradient descent with

$$w_{11} \rightarrow w_{11} - \alpha \frac{\partial L}{\partial w_{11}}$$

to find optimal value of w_{11} that gives the least error

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -x_1 w_1 a_1(1-a_1)(y - \hat{y})$$

Perform gradient descent with

$w_{11} \rightarrow w_{11} - \alpha$

to find optimal value of w_{11} that gives the least error

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial z_1}{\partial w_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{11}} = x_1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

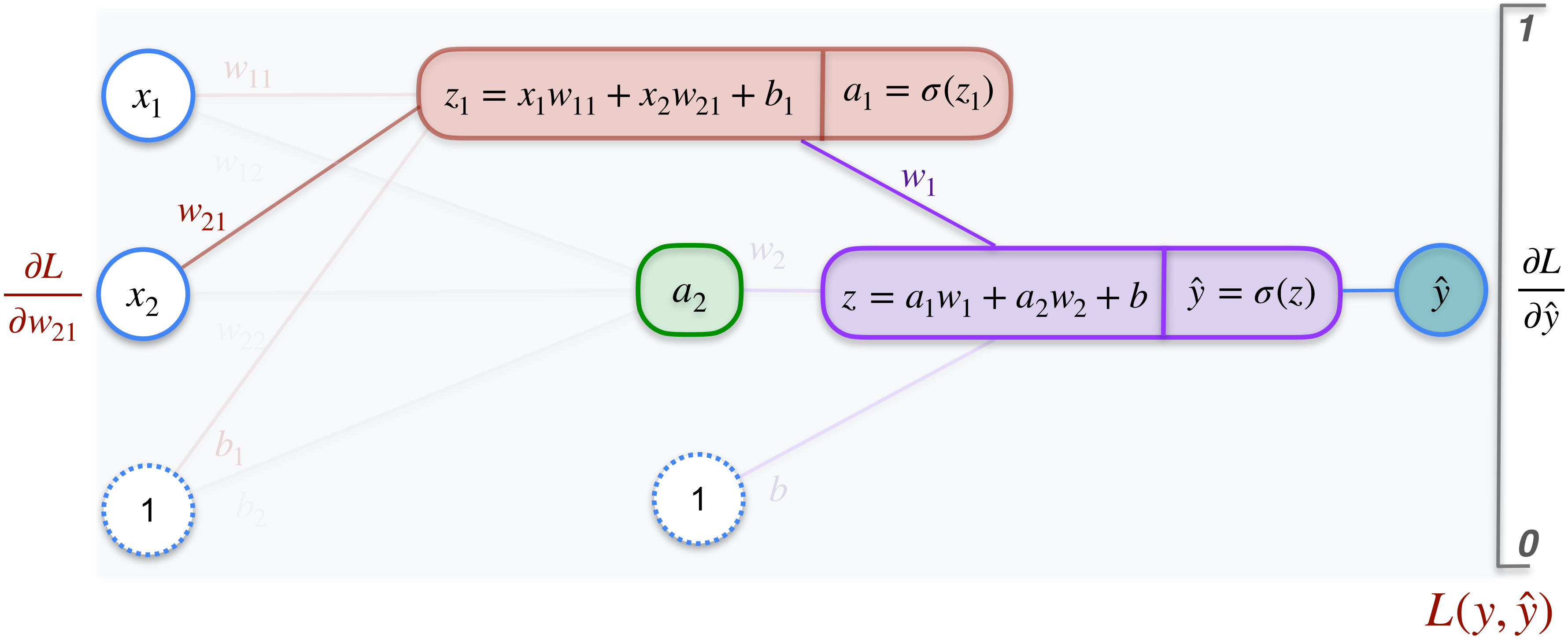
$$= -x_1 w_1 a_1 (1-a_1)(y - \hat{y})$$

Perform gradient descent with

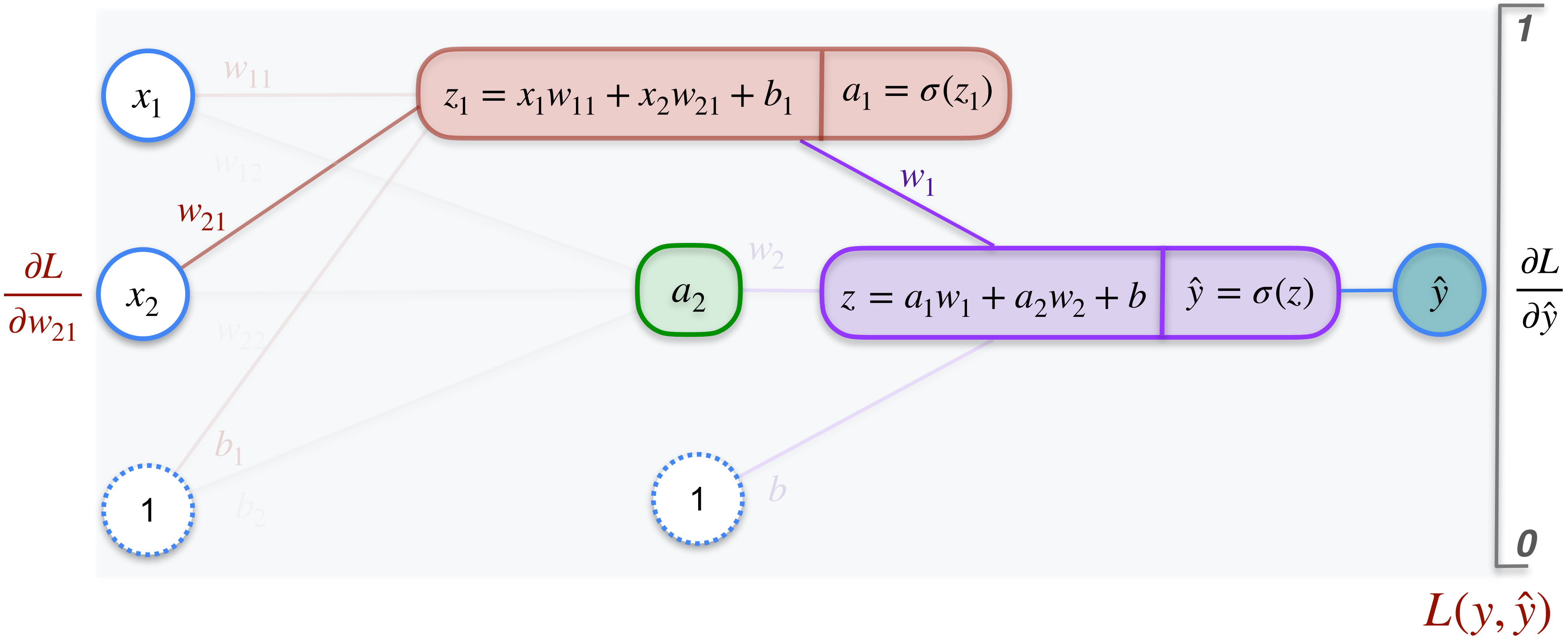
$$w_{11} \rightarrow w_{11} - \alpha \text{-} x_1 w_1 a_1 (1-a_1)(y - \hat{y})$$

