



DeepLearning.AI

# Optimization in Neural Networks and Newton's Method

---

## Gradient Descent and Backpropagation

# Back Propagation Introduction

# Back Propagation Introduction



# Back Propagation Introduction



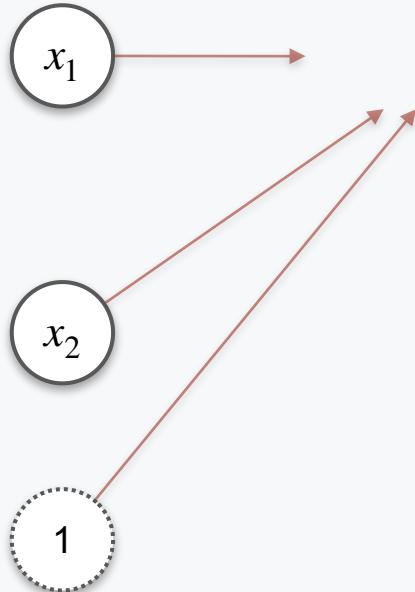
# Back Propagation Introduction

$x_1$

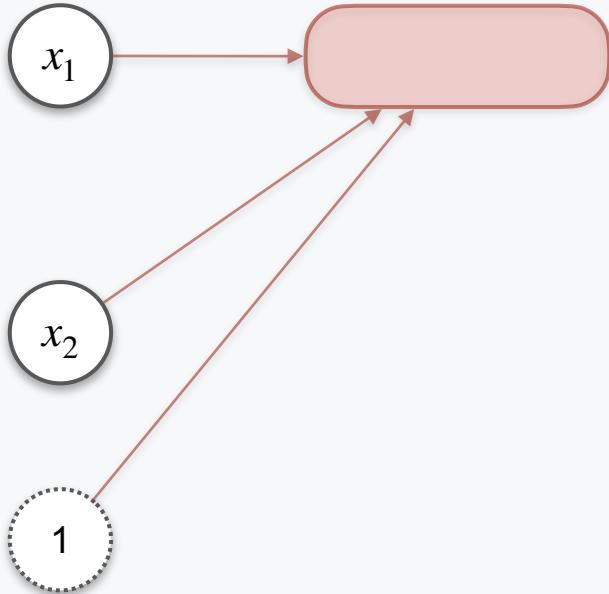
$x_2$

1

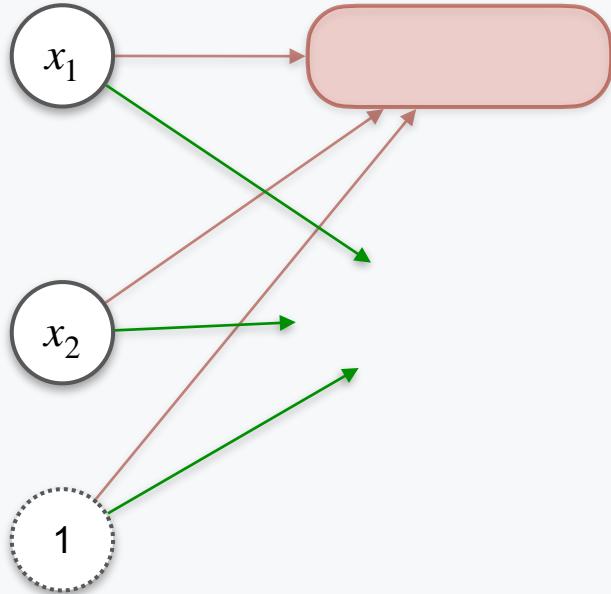
# Back Propagation Introduction



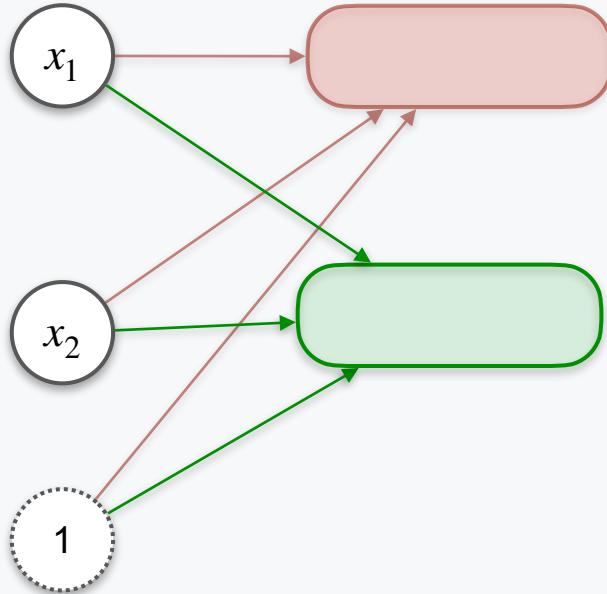
# Back Propagation Introduction



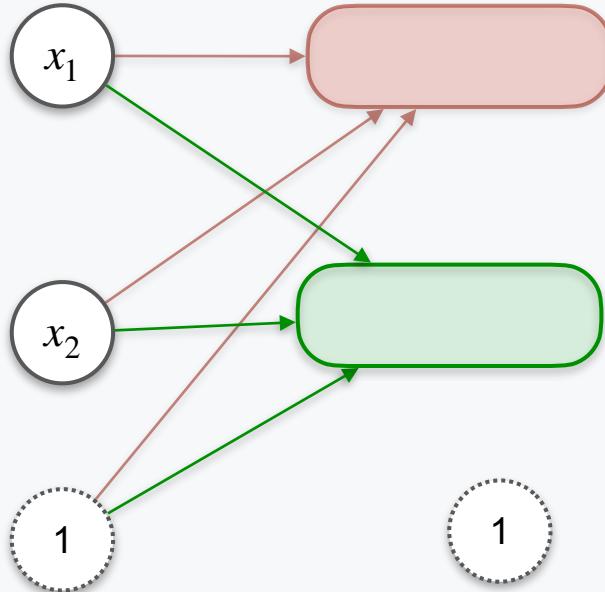
# Back Propagation Introduction



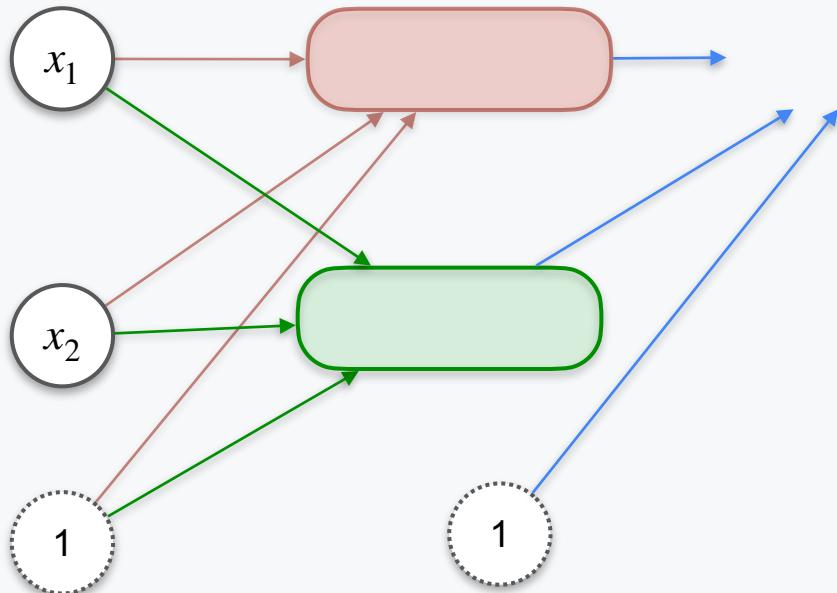
# Back Propagation Introduction



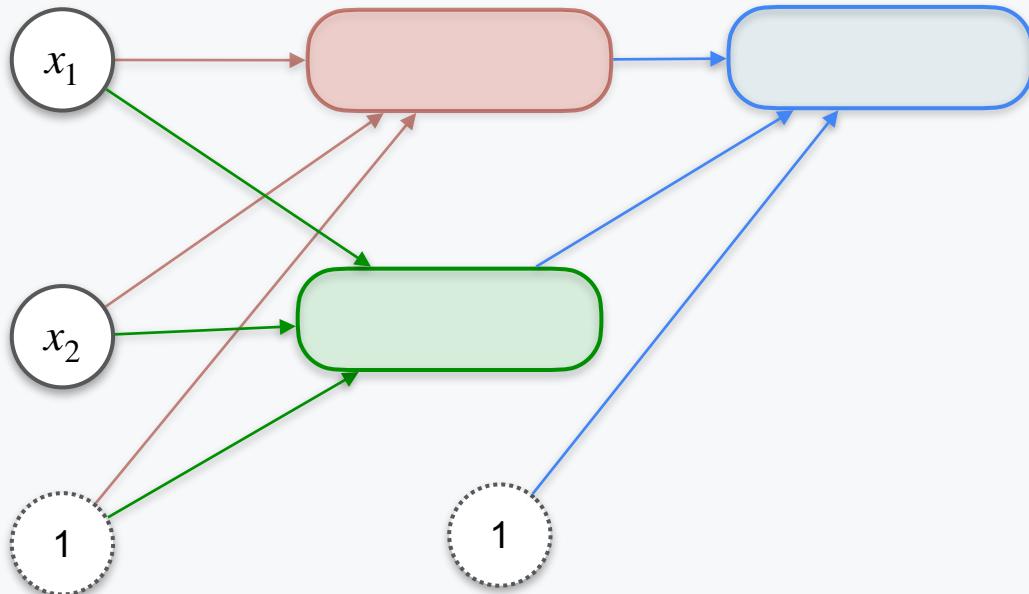
# Back Propagation Introduction



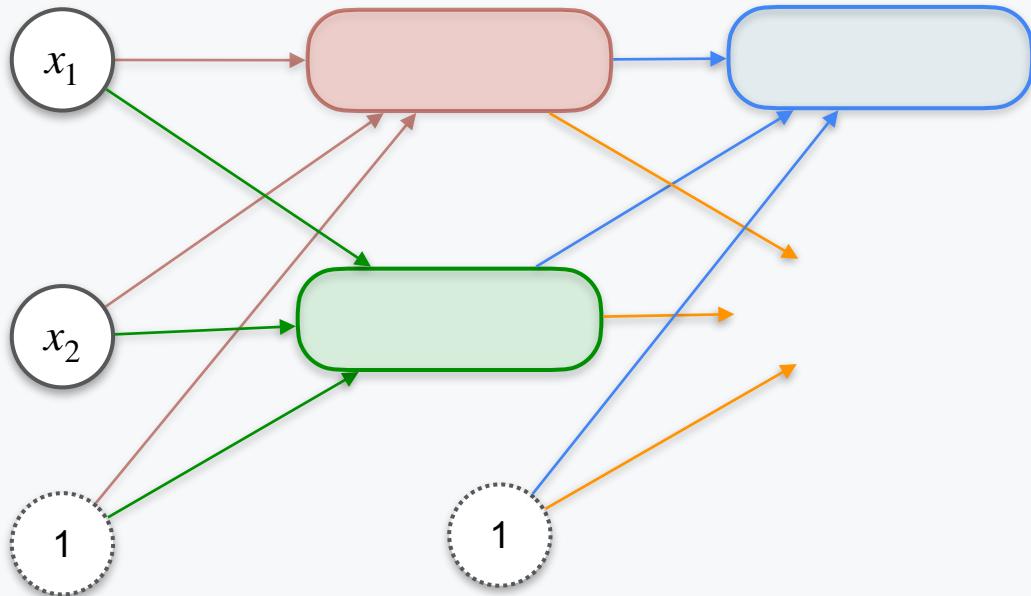
# Back Propagation Introduction



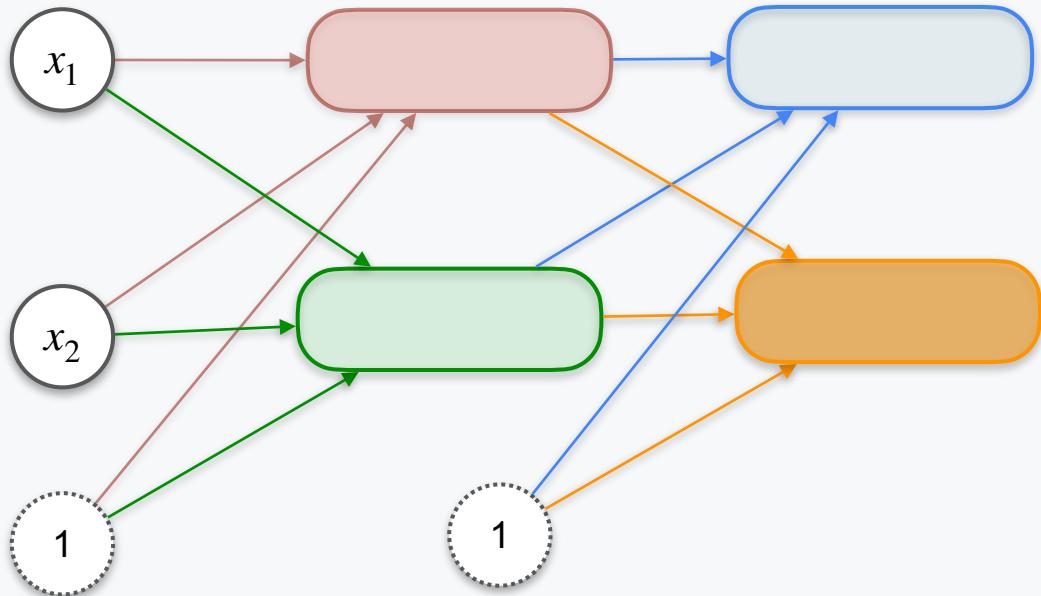
# Back Propagation Introduction



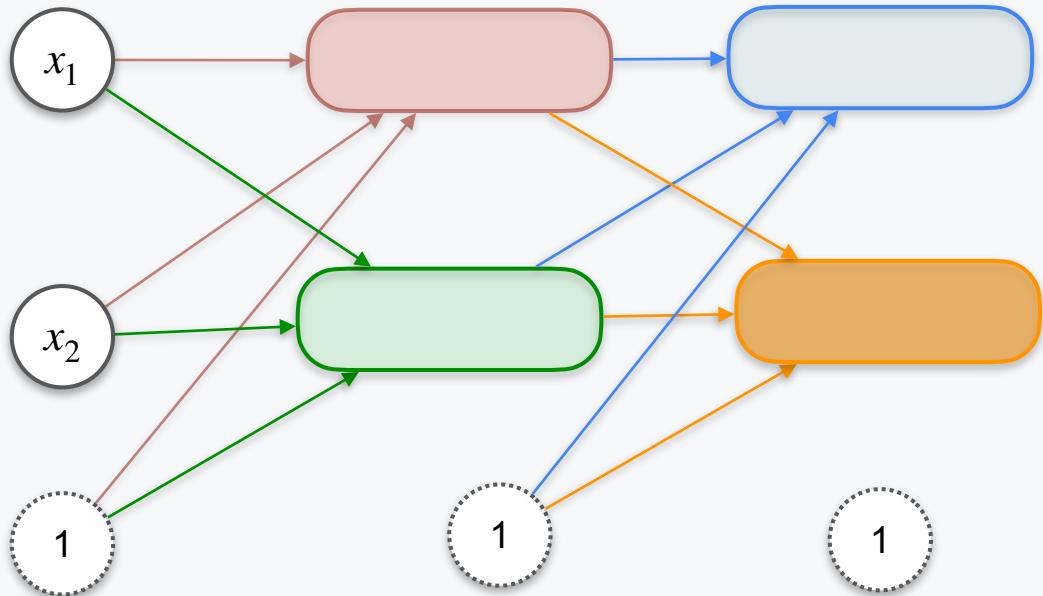
# Back Propagation Introduction



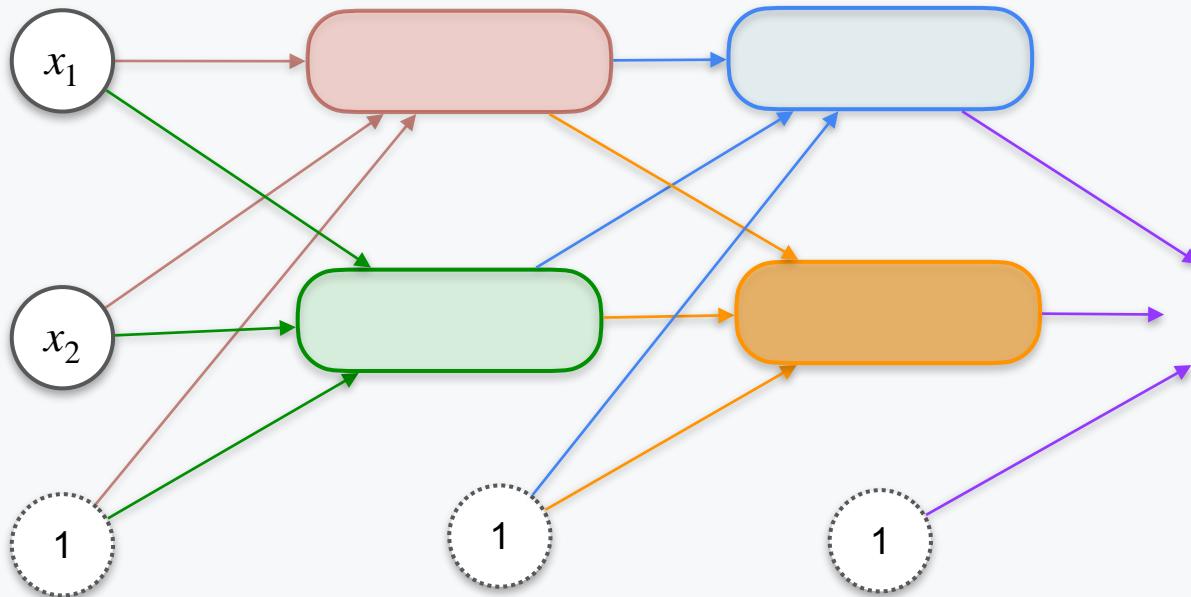
# Back Propagation Introduction



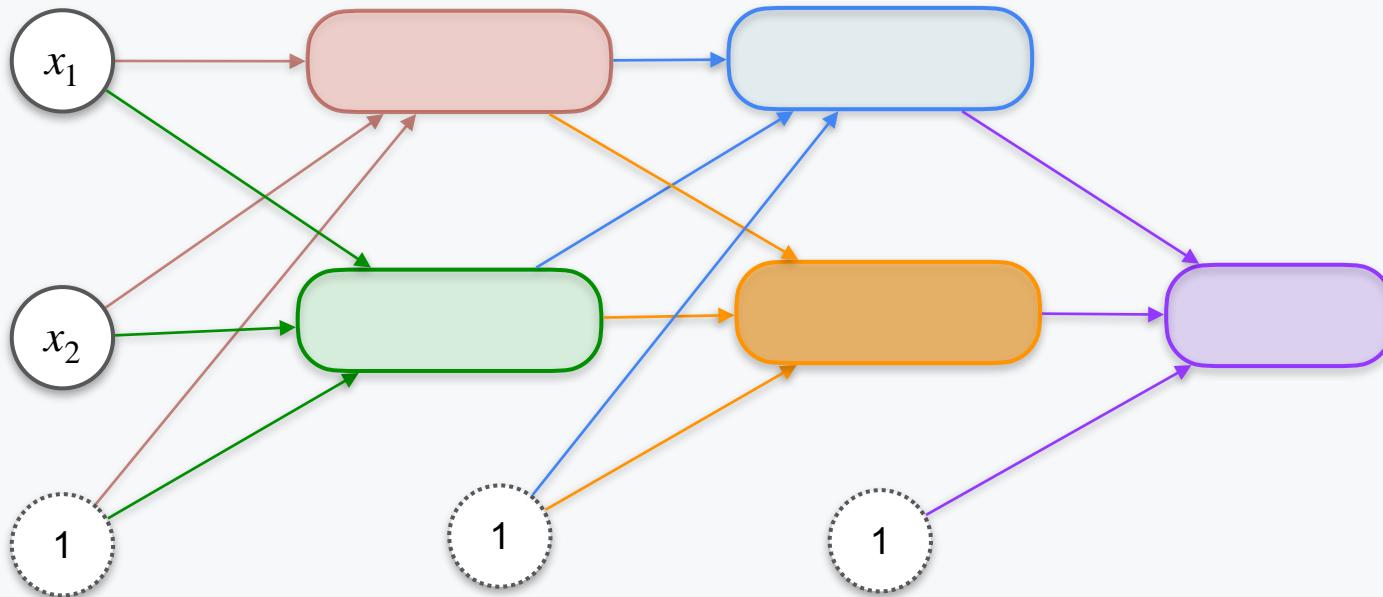
# Back Propagation Introduction



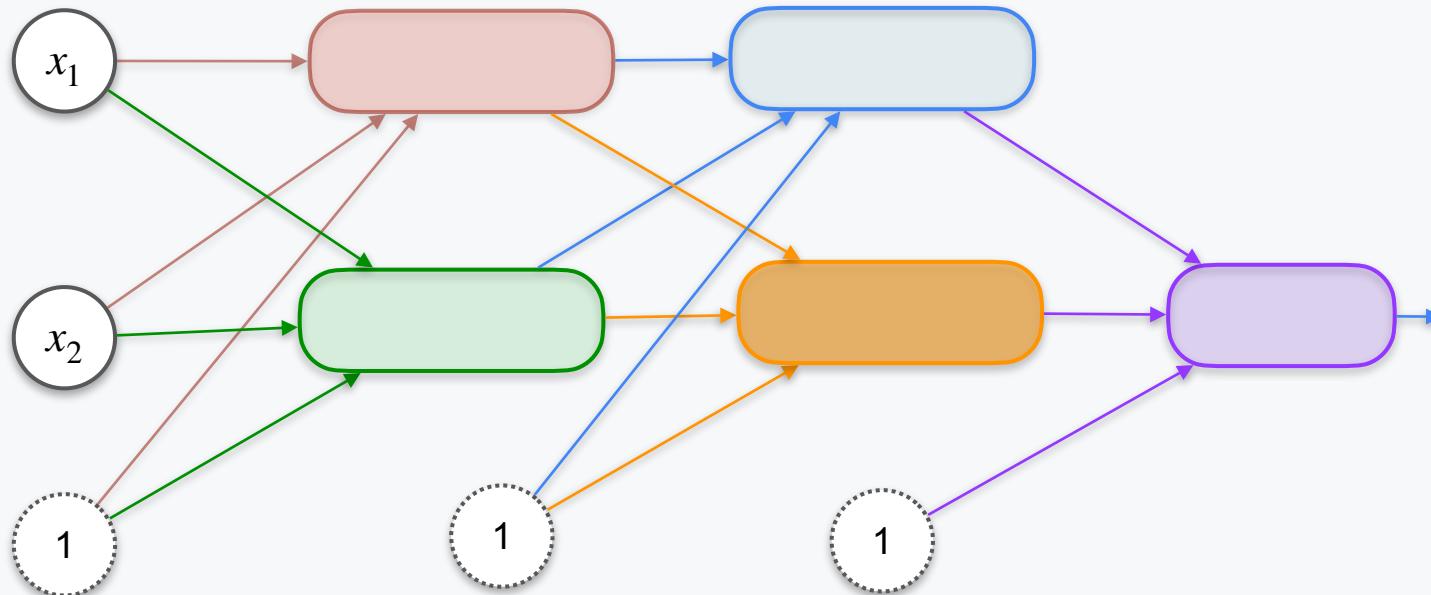
# Back Propagation Introduction



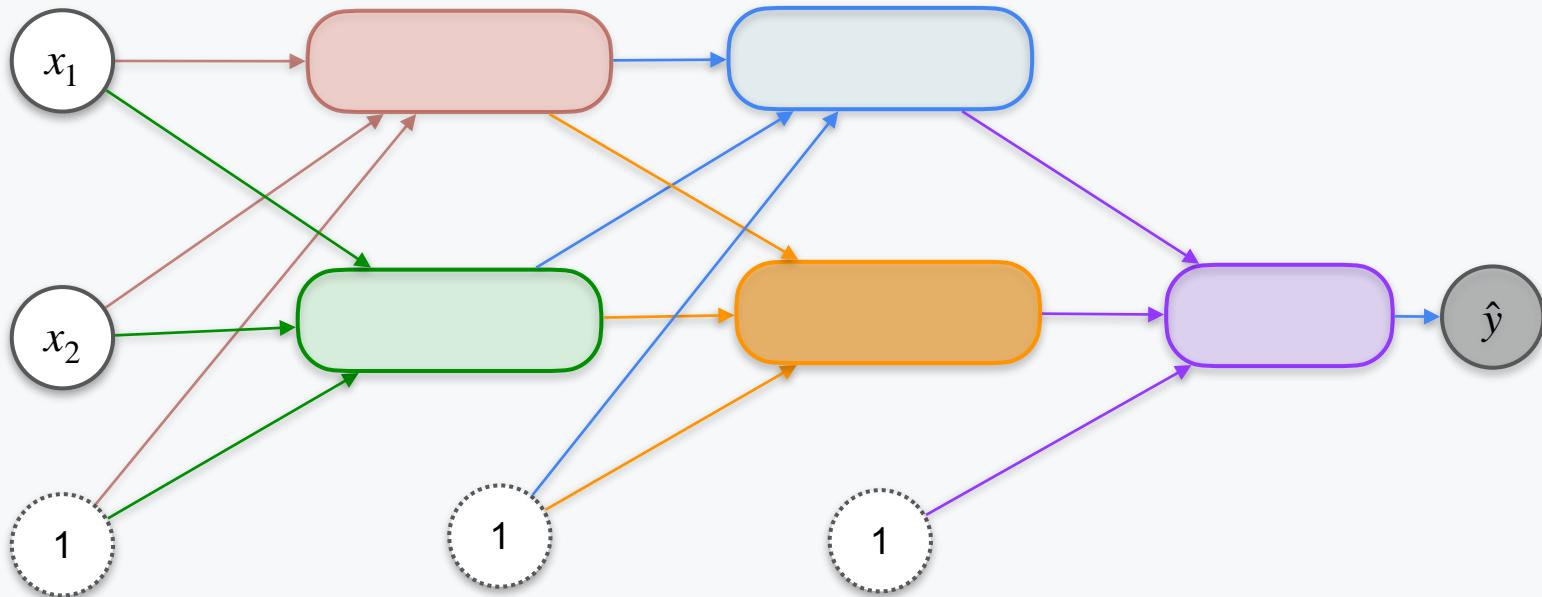
# Back Propagation Introduction



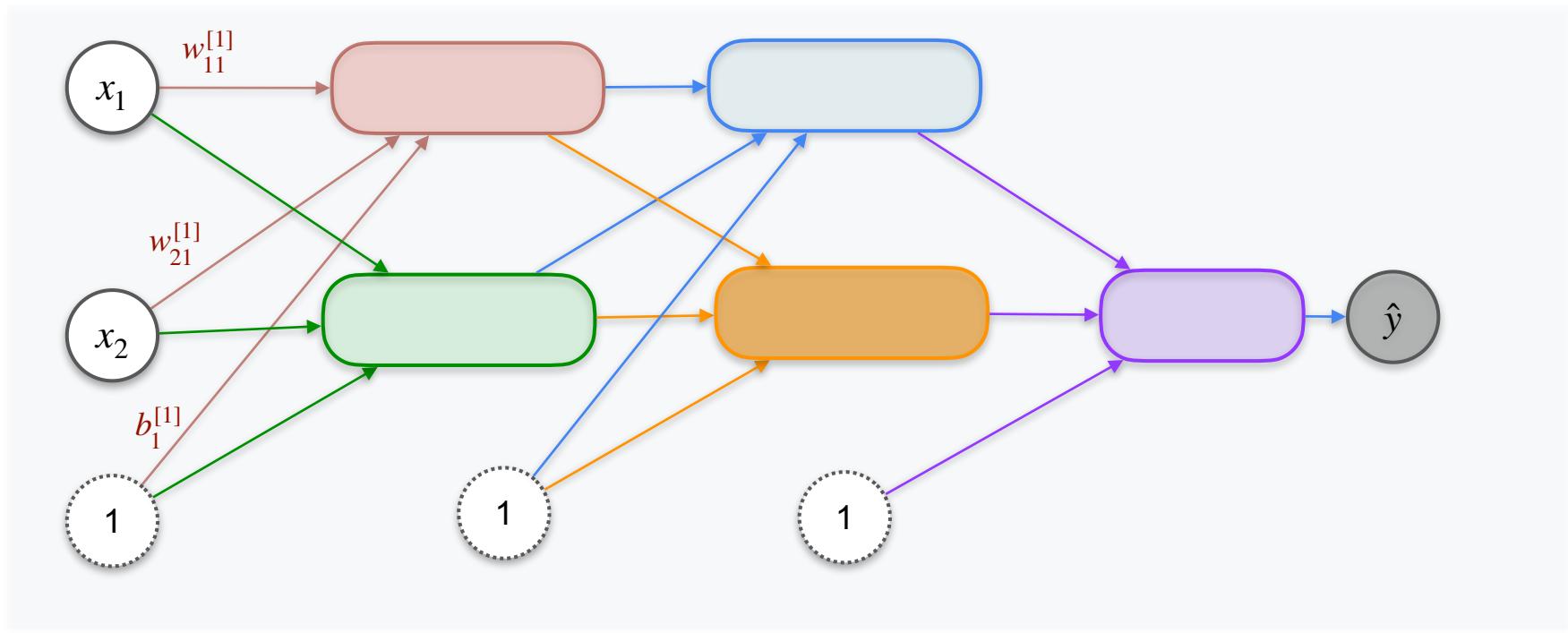