to find optimal value of w_{11} that gives the least error

2,2,1 Neural Network

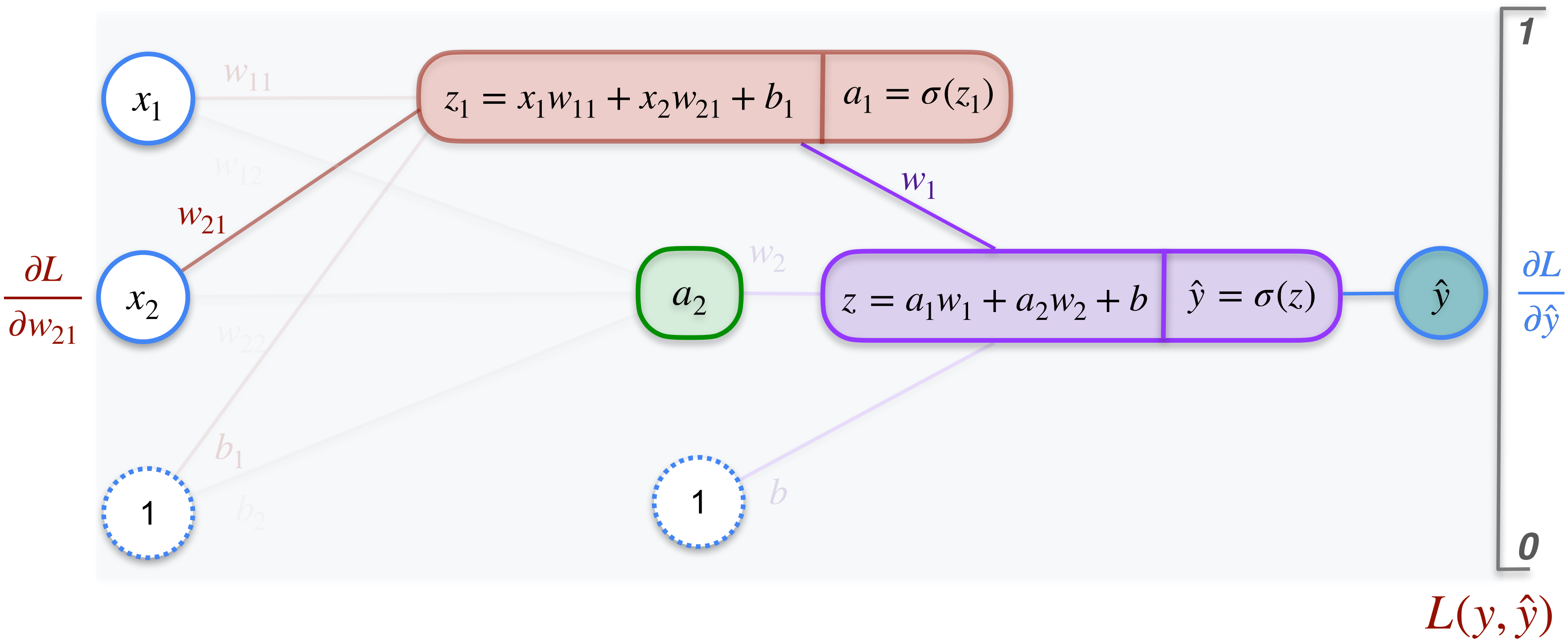


2,2,1 Neural Network

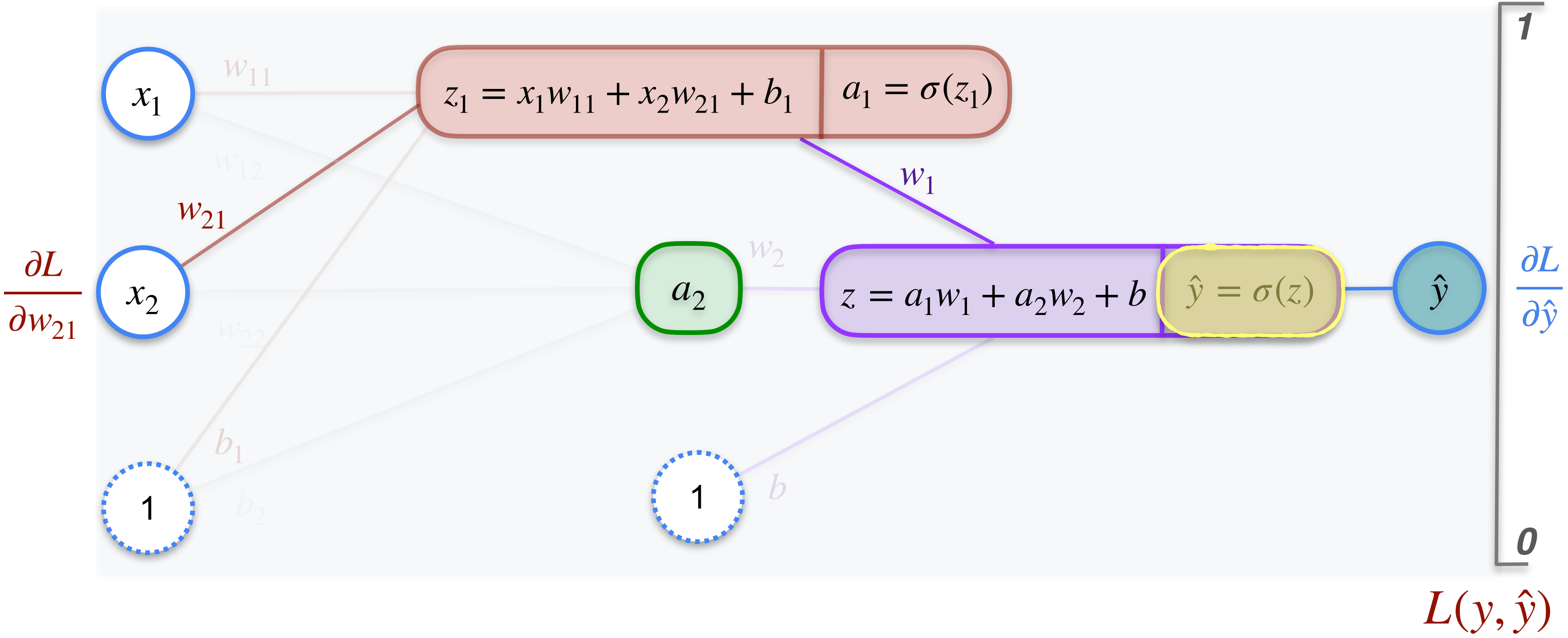


DeepLearning.AI

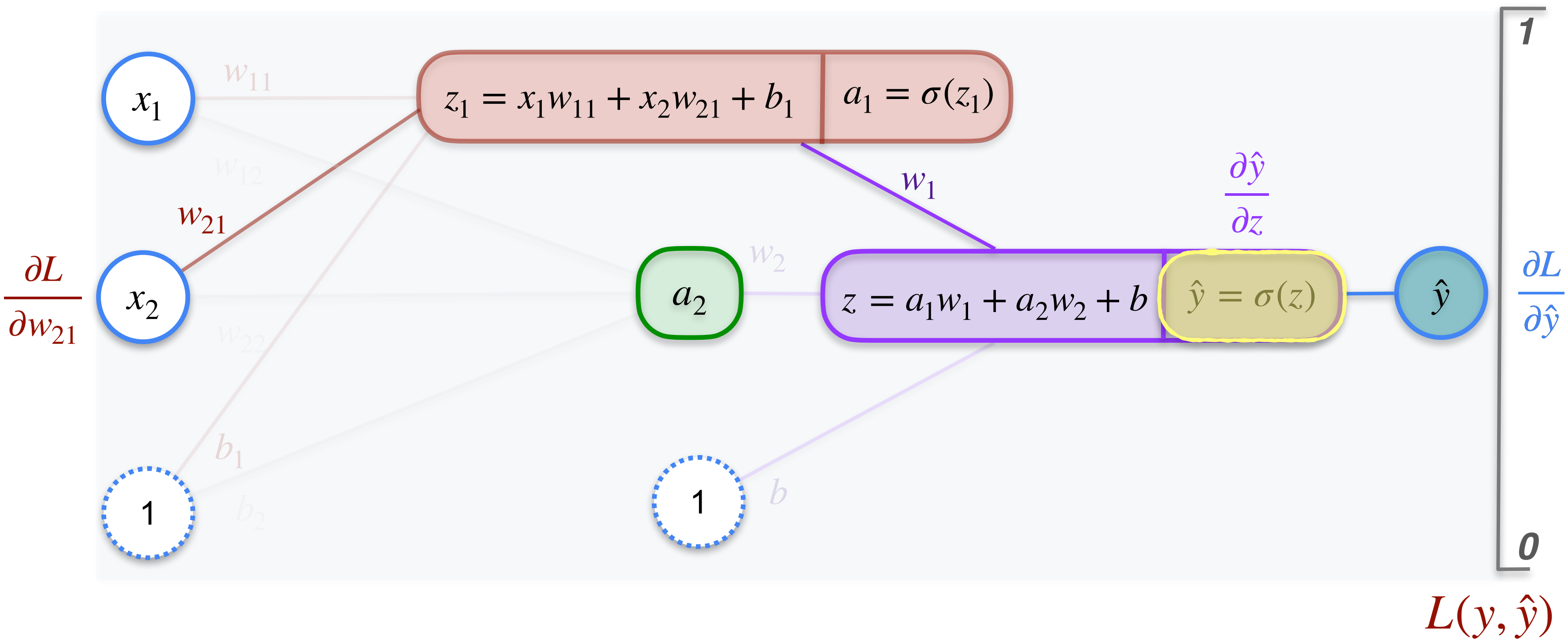
2,2,1 Neural Network



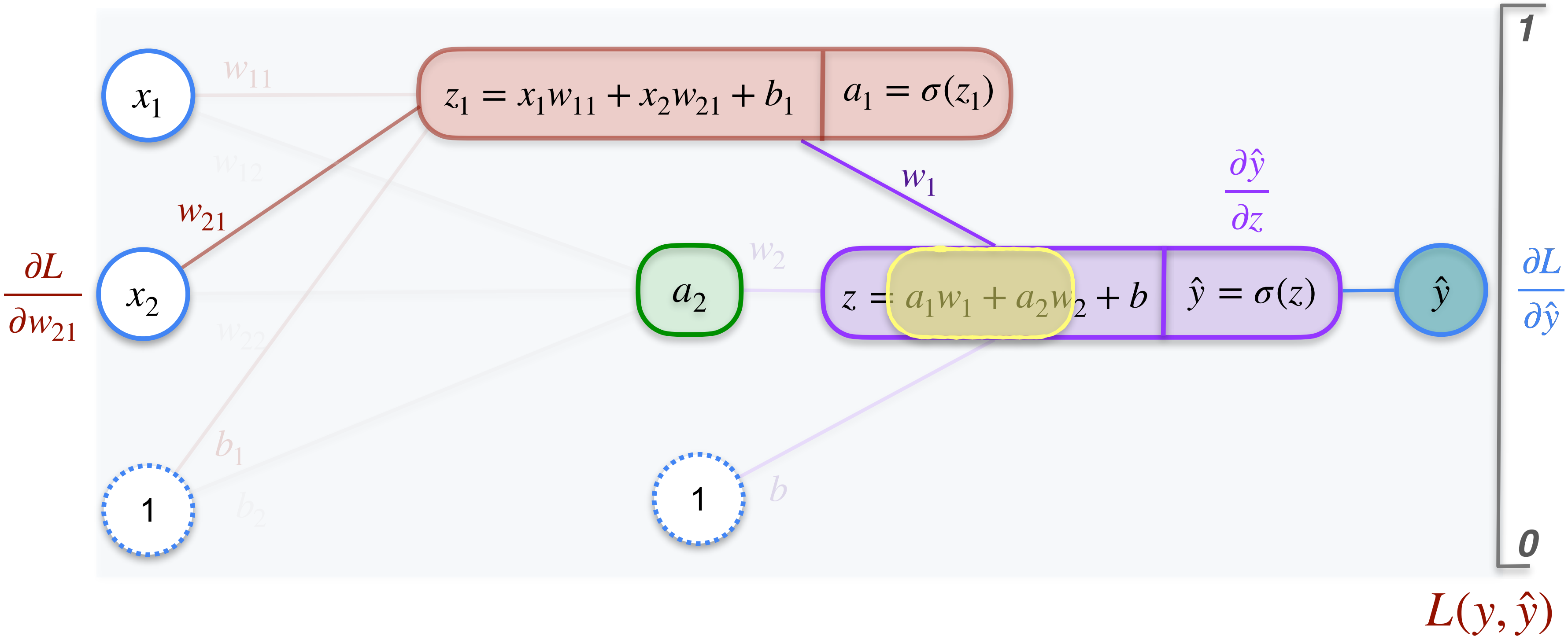
2,2,1 Neural Network



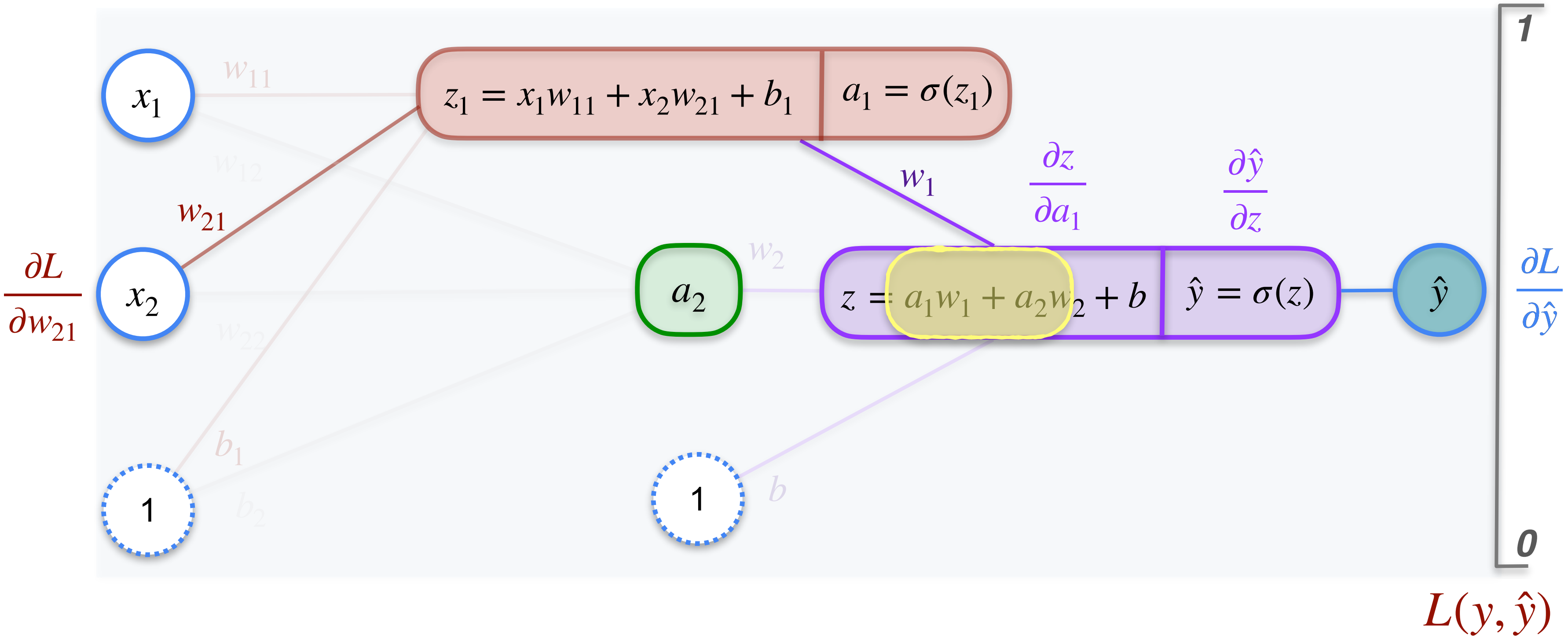
2,2,1 Neural Network



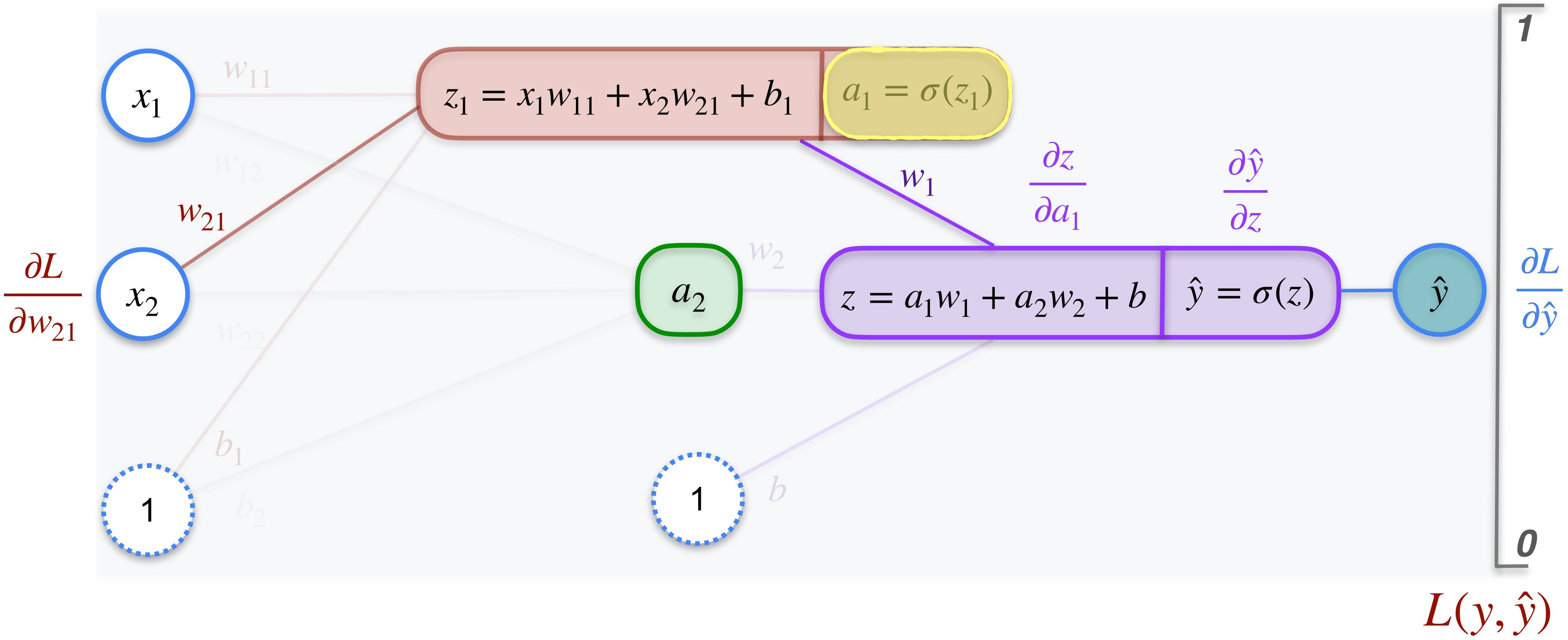
2,2,1 Neural Network



2,2,1 Neural Network

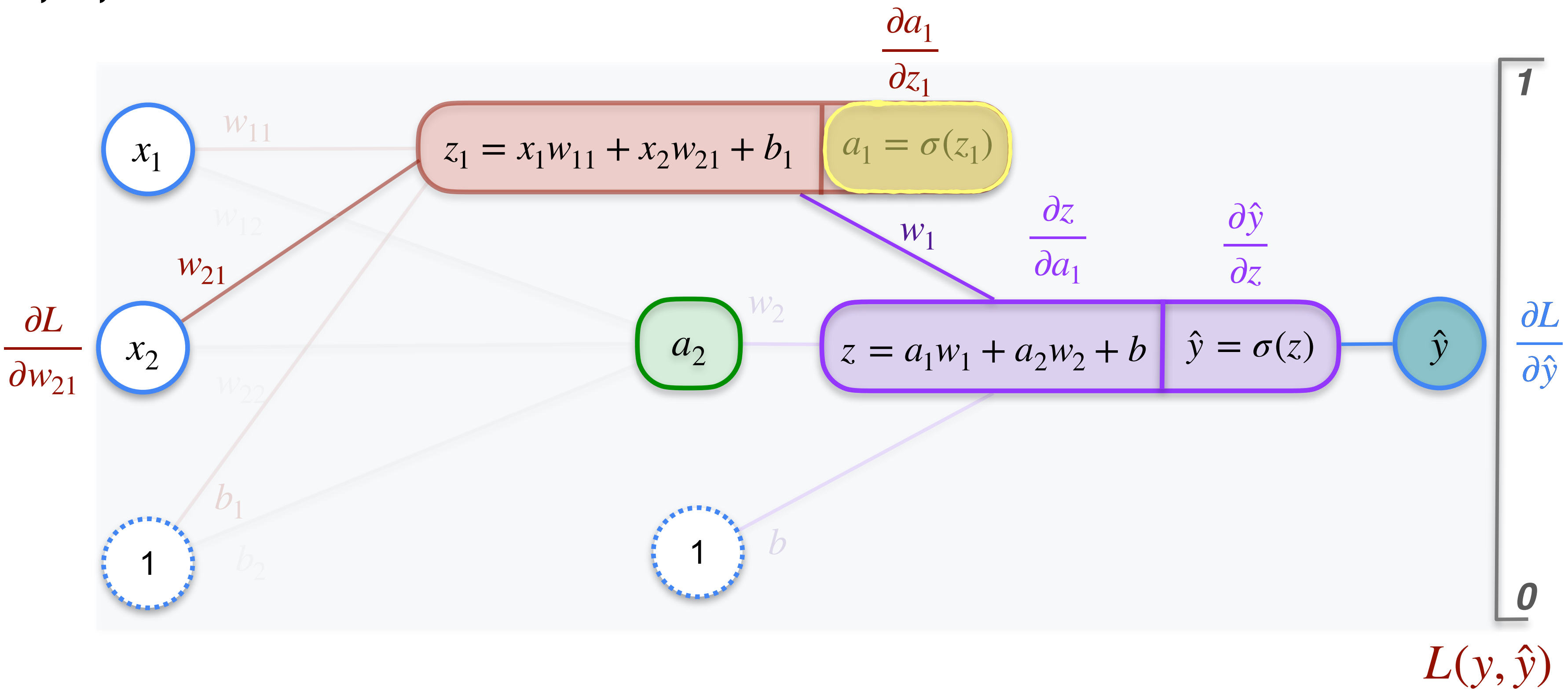


2,2,1 Neural Network

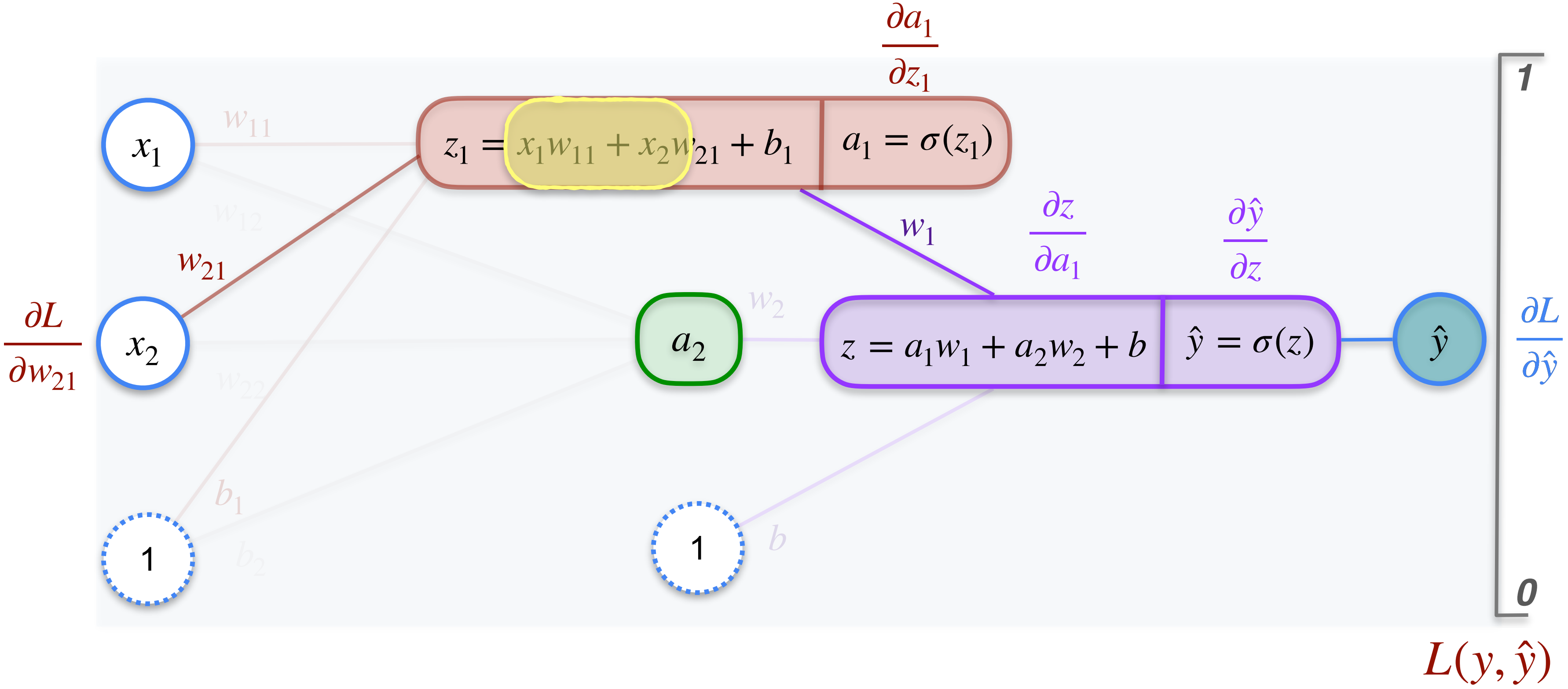


DeepLearning.AI

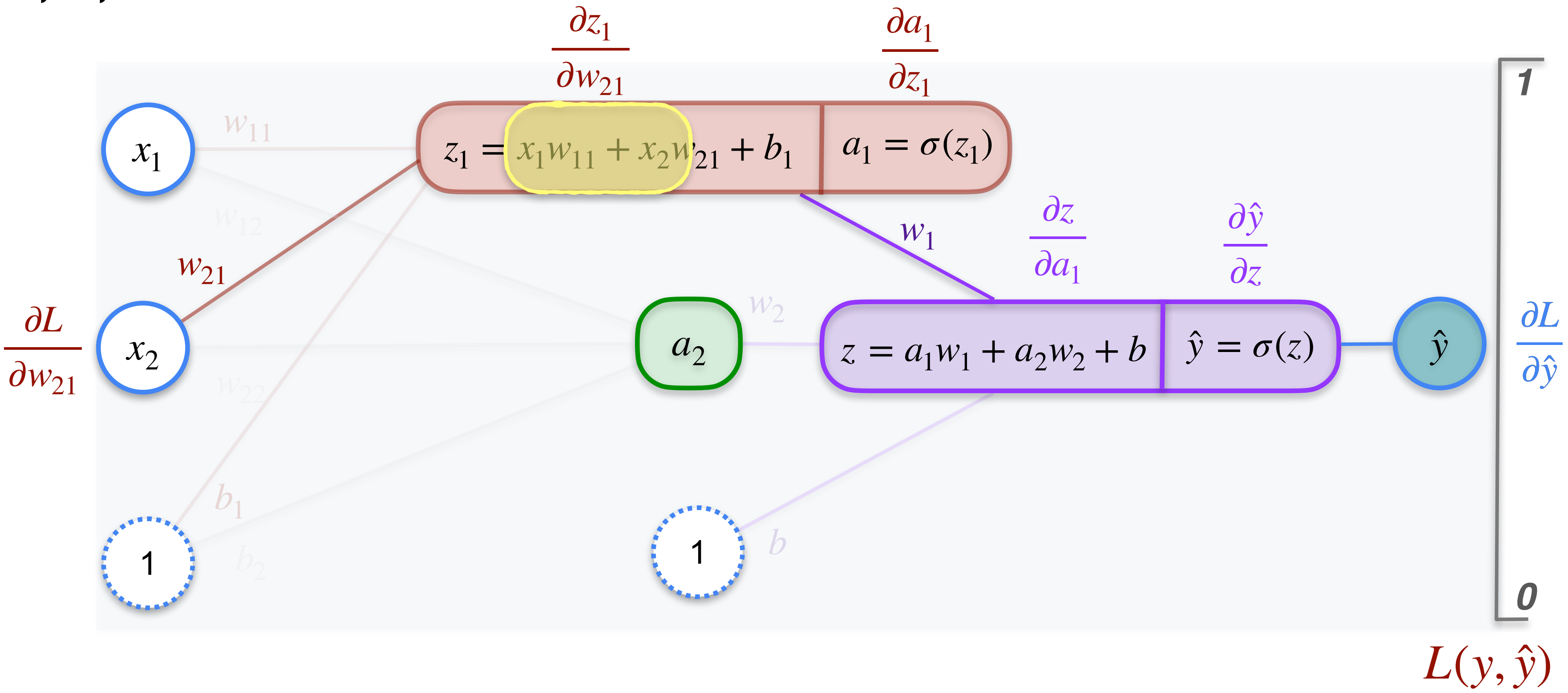
2,2,1 Neural Network



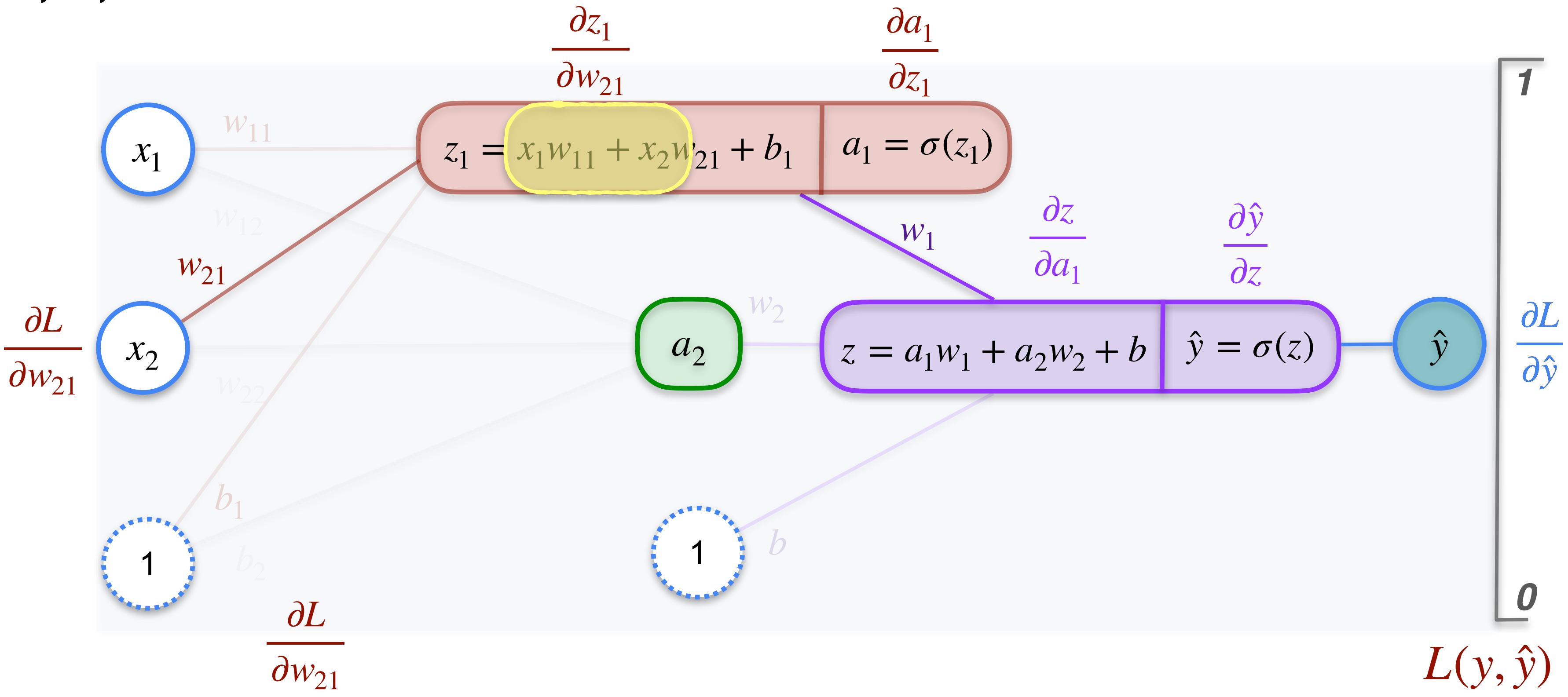
2,2,1 Neural Network



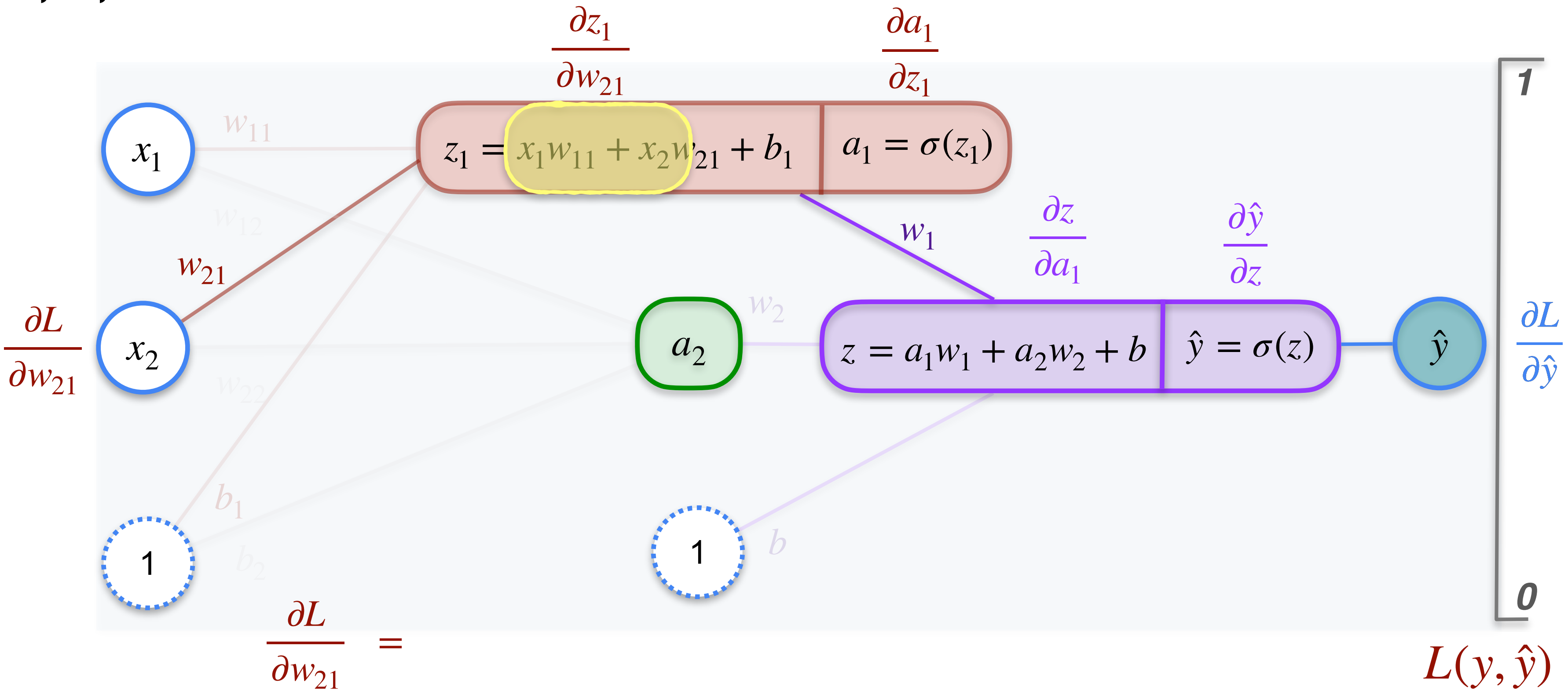
2,2,1 Neural Network



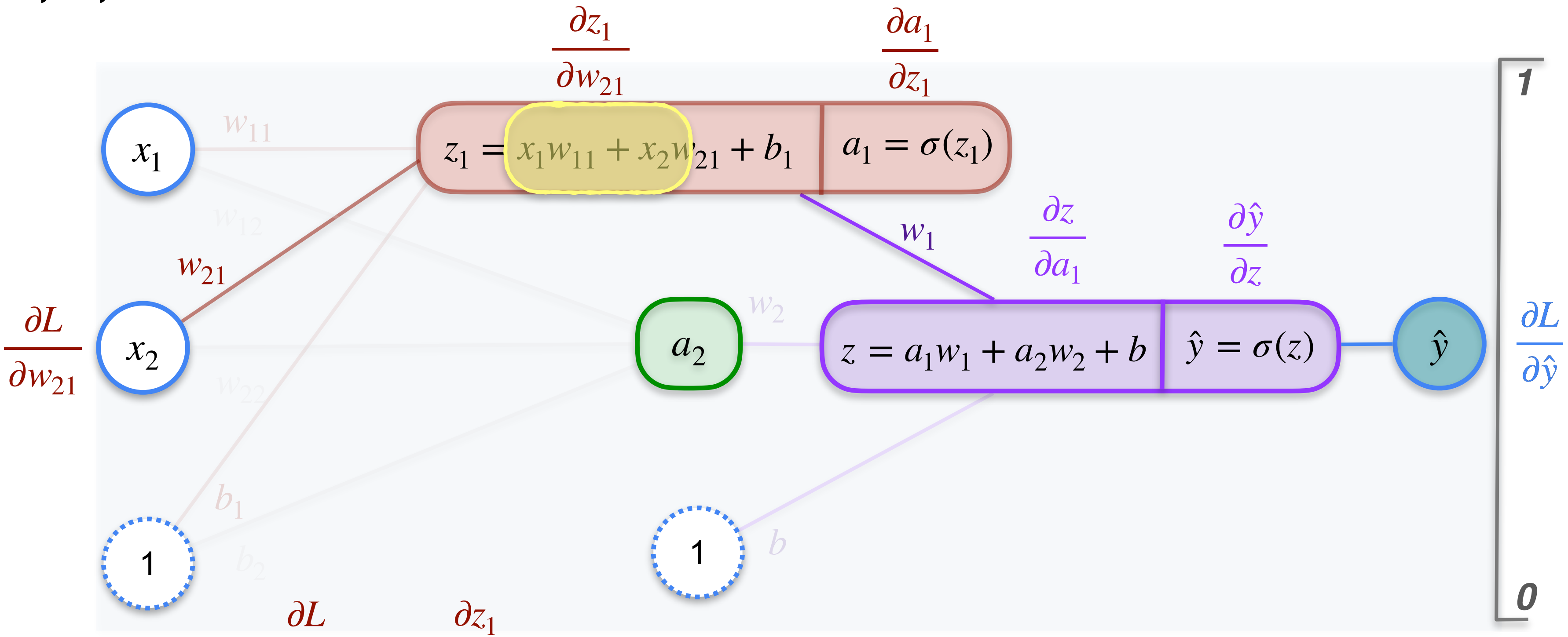
2,2,1 Neural Network



2,2,1 Neural Network

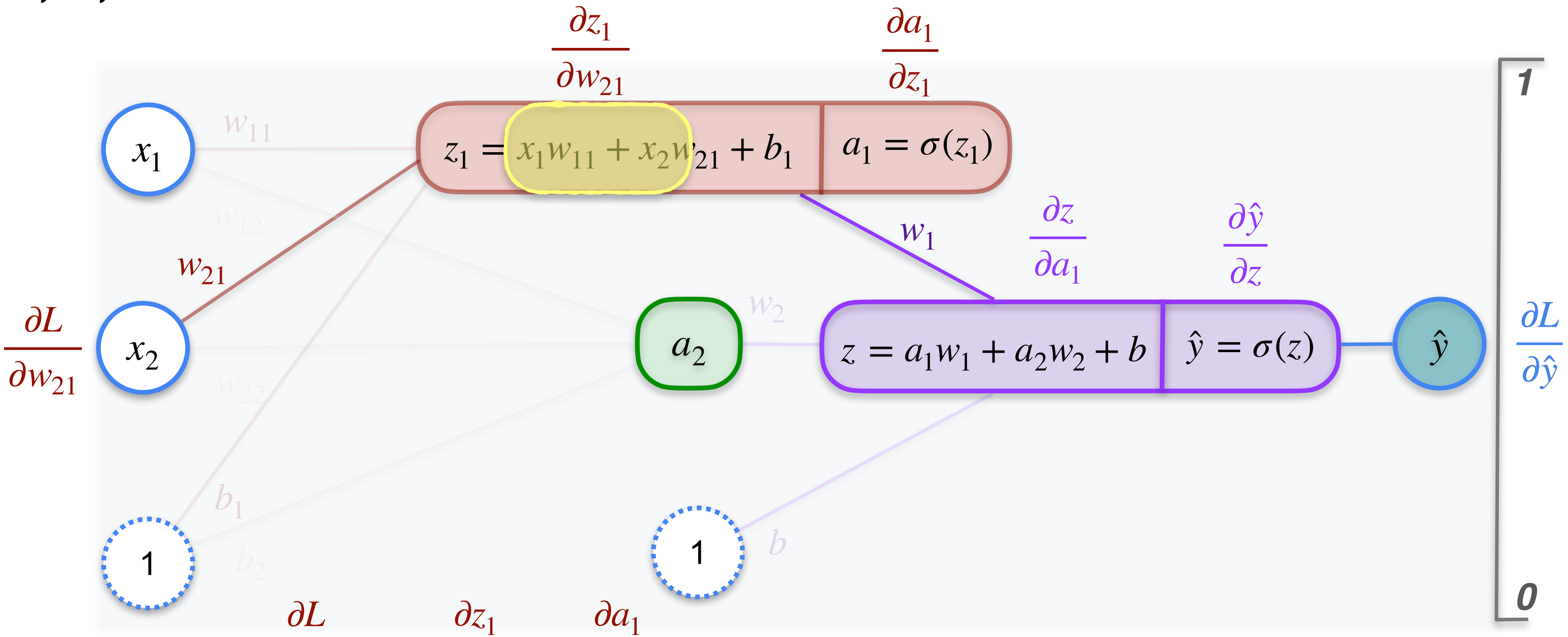


2,2,1 Neural Network



$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}}$$

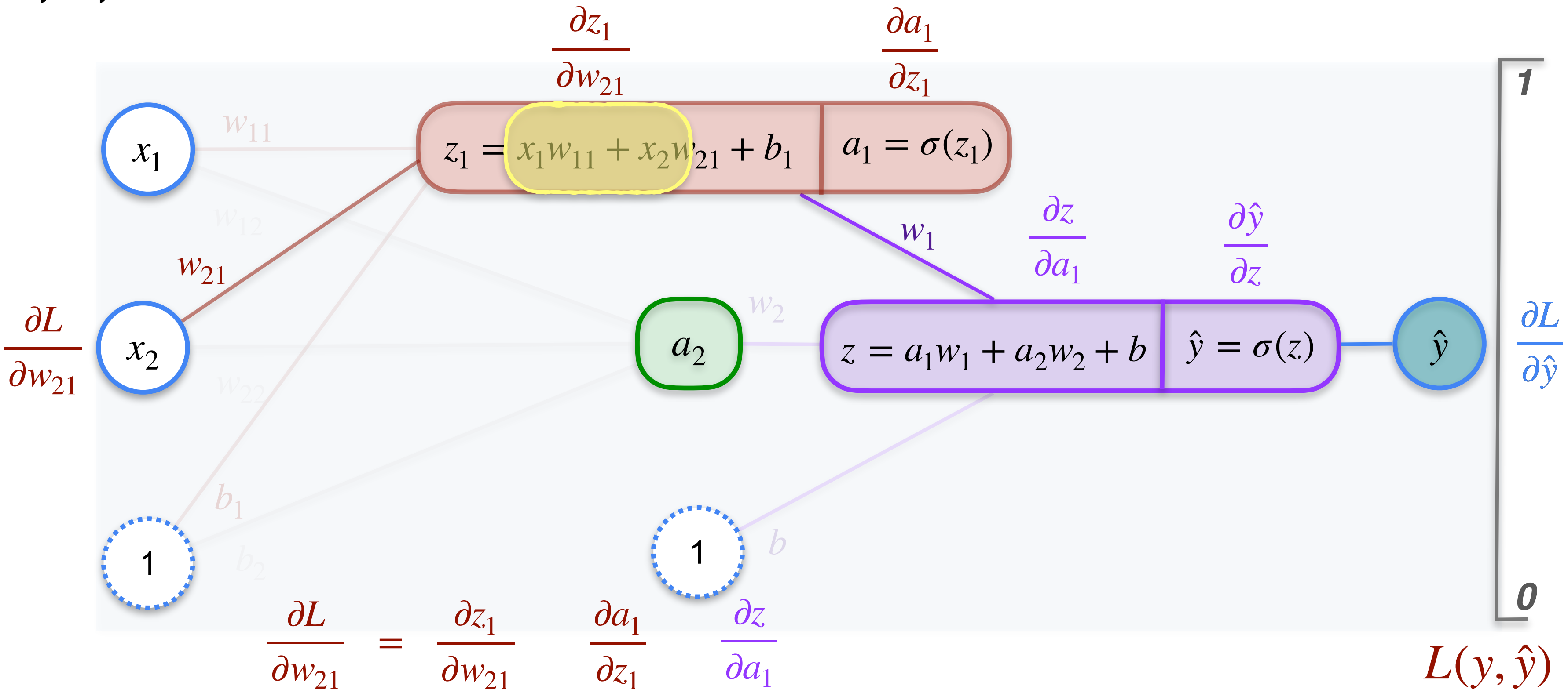
2,2,1 Neural Network



$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \quad \frac{\partial a_1}{\partial z_1}$$

DeepLearning.AI

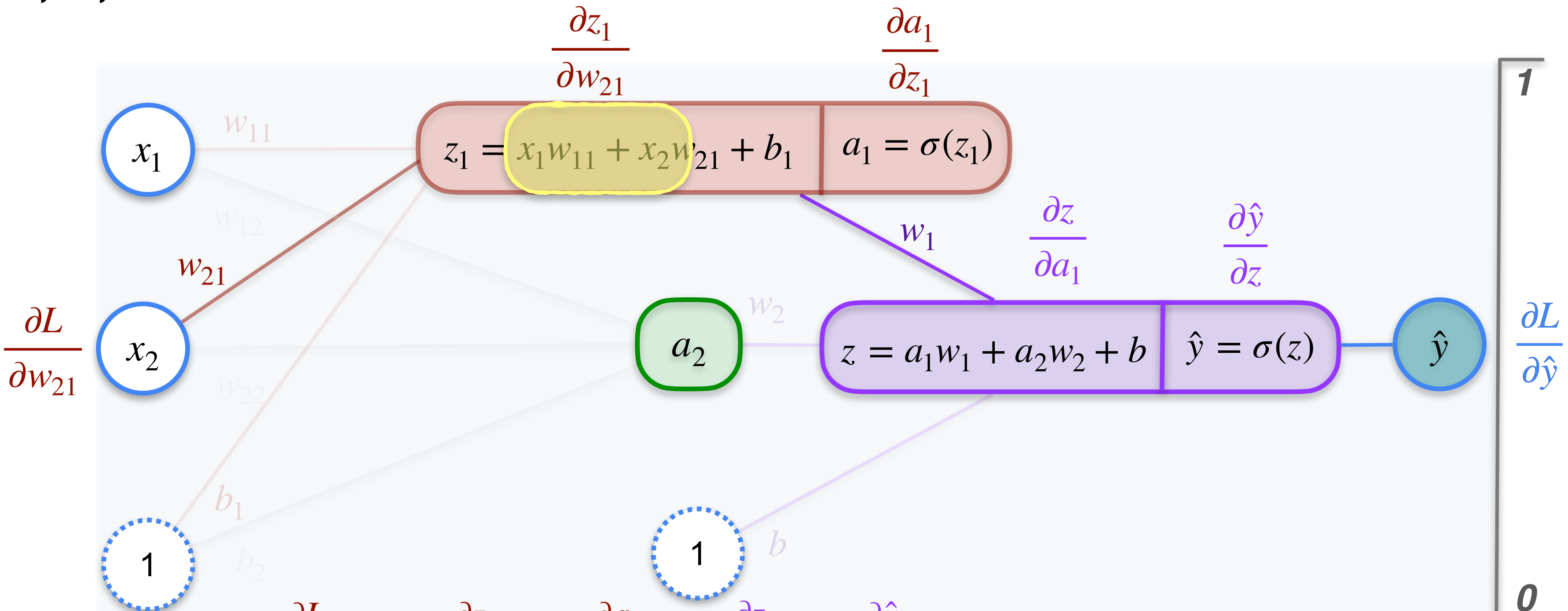
2,2,1 Neural Network



$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1}$$

DeepLearning.AI

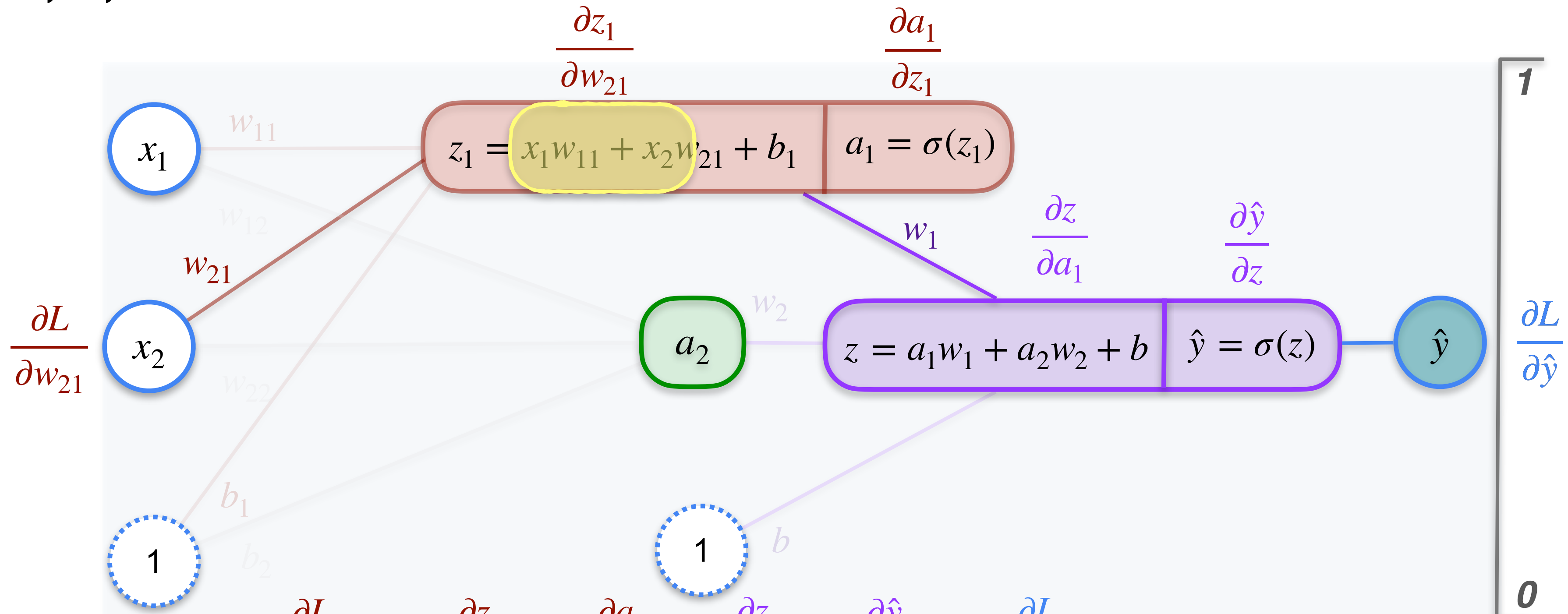
2,2,1 Neural Network



$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1} \quad \frac{\partial \hat{y}}{\partial z}$$