# Back Propagation Introduction



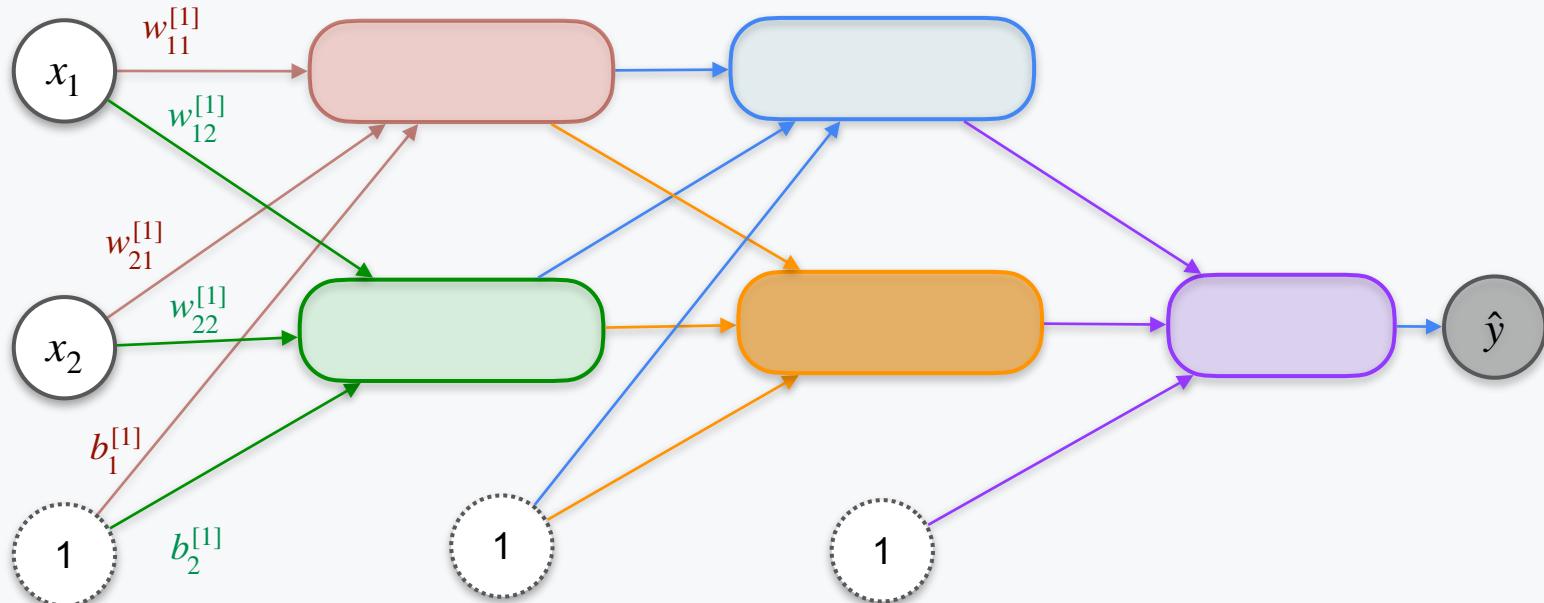
# Back Propagation Introduction



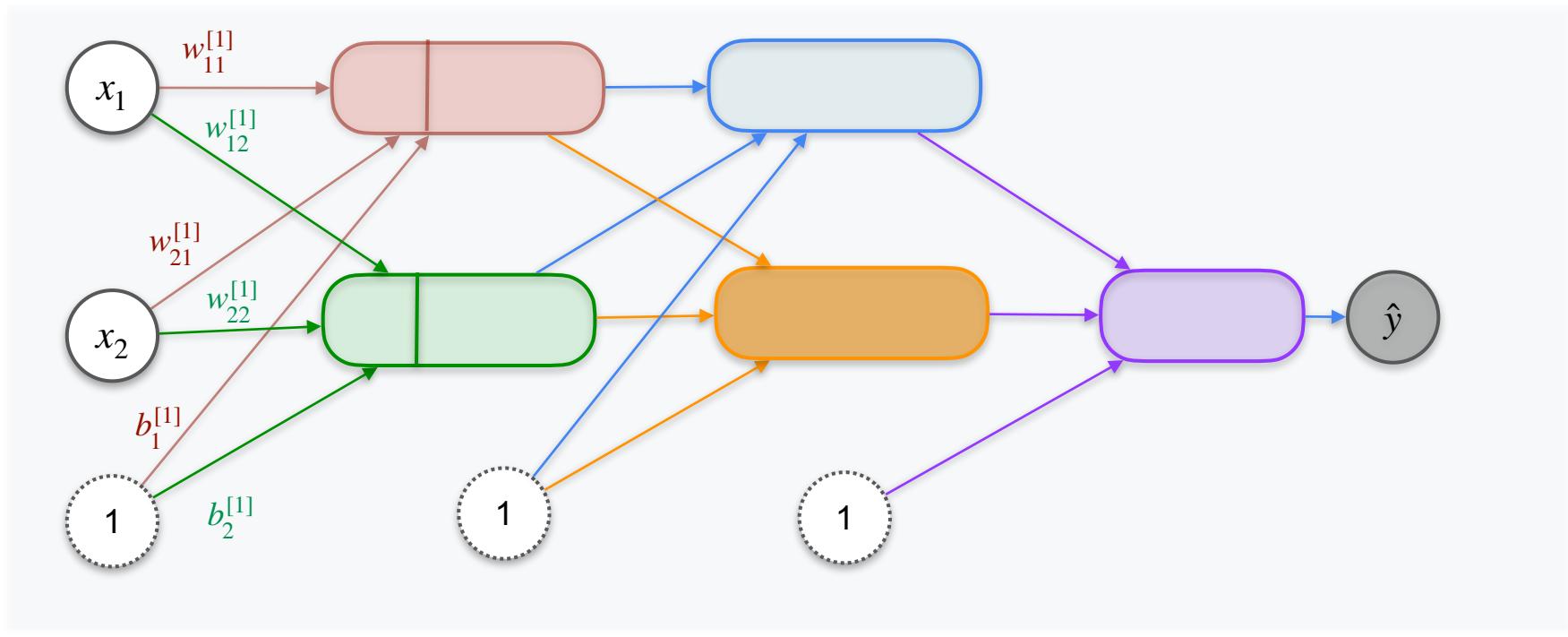
# Back Propagation Introduction



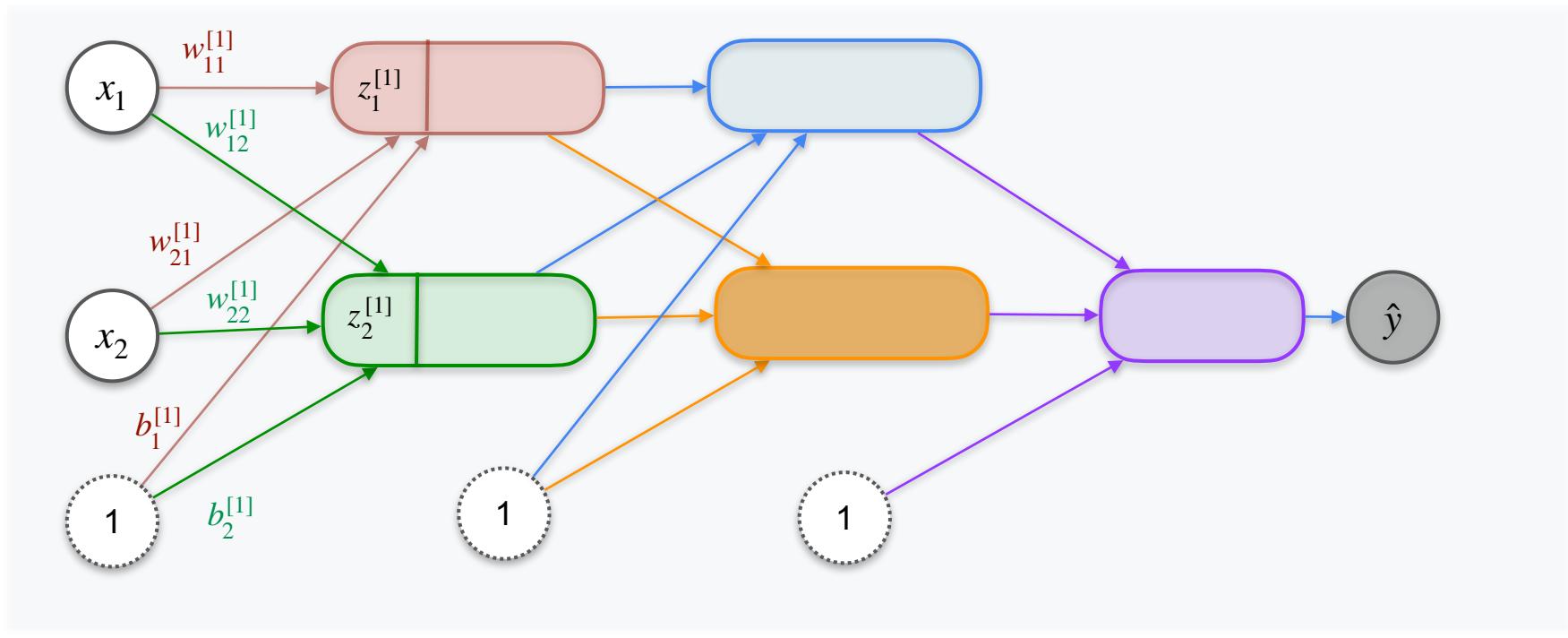
# Back Propagation Introduction



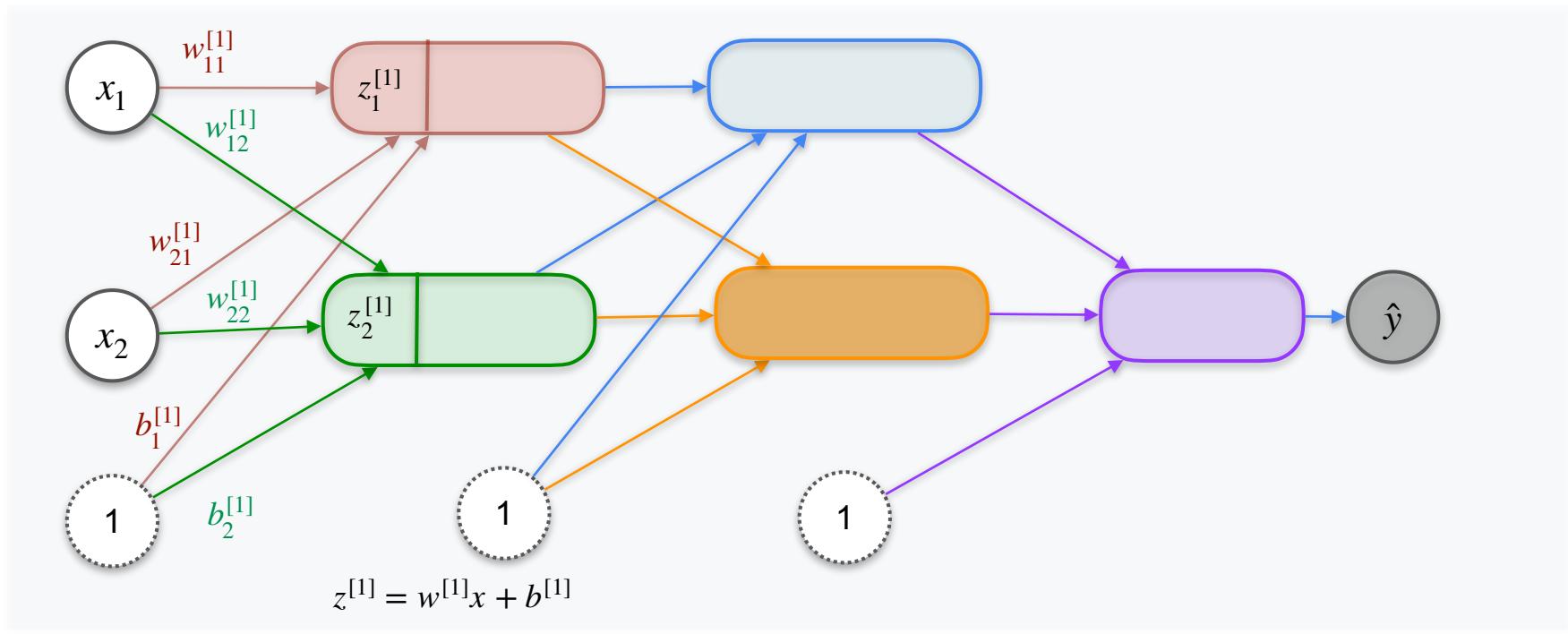
# Back Propagation Introduction



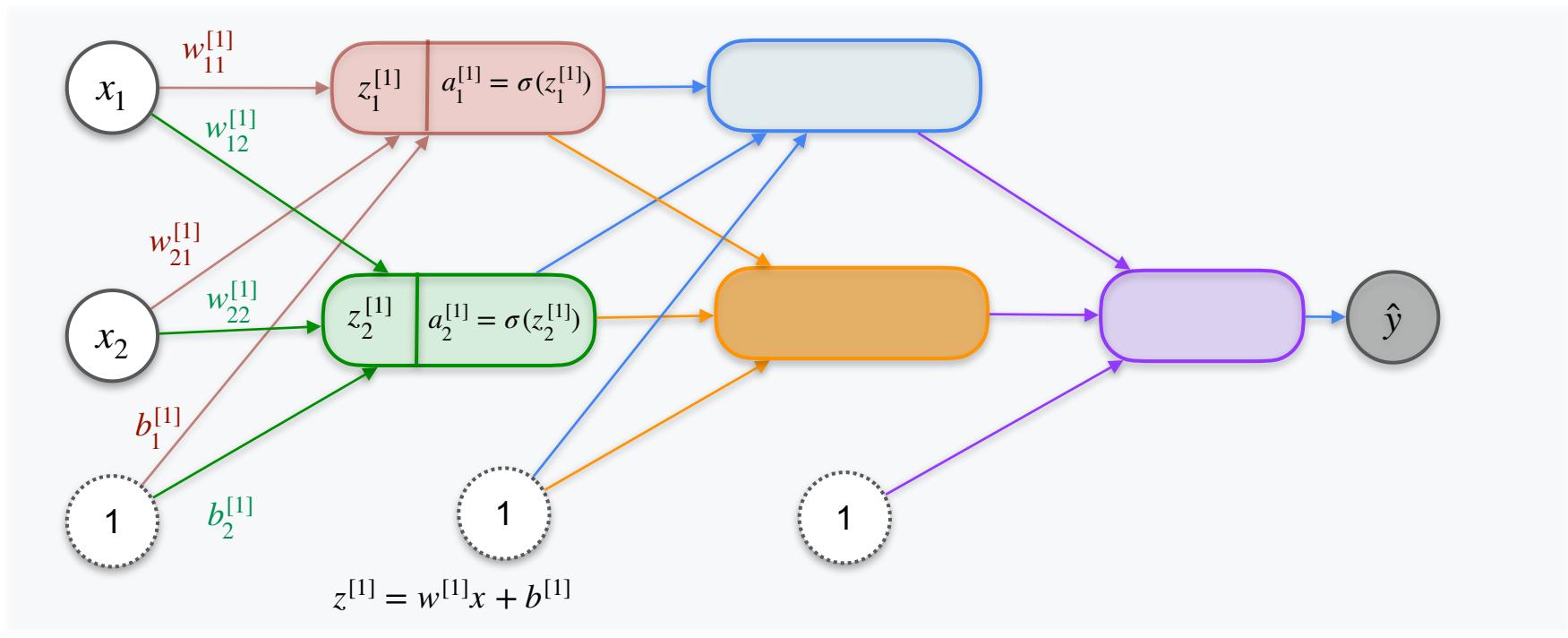
# Back Propagation Introduction



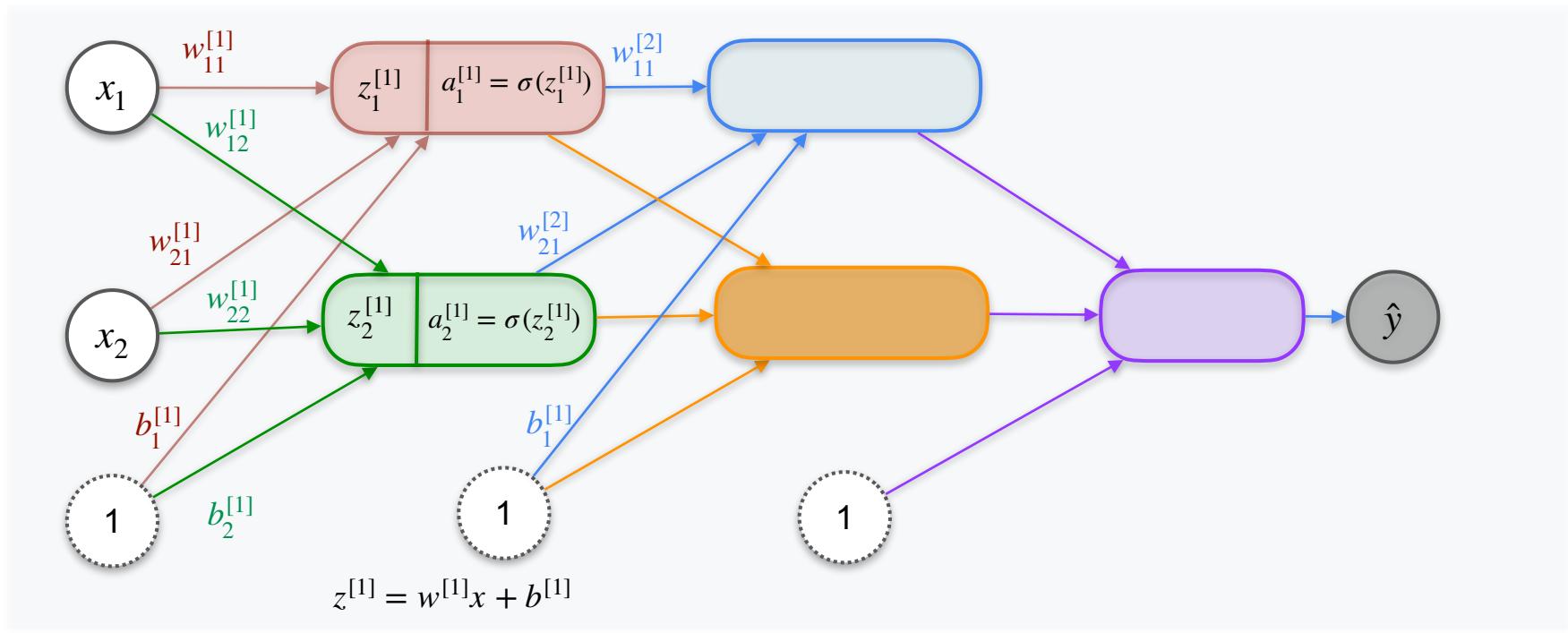
# Back Propagation Introduction



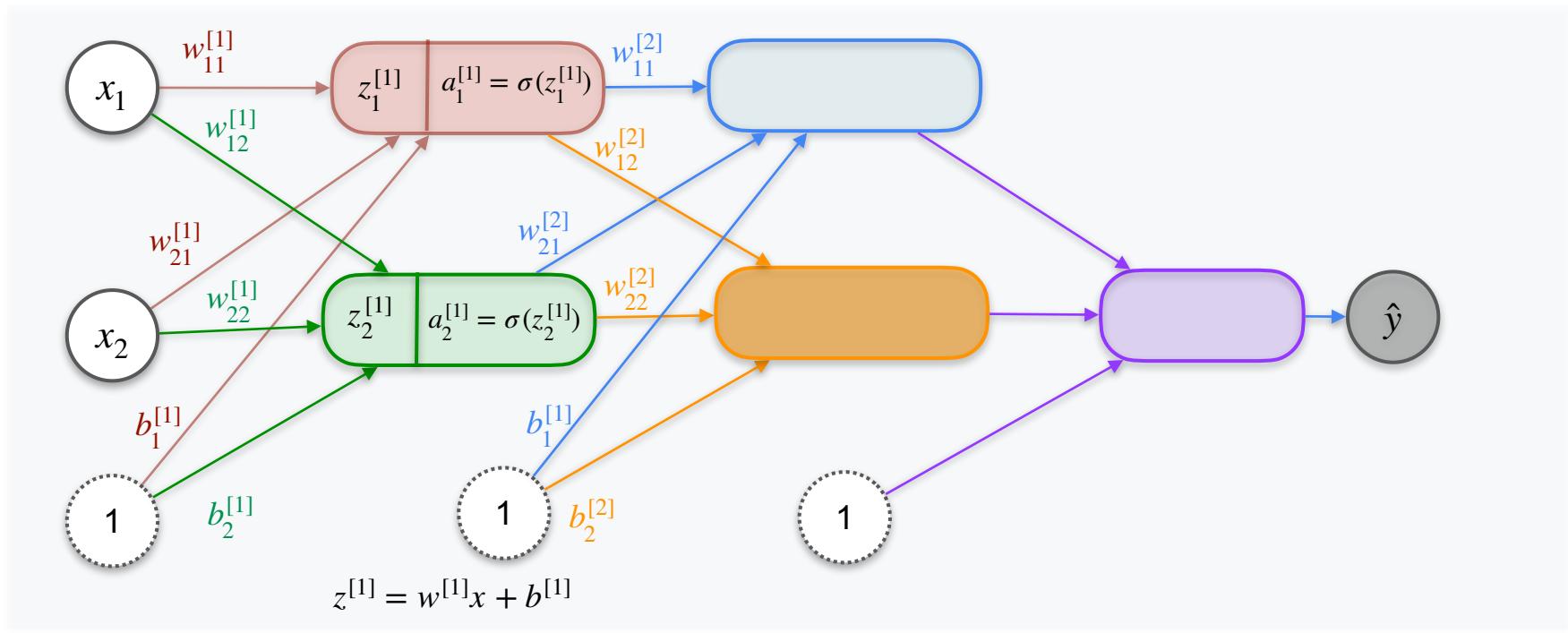
# Back Propagation Introduction



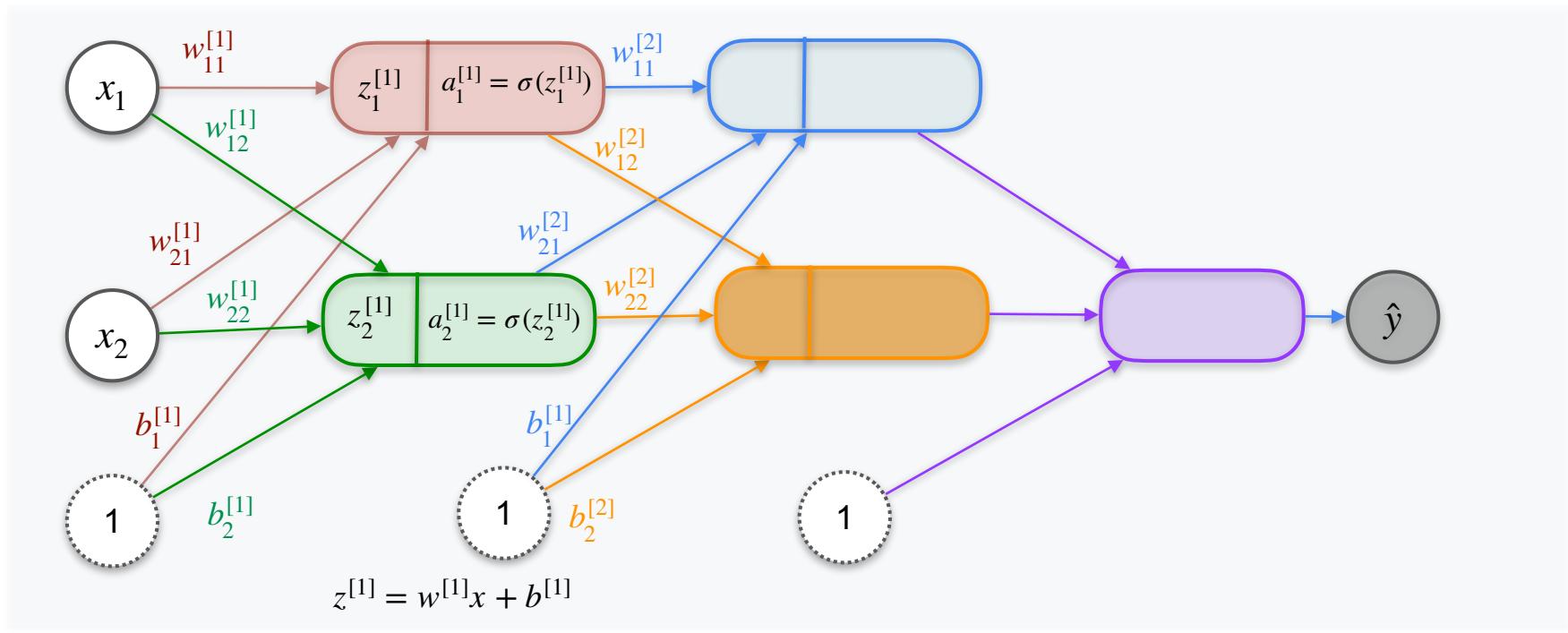
# Back Propagation Introduction



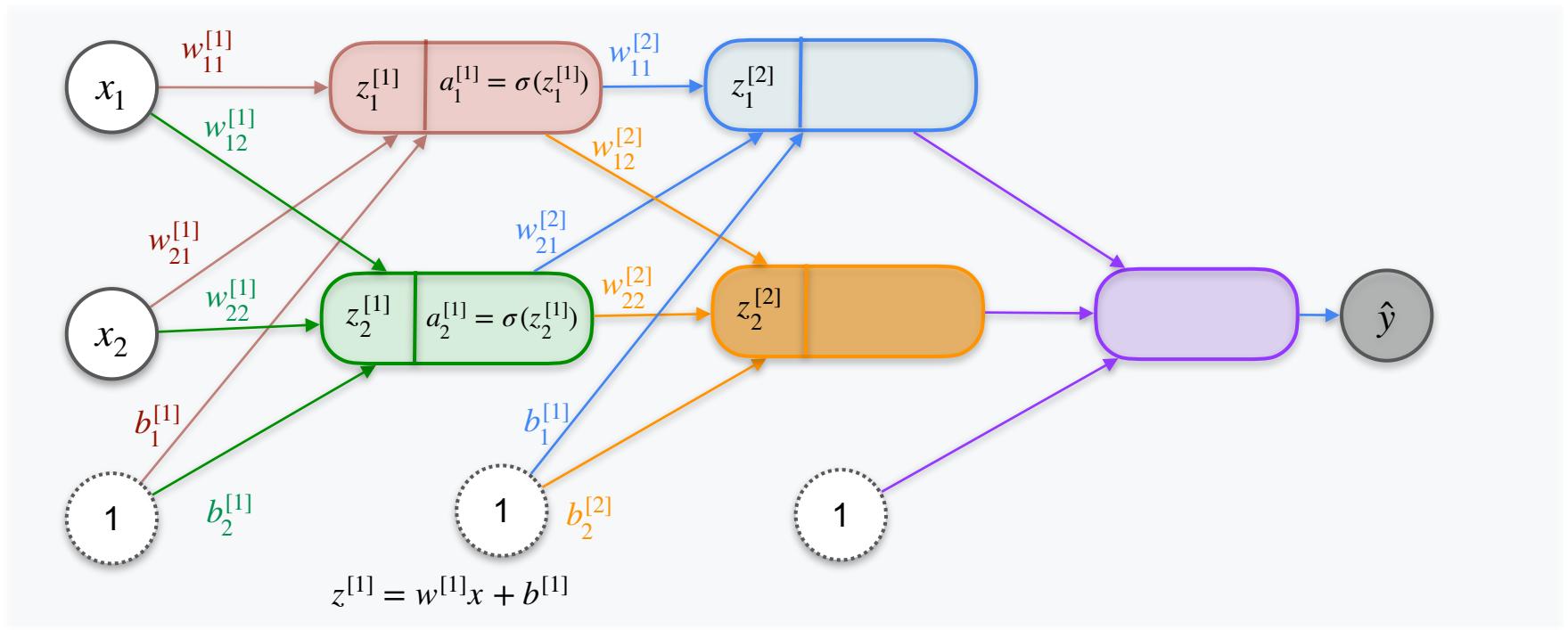
# Back Propagation Introduction



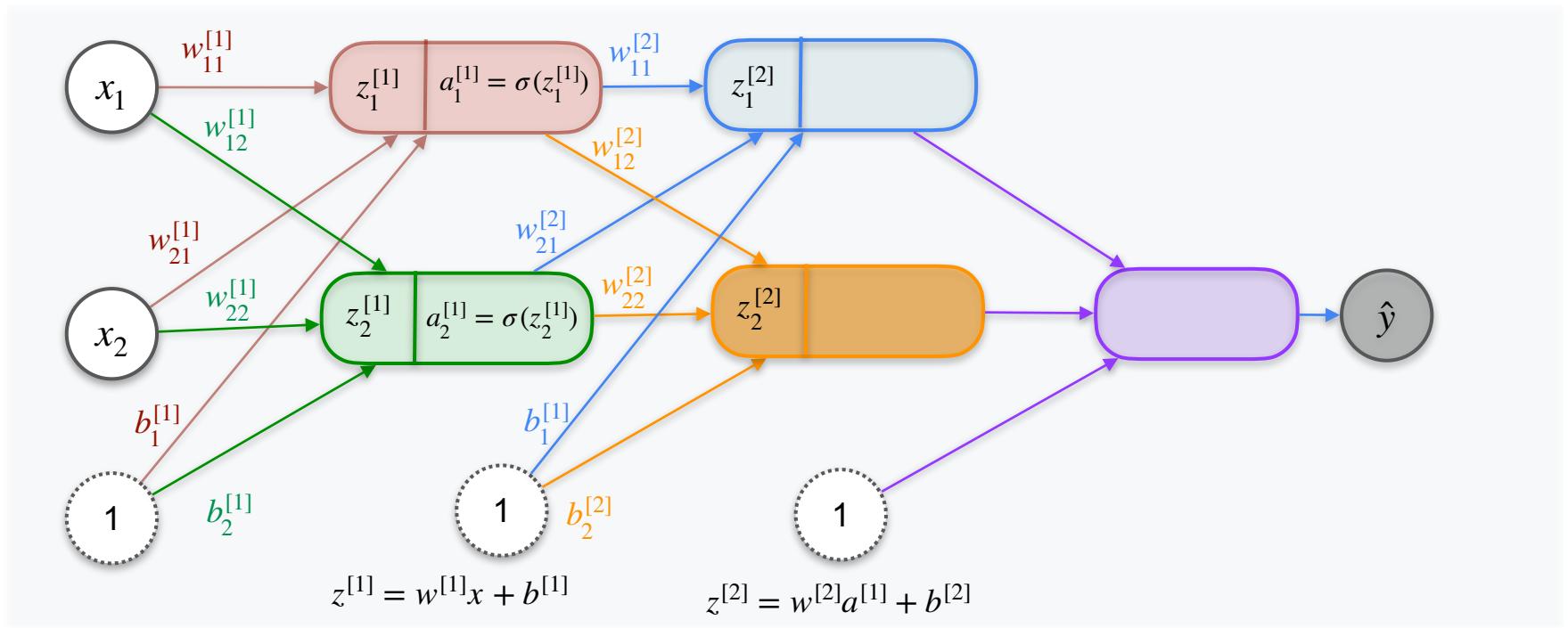
# Back Propagation Introduction



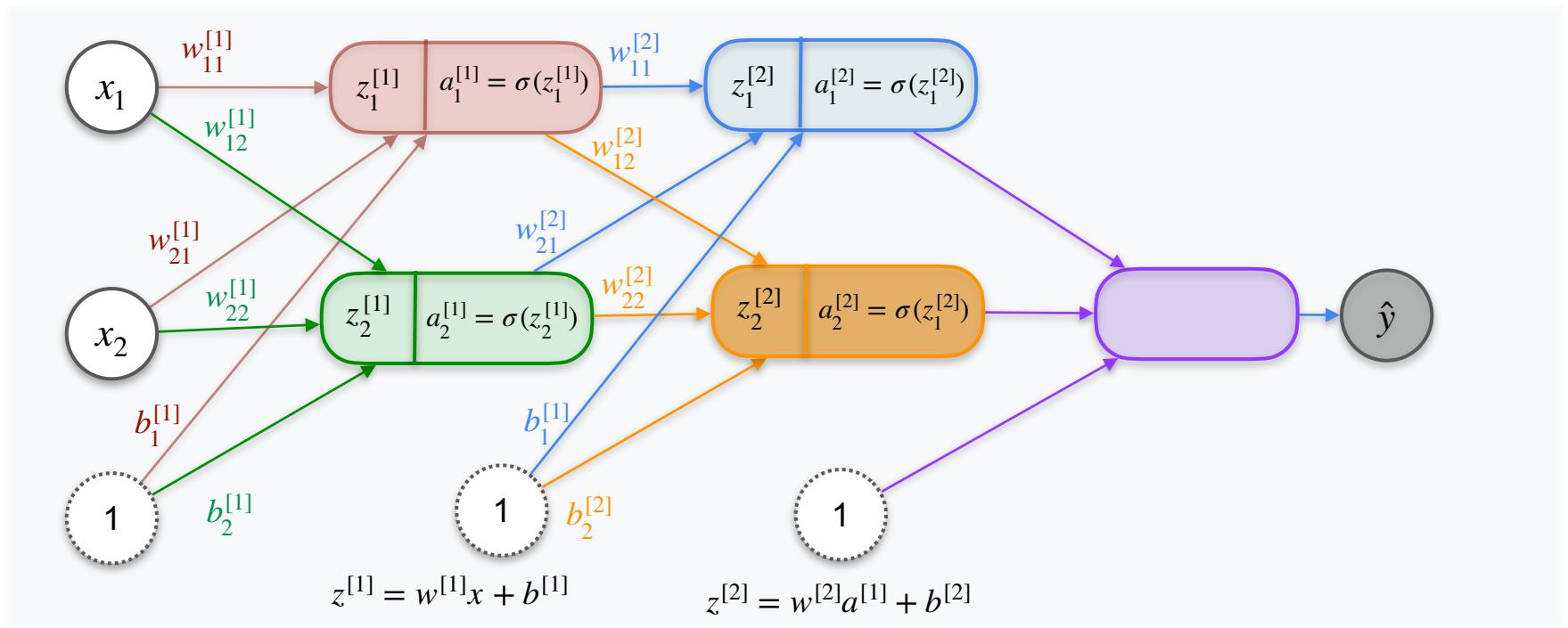
# Back Propagation Introduction



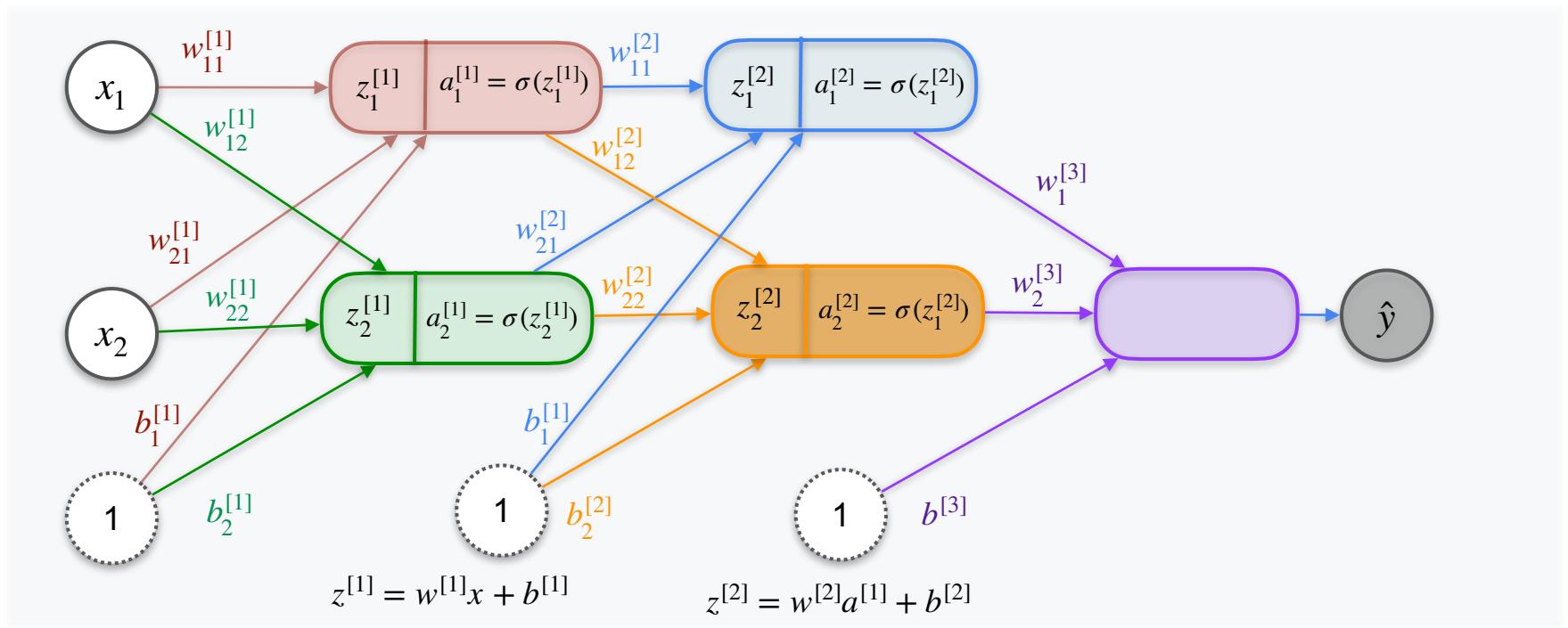
# Back Propagation Introduction



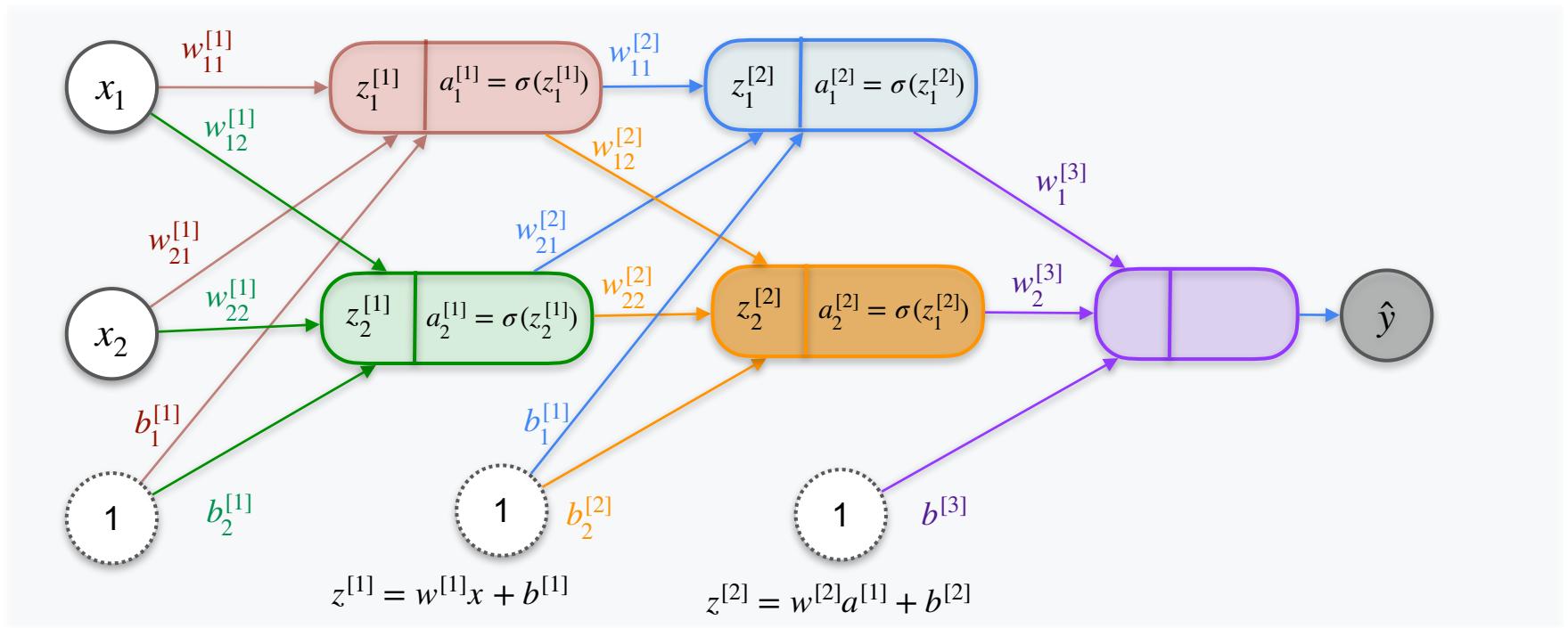
# Back Propagation Introduction



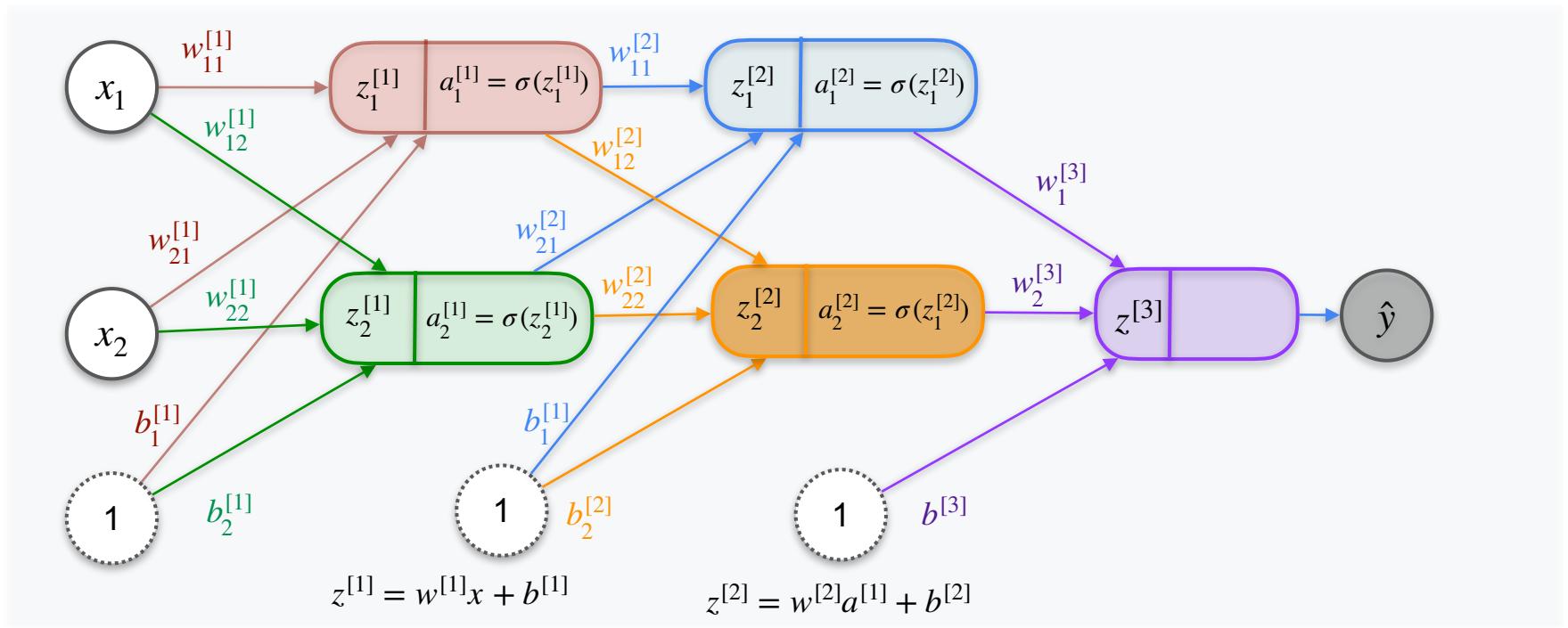
# Back Propagation Introduction



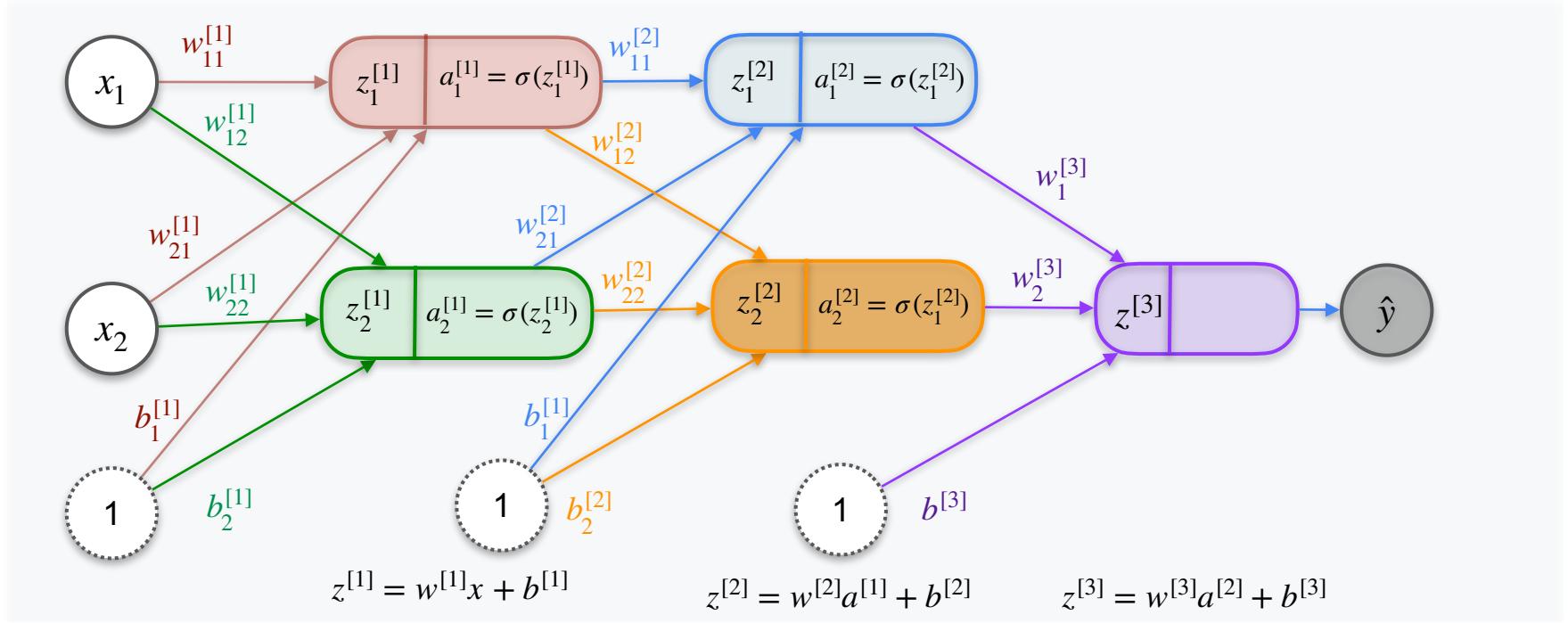
# Back Propagation Introduction



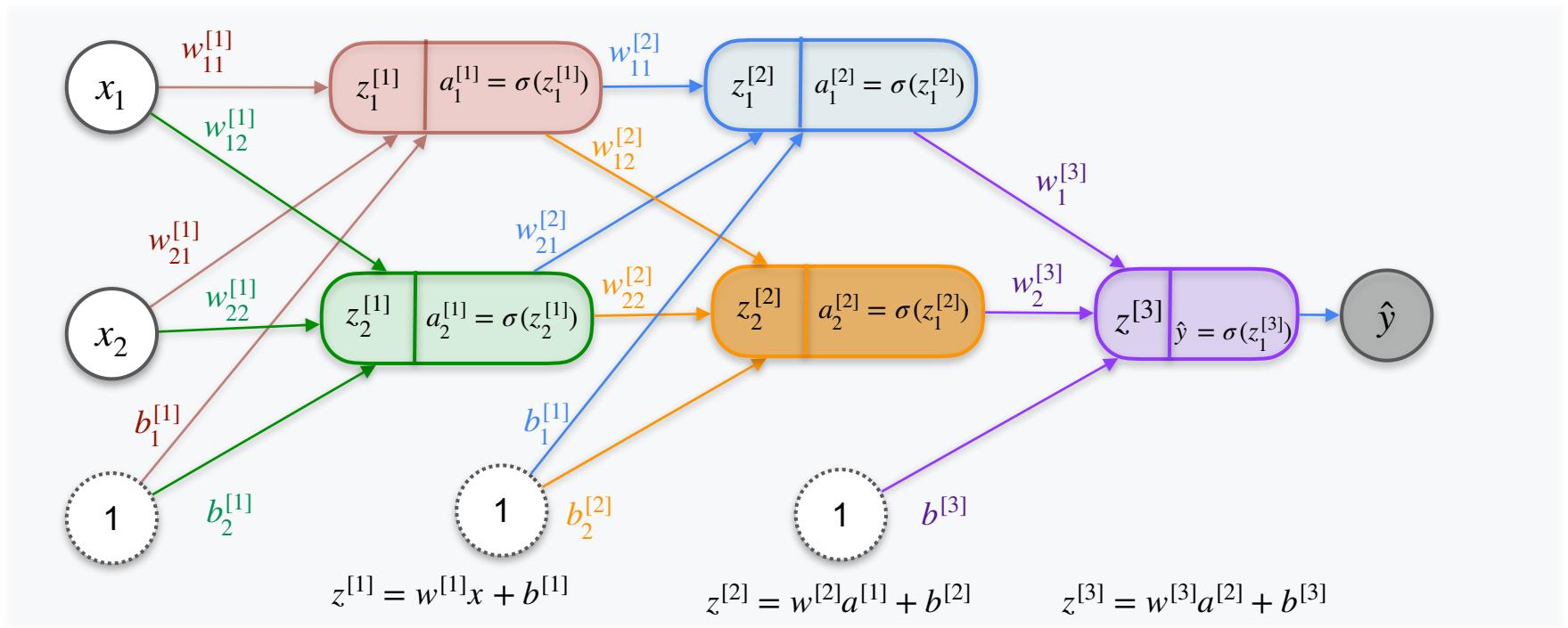