$L(y, \hat{y})$

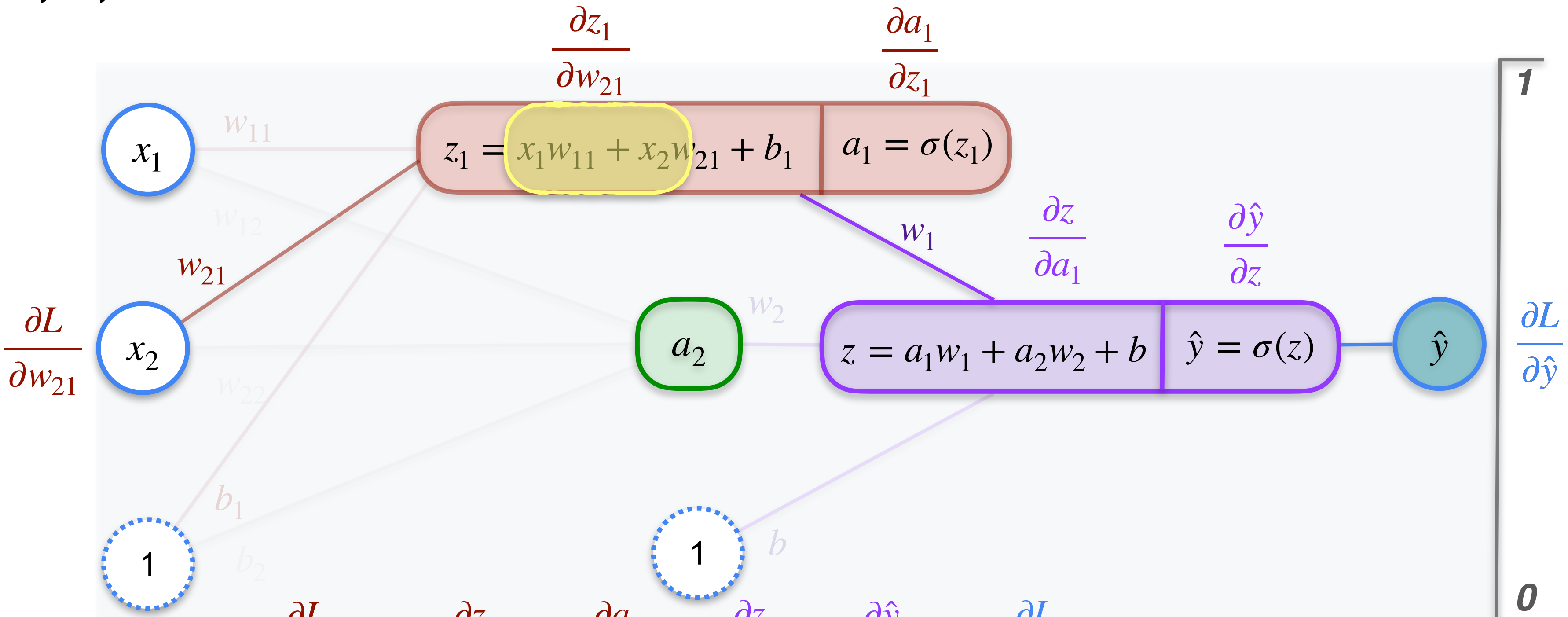
2,2,1 Neural Network



$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1} \quad \frac{\partial \hat{y}}{\partial z} \quad \frac{\partial L}{\partial \hat{y}}$$

$L(y, \hat{y})$

2,2,1 Neural Network



$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$L(y, \hat{y})$

2,2,1 Neural Network

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}}$$

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

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} =$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} =$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2$$

DeepLearning.AI

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1 - a_1)$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1-a_1)$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1 - a_1)$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1 - a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1 - a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1-a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \boxed{\frac{\partial \hat{y}}{\partial z} \cdot} \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1-a_1) \quad w_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1 - a_1) \quad w_1 \quad \hat{y}(1 - \hat{y})$$

DeepLearning.AI

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1-a_1) \quad w_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \quad a_1(1-a_1) \quad w_1 \quad \hat{y}(1-\hat{y}) \quad \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1-a_1) \cdot w_1 \cdot \hat{y}(1-\hat{y}) \cdot \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

DeepLearning.AI

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -x_2 w_1 a_1 (1-a_1)(y - \hat{y})$$

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -x_2 w_1 a_1 (1-a_1)(y - \hat{y})$$

Perform gradient descent with

to find optimal value of w_{21} that gives the least error

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -x_2 w_1 a_1 (1-a_1)(y - \hat{y})$$

Perform gradient descent with

$$w_{21} \rightarrow w_{21} - \alpha \frac{\partial L}{\partial w_{21}}$$

to find optimal value of w_{21} that gives the least error

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -x_2w_1a_1(1-a_1)(y - \hat{y})$$

Perform gradient descent with

$$w_{21} \rightarrow w_{21} - \alpha$$

to find optimal value of w_{21} that gives the least error

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial z_1}{\partial w_{21}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_{21}} = x_2 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

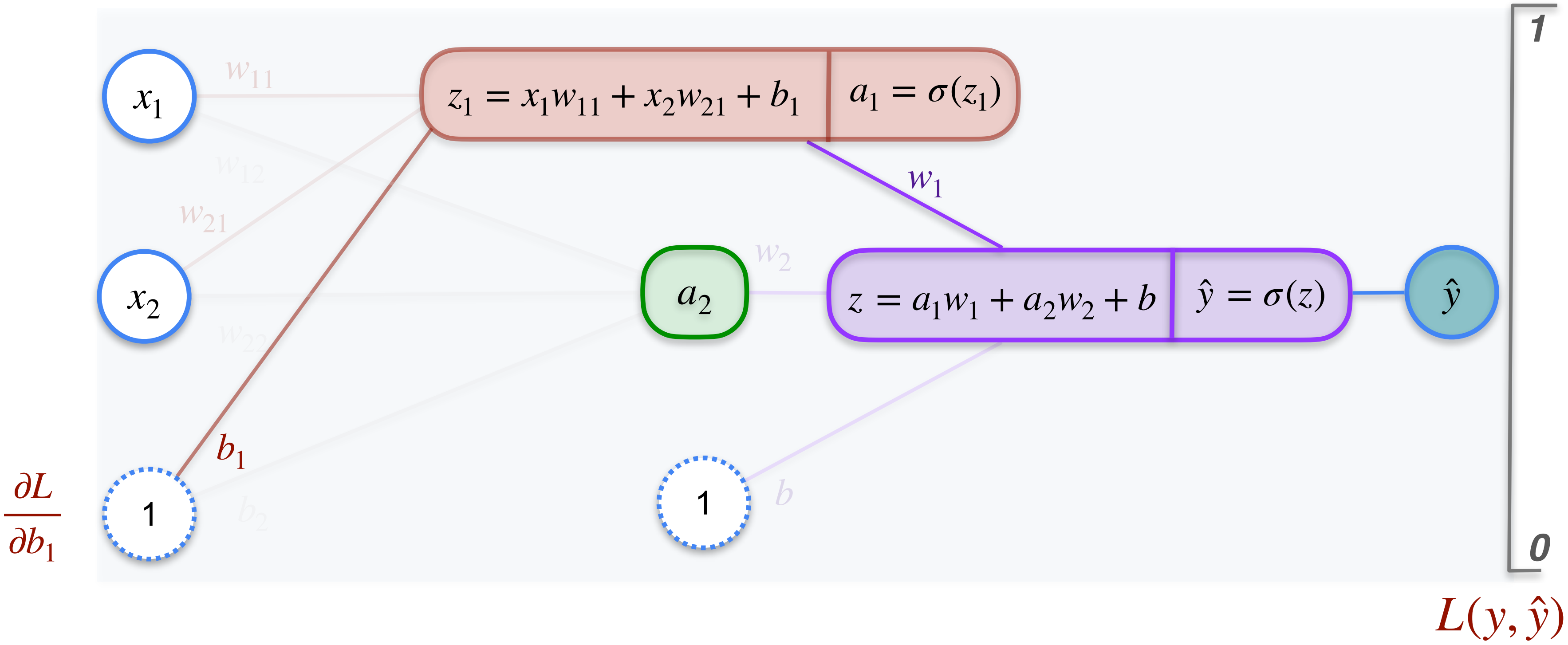
$$= -x_2w_1a_1(1-a_1)(y - \hat{y})$$

Perform gradient descent with

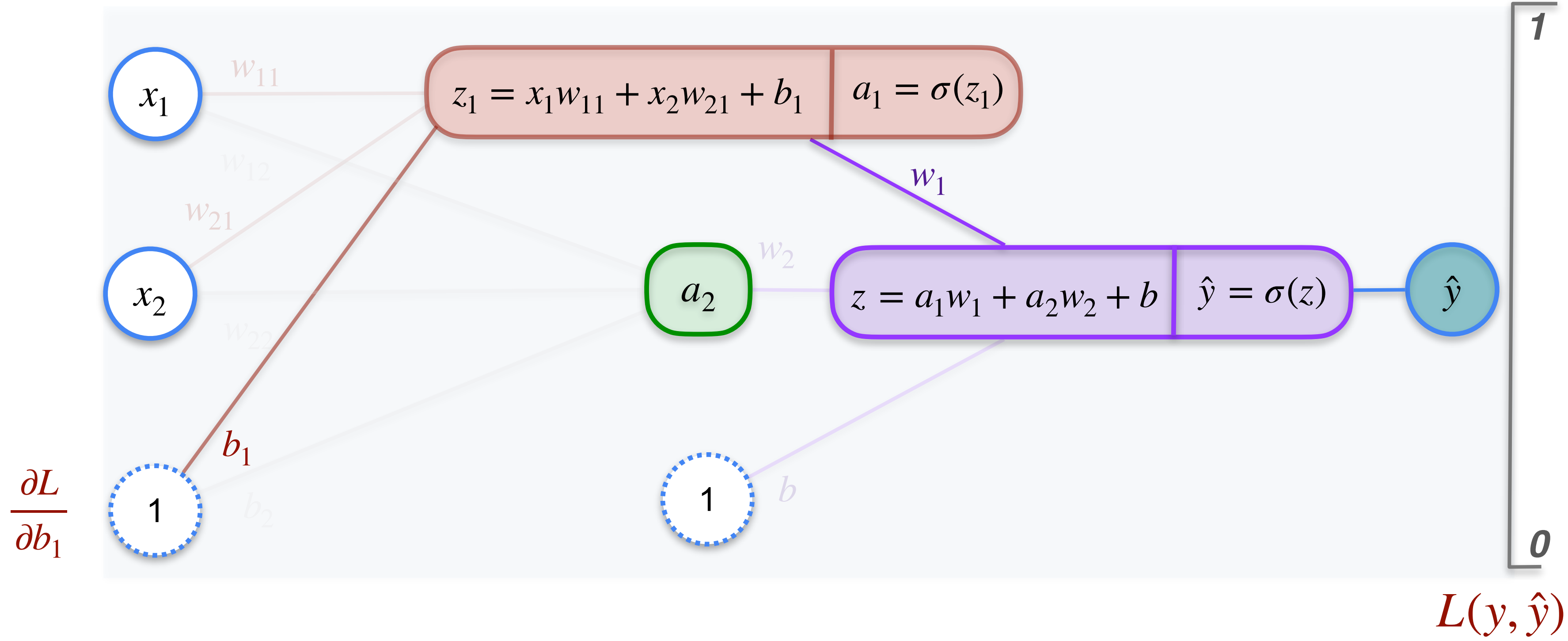
$$w_{21} \rightarrow w_{21} - \alpha \text{-} x_2w_1a_1(1-a_1)(y - \hat{y})$$