# Back Propagation Introduction



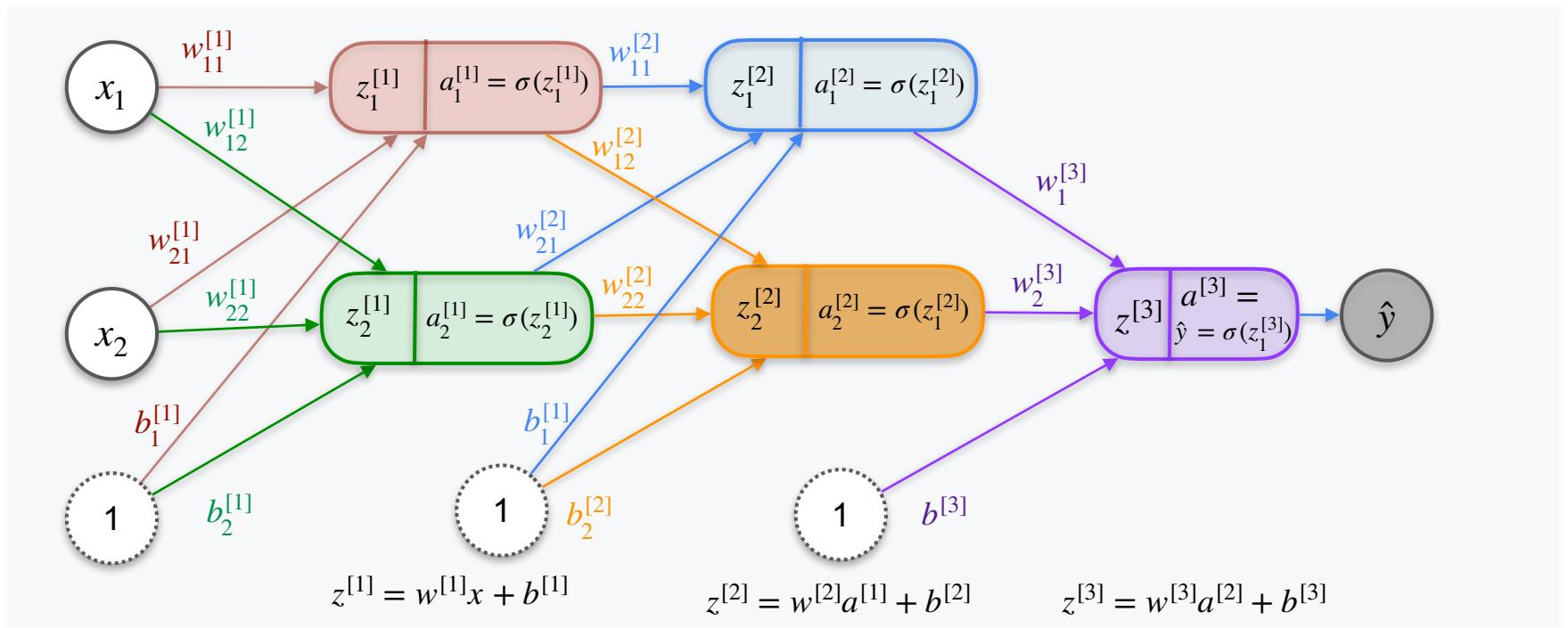
# Back Propagation Introduction



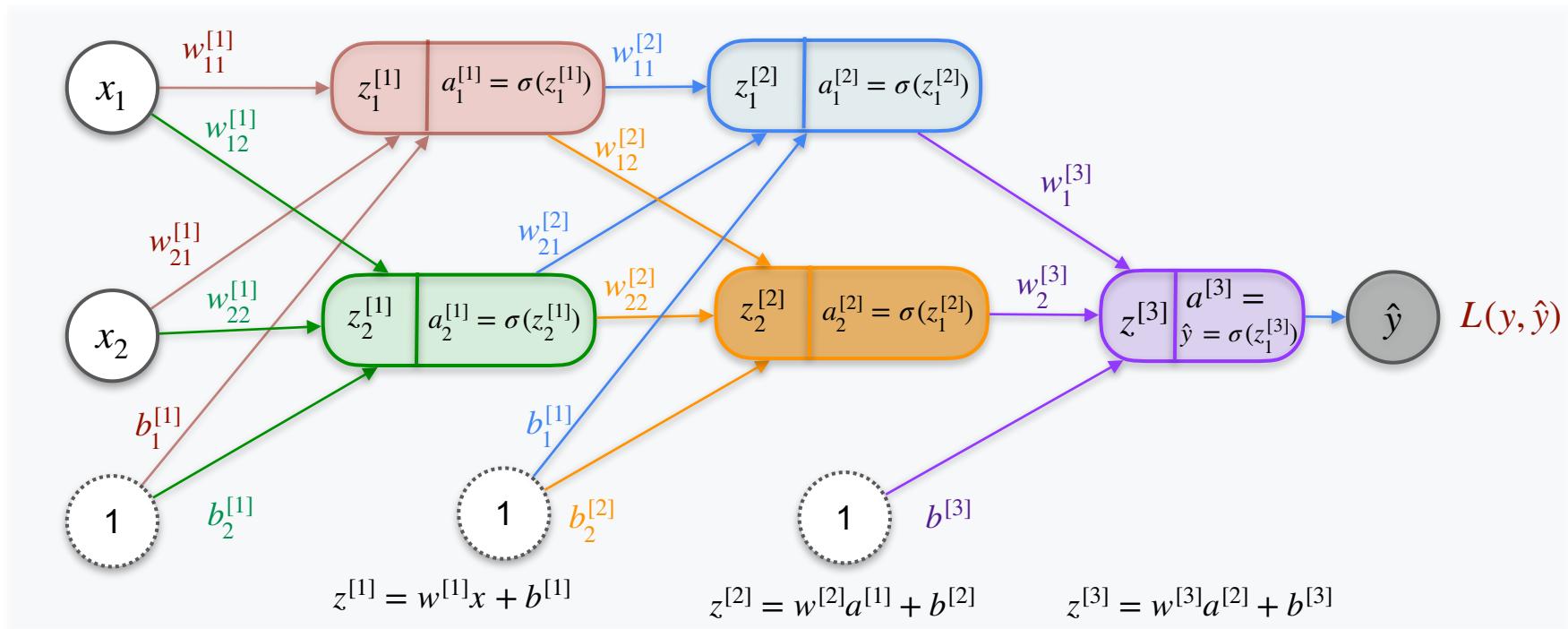
# Back Propagation Introduction



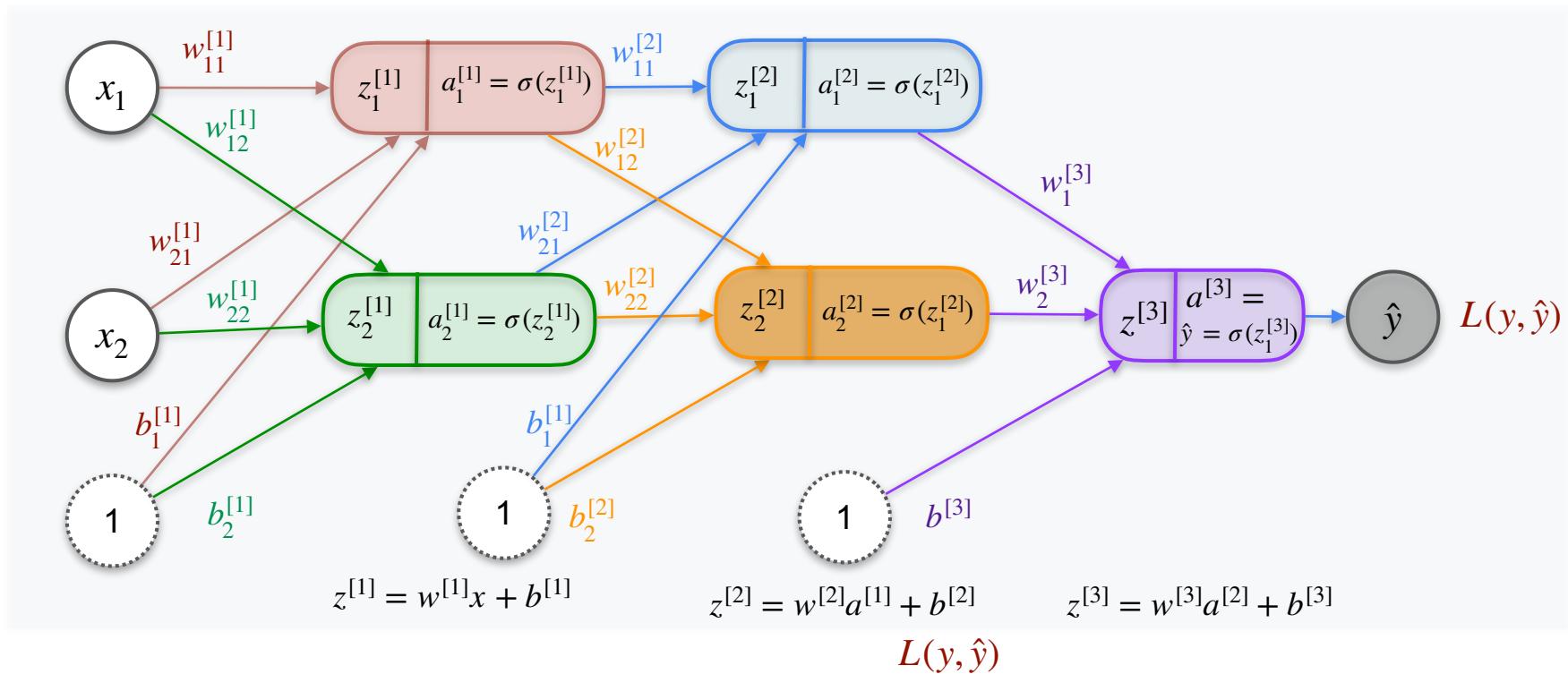
# Back Propagation Introduction



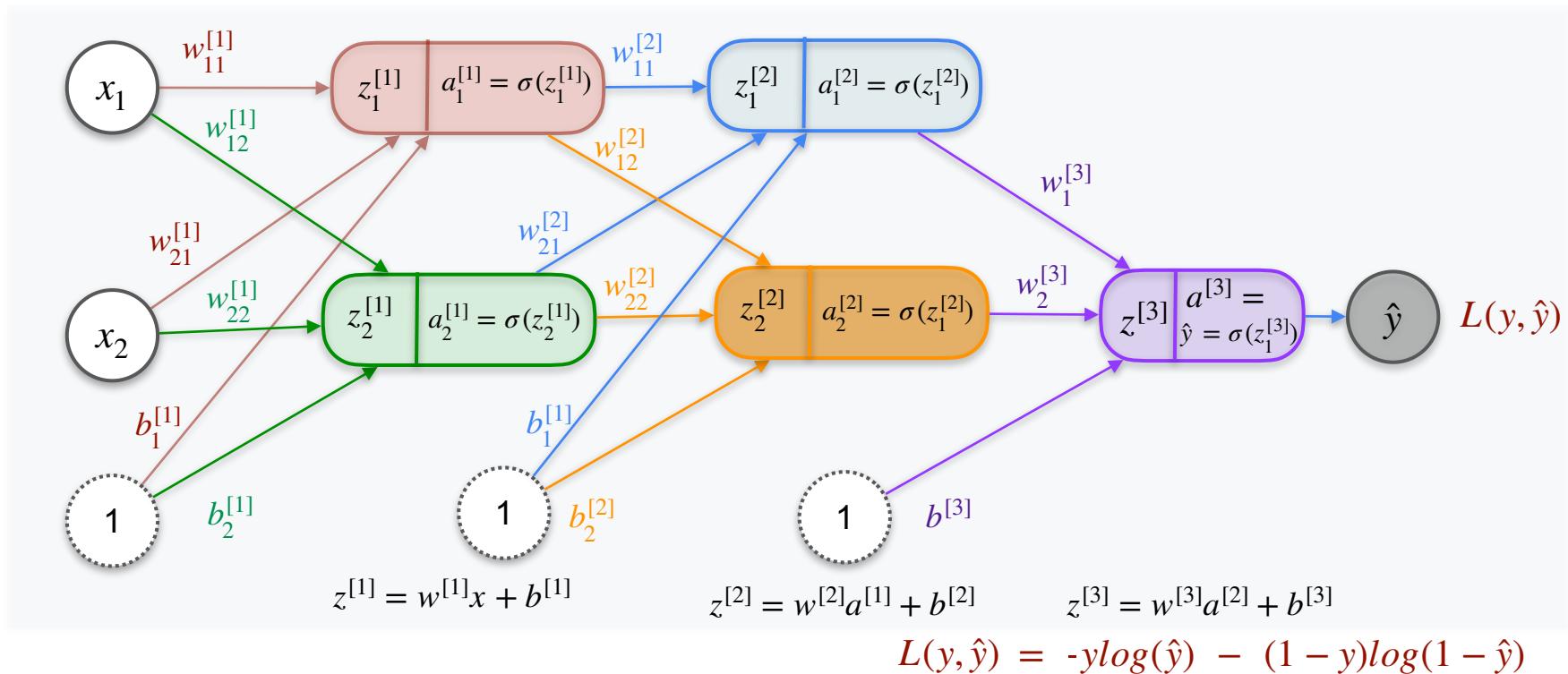
# Back Propagation Introduction



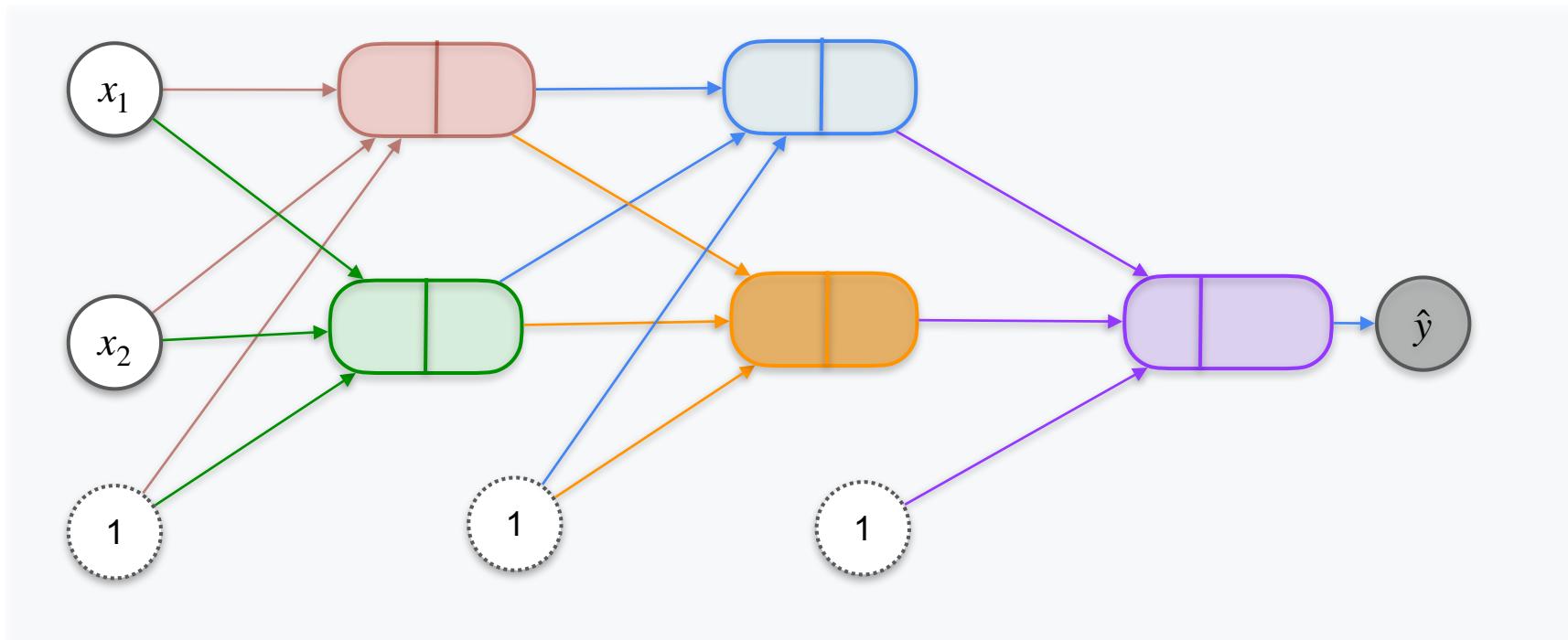
# Back Propagation Introduction



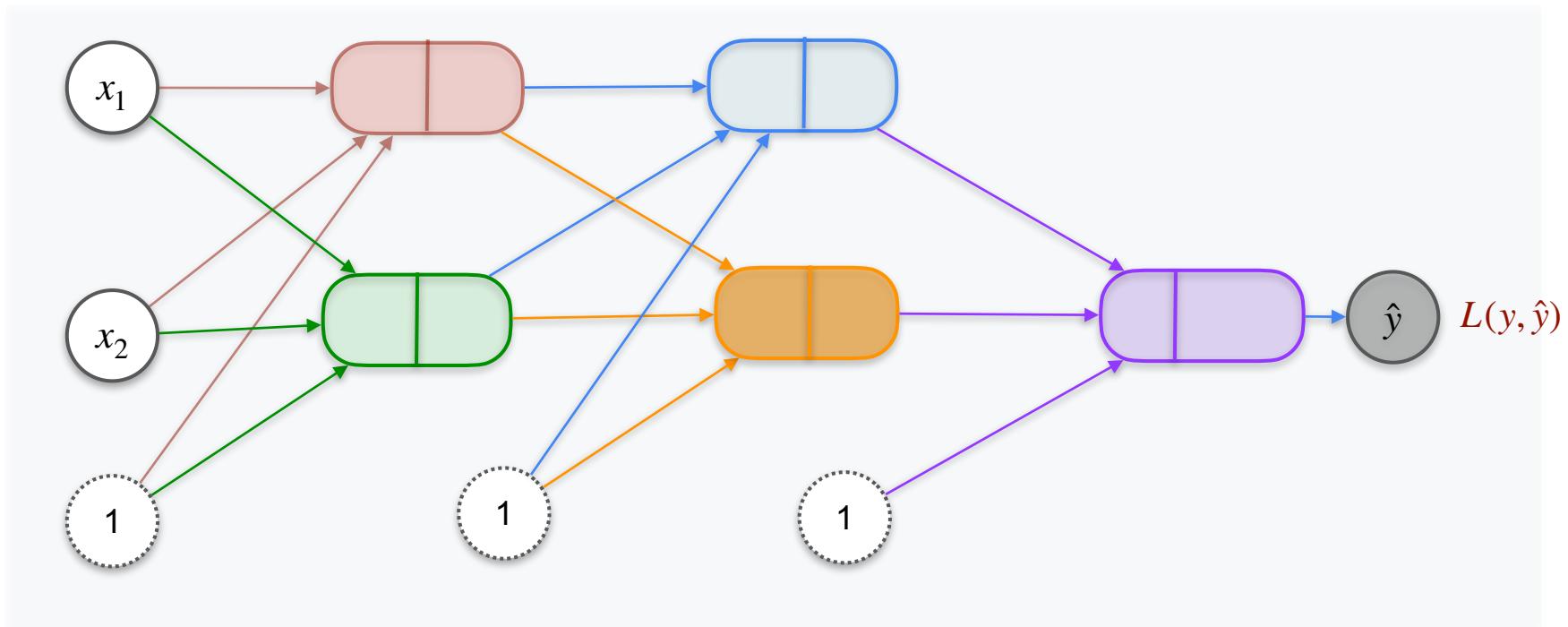
# Back Propagation Introduction



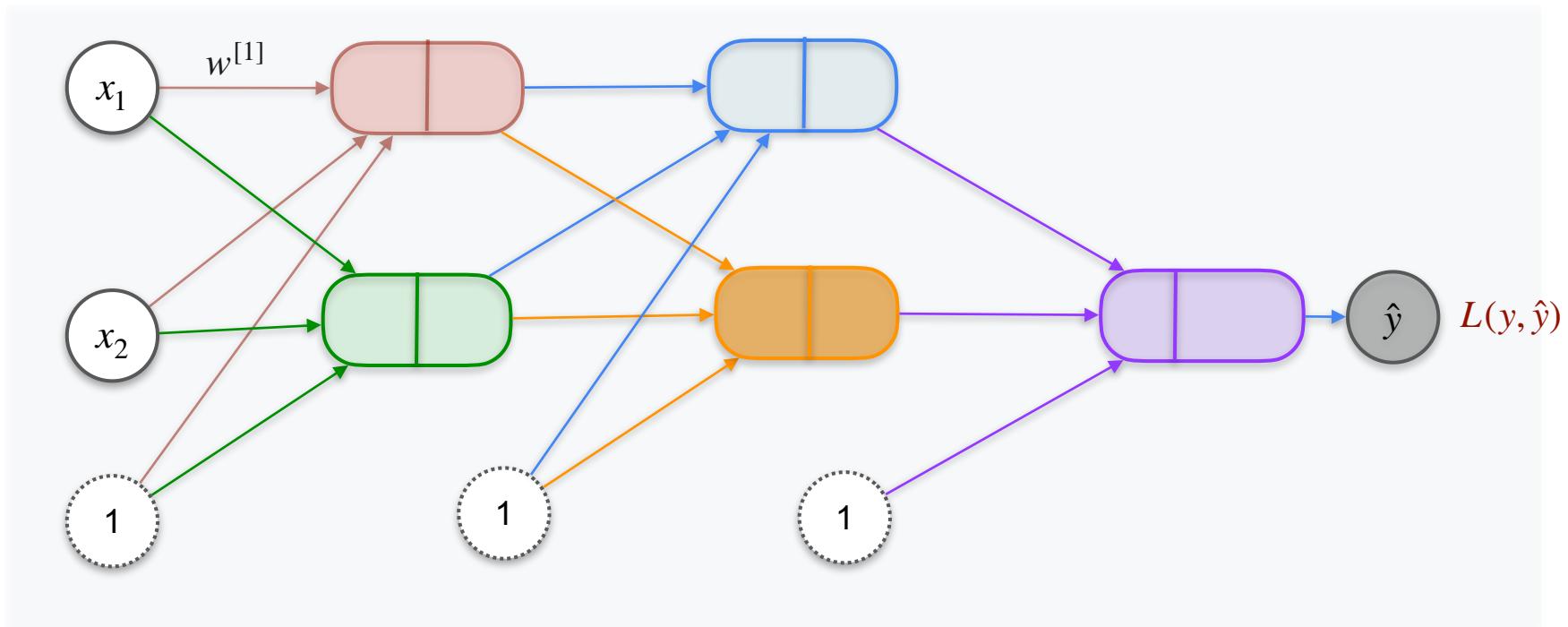
# Back Propagation Introduction



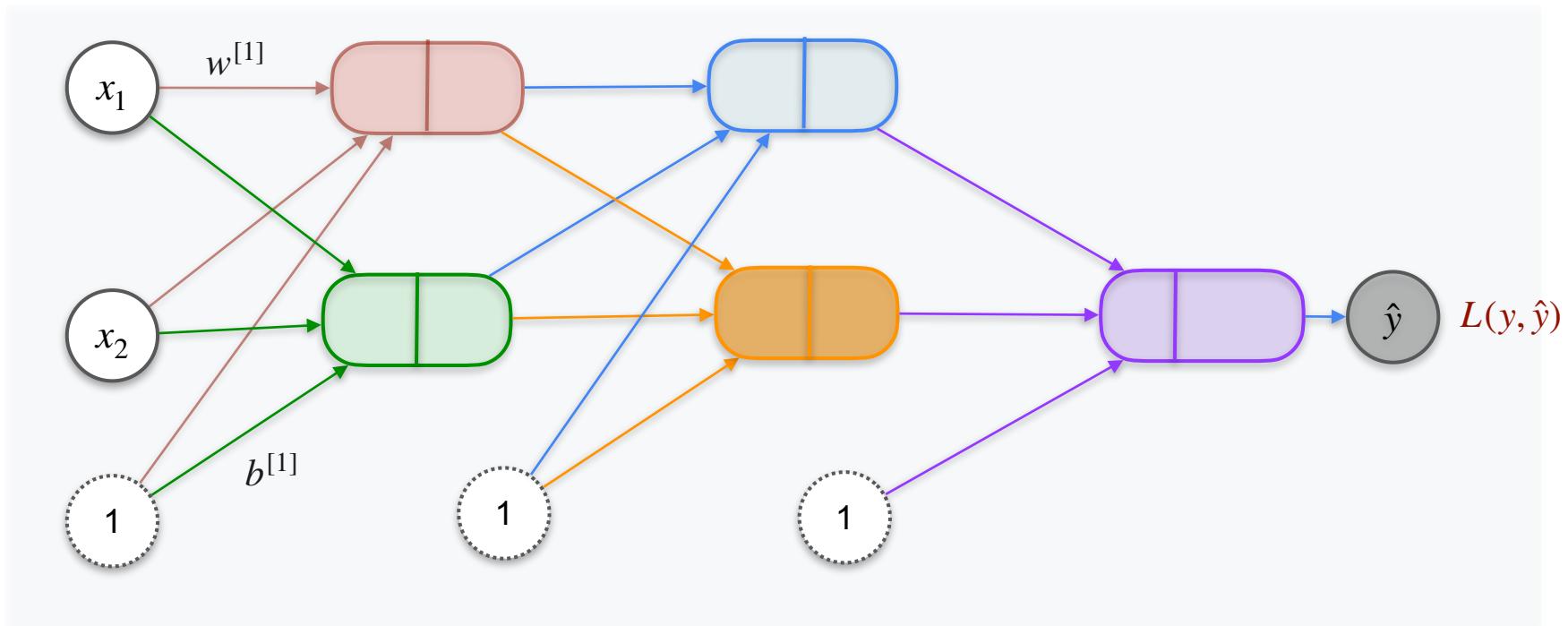
# Back Propagation Introduction



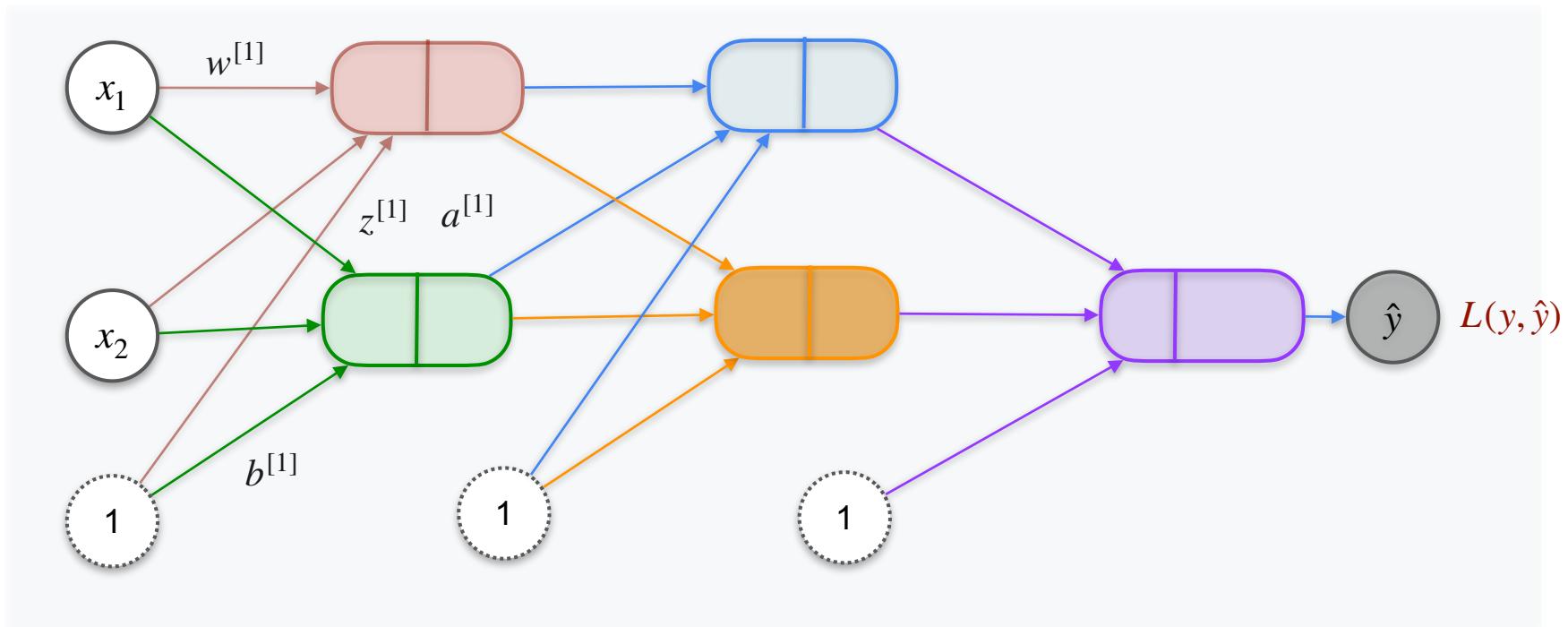
# Back Propagation Introduction



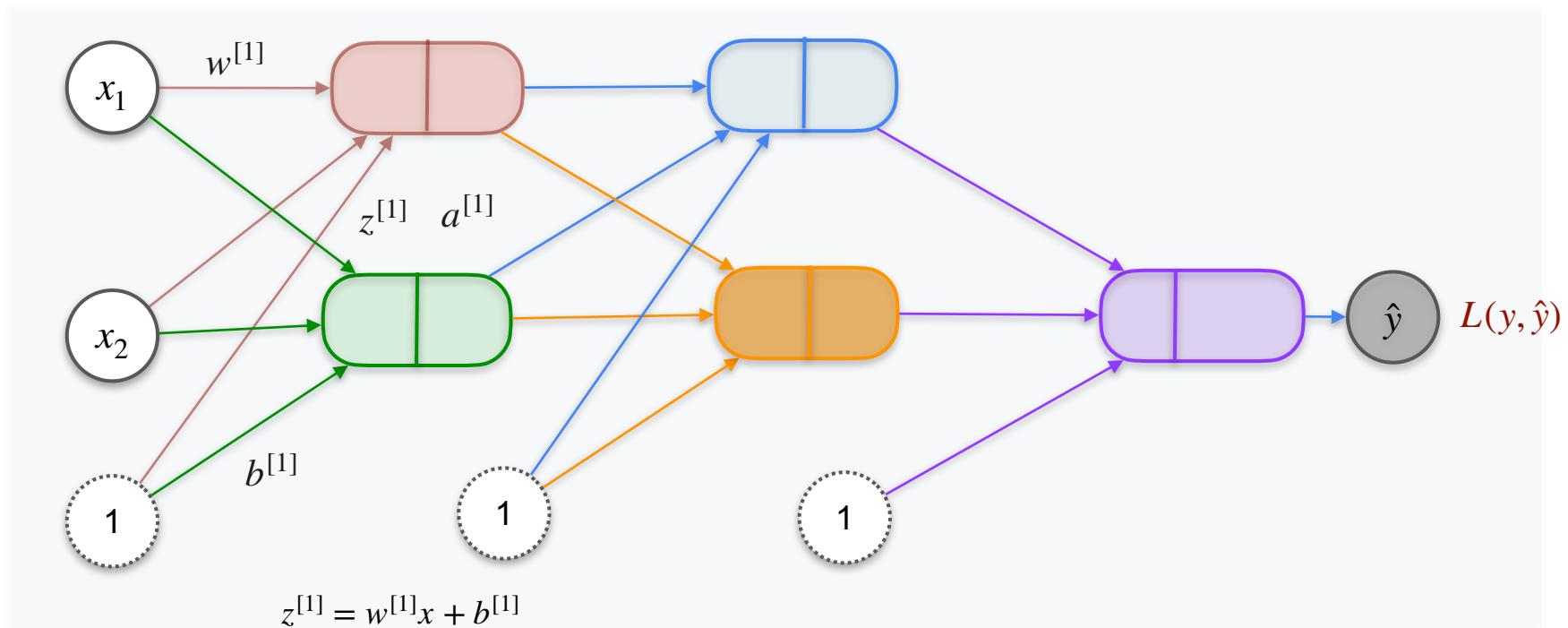
# Back Propagation Introduction



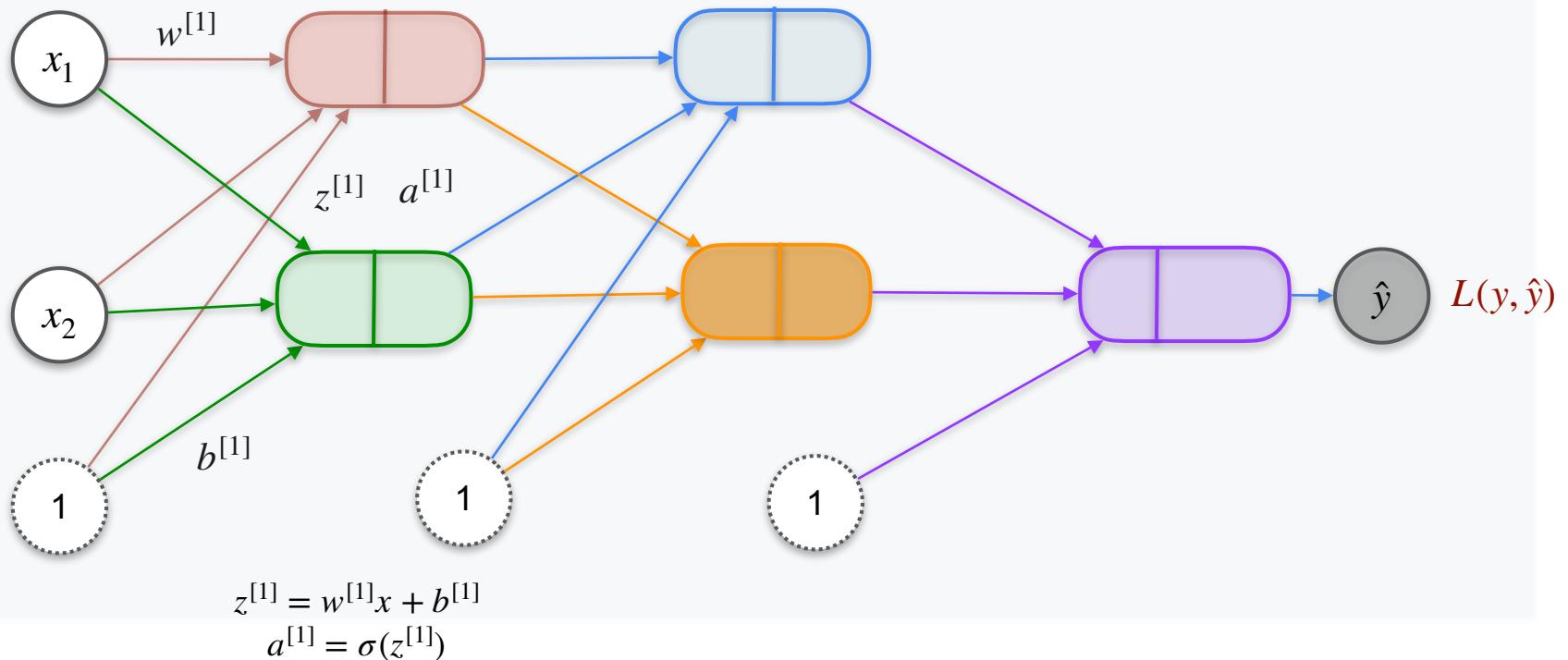
# Back Propagation Introduction



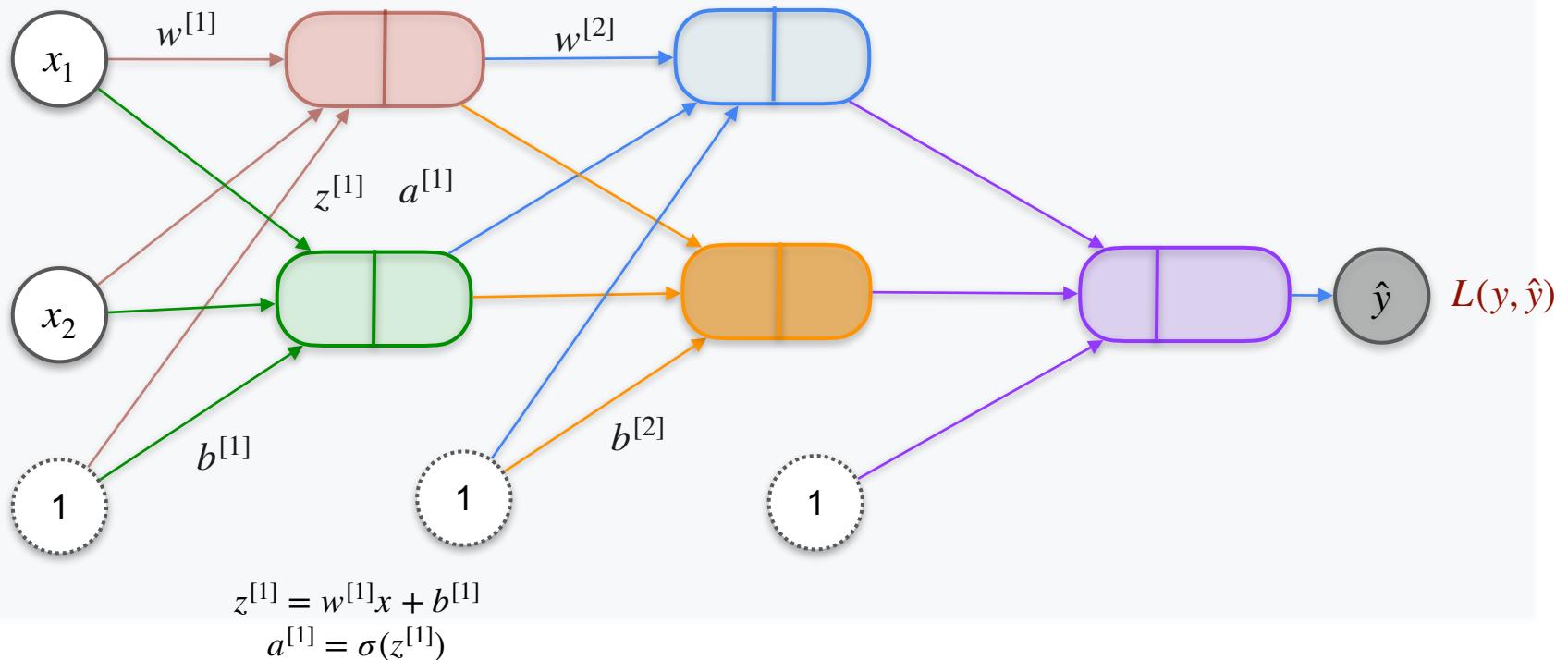
# Back Propagation Introduction



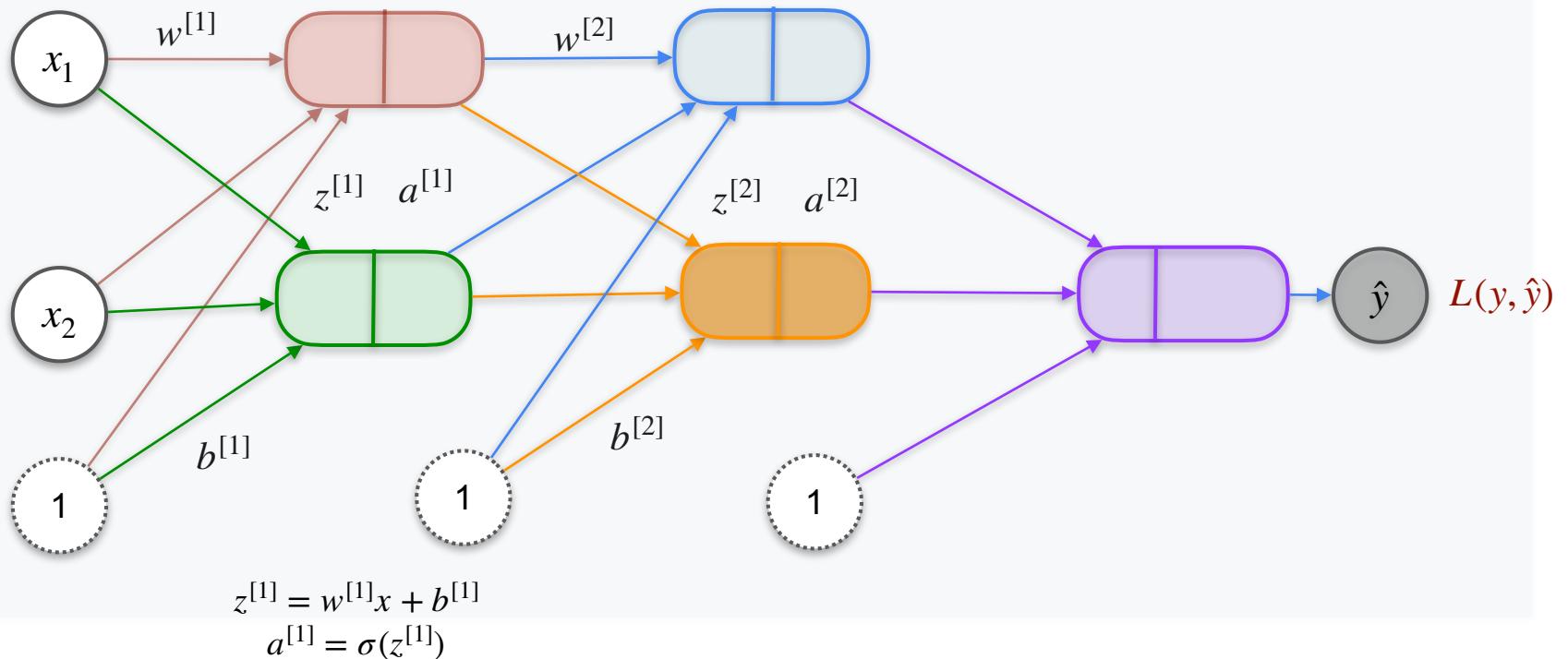
# Back Propagation Introduction



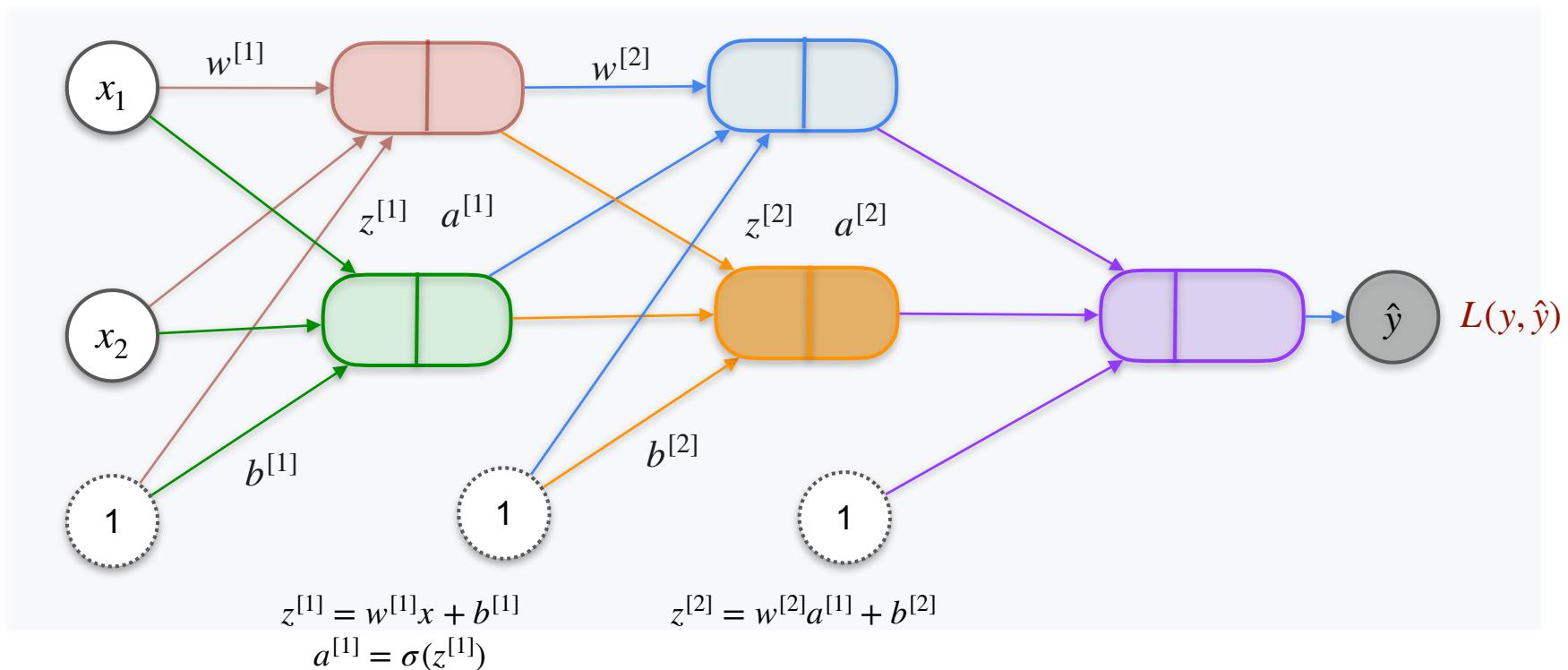
# Back Propagation Introduction



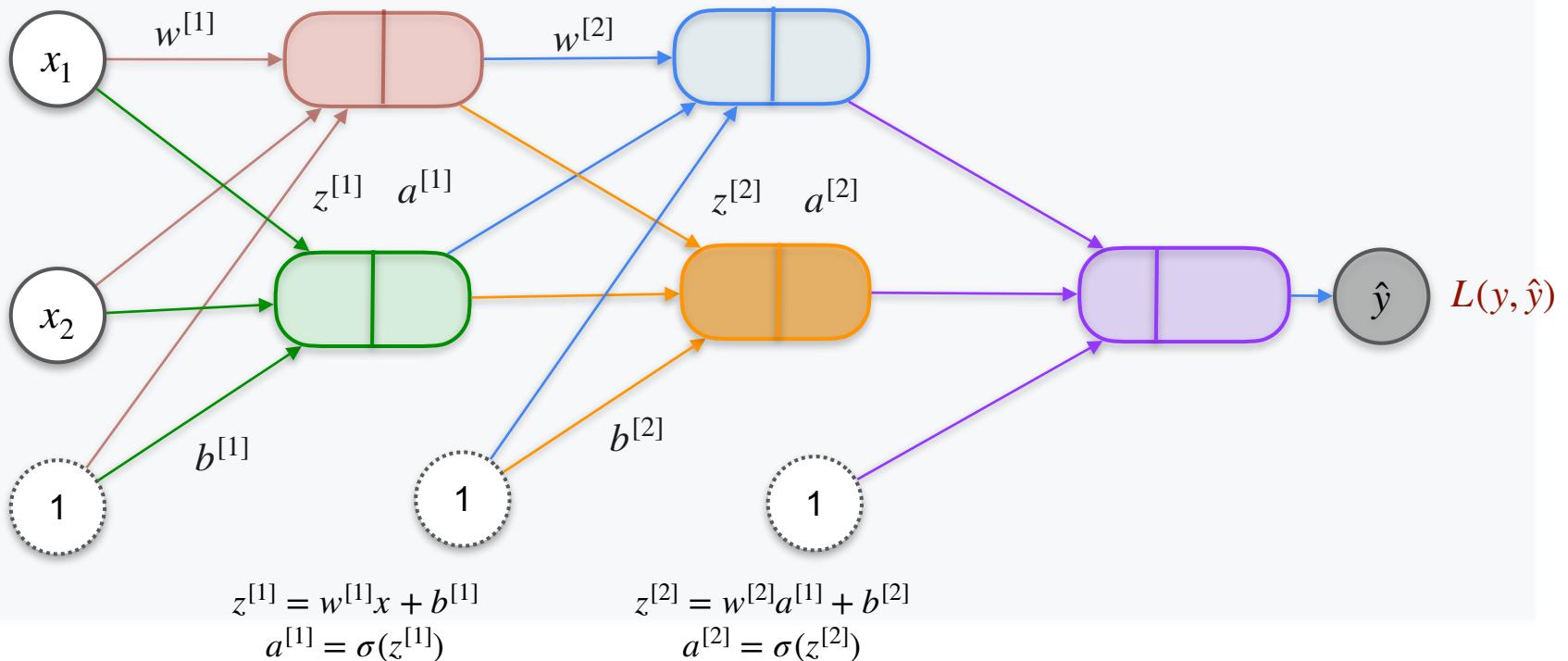
# Back Propagation Introduction



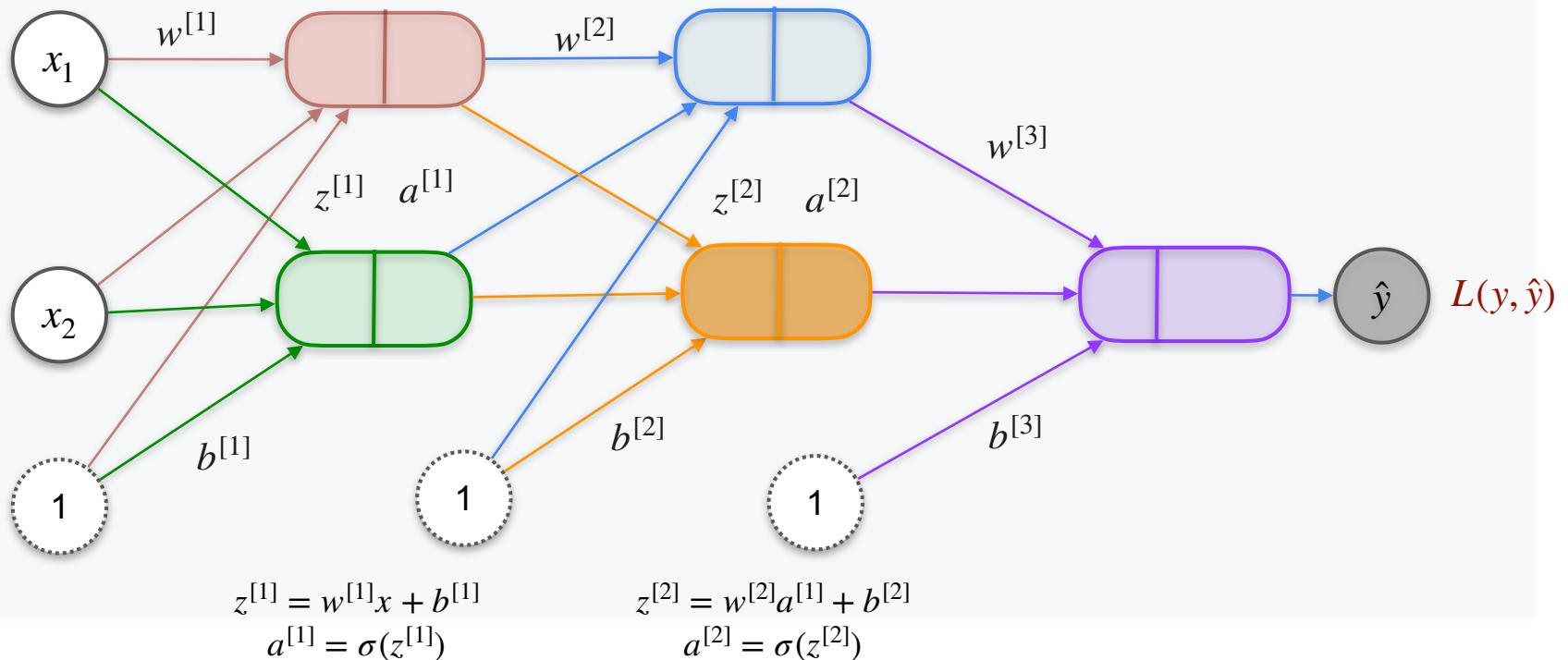
# Back Propagation Introduction



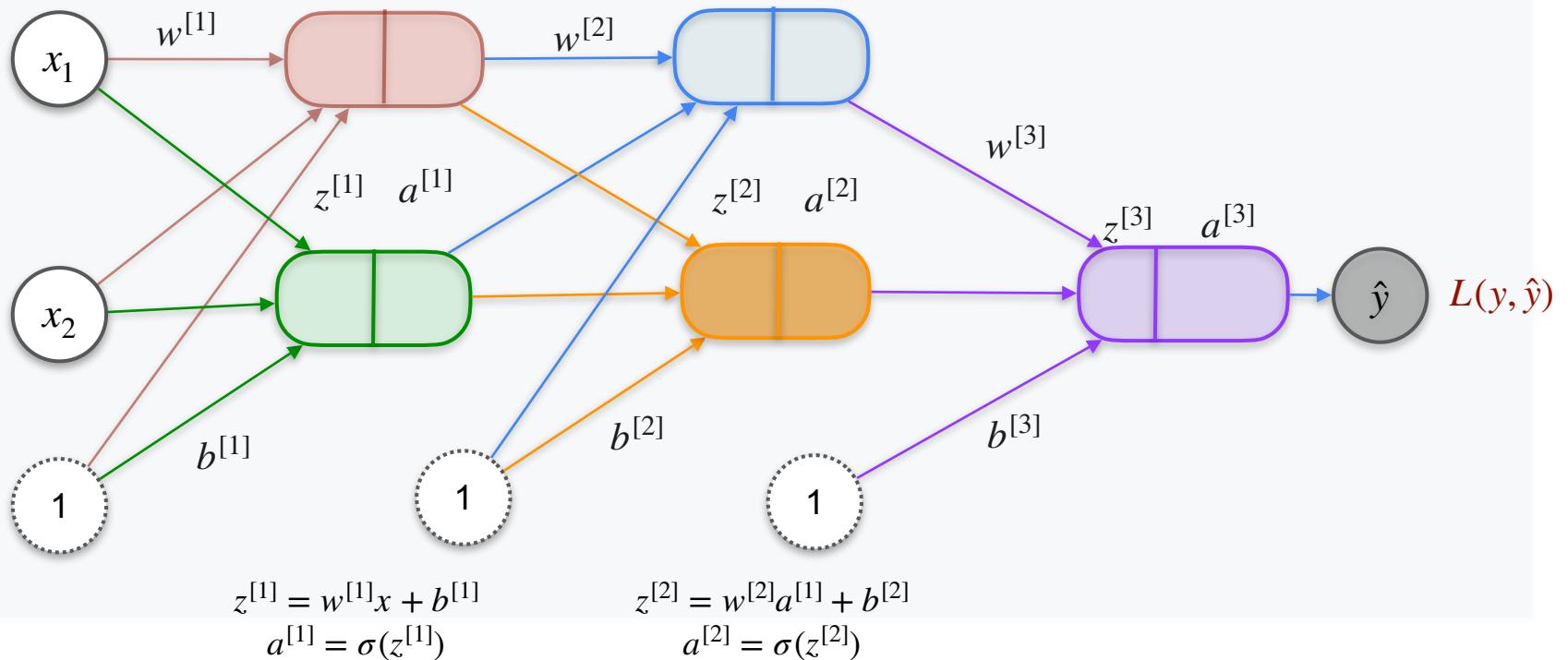