to find optimal value of w_{21} that gives the least error

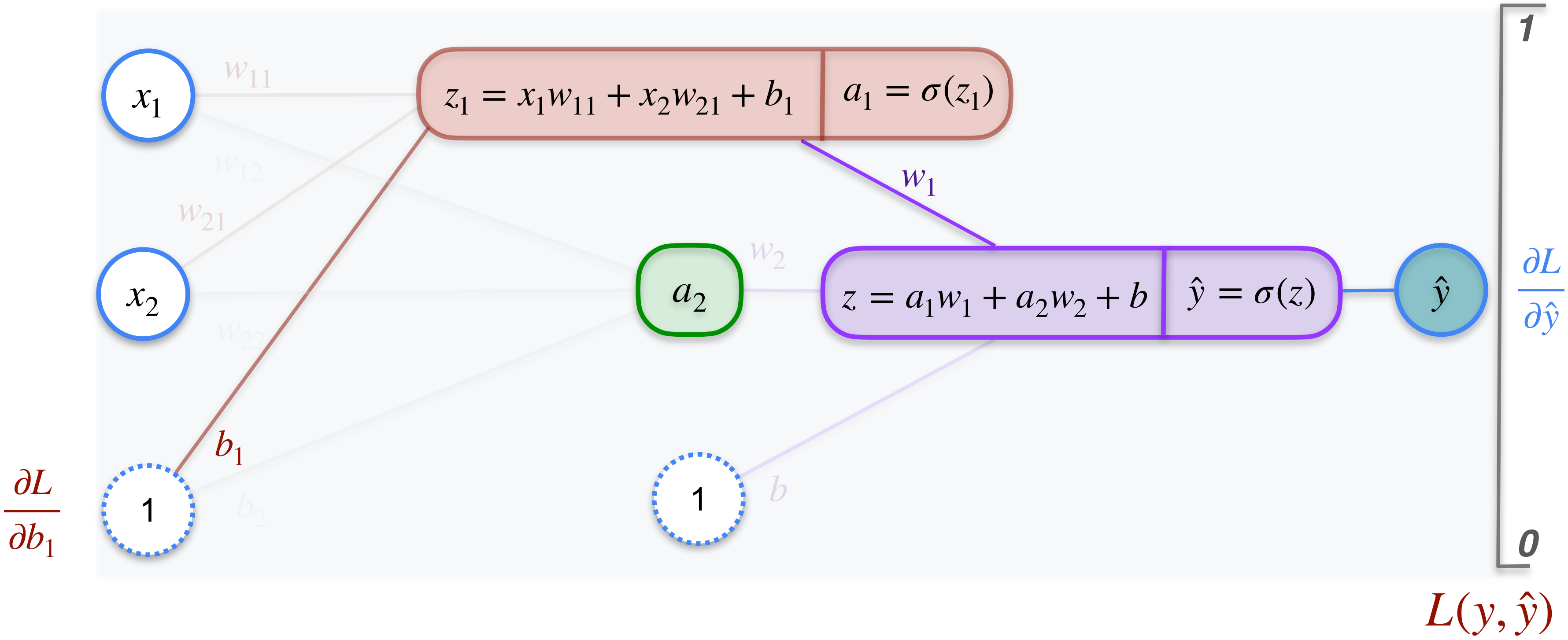
2,2,1 Neural Network



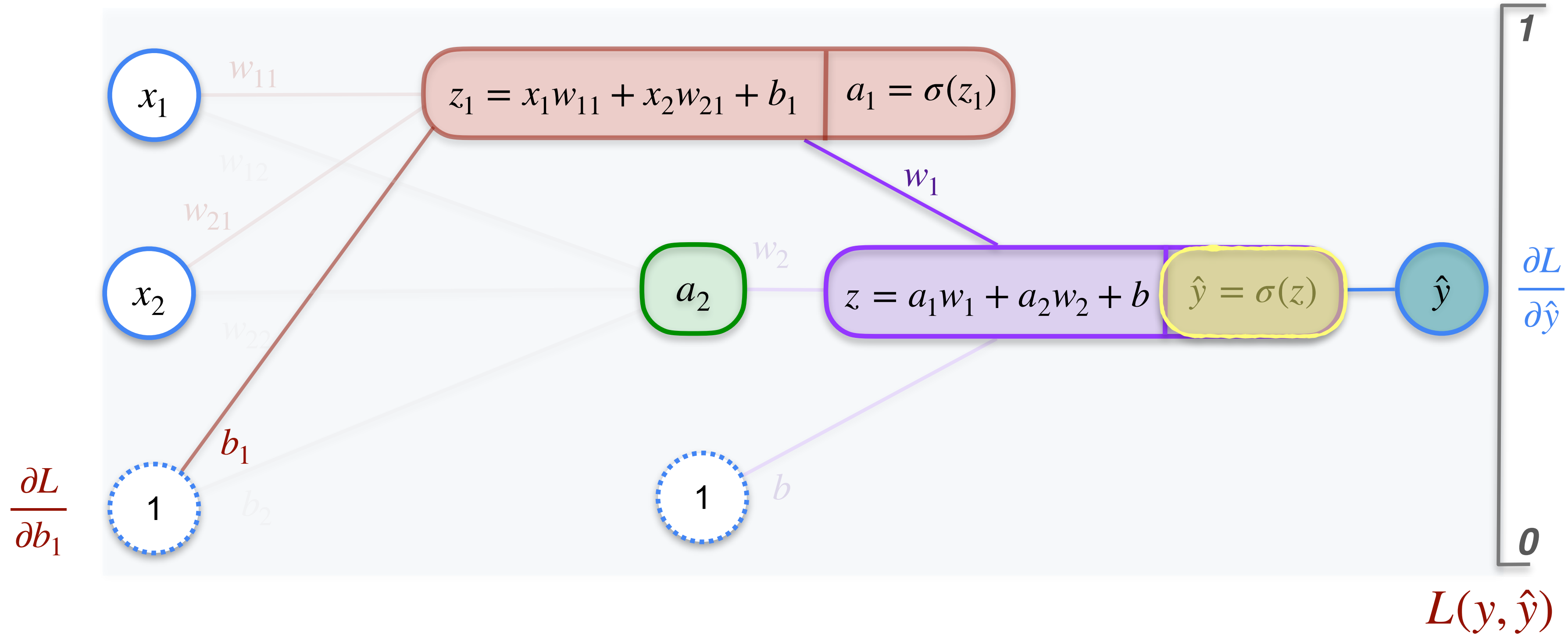
2,2,1 Neural Network



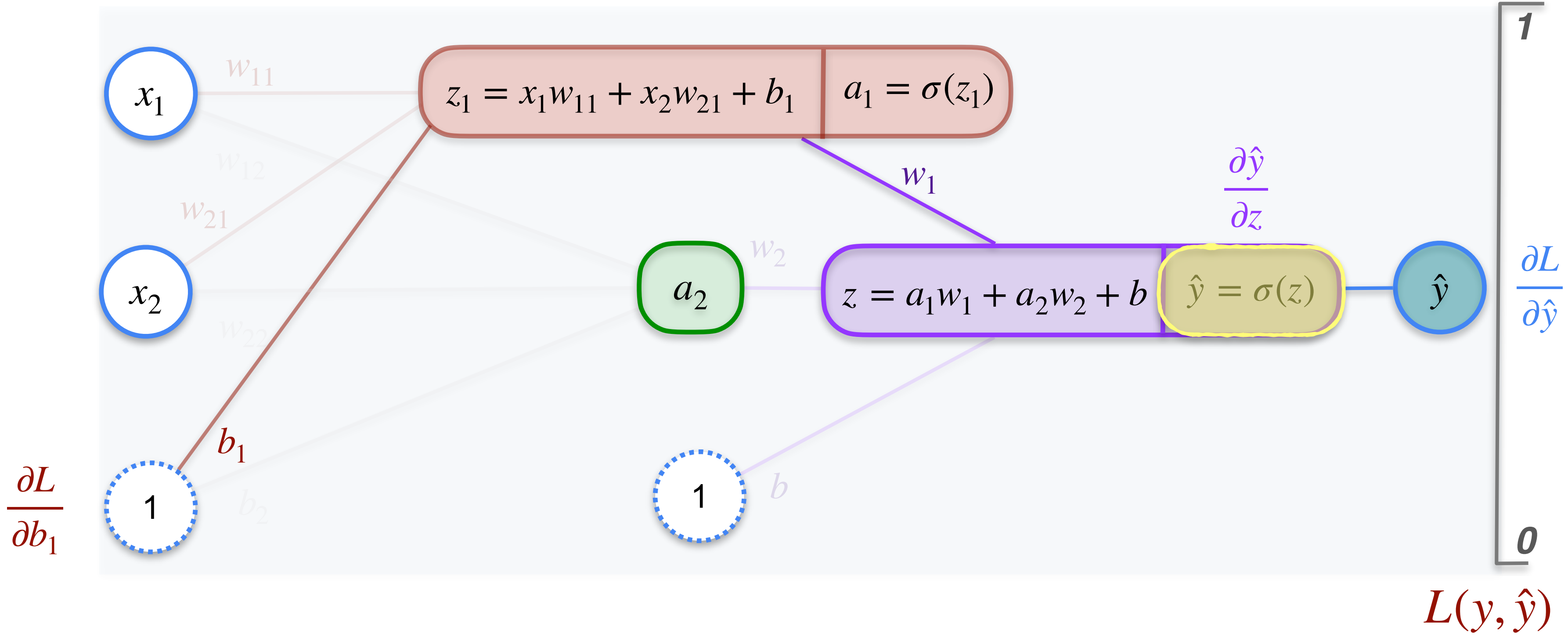
2,2,1 Neural Network



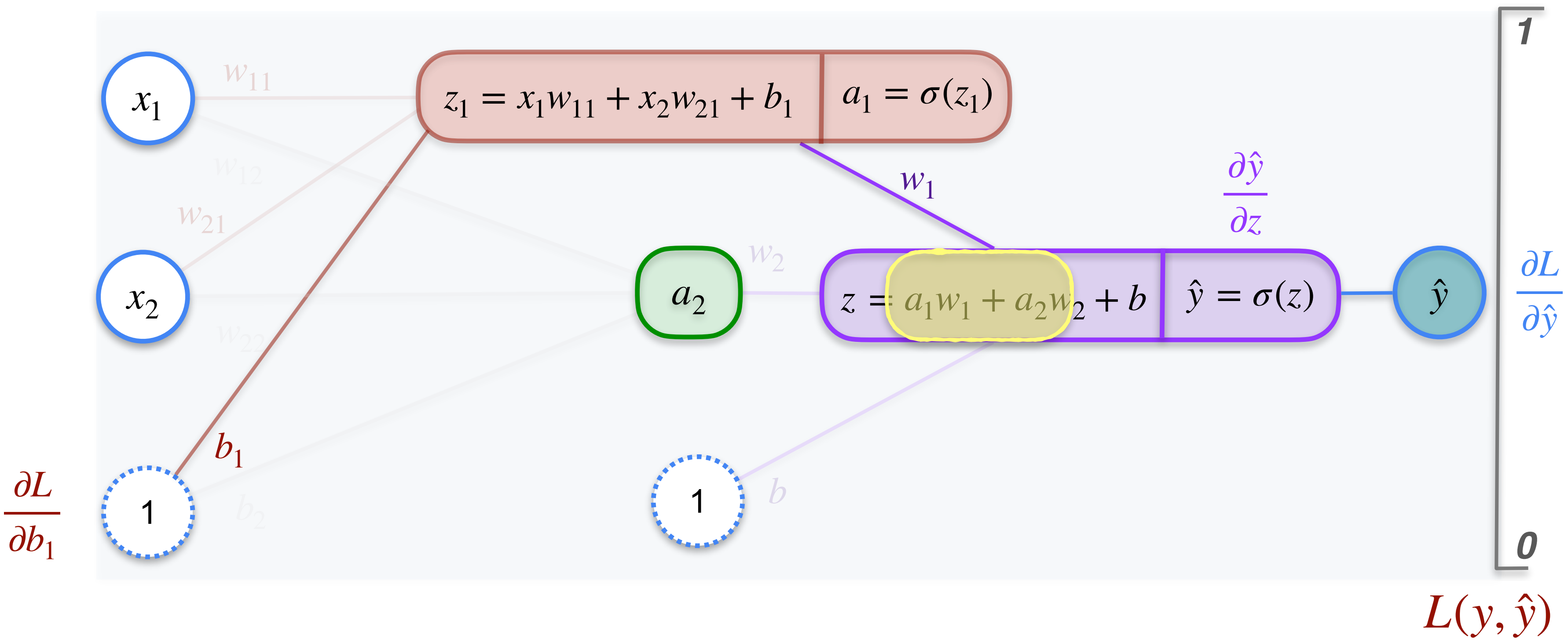
2,2,1 Neural Network



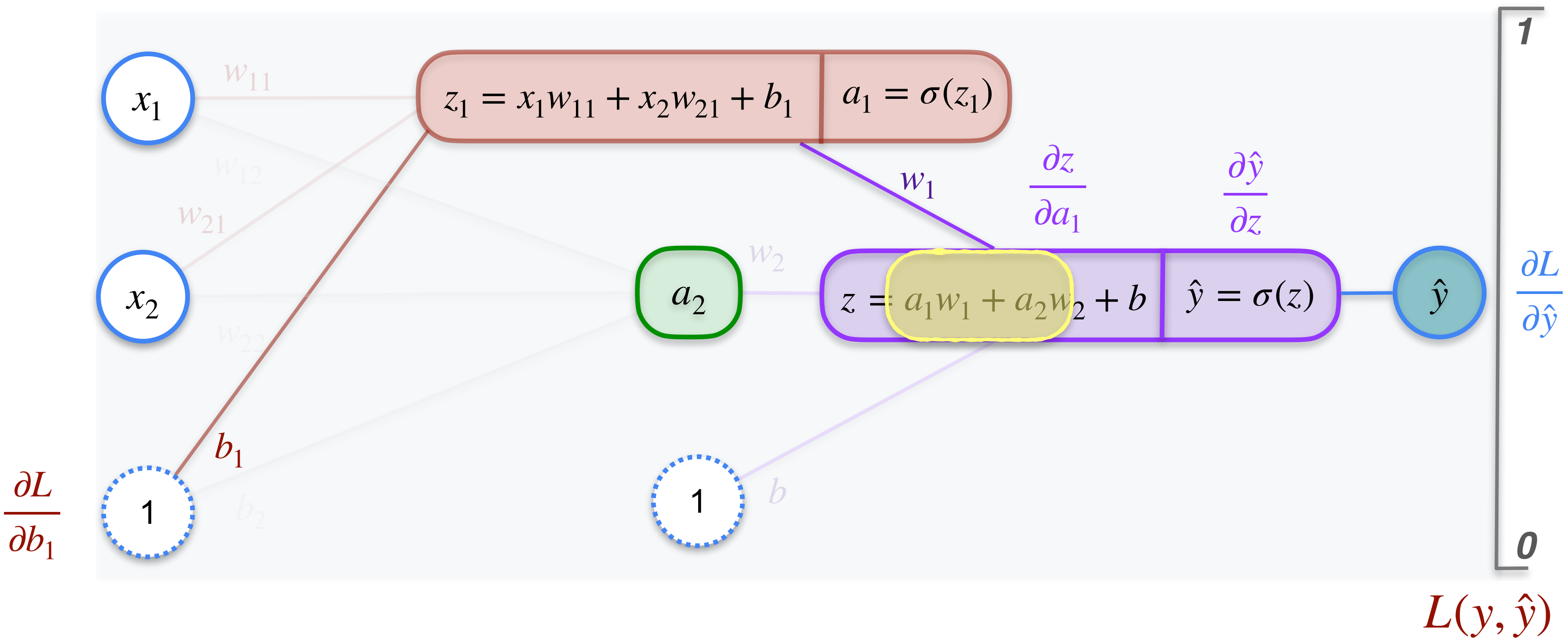
2,2,1 Neural Network



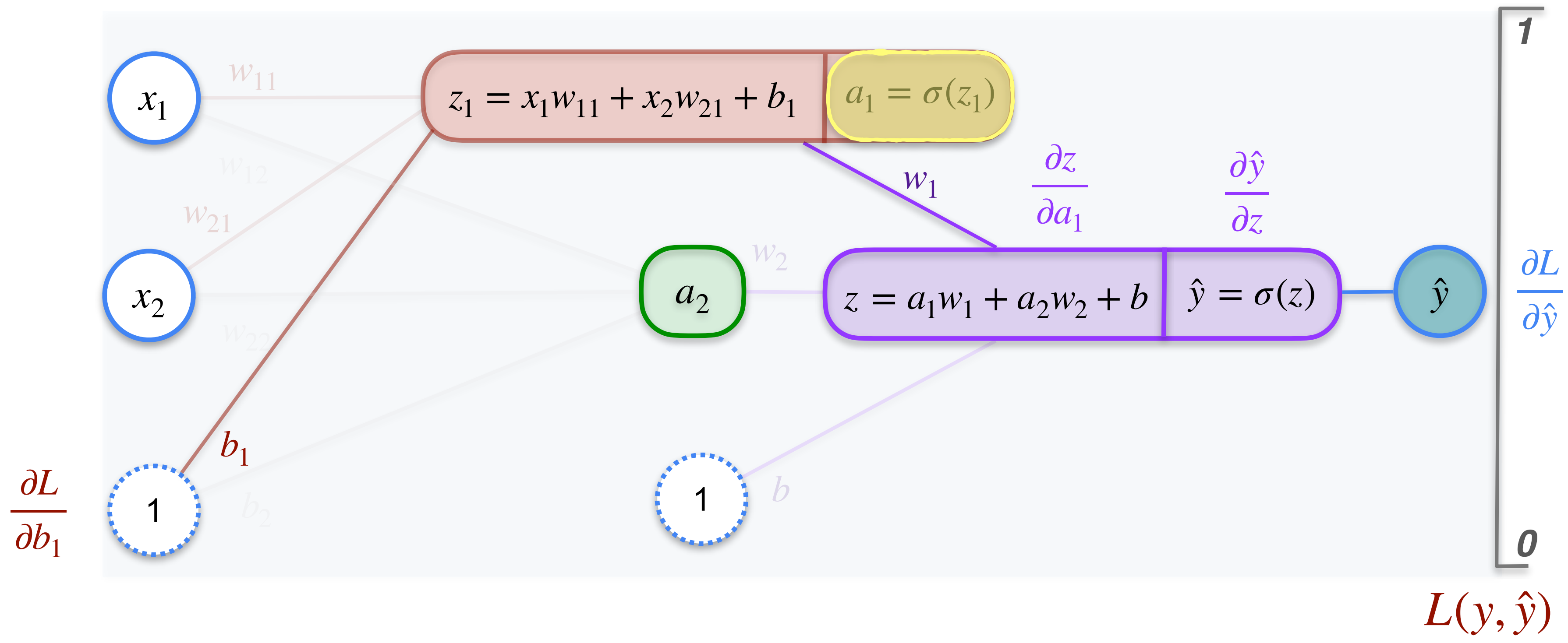
2,2,1 Neural Network



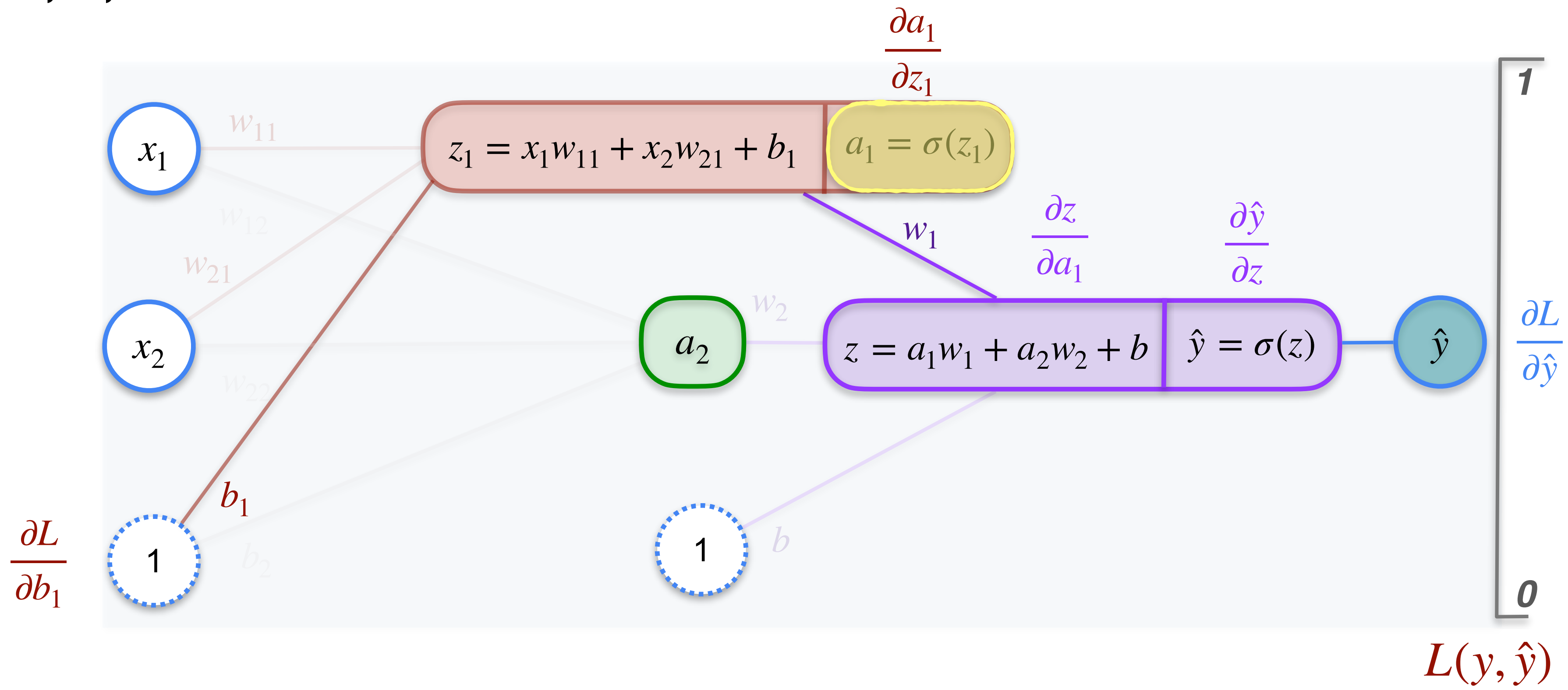
2,2,1 Neural Network



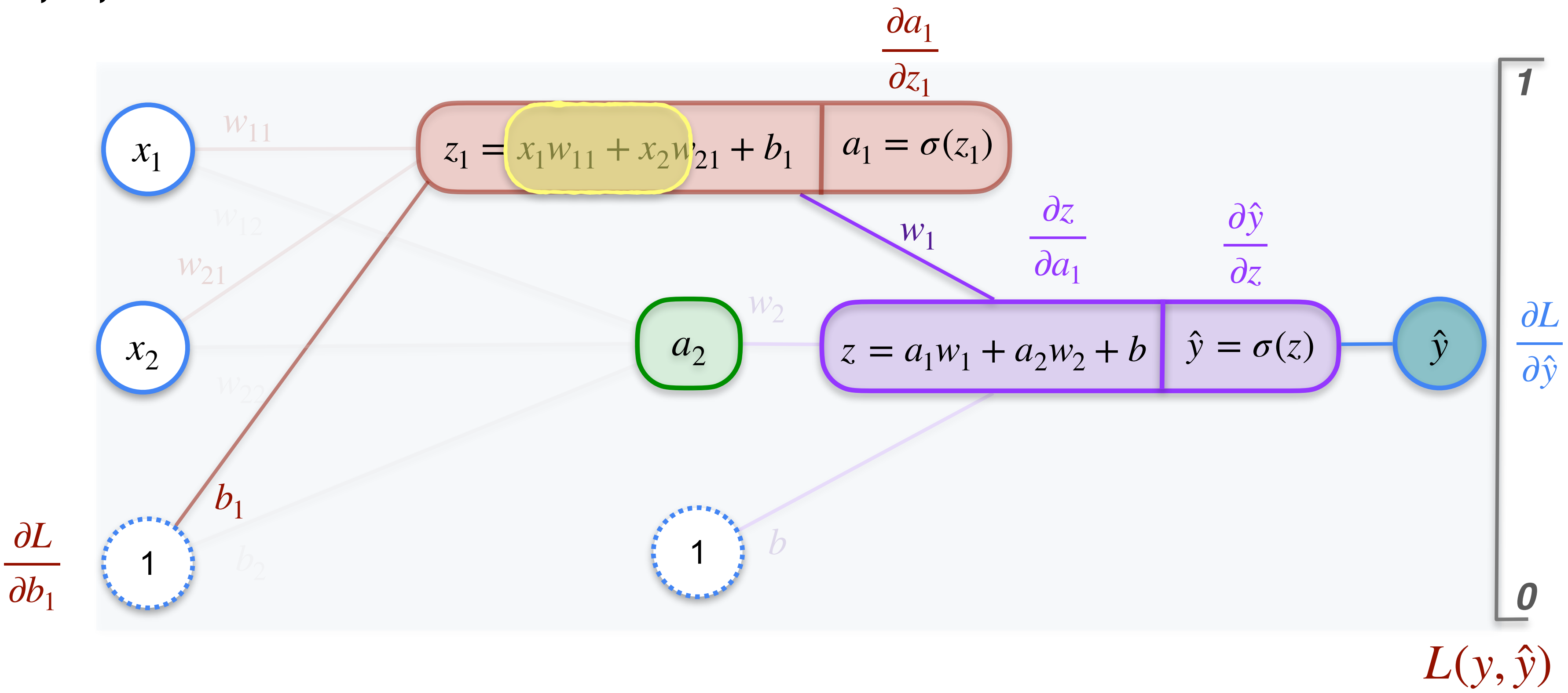
2,2,1 Neural Network



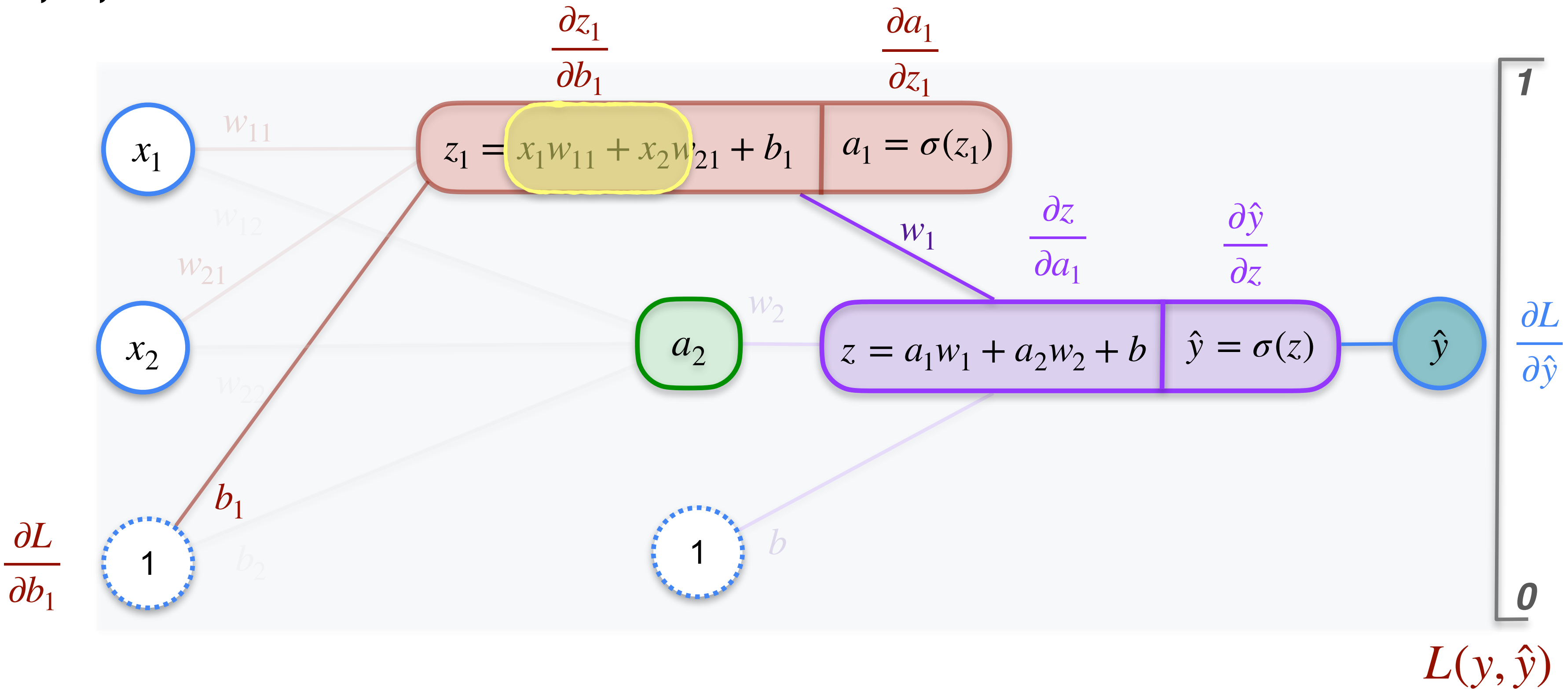
2,2,1 Neural Network



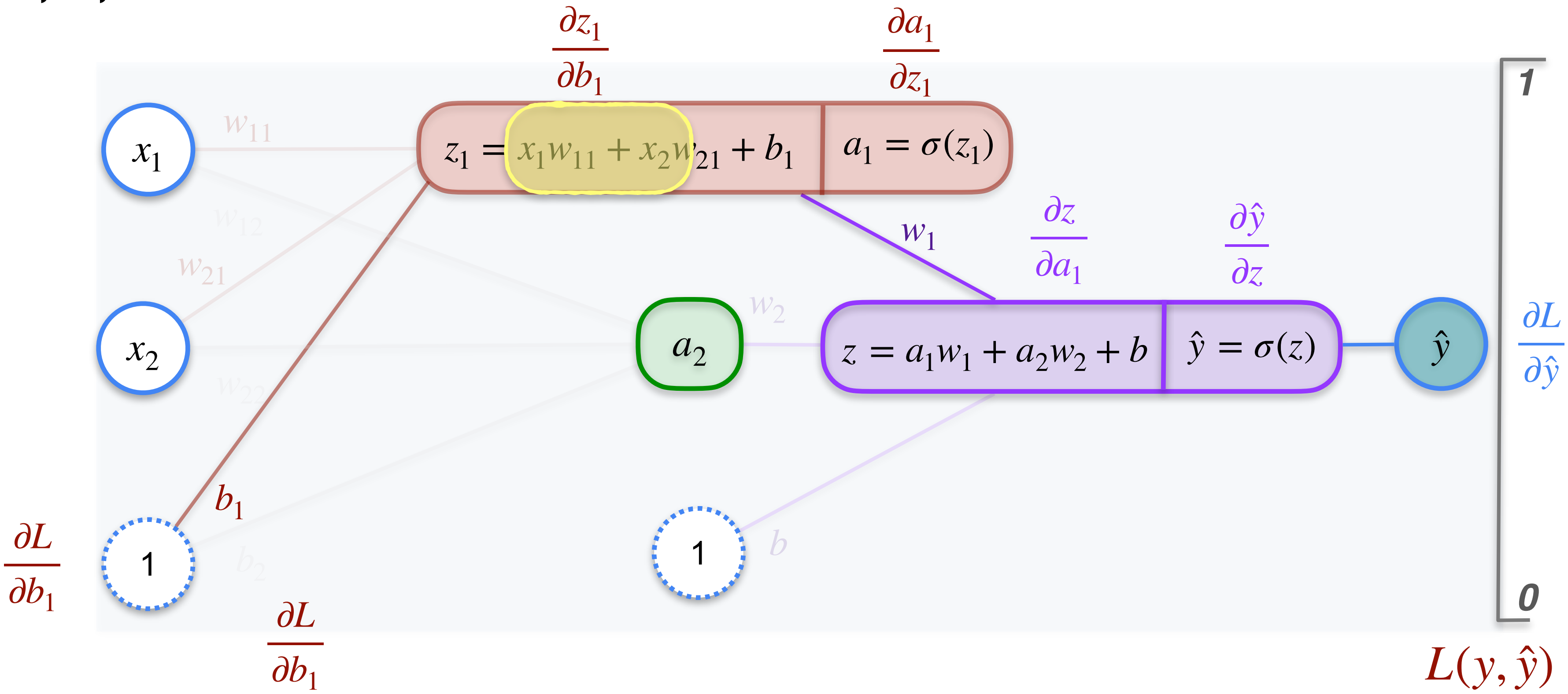
2,2,1 Neural Network



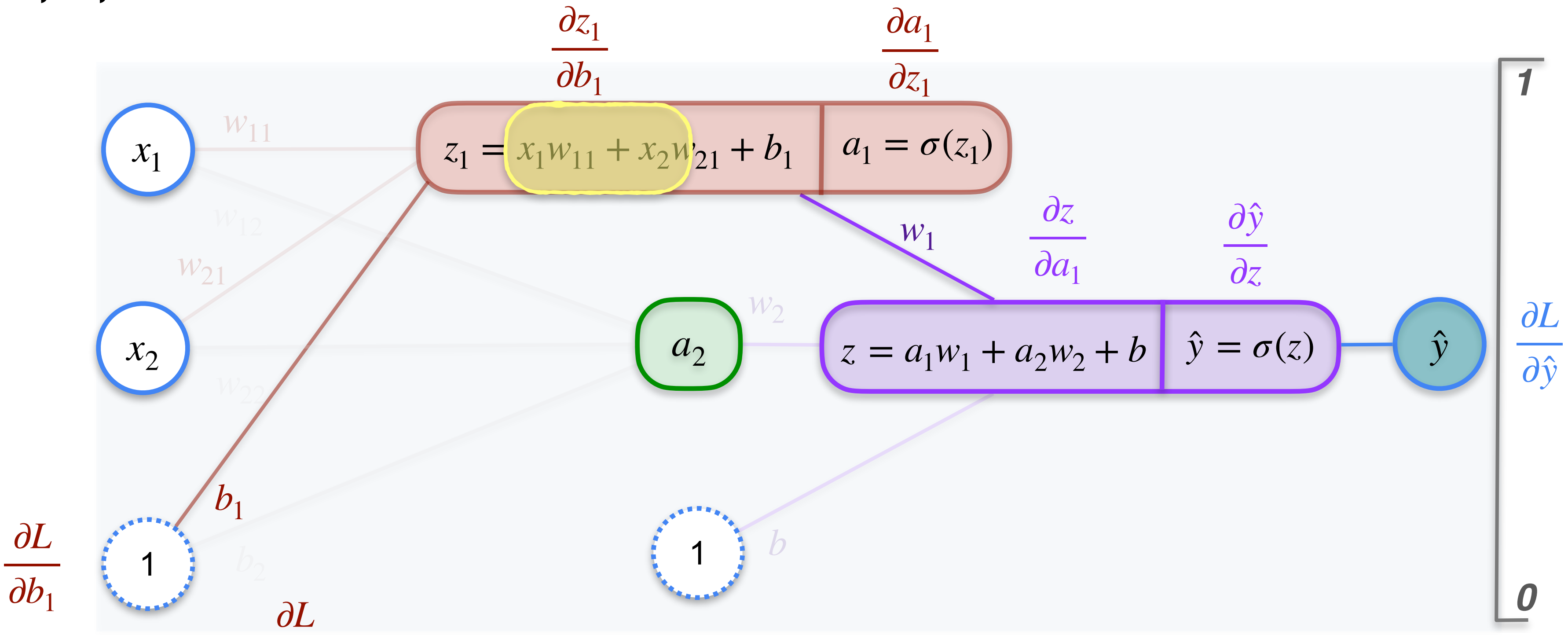
2,2,1 Neural Network



2,2,1 Neural Network

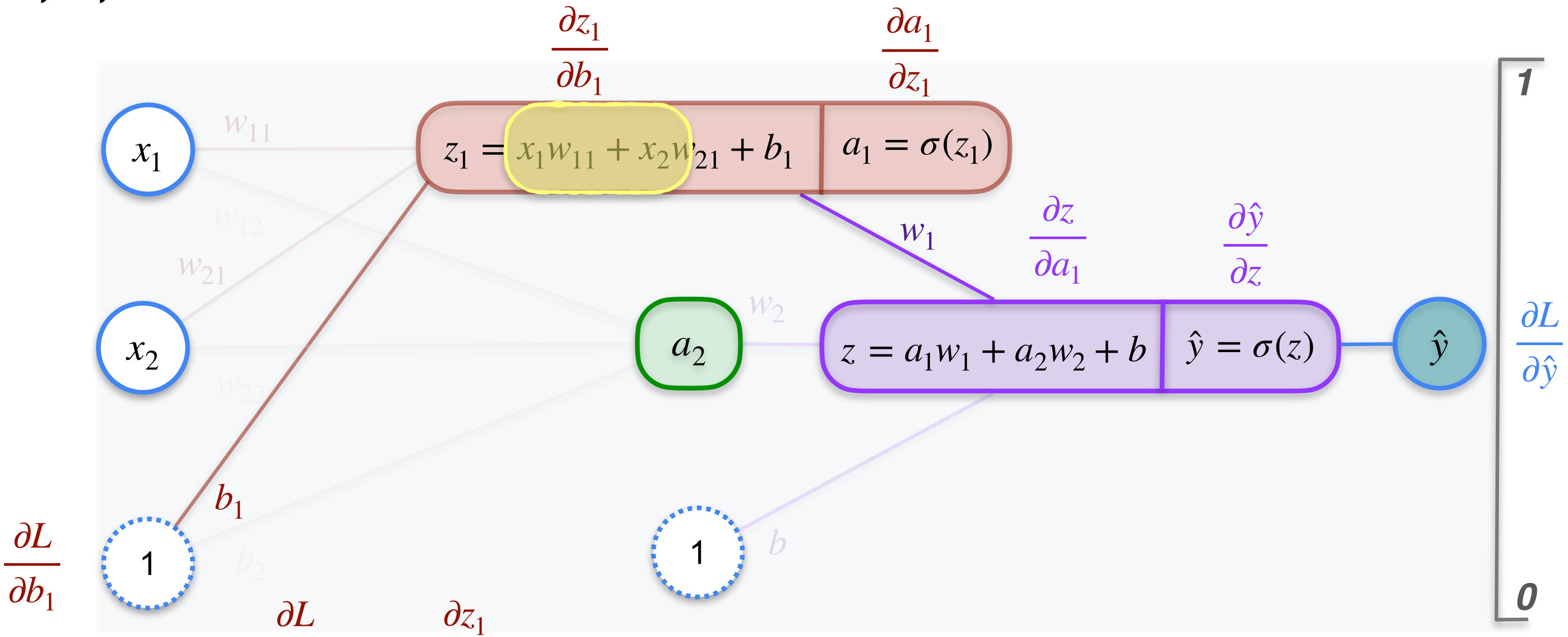


2,2,1 Neural Network



$$\frac{\partial L}{\partial b_1} =$$

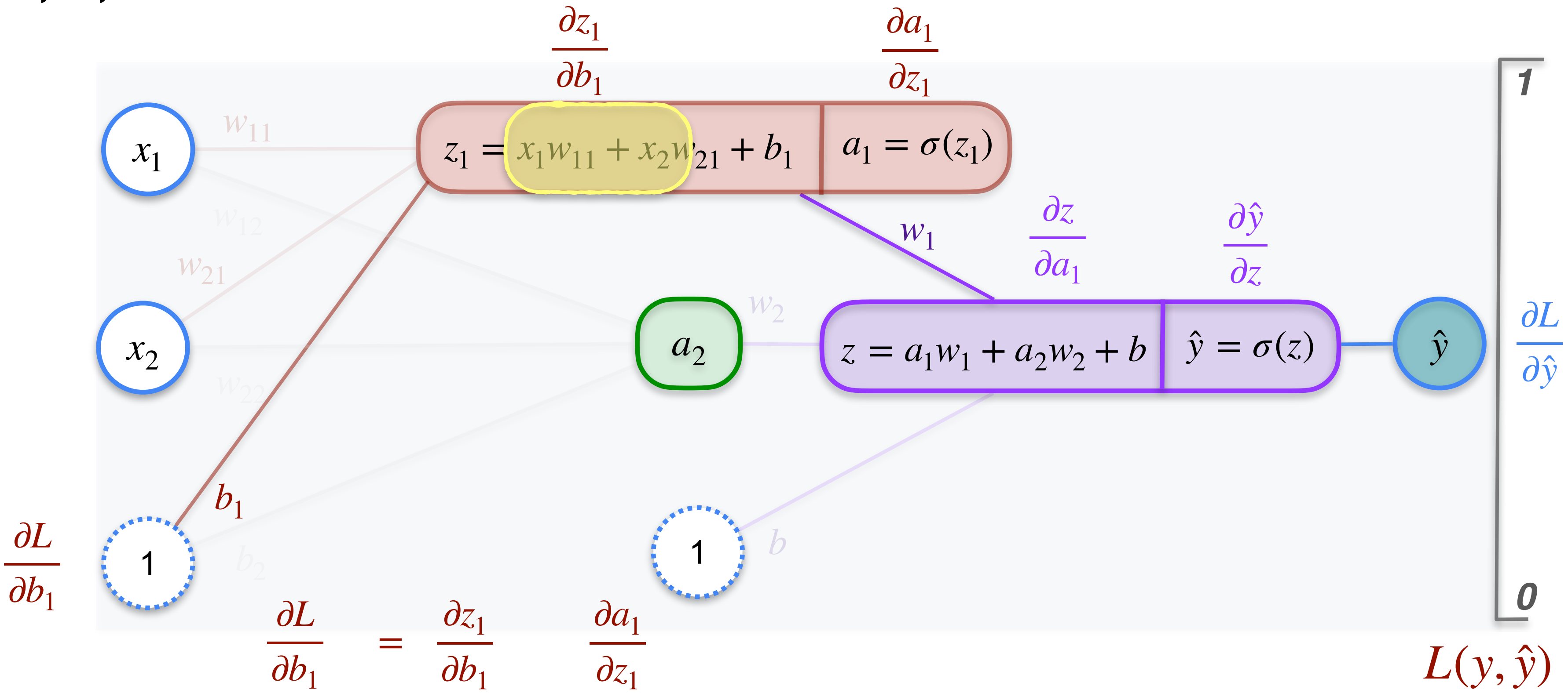
2,2,1 Neural Network



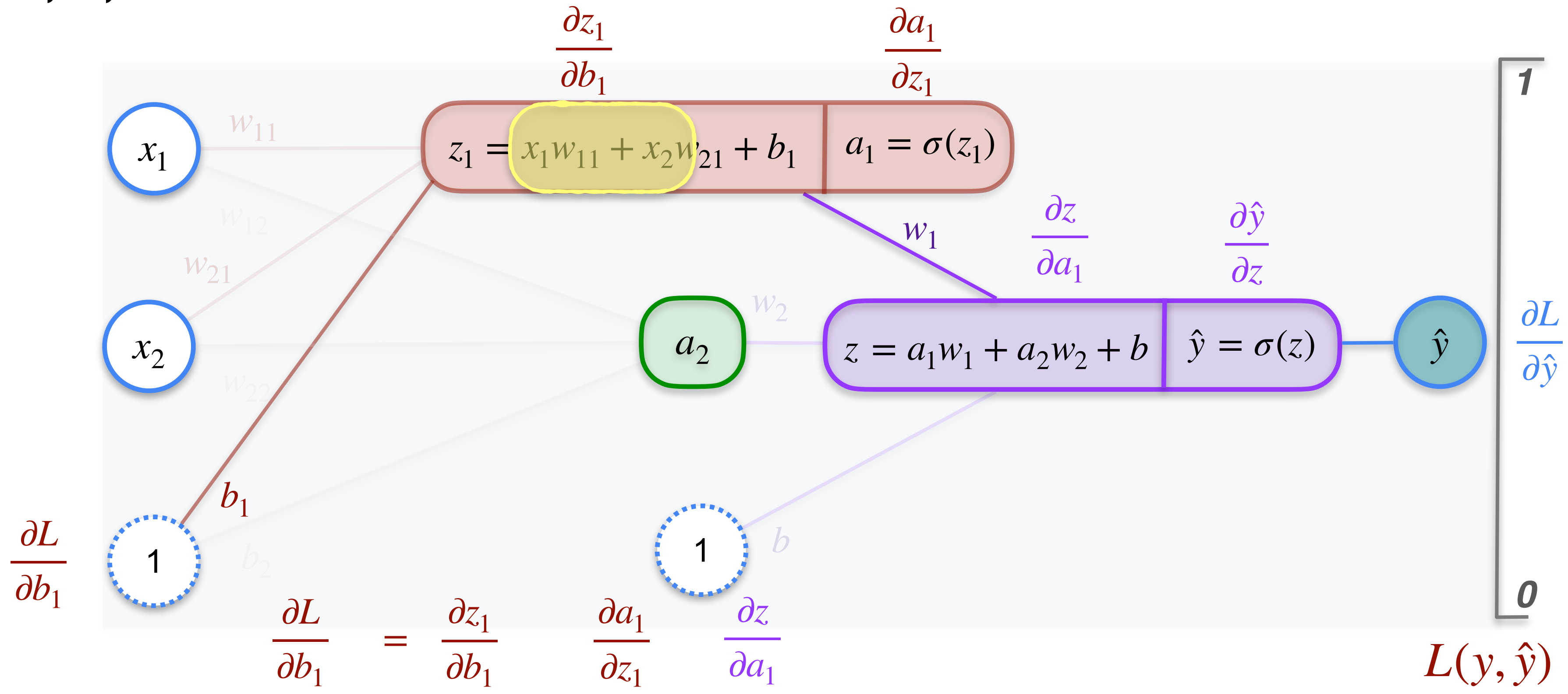
$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1}$$