# Back Propagation Introduction



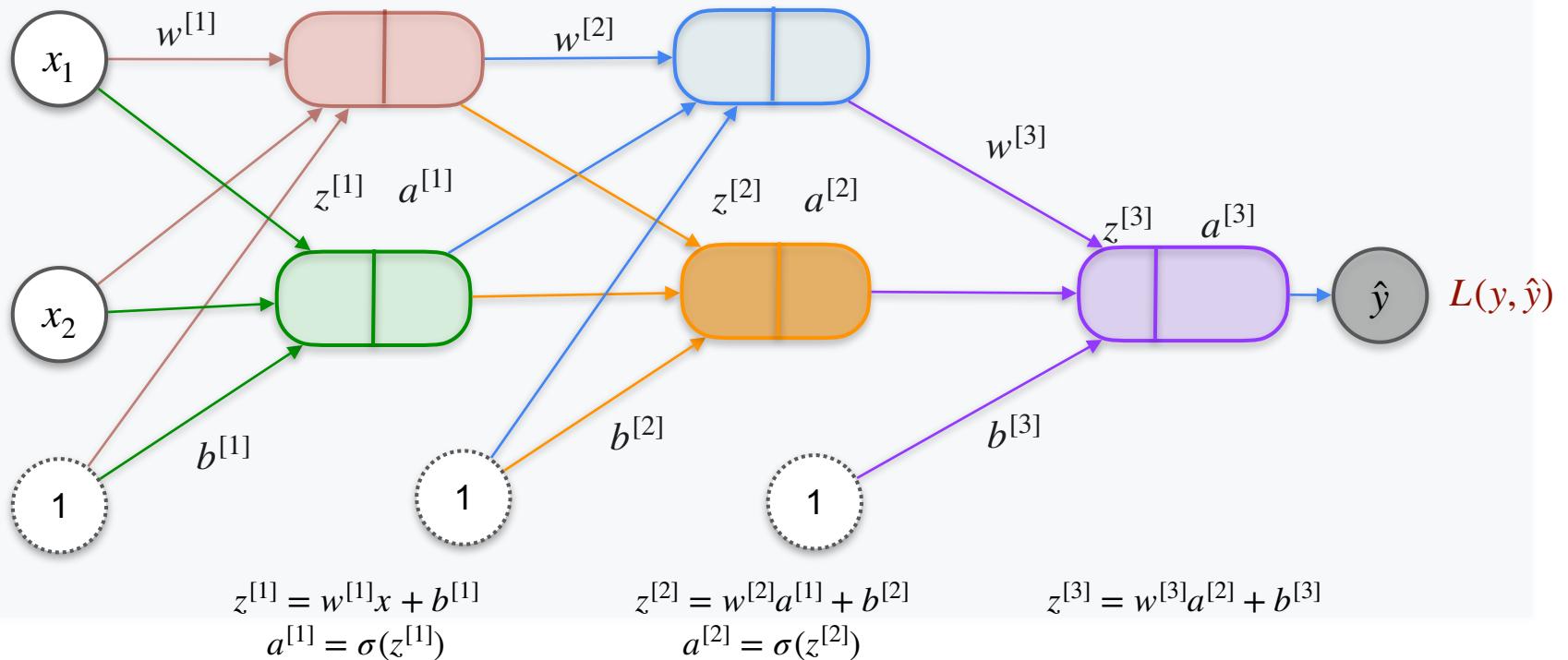
# Back Propagation Introduction



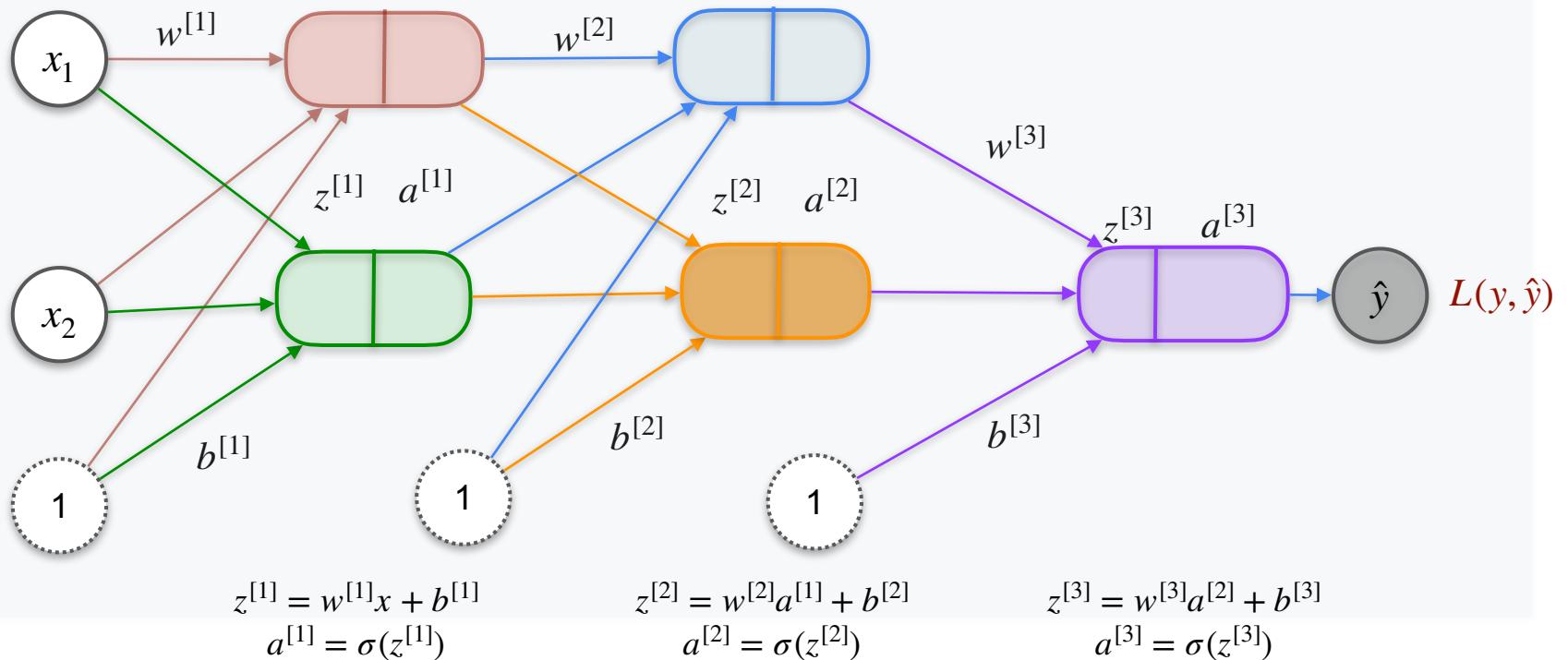
# Back Propagation Introduction



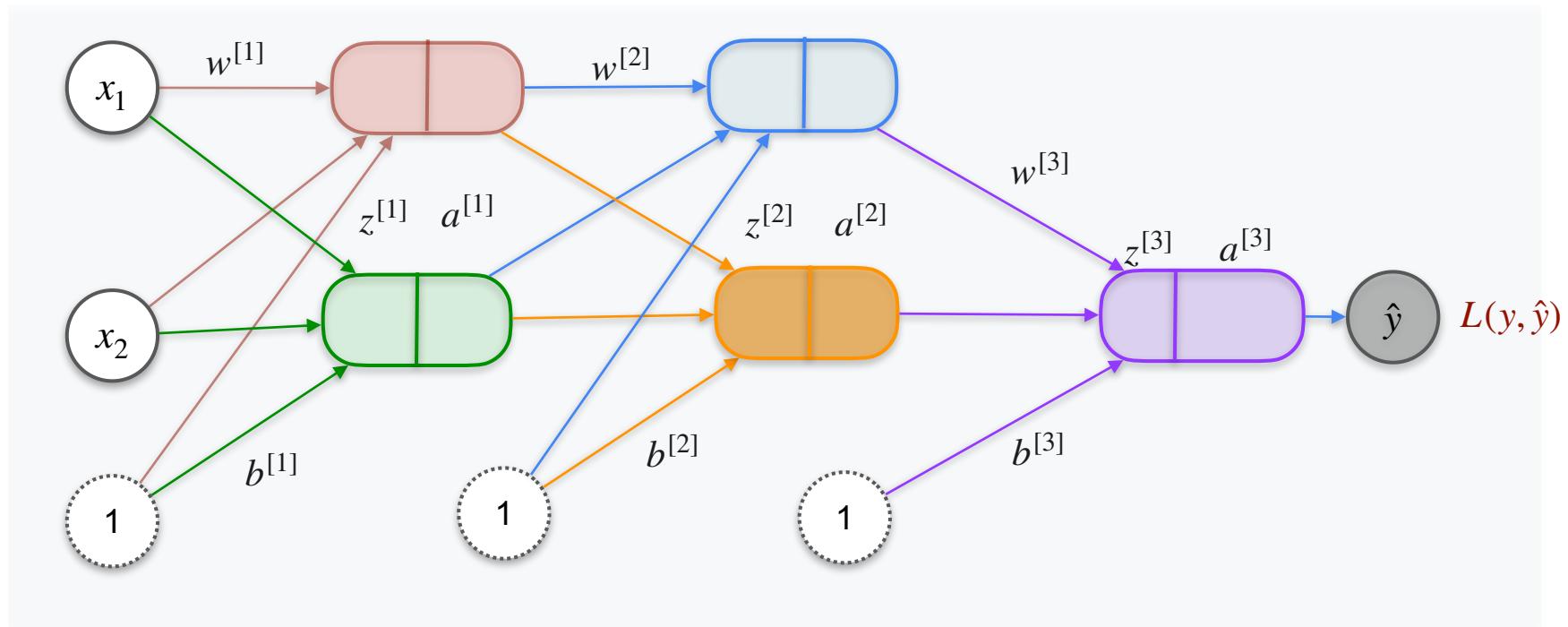
# Back Propagation Introduction



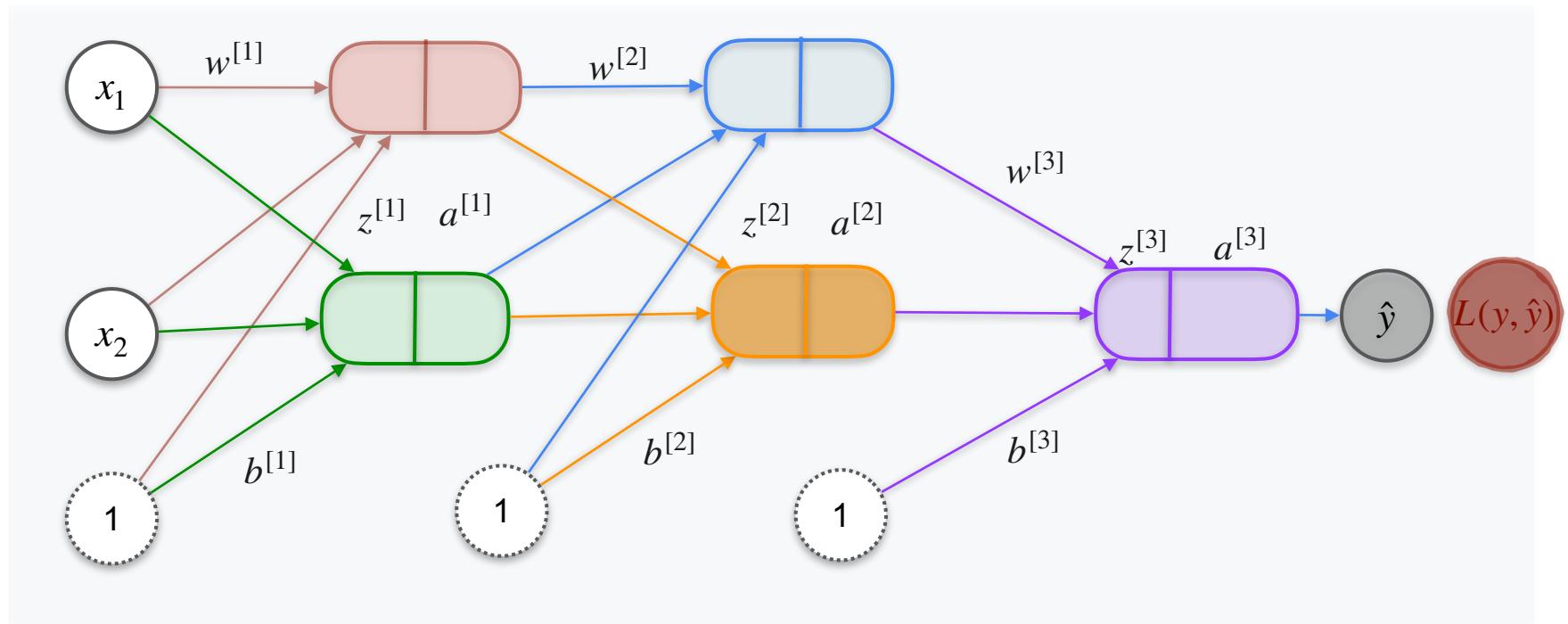
# Back Propagation Introduction



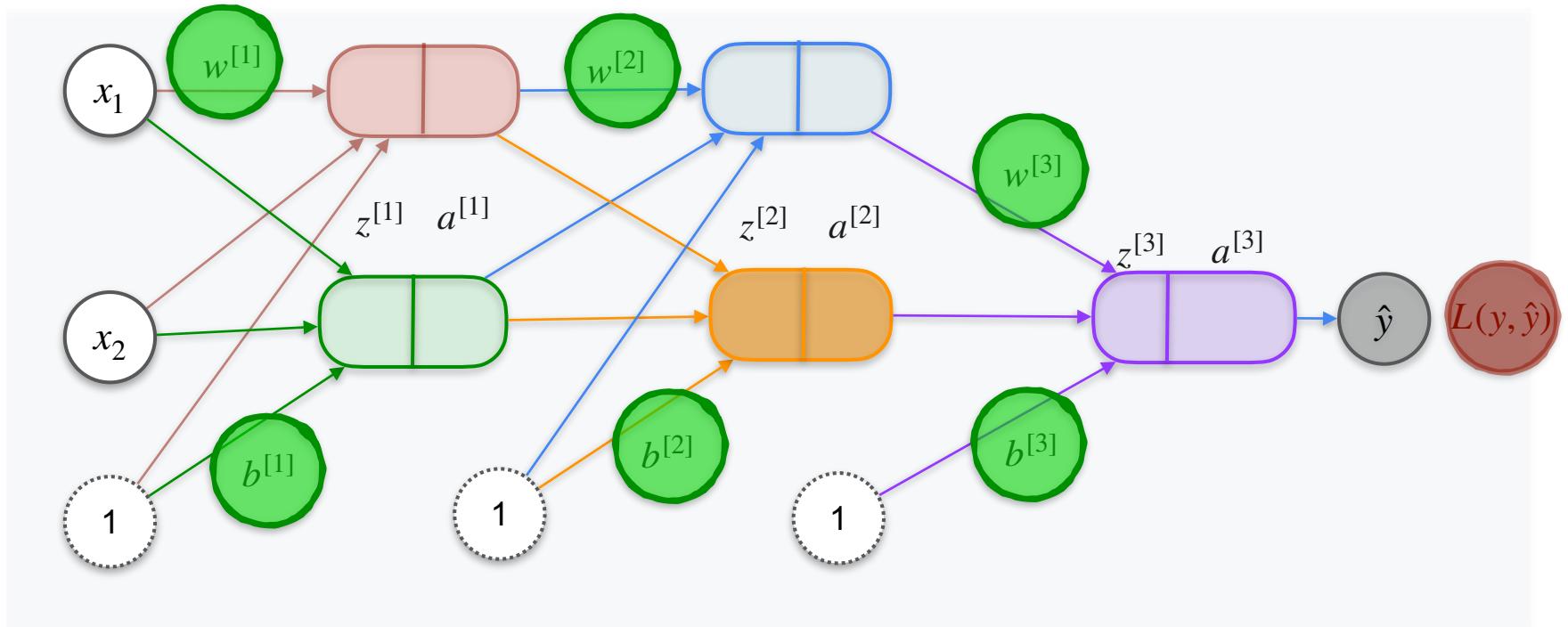
# Back Propagation Introduction



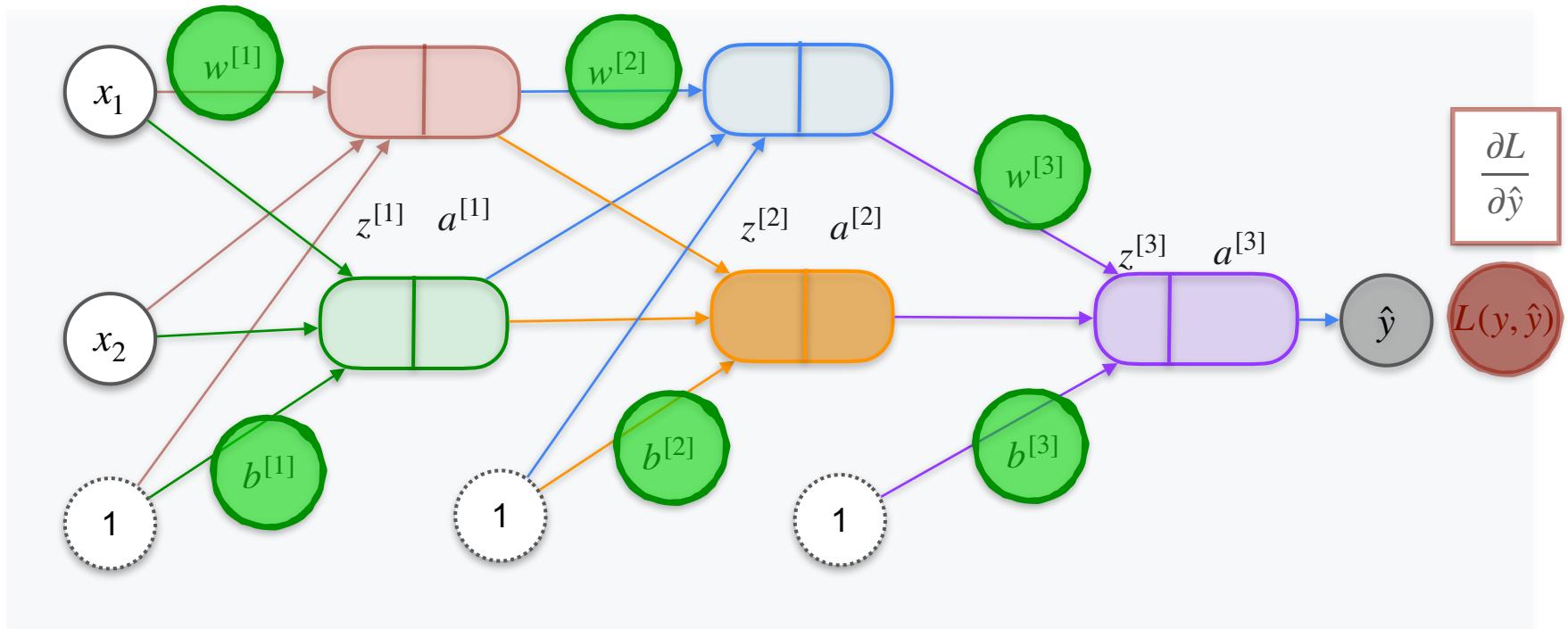
# Back Propagation Introduction



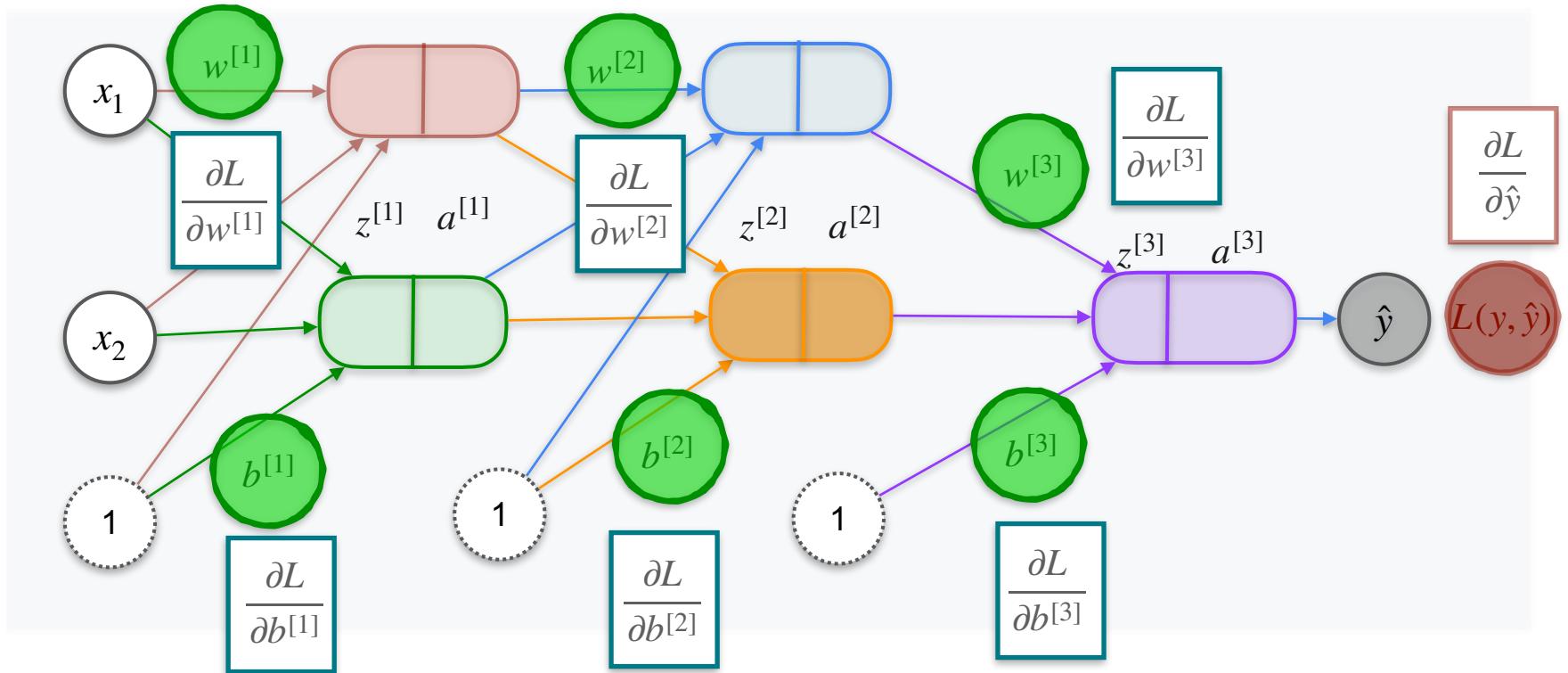
# Back Propagation Introduction



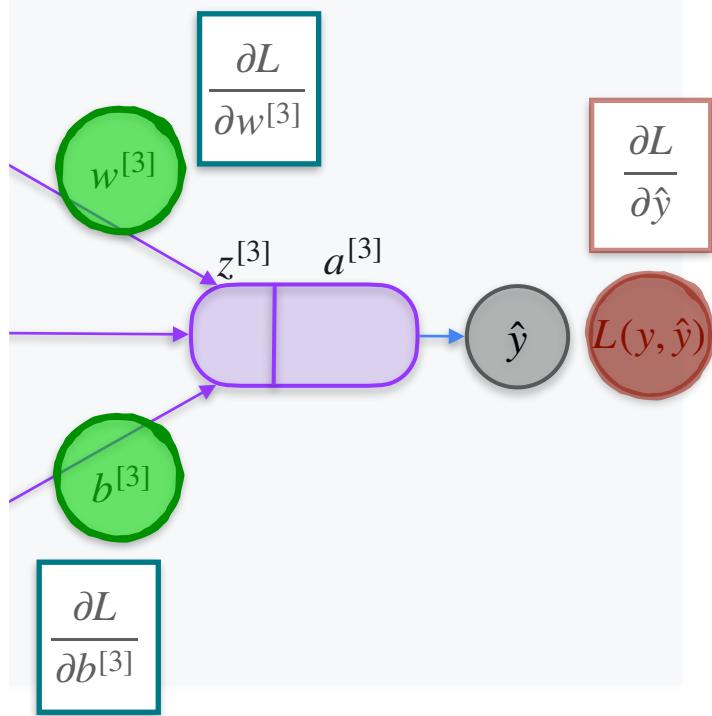
# Back Propagation Introduction



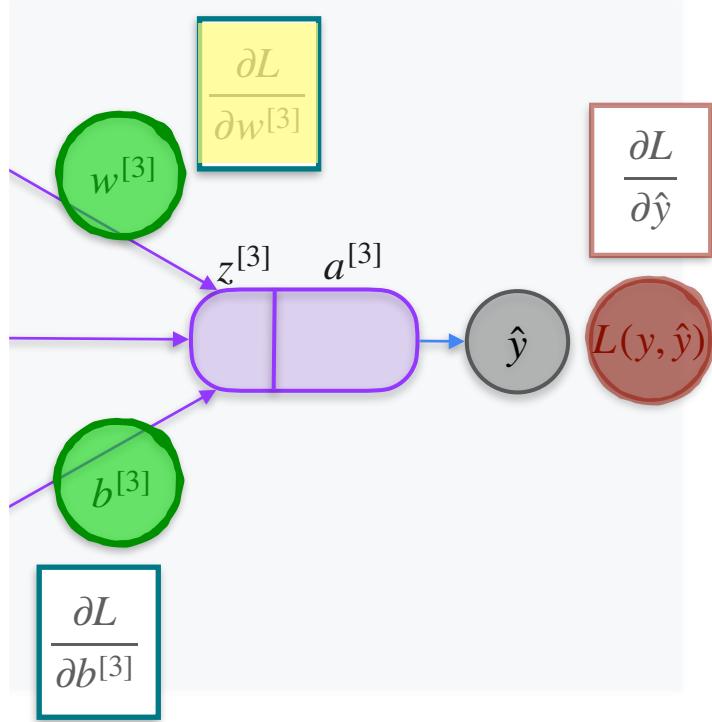