$L(y, \hat{y})$

2,2,1 Neural Network

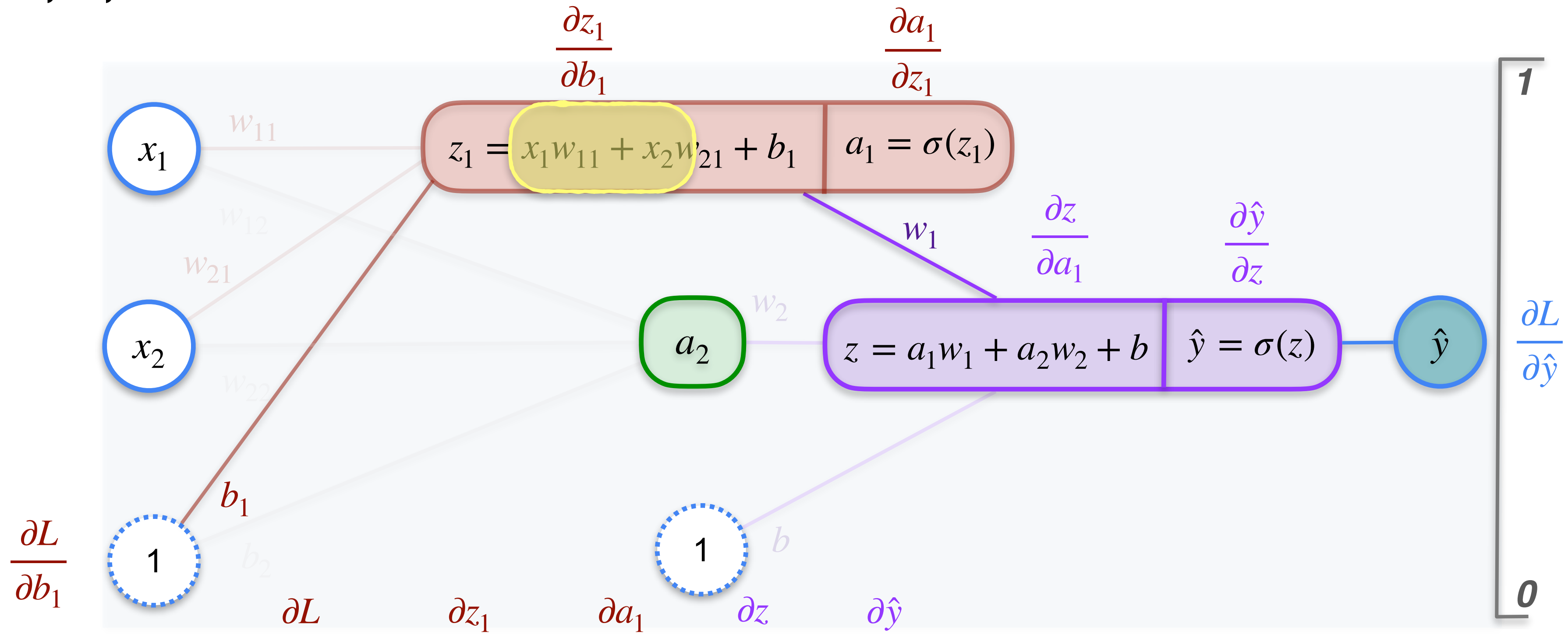


2,2,1 Neural Network



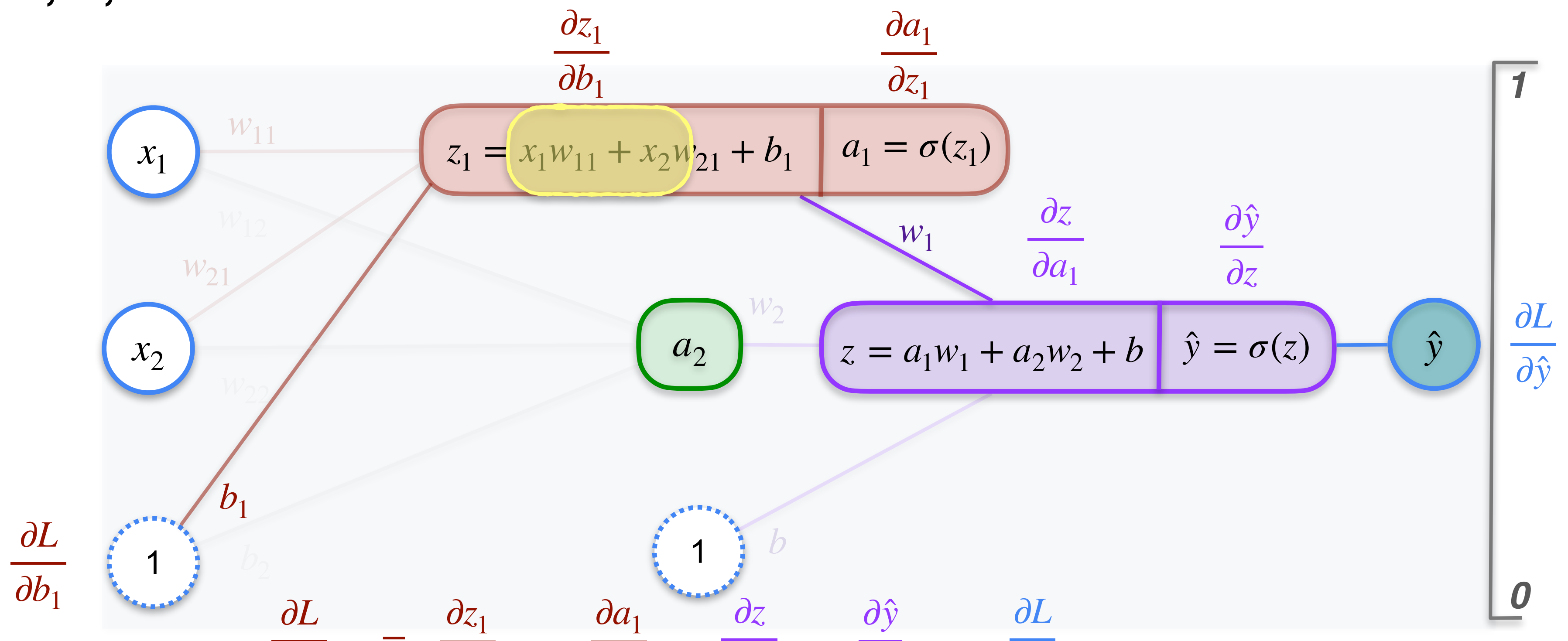
$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1}$$

2,2,1 Neural Network



$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1} \quad \frac{\partial \hat{y}}{\partial z}$$

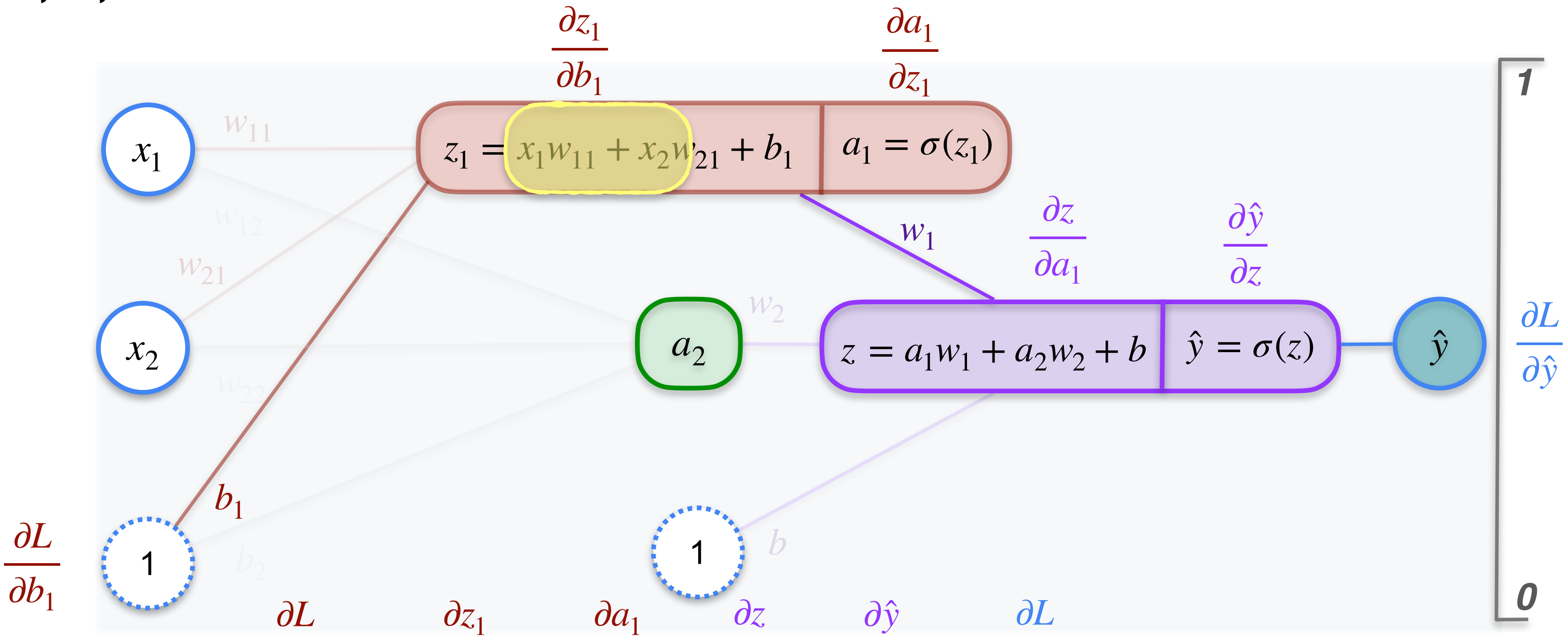
2,2,1 Neural Network



$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \quad \frac{\partial a_1}{\partial z_1} \quad \frac{\partial z}{\partial a_1} \quad \frac{\partial \hat{y}}{\partial z} \quad \frac{\partial L}{\partial \hat{y}}$$

$L(y, \hat{y})$

2,2,1 Neural Network



$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$L(y, \hat{y})$

2,2,1 Neural Network

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1}$$

DeepLearning.AI

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} =$$

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} =$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} =$$

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1$$

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \boxed{\frac{\partial a_1}{\partial z_1}} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1-a_1)$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1-a_1)$$

DeepLearning.AI

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1-a_1)$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1 - a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1-a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1-a_1) \quad w_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1-a_1) \quad w_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1 - a_1) \quad w_1 \quad \hat{y}(1 - \hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1-a_1) \quad w_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b_1} = 1 \quad a_1(1 - a_1) \quad w_1 \quad \hat{y}(1 - \hat{y}) \quad \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1 - a_1) \cdot w_1 \cdot \hat{y}(1 - \hat{y}) \cdot \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1w_{11} + x_2w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1w_{11} + x_2w_{21} + b_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -w_1a_1(1-a_1)(y - \hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -w_1 a_1 (1-a_1)(y - \hat{y})$$

Perform gradient descent with

to find optimal value of b_1 that gives the least error

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$a_1 = \sigma(z_1)$$

$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -w_1 a_1(1-a_1)(y - \hat{y})$$

Perform gradient descent with

$$b_1 \rightarrow b_1 - \alpha \frac{\partial L}{\partial b_1}$$

to find optimal value of b_1 that gives the least error

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1 - a_1) \cdot w_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

$$= -w_1 a_1 (1 - a_1)(y - \hat{y})$$

Perform gradient descent with

$$b_1 \rightarrow b_1 - \alpha$$

to find optimal value of b_1 that gives the least error

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

$$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b_1} = 1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

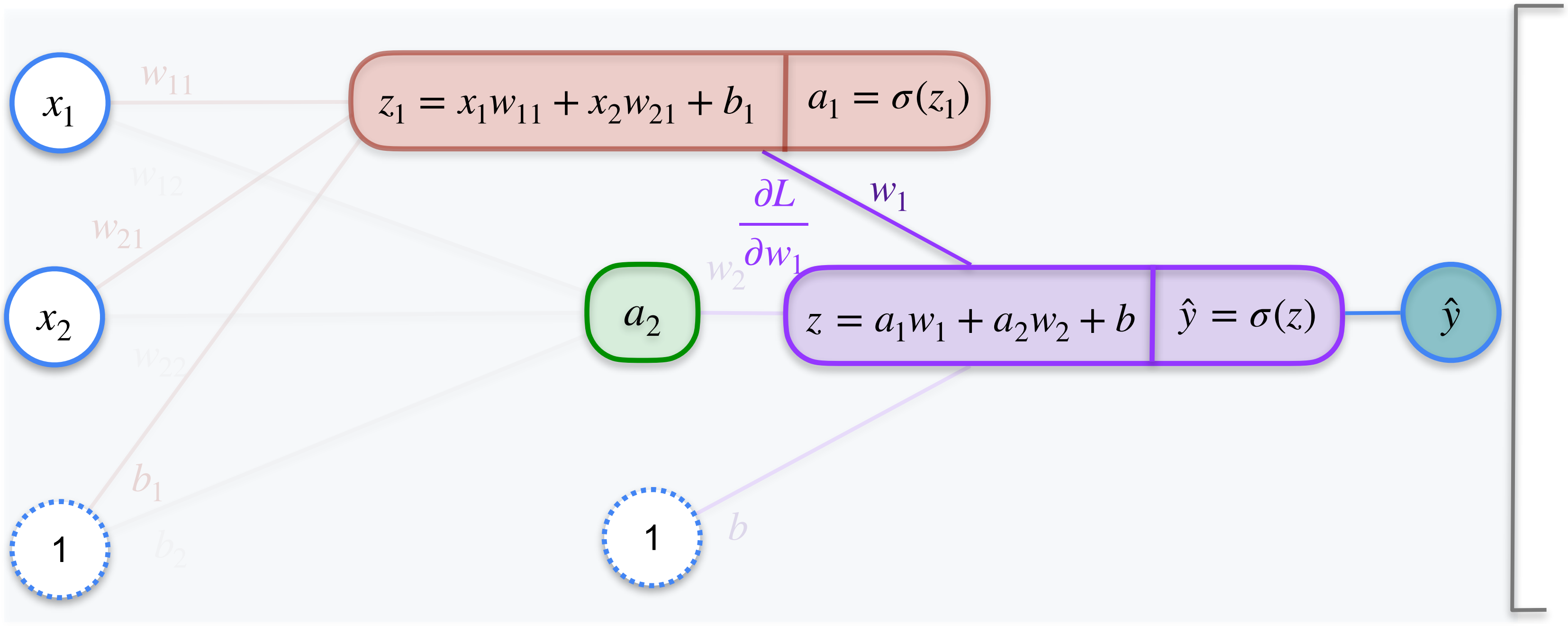
$$= -w_1 a_1 (1-a_1)(y - \hat{y})$$

Perform gradient descent with

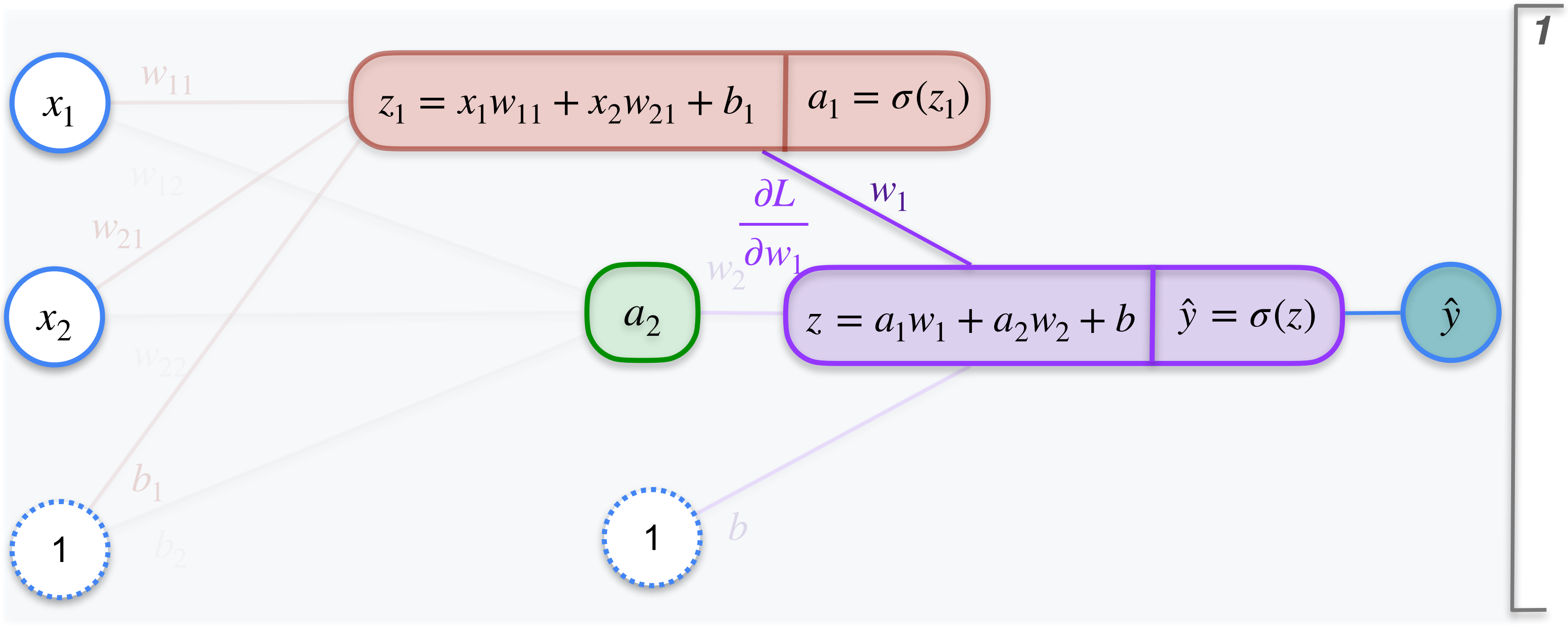
$$b_1 \rightarrow b_1 - \alpha(-w_1 a_1 (1-a_1)(y - \hat{y}))$$

to find optimal value of b_1 that gives the least error

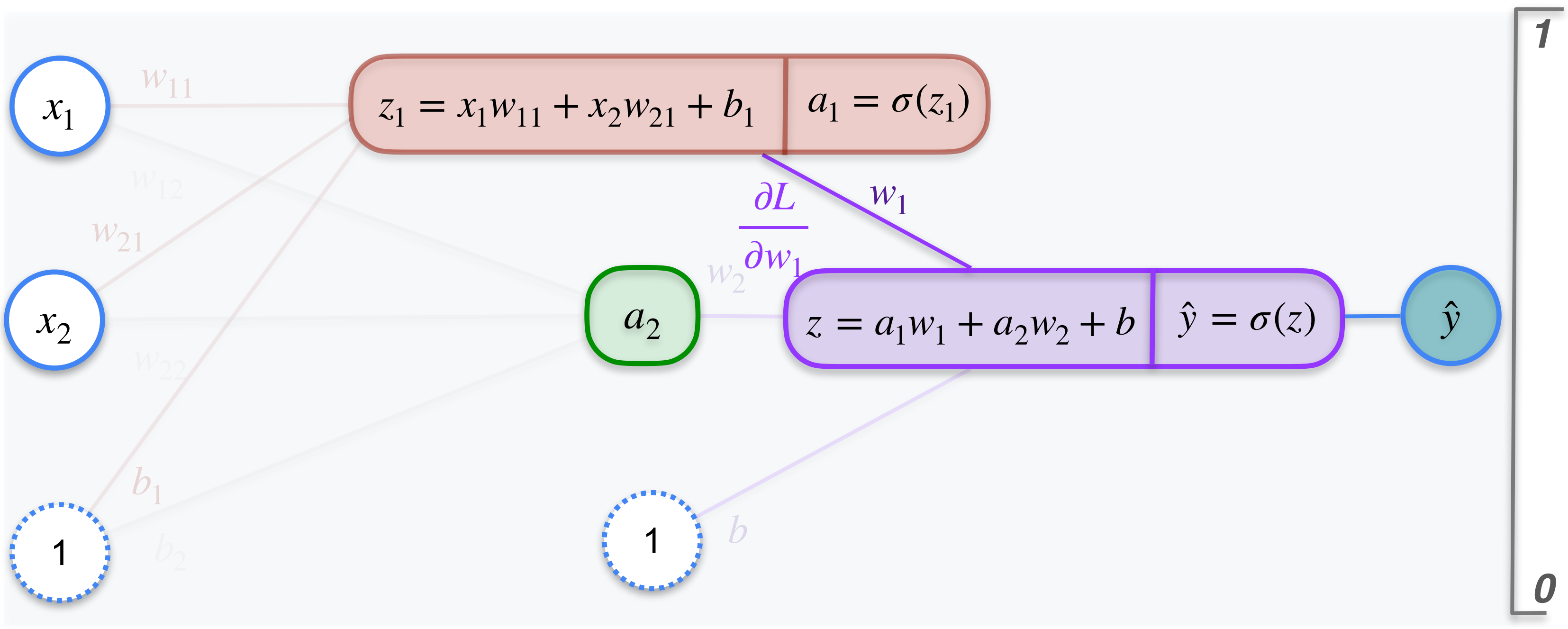
2,2,1 Neural Network



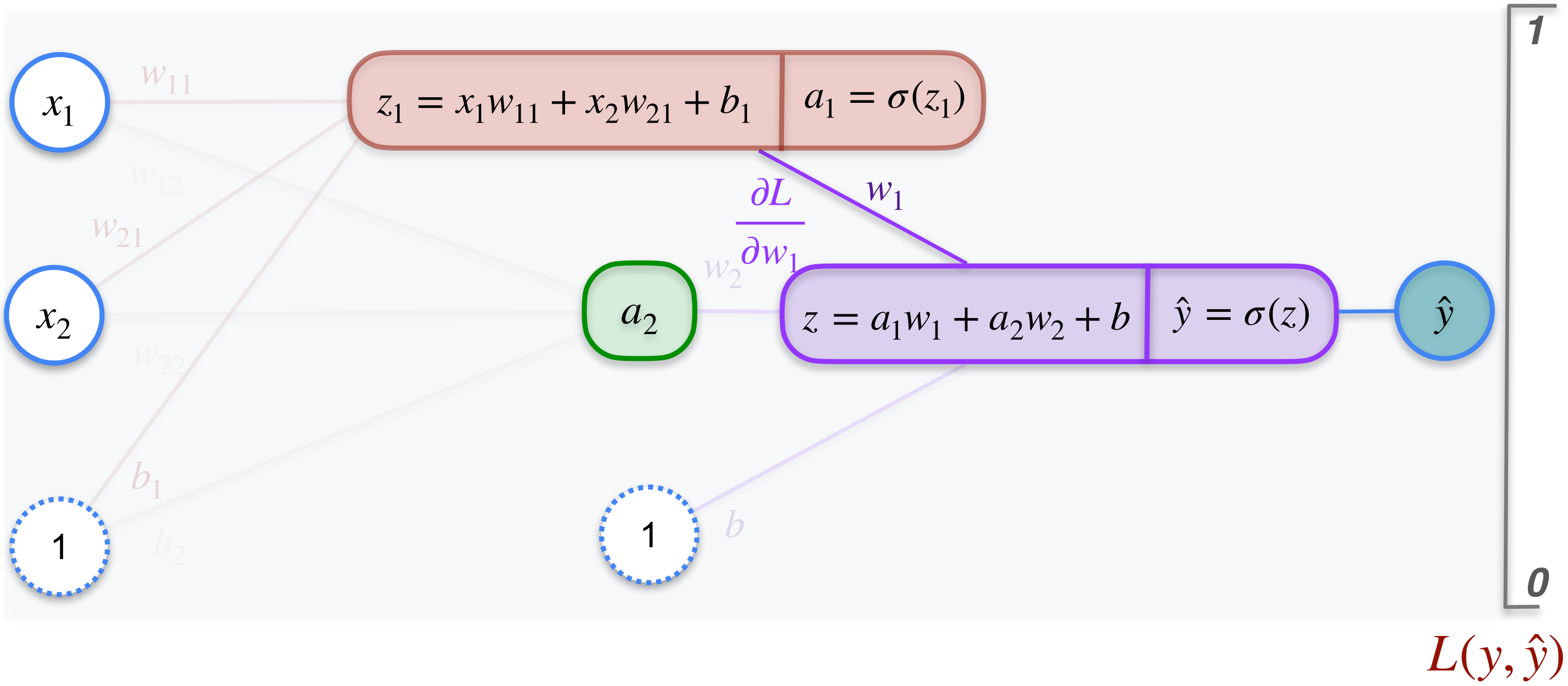
2,2,1 Neural Network



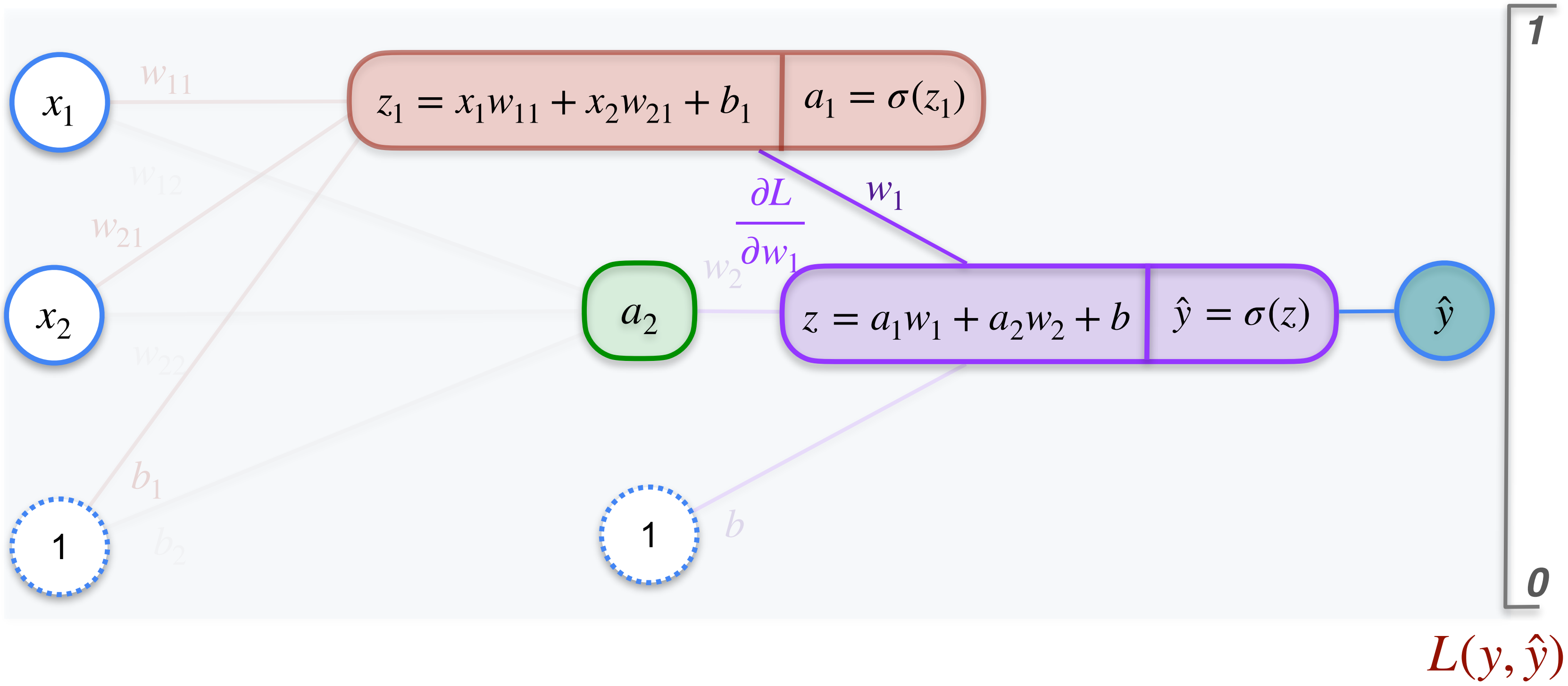
2,2,1 Neural Network



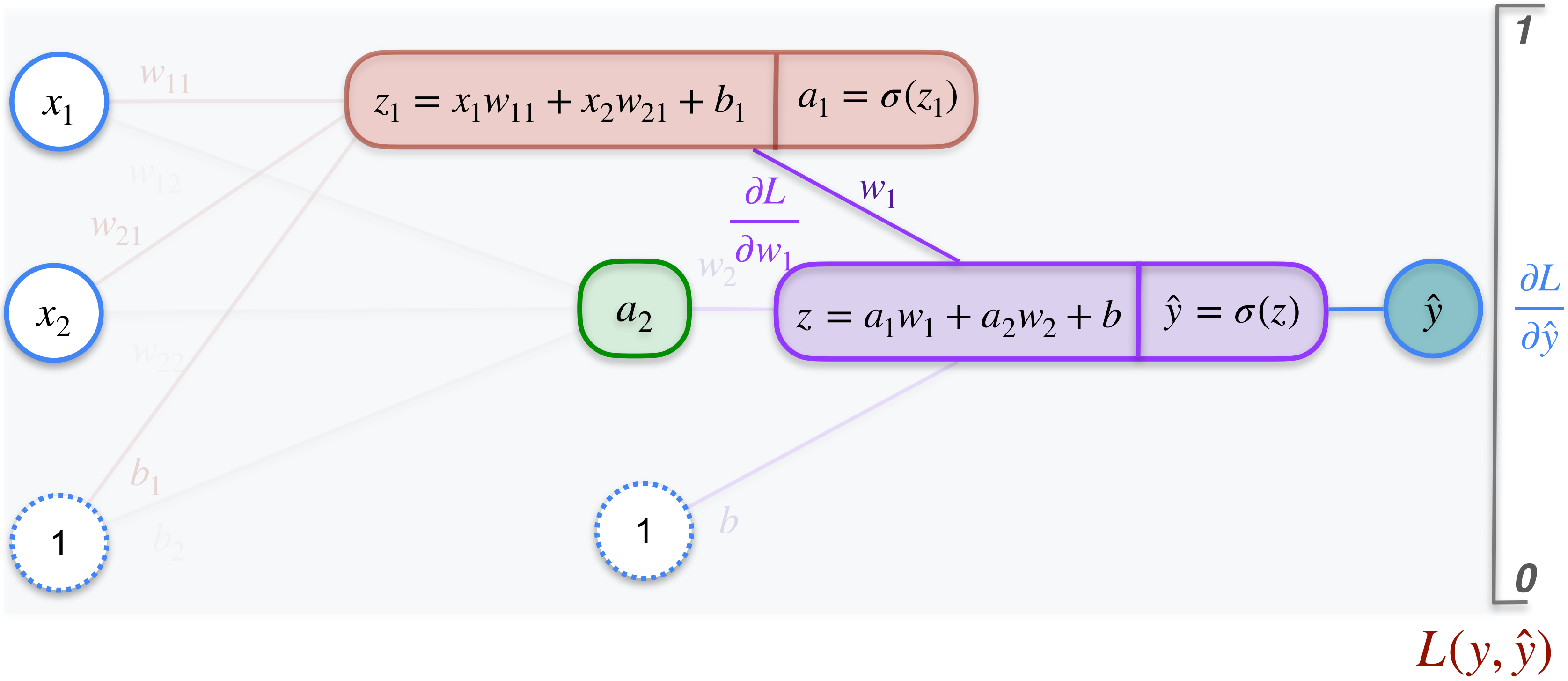
2,2,1 Neural Network



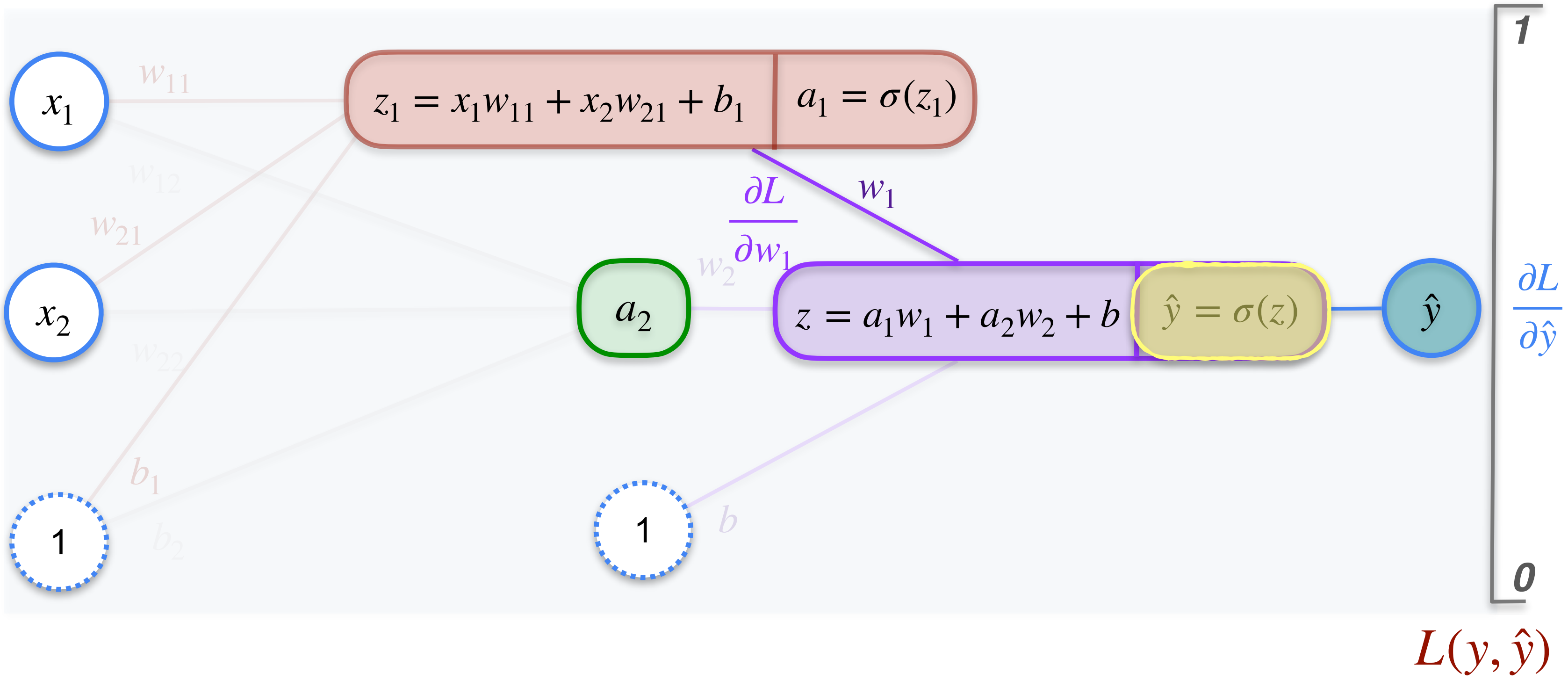
2,2,1 Neural Network



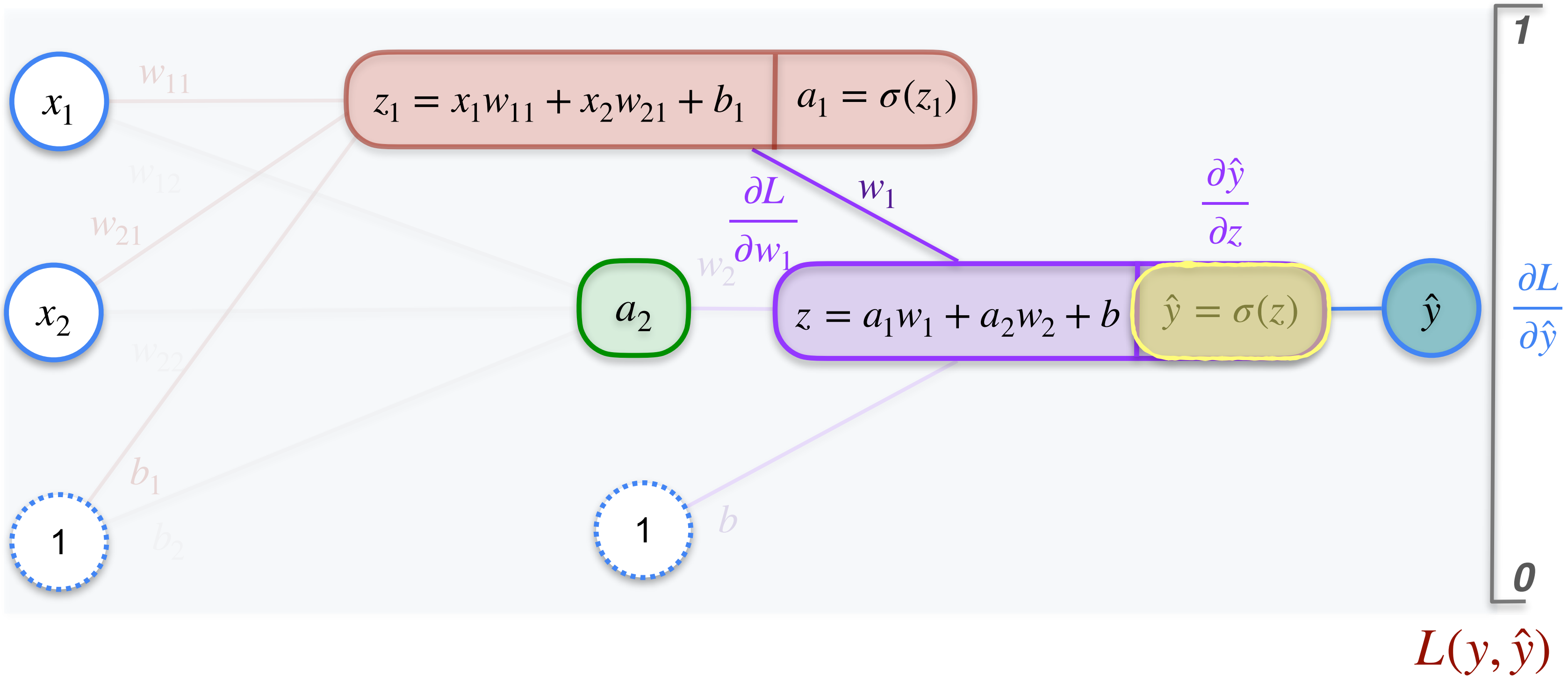
2,2,1 Neural Network



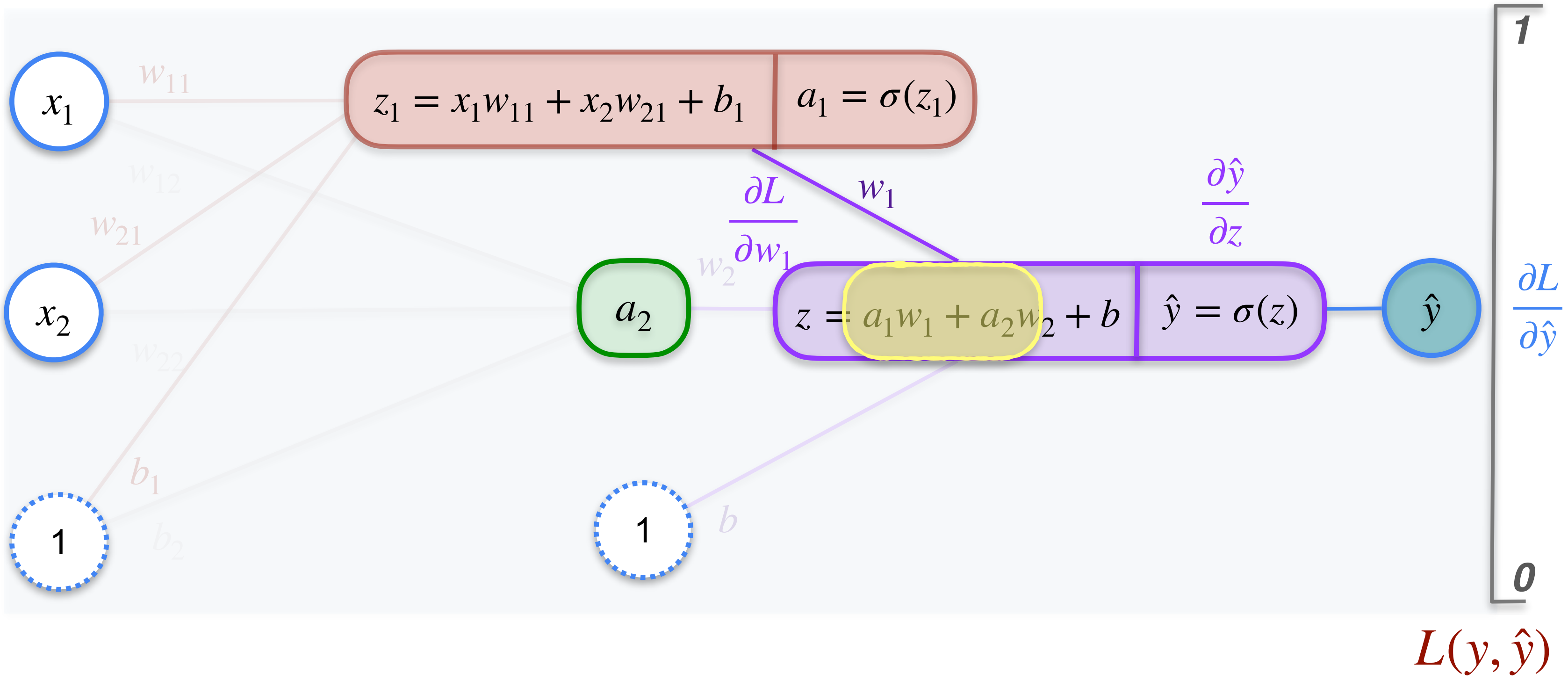
2,2,1 Neural Network



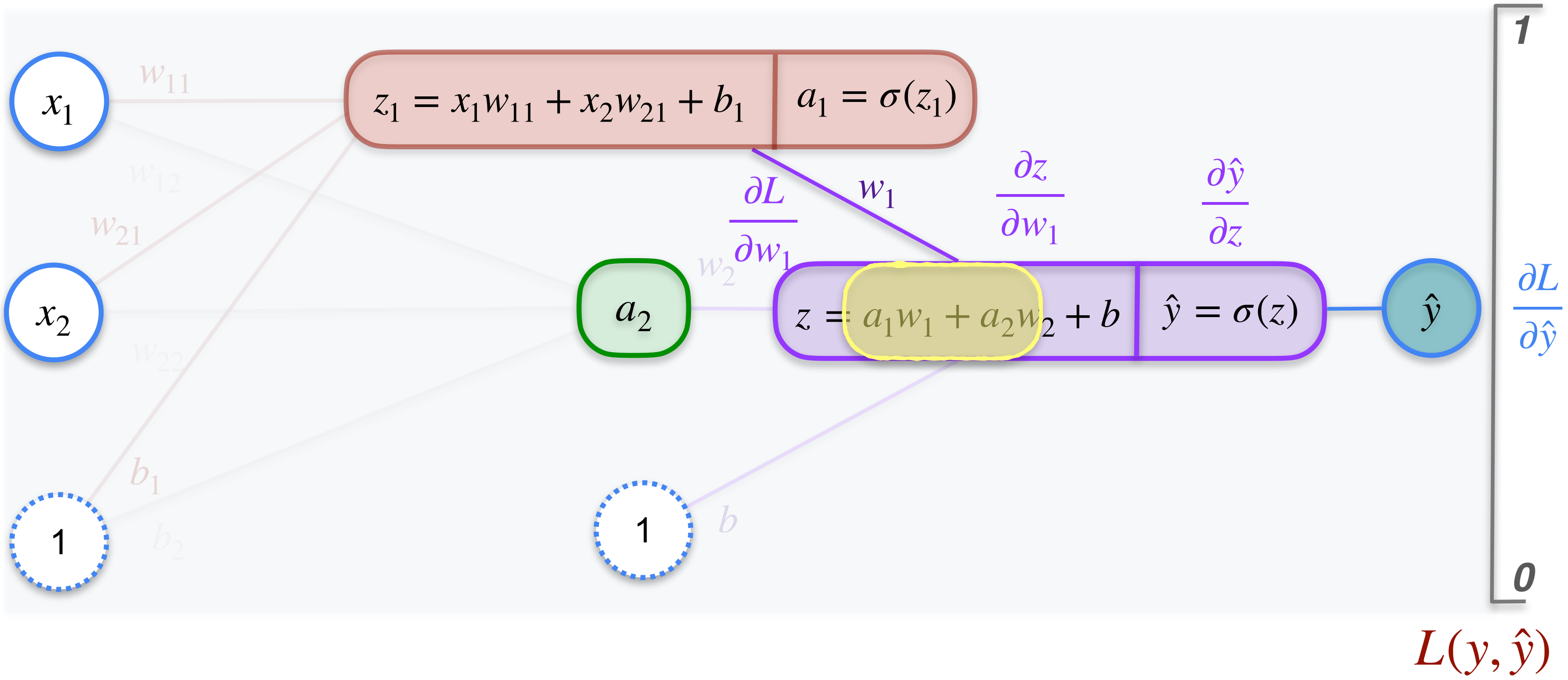
2,2,1 Neural Network



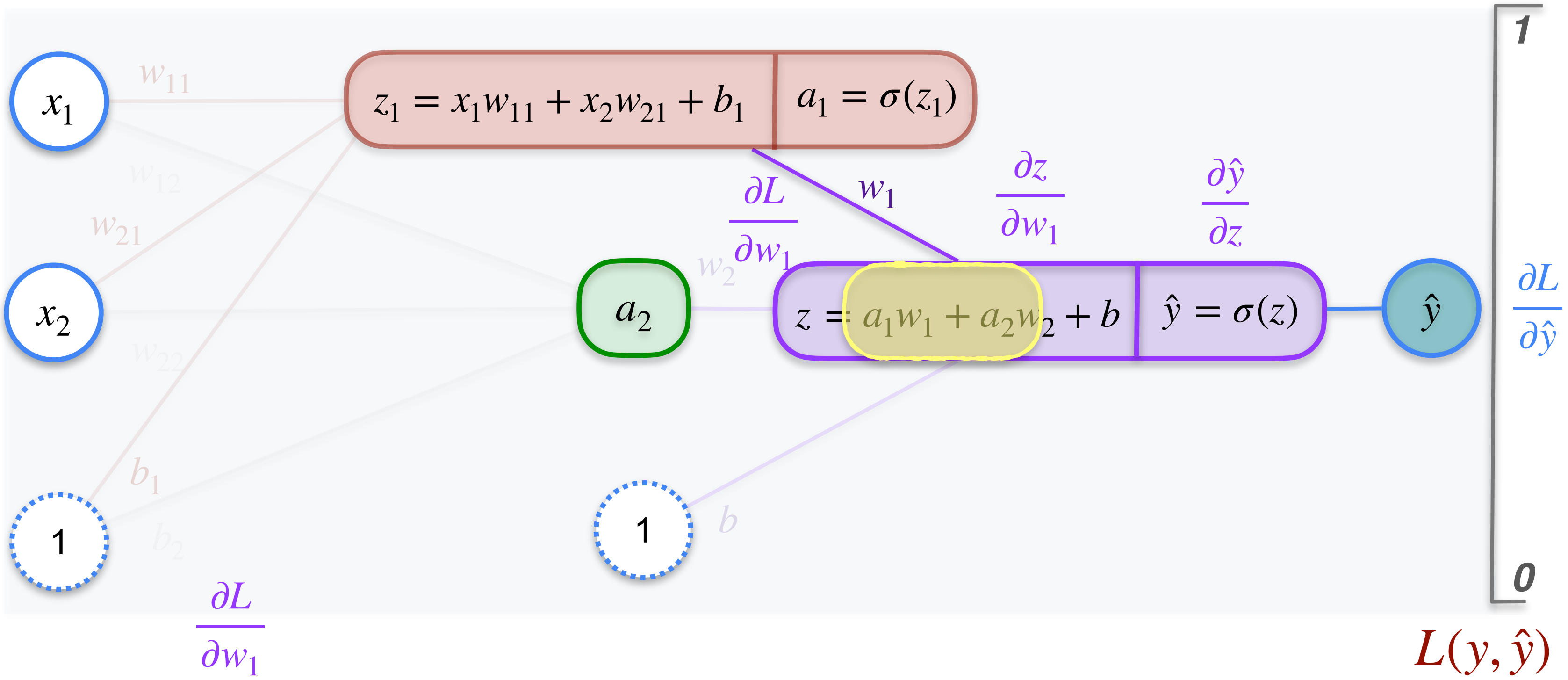
2,2,1 Neural Network



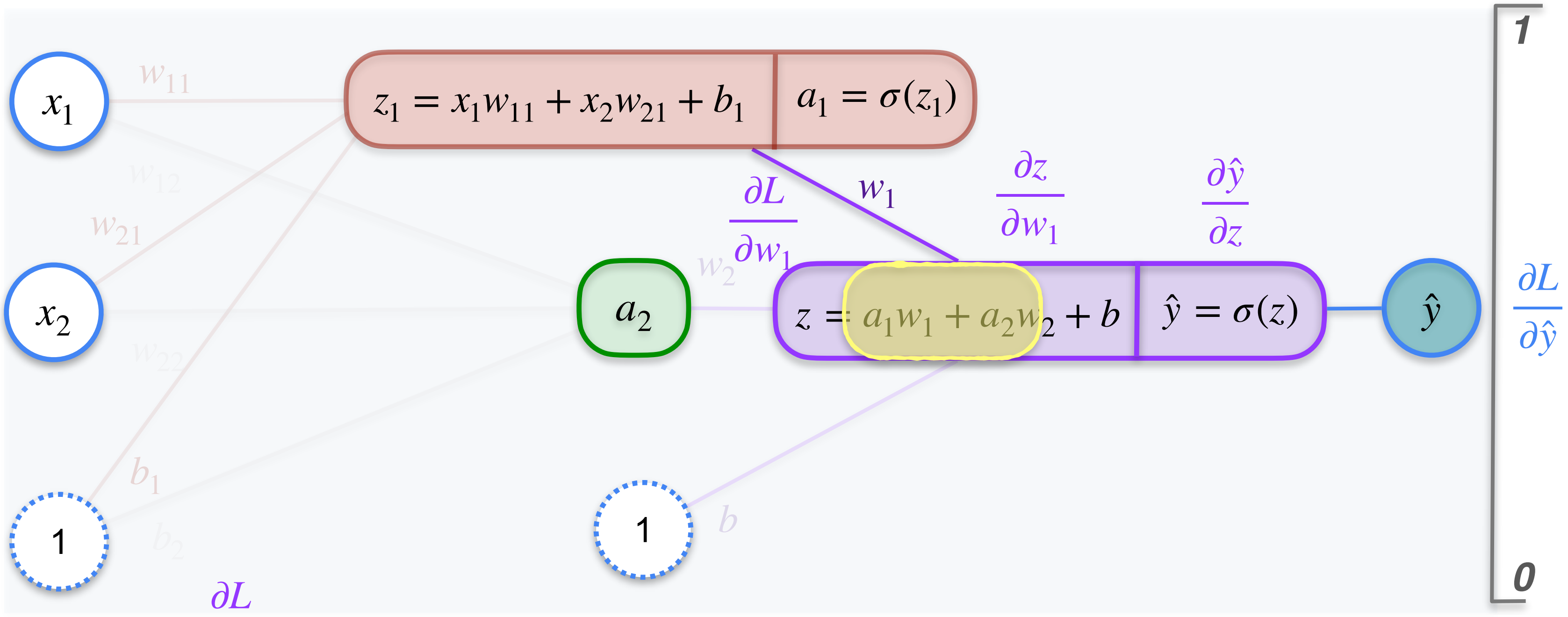
2,2,1 Neural Network



2,2,1 Neural Network



2,2,1 Neural Network

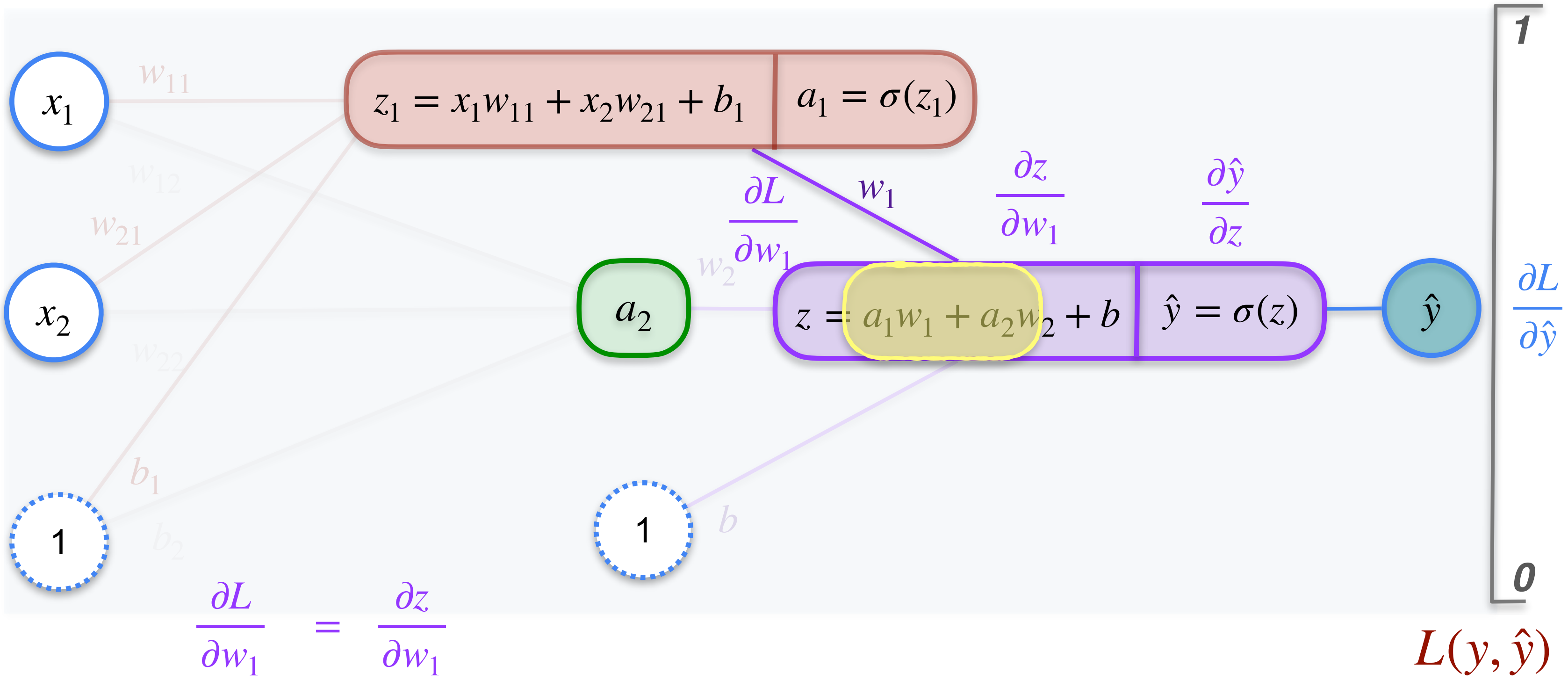


$$\frac{\partial L}{\partial w_1} =$$

$L(y, \hat{y})$

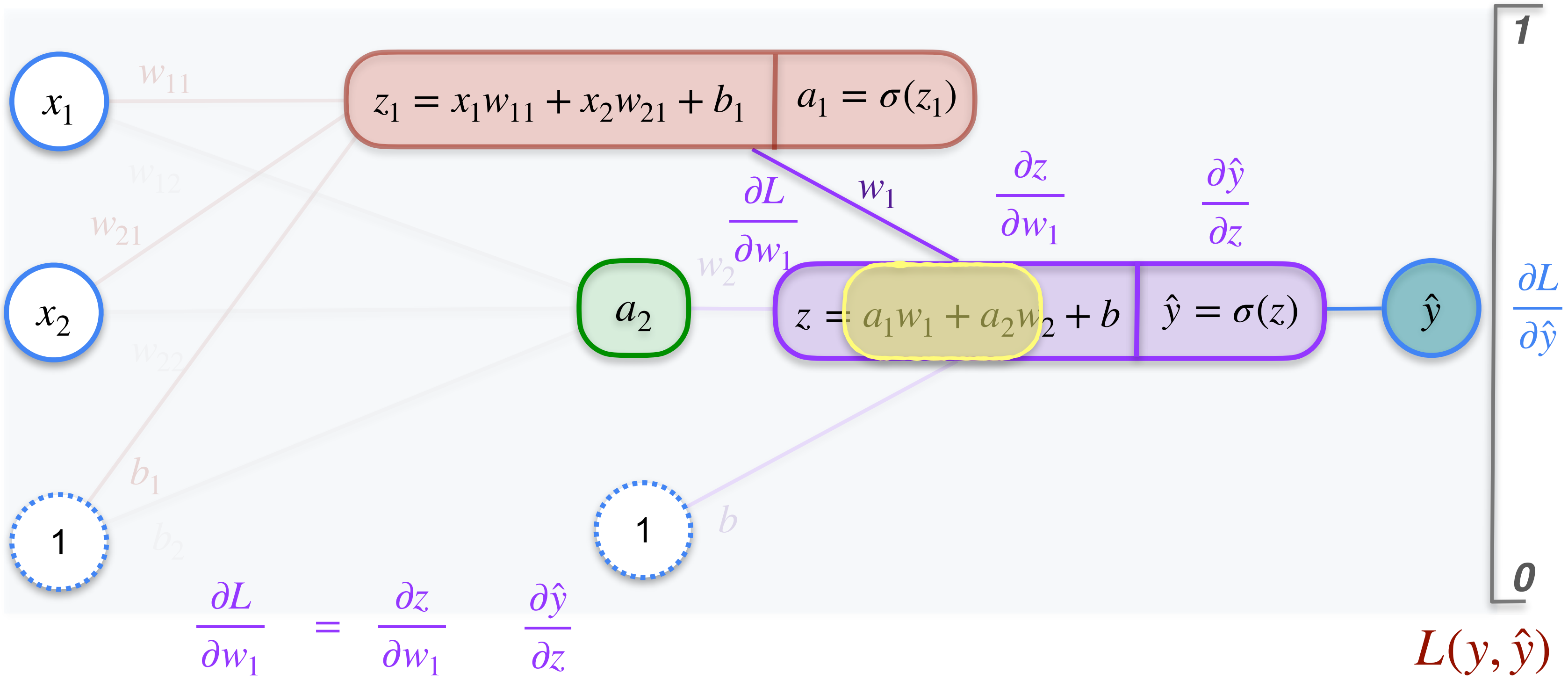
DeepLearning.AI

2,2,1 Neural Network



$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1}$$

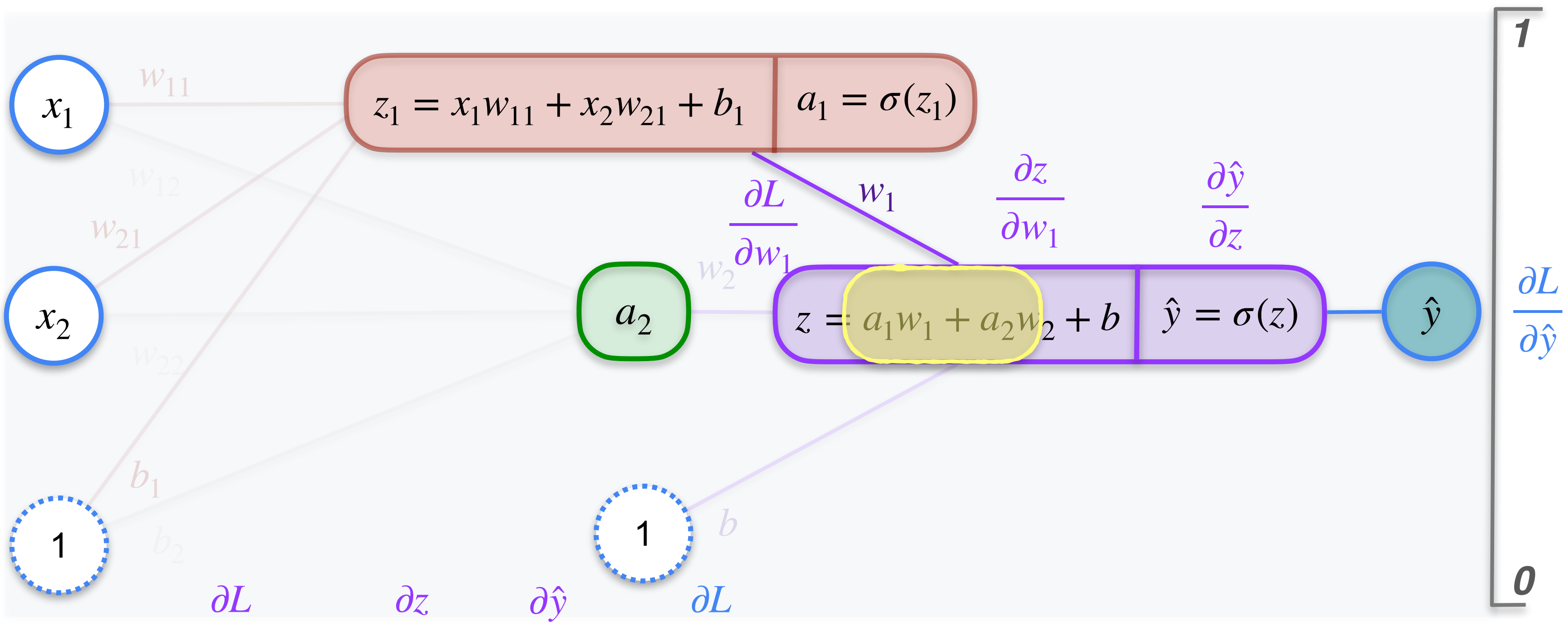
2,2,1 Neural Network



$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \quad \frac{\partial \hat{y}}{\partial z}$$

DeepLearning.AI

2,2,1 Neural Network



$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

2,2,1 Neural Network

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

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

$z = a_1 w_1 + a_2 w_2 + b$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

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

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

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

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

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

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

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

2,2,1 Neural Network

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

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

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \boxed{\frac{\partial \hat{y}}{\partial z}} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1$$

2,2,1 Neural Network

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

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

$z = a_1w_1 + a_2w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

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

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

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

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

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

$$z = a_1w_1 + a_2w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \quad \hat{y}(1-\hat{y})$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1 - y) log(1 - \hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \quad \hat{y}(1 - \hat{y}) \quad \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1 - y) log(1 - \hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \hat{y}(1 - \hat{y}) \cdot \frac{-(y - \hat{y})}{\hat{y}(1 - \hat{y})}$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1-y) log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1-y) log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -a_1(y - \hat{y})$$

2,2,1 Neural Network

$L(y, \hat{y}) = -ylog(\hat{y}) - (1 - y)log(1 - \hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

$$= -a_1(y - \hat{y})$$

to find optimal value of w_1 that gives the least error

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -a_1(y - \hat{y})$$

Perform gradient descent with

to find optimal value of w_1 that gives the least error

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1 - y)\log(1 - \hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

$$= -a_1(y - \hat{y})$$

Perform gradient descent with

$$w_1 \rightarrow w_1 - \alpha \frac{\partial L}{\partial w_1}$$

to find optimal value of w_1 that gives the least error

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -a_1(y - \hat{y})$$

Perform gradient descent with

$$w_1 \rightarrow w_1 - \alpha$$

to find optimal value of w_1 that gives the least error

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1w_1 + a_2w_2 + b$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial z}{\partial w_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial w_1} = a_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -a_1(y - \hat{y})$$

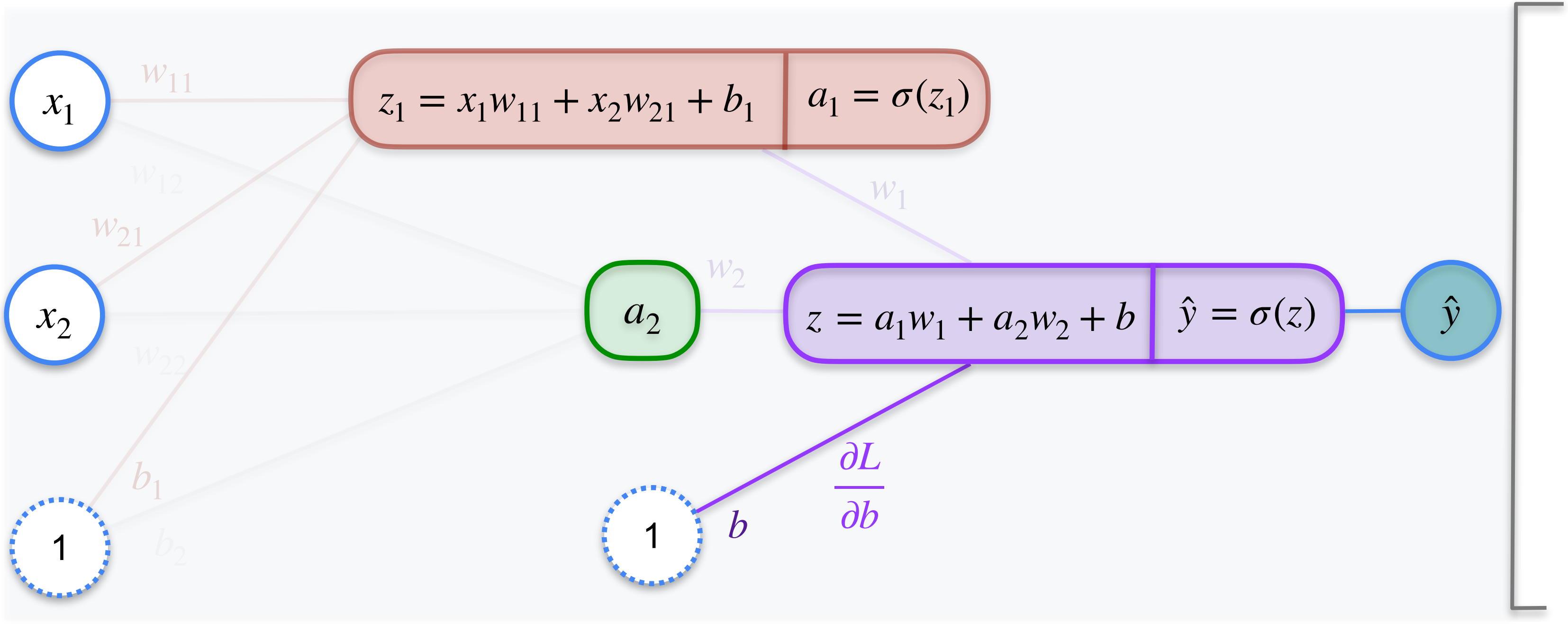
Perform gradient descent with

$$w_1 \rightarrow w_1 - \alpha(-a_1(y - \hat{y}))$$

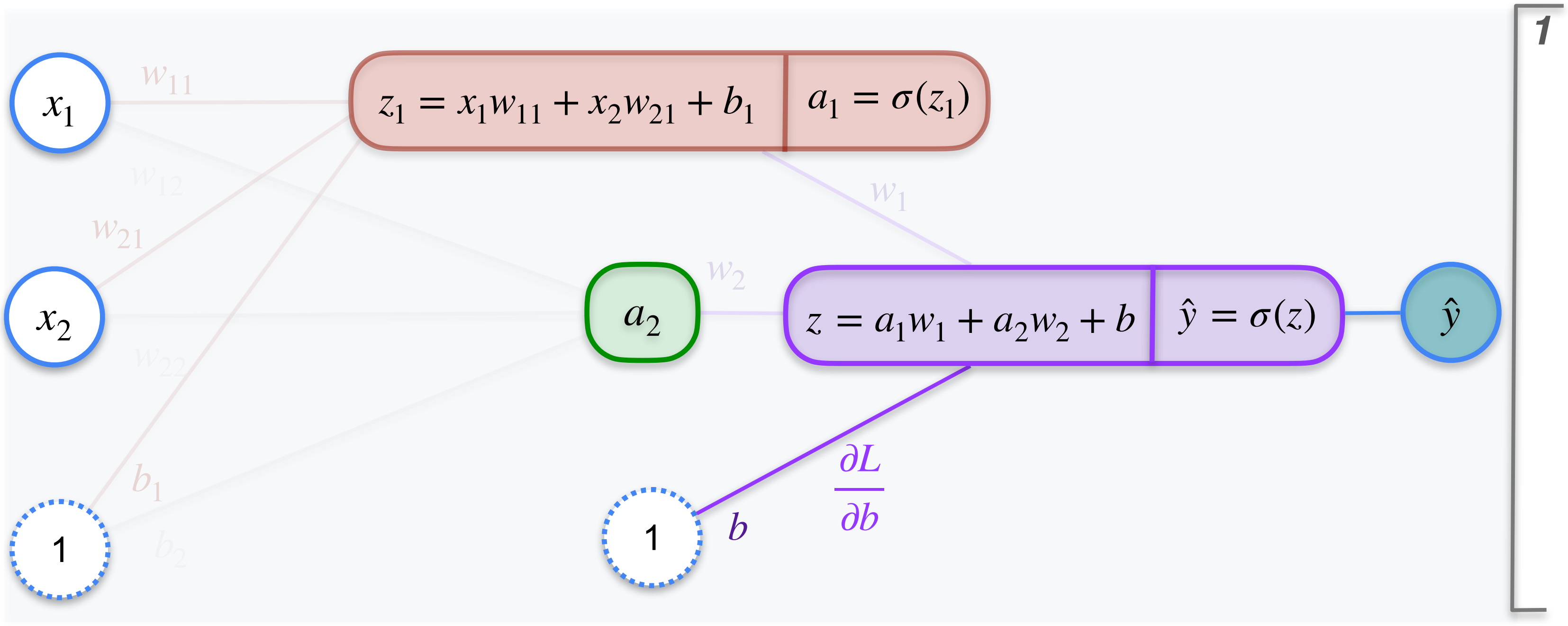
to find optimal value of w_1 that gives the least error

DeepLearning.AI

2,2,1 Neural Network

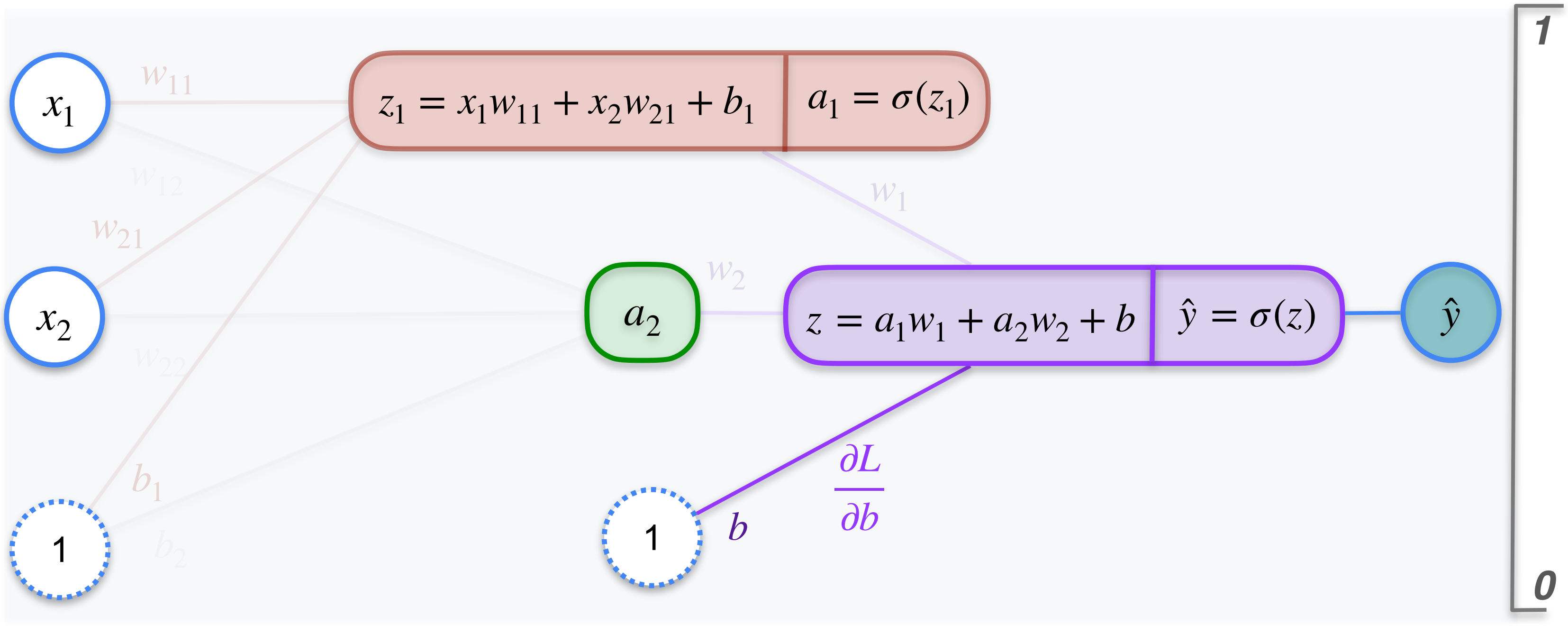


2,2,1 Neural Network

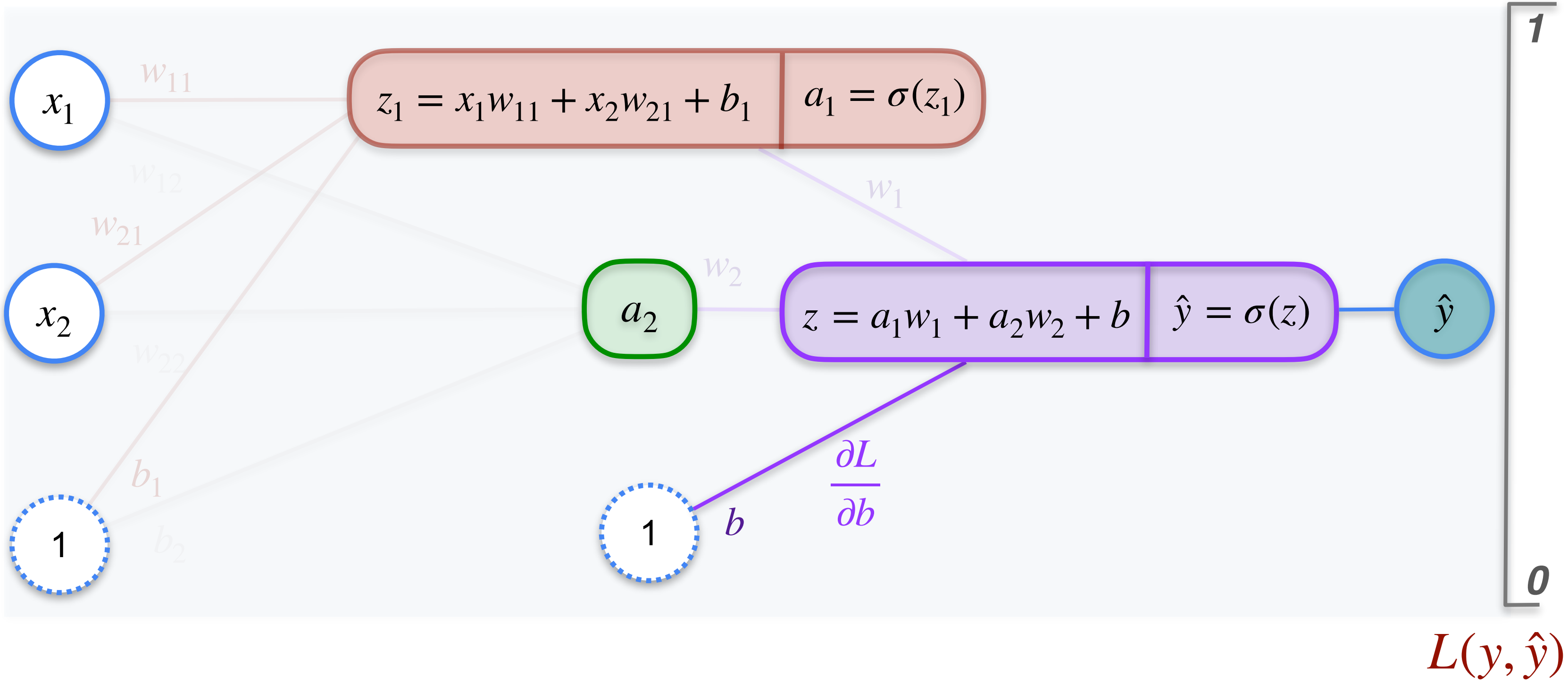


DeepLearning.AI

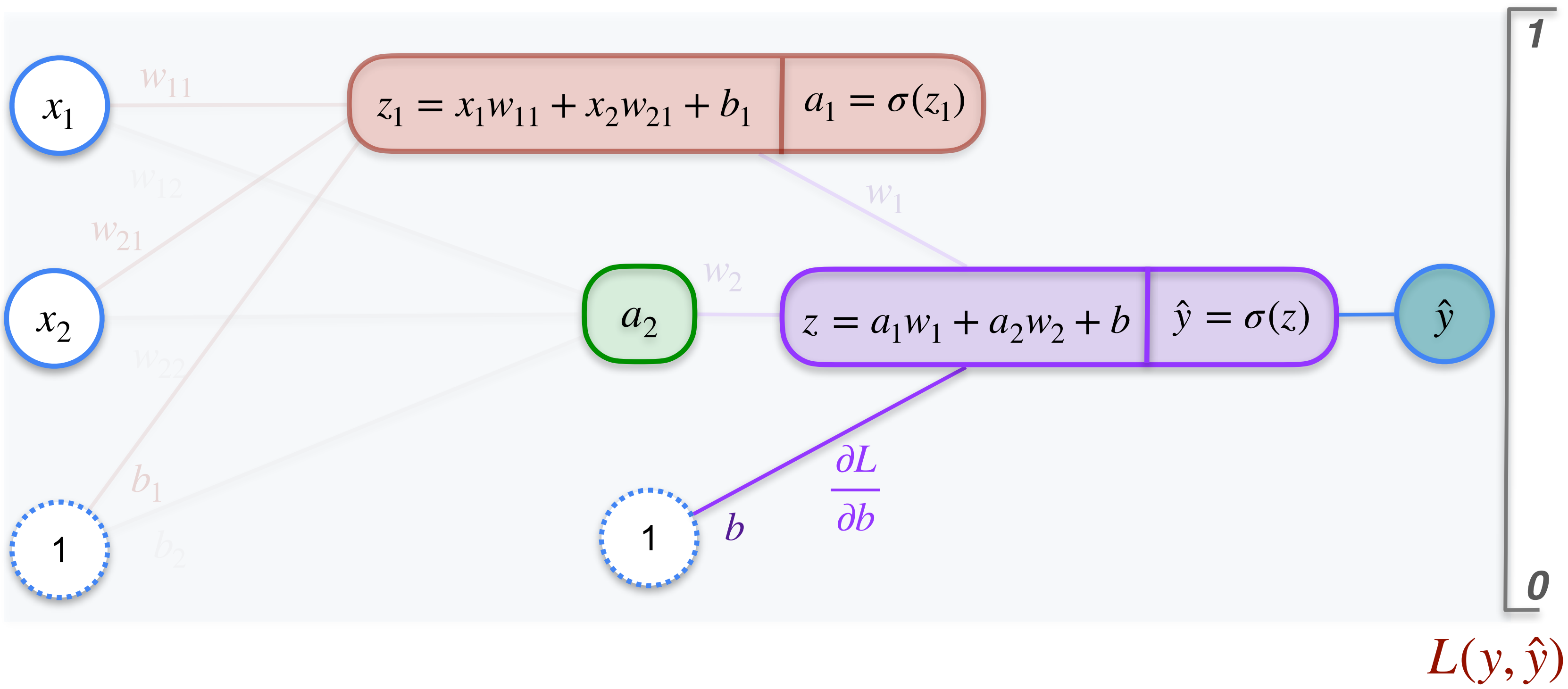
2,2,1 Neural Network



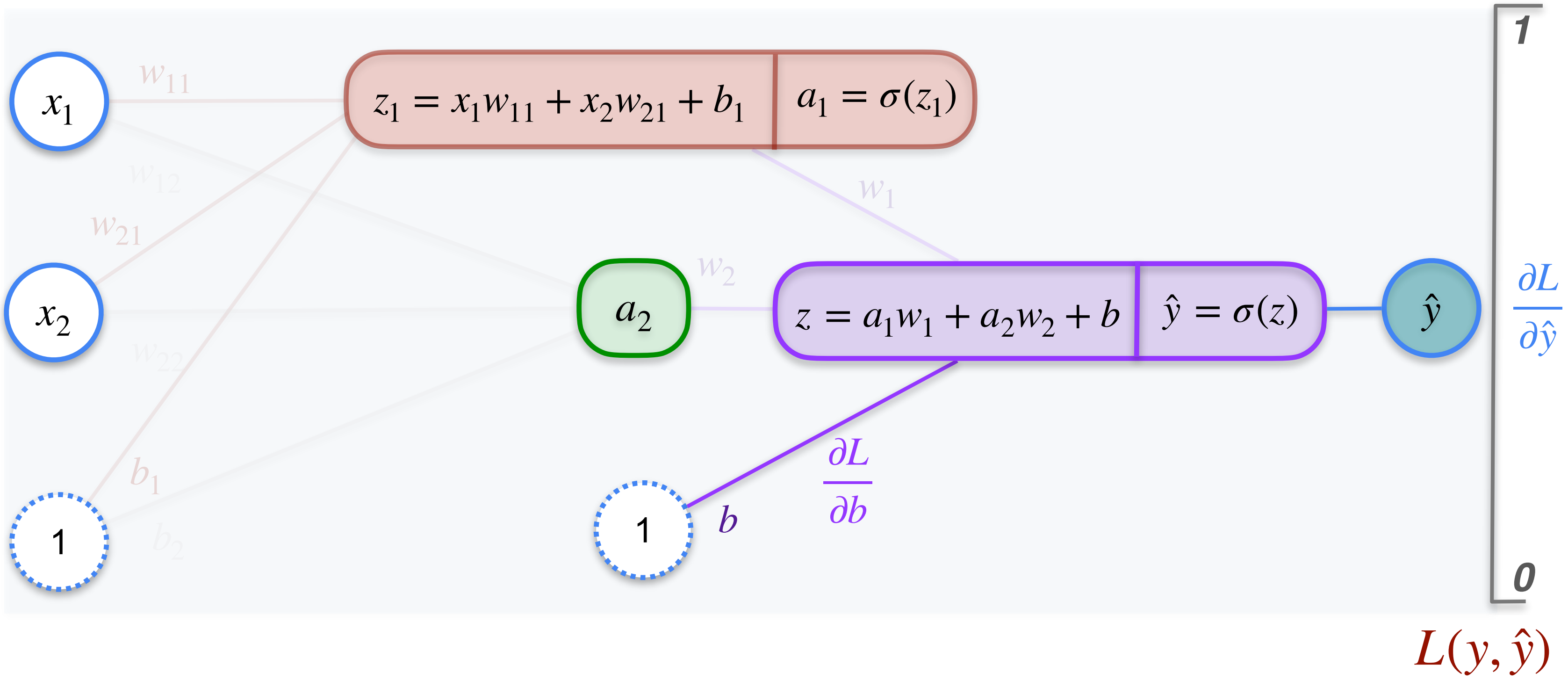
2,2,1 Neural Network



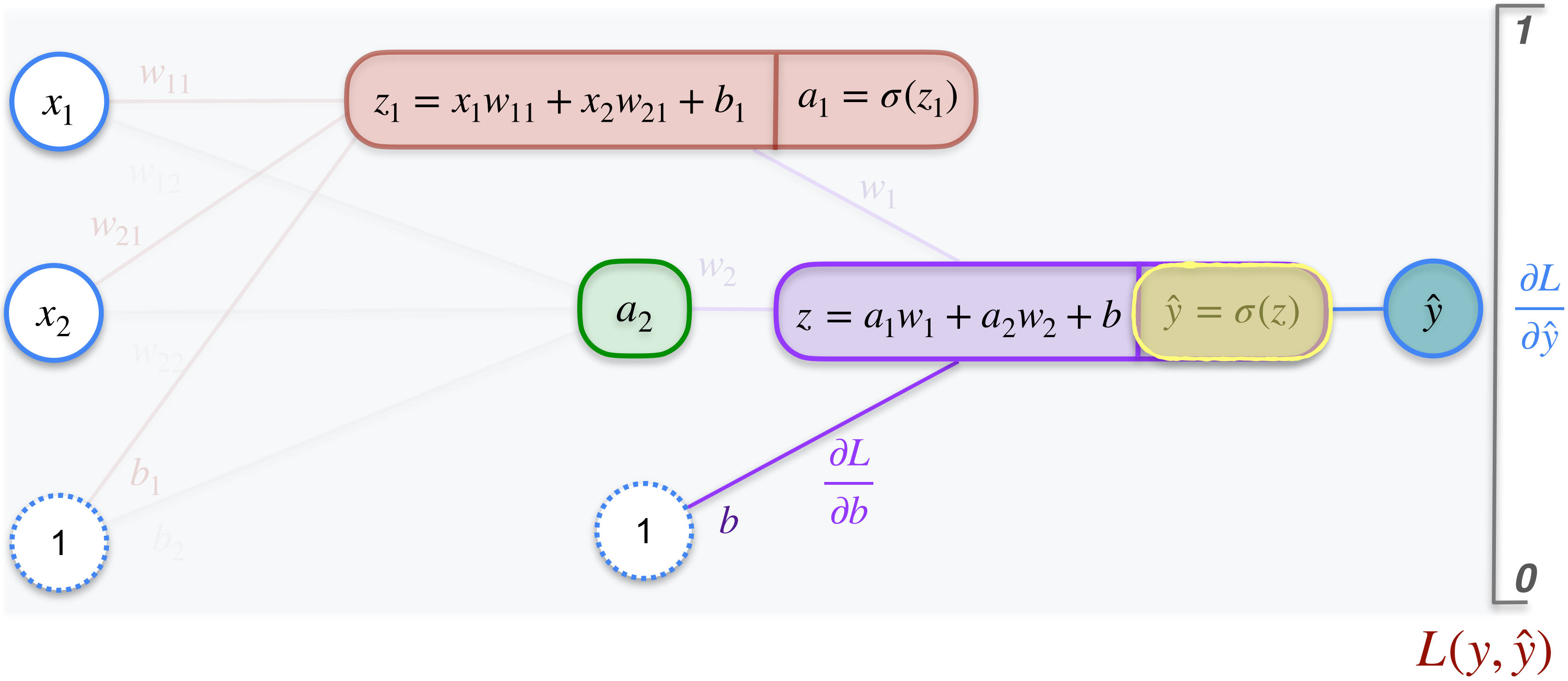
2,2,1 Neural Network



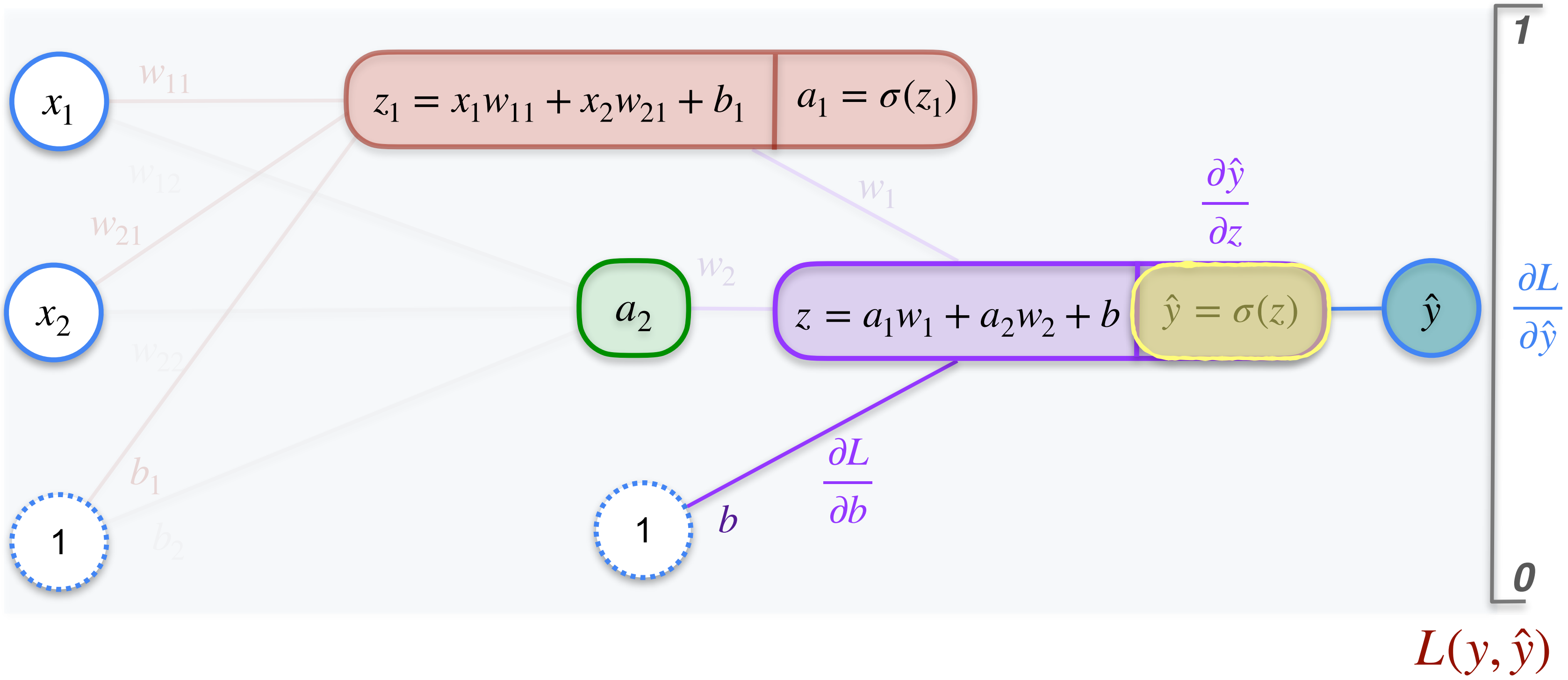
2,2,1 Neural Network



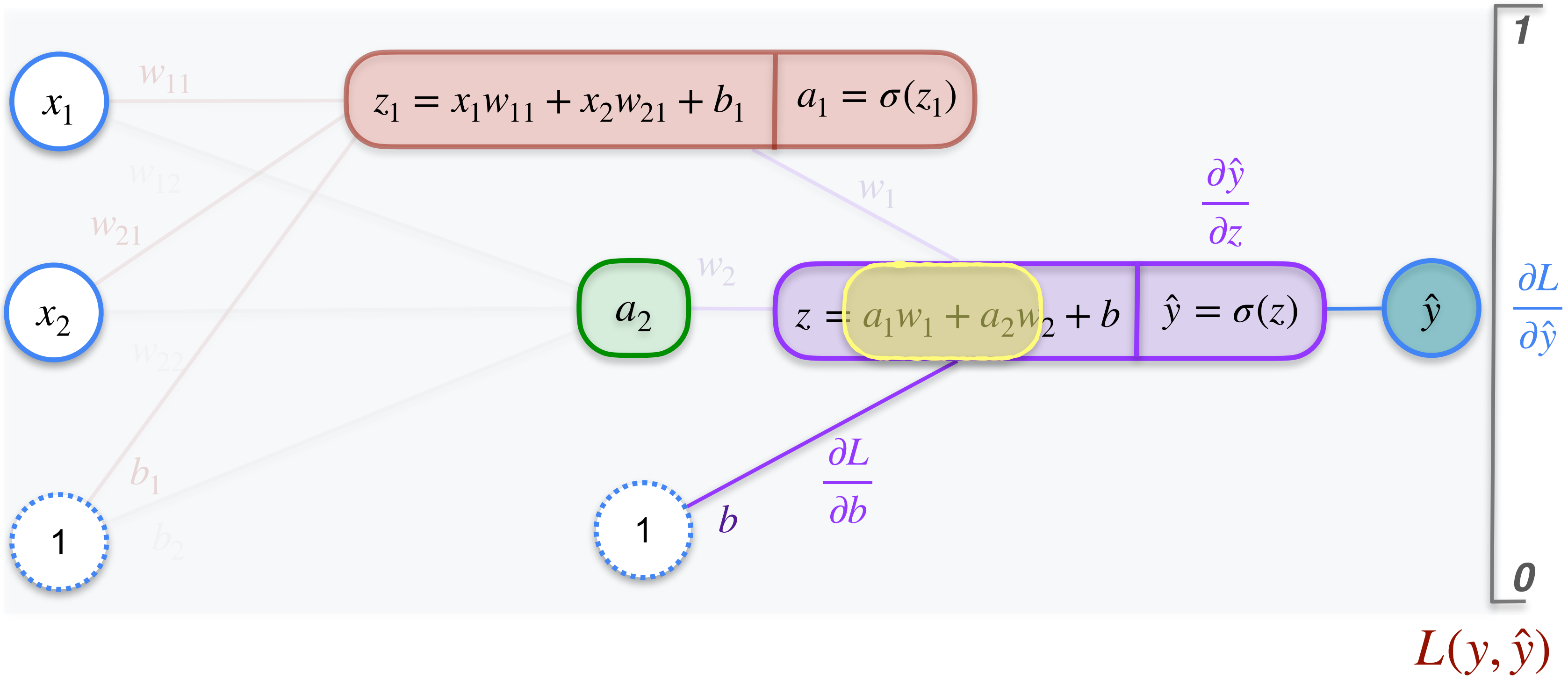
2,2,1 Neural Network



2,2,1 Neural Network

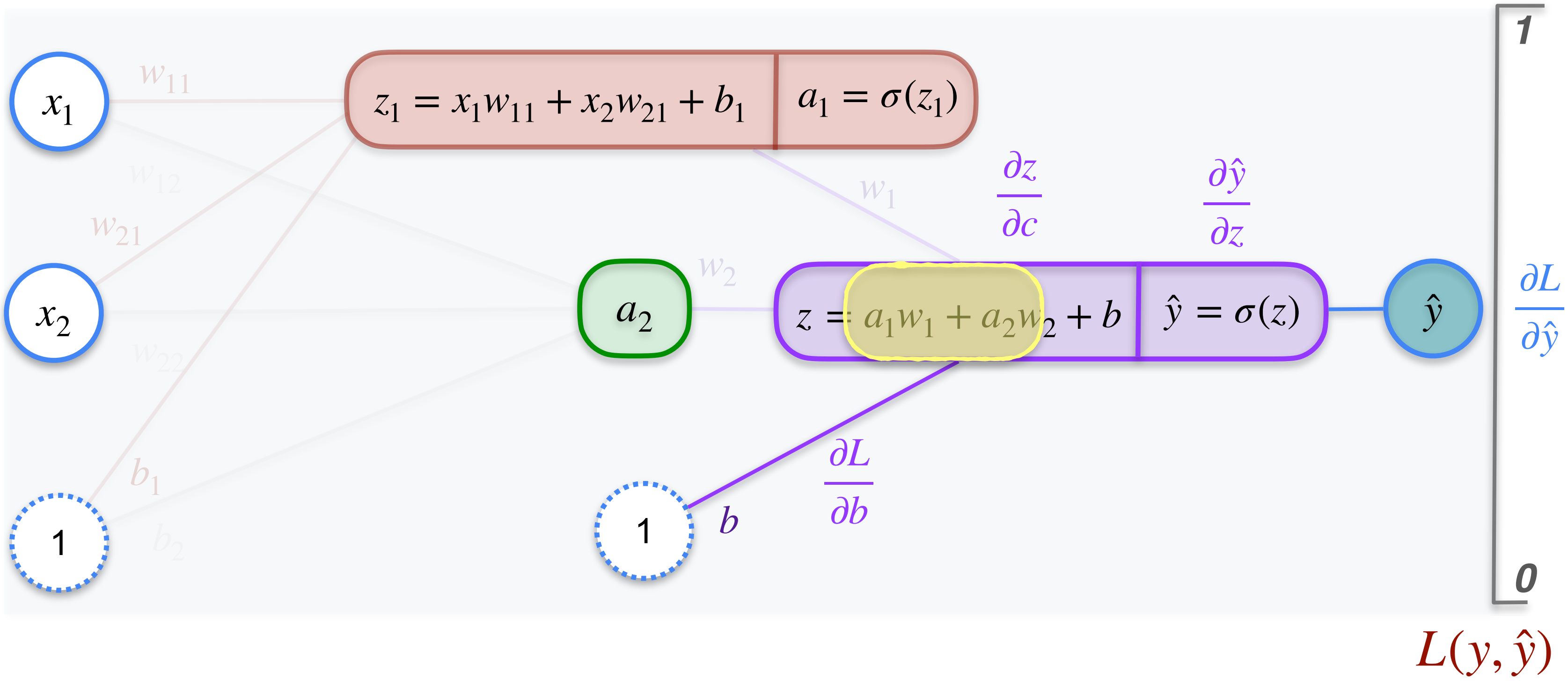


2,2,1 Neural Network

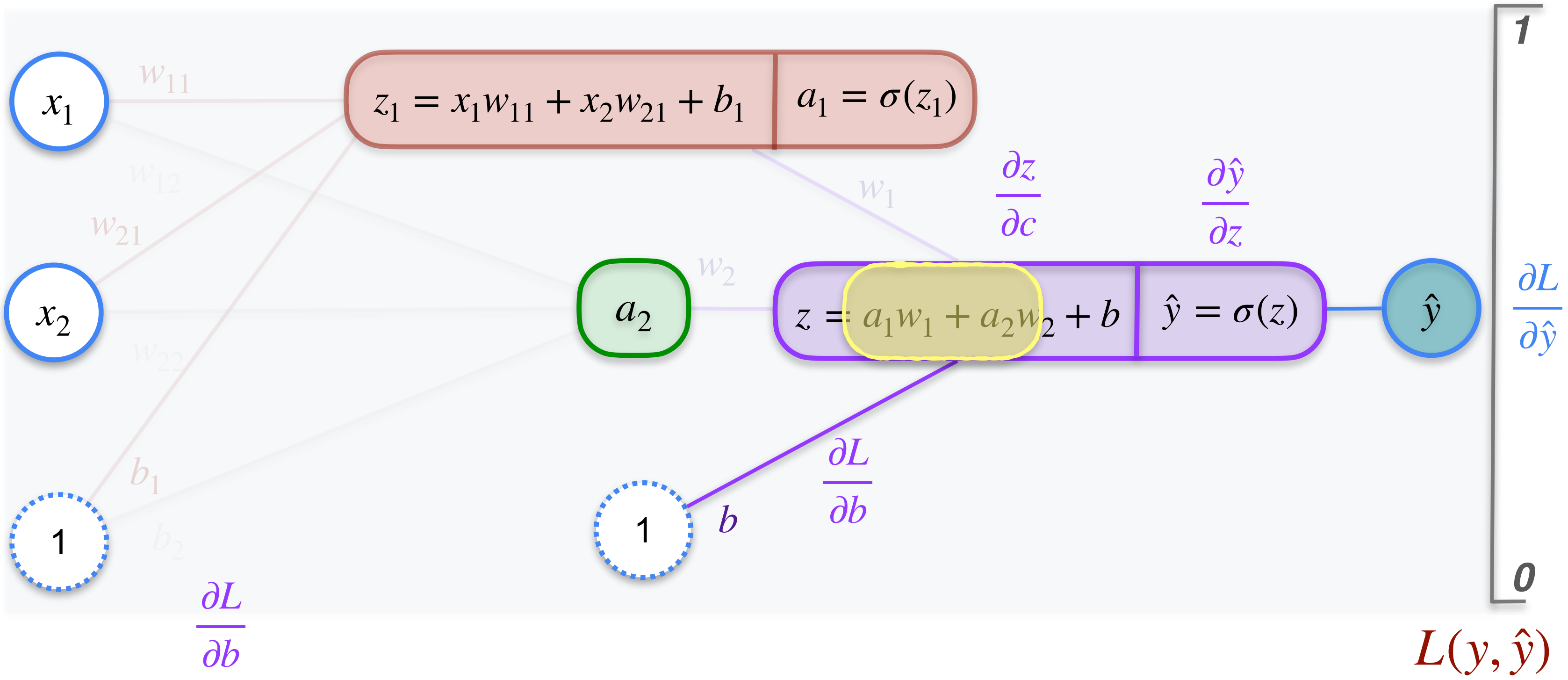


DeepLearning.AI

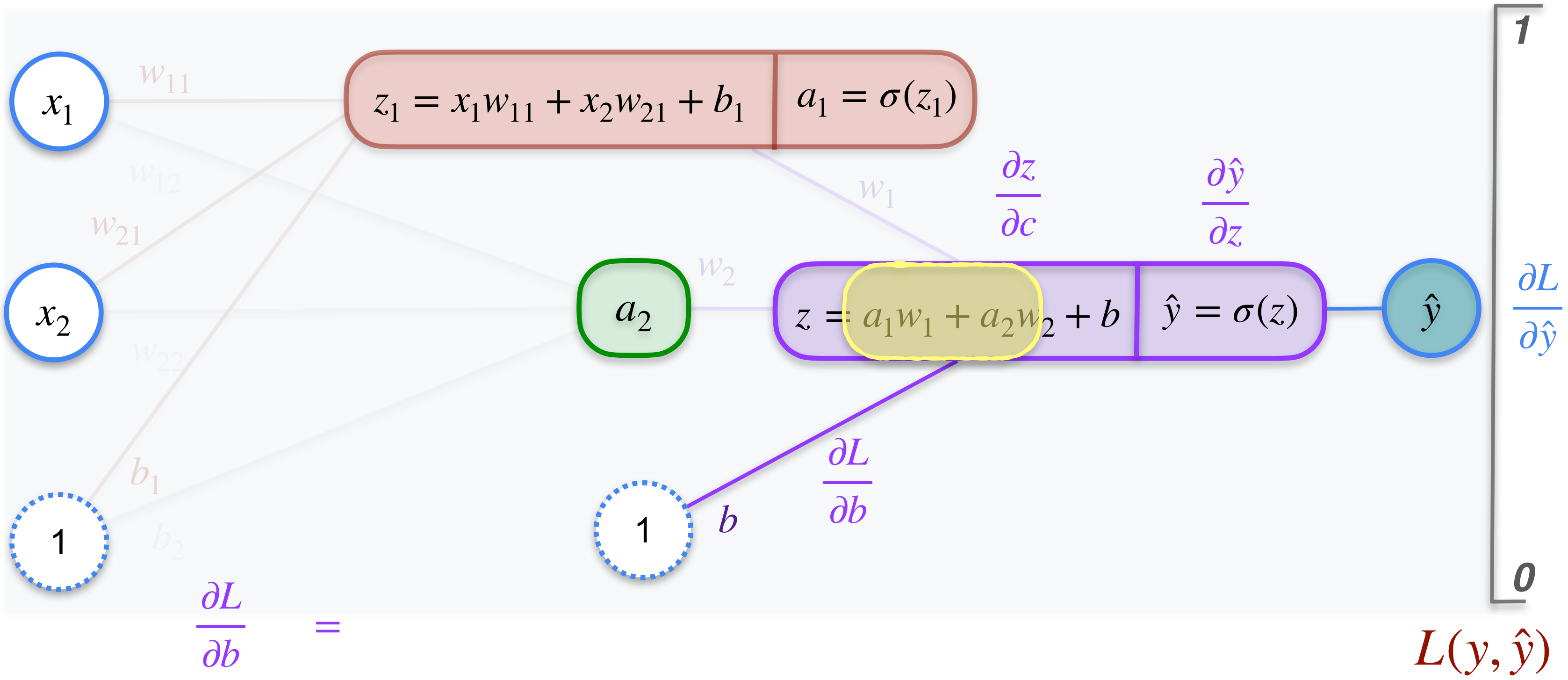
2,2,1 Neural Network



2,2,1 Neural Network

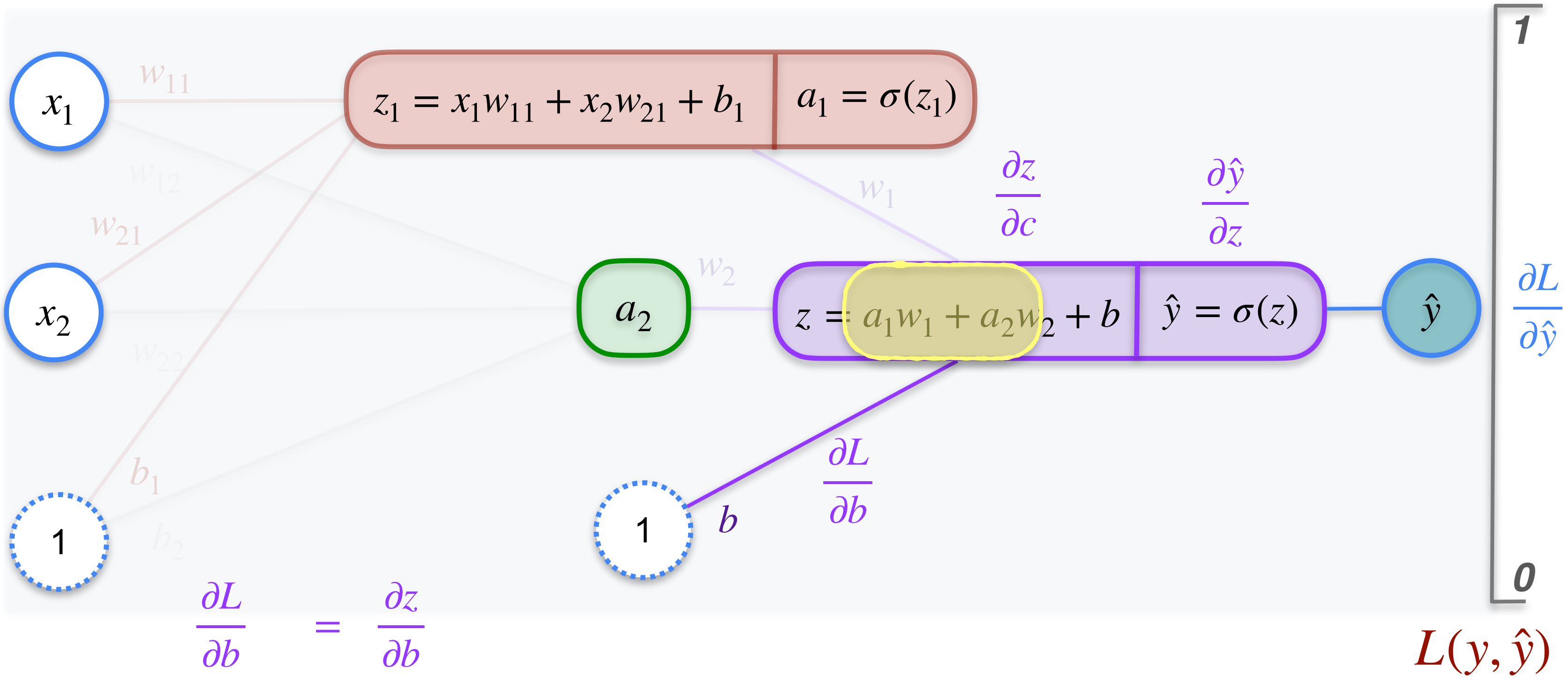


2,2,1 Neural Network



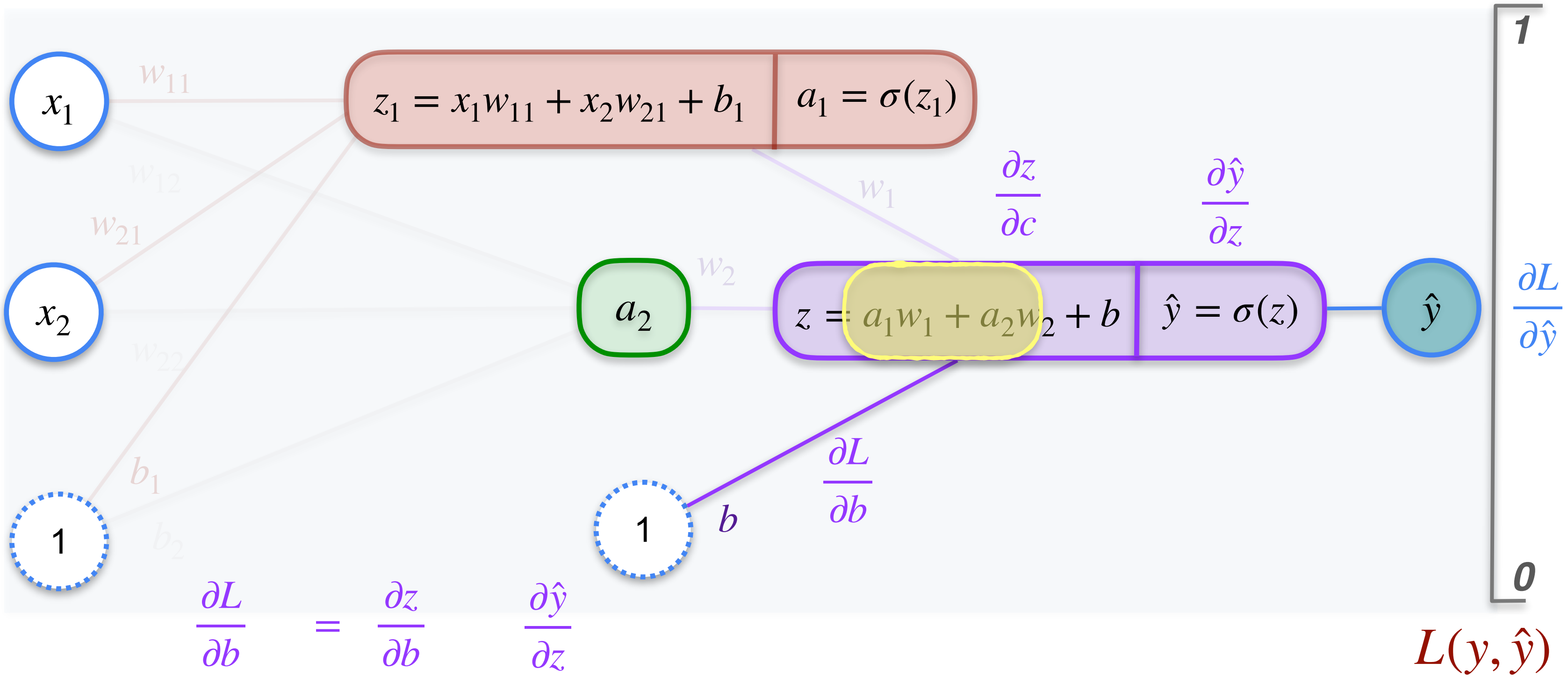
$$\frac{\partial L}{\partial b} =$$

2,2,1 Neural Network



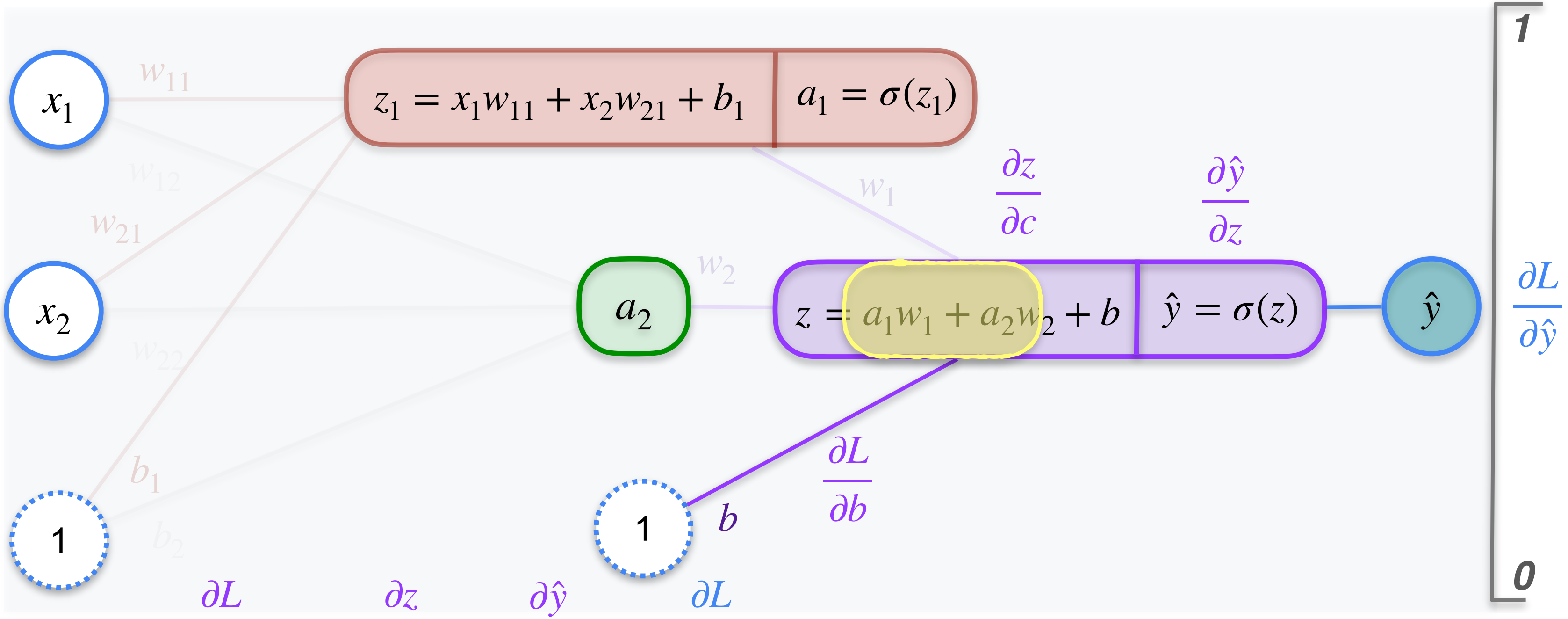
$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b}$$

2,2,1 Neural Network



$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \quad \frac{\partial \hat{y}}{\partial z}$$

2,2,1 Neural Network



$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$L(y, \hat{y})$

2,2,1 Neural Network

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1w_1 + a_2w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1w_1 + a_2w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b}$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} =$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1-y) log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} =$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1-y) log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} =$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1 - y) log(1 - \hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y log(\hat{y}) - (1-y) log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1w_1 + a_2w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1 - y) log(1 - \hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \boxed{\frac{\partial \hat{y}}{\partial z}} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1 \quad \hat{y}(1 - \hat{y})$$

DeepLearning.AI

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1 - y) log(1 - \hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1 \quad \hat{y}(1 - \hat{y})$$

2,2,1 Neural Network

$L(y, \hat{y}) = -ylog(\hat{y}) - (1 - y)log(1 - \hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1 \quad \hat{y}(1 - \hat{y})$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -ylog(\hat{y}) - (1-y)log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1w_1 + a_2w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1 \quad \hat{y}(1-\hat{y}) \quad \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y log(\hat{y}) - (1-y) log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \hat{y}(1-\hat{y}) \cdot \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

2,2,1 Neural Network

$L(y, \hat{y}) = -ylog(\hat{y}) - (1-y)log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\hat{y}(1-\hat{y})}$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y log(\hat{y}) - (1 - y) log(1 - \hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\,log(\hat{y}) - (1-y)log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -(y - \hat{y})$$

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -(y - \hat{y})$$

to find optimal value of b that gives the least error

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -(y - \hat{y})$$

Perform gradient descent with

to find optimal value of b that gives the least error

2,2,1 Neural Network

$$L(y, \hat{y}) = -y log(\hat{y}) - (1 - y) log(1 - \hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1 - \hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1 - \hat{y})}}$$

$$= -(y - \hat{y})$$

Perform gradient descent with

$$b \rightarrow b - \alpha \frac{\partial L}{\partial b}$$

to find optimal value of b that gives the least error

2,2,1 Neural Network

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

$$\hat{y} = \sigma(z)$$

$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -(y - \hat{y})$$

Perform gradient descent with

$$b \rightarrow b - \alpha$$

to find optimal value of b that gives the least error

2,2,1 Neural Network

$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

$\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \boxed{\frac{\partial L}{\partial \hat{y}}}$$

$$\frac{\partial L}{\partial b} = 1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{-(y - \hat{y})}{\cancel{\hat{y}(1-\hat{y})}}$$

$$= -(y - \hat{y})$$

Perform gradient descent with

$$b \rightarrow b - \alpha(-(y - \hat{y}))$$

to find optimal value of b that gives the least error