# Back Propagation Introduction



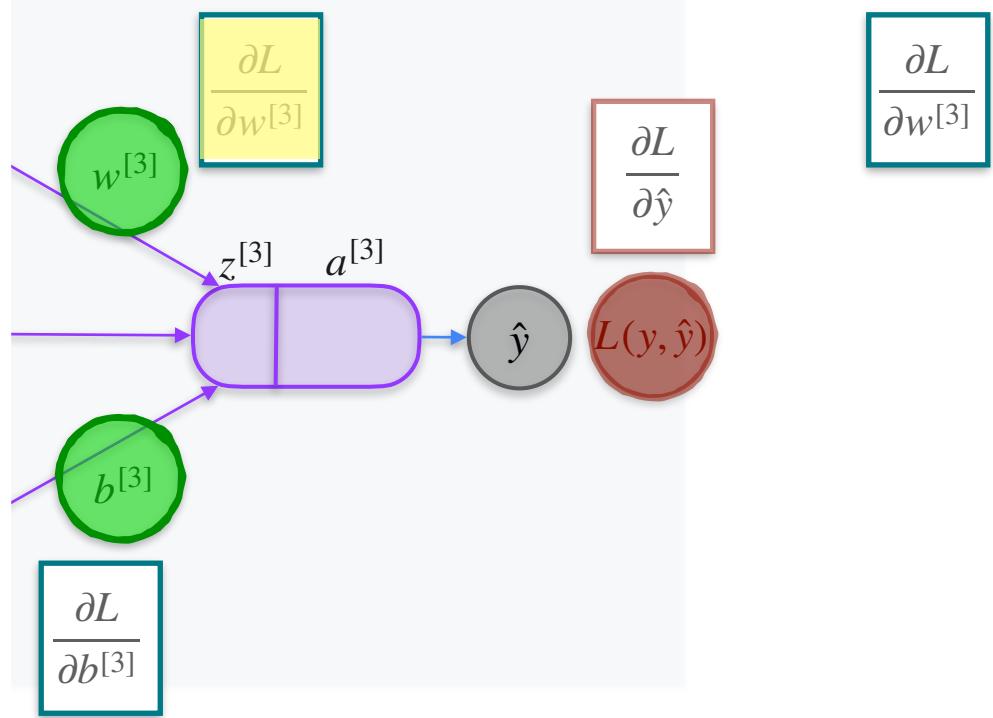
# Back Propagation Introduction



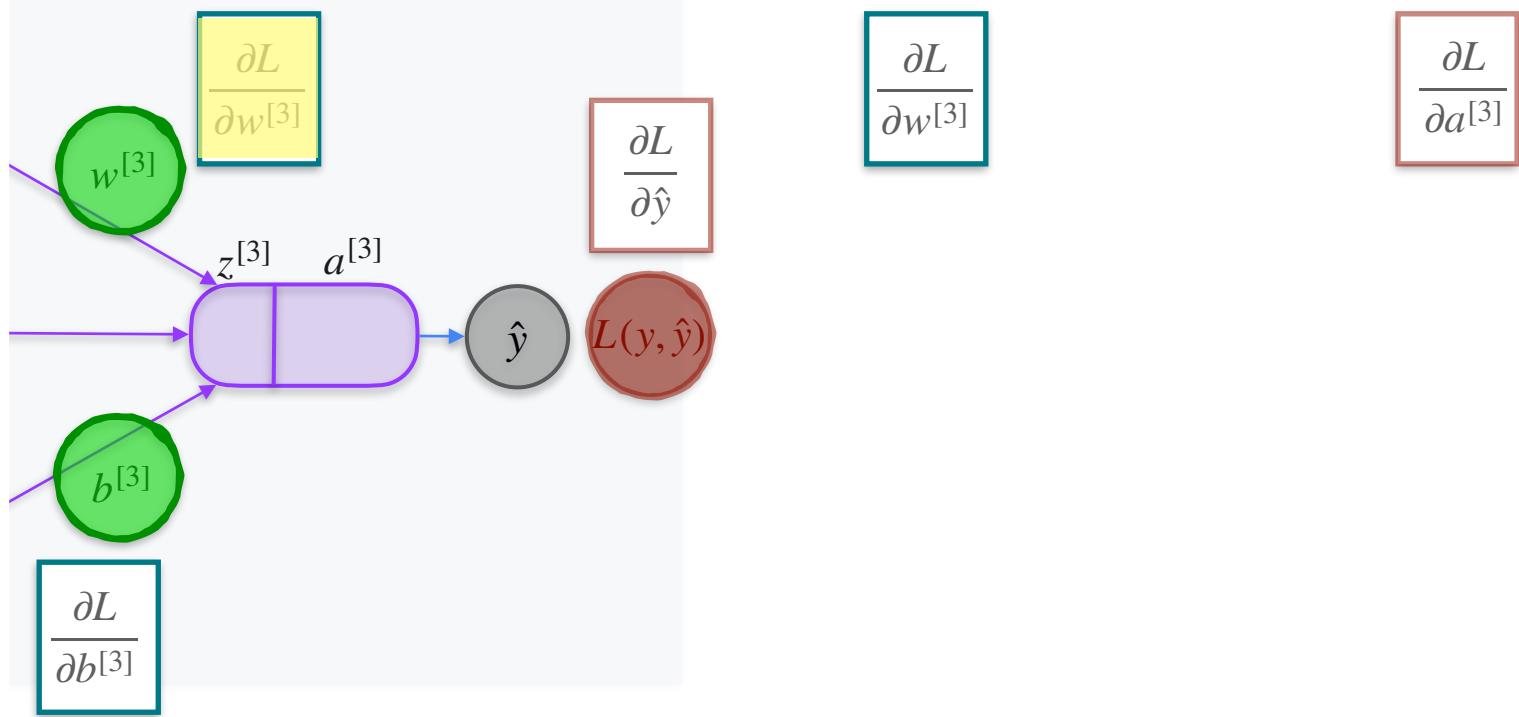
# Back Propagation Introduction



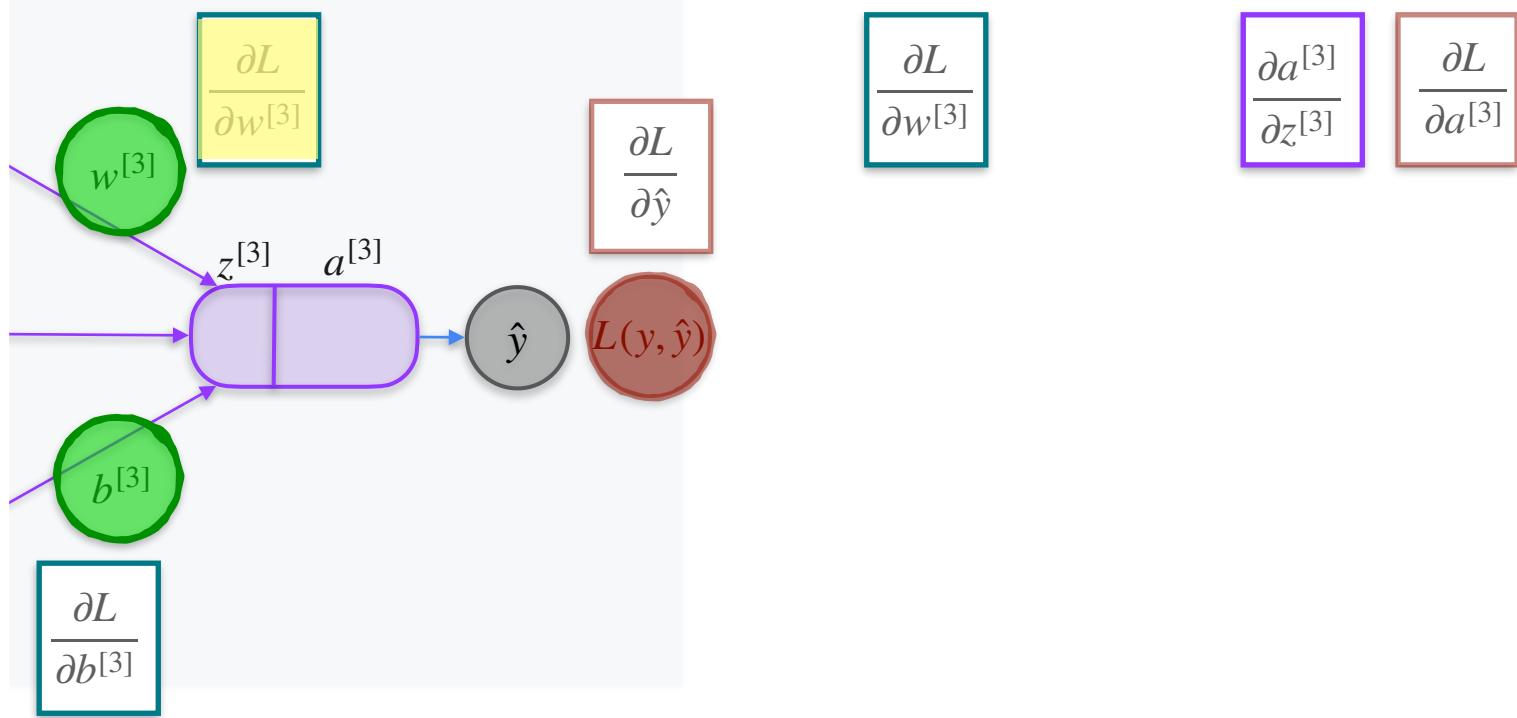
# Back Propagation Introduction



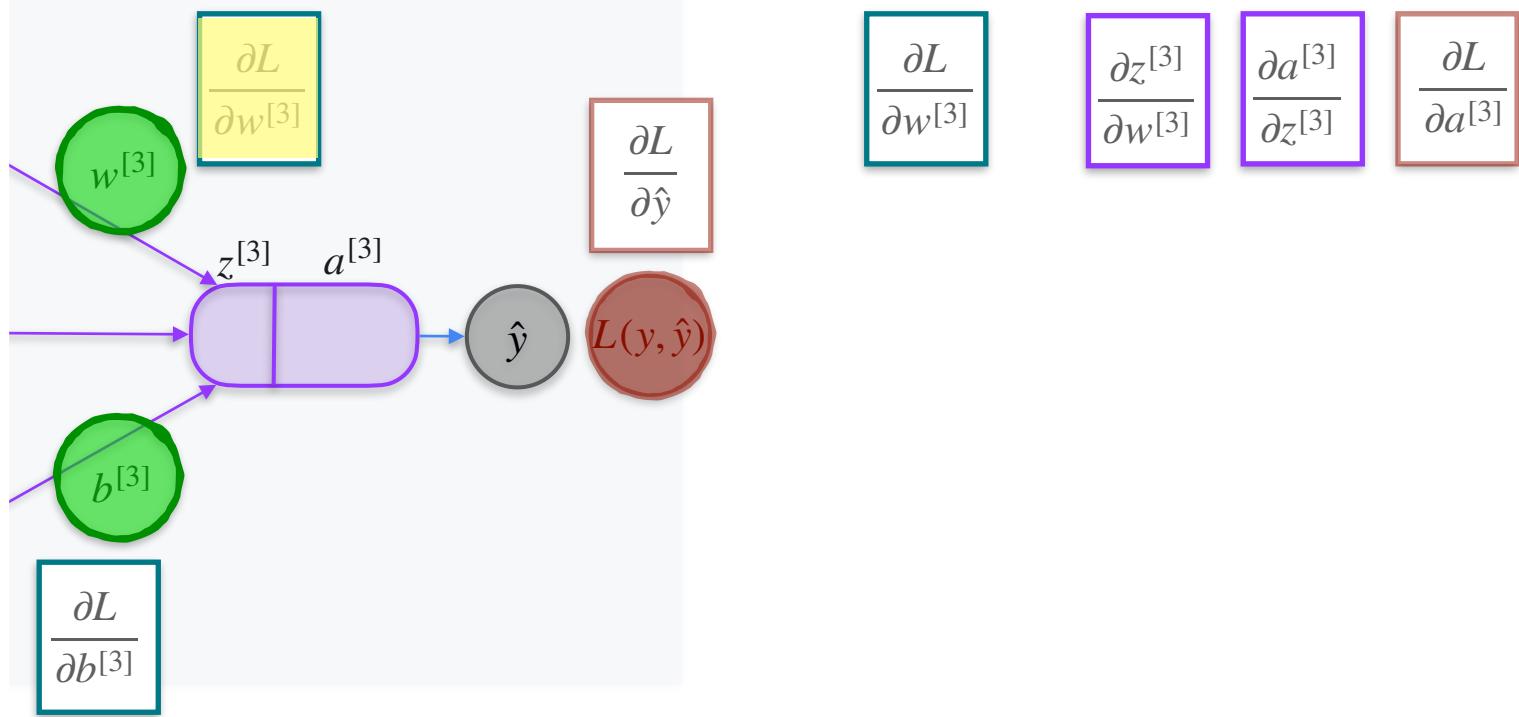
# Back Propagation Introduction



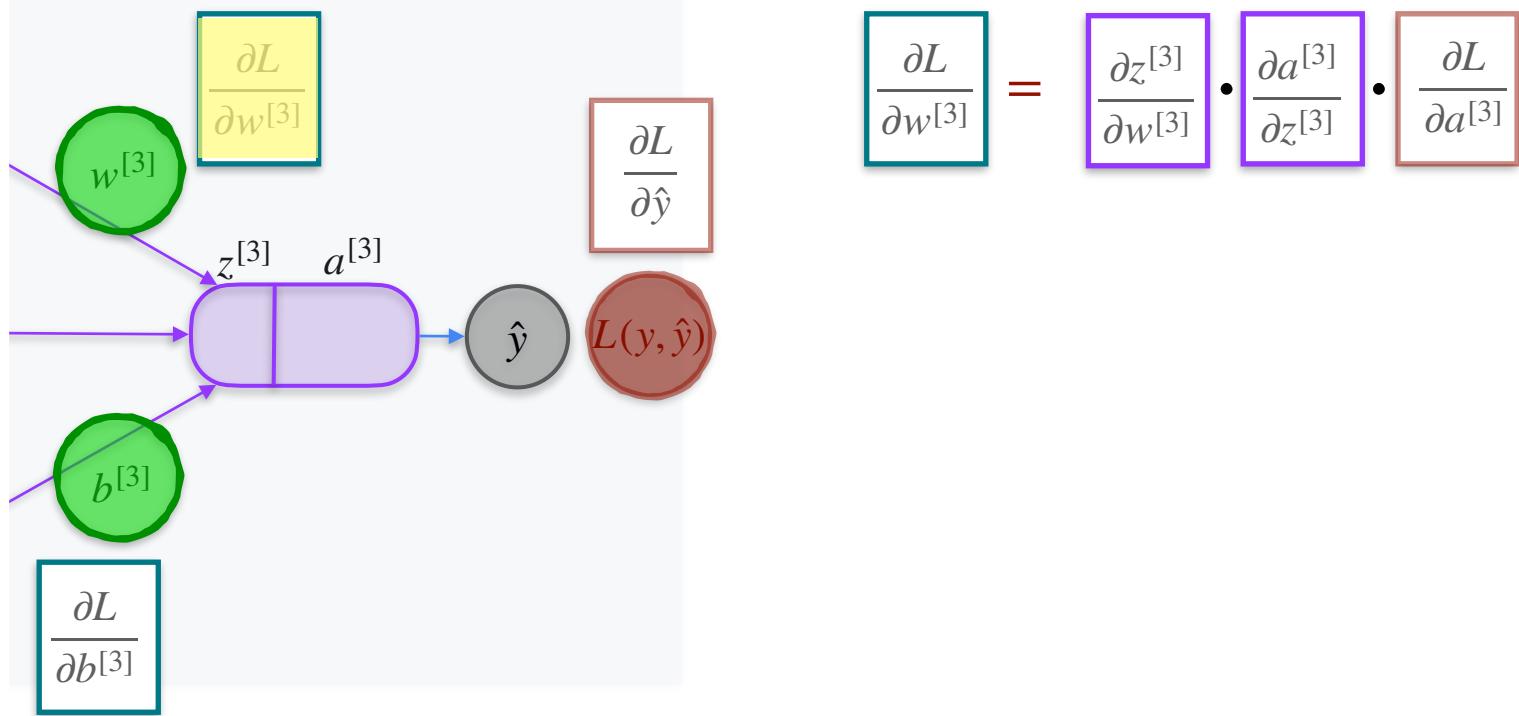
# Back Propagation Introduction



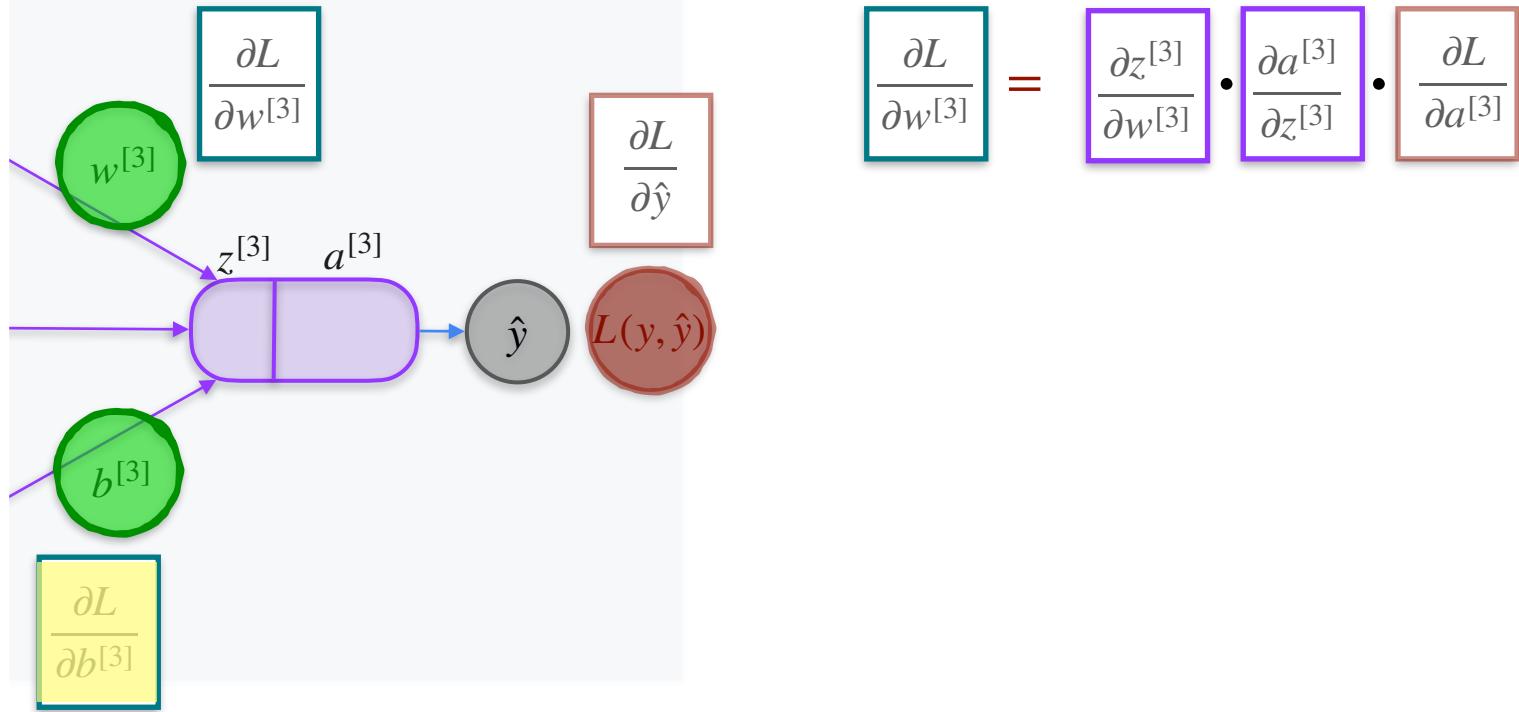
# Back Propagation Introduction



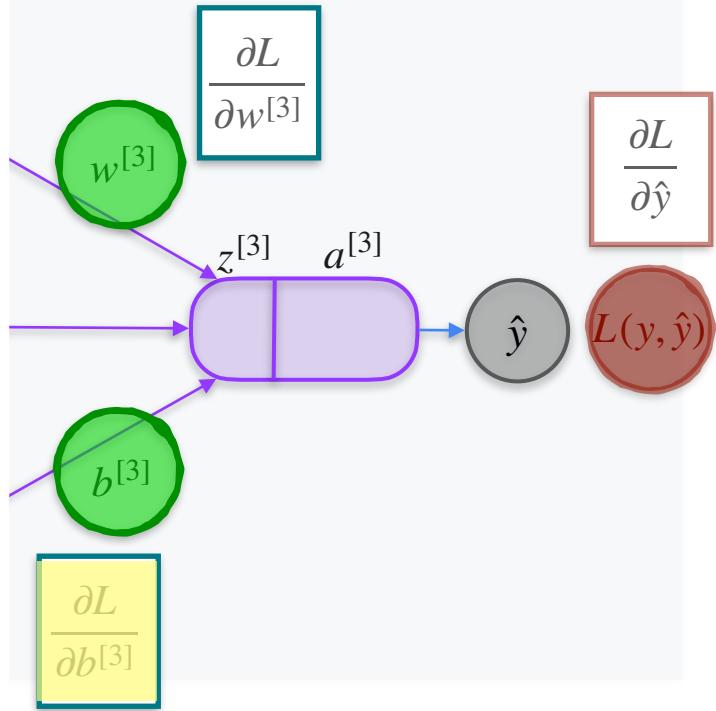
# Back Propagation Introduction



# Back Propagation Introduction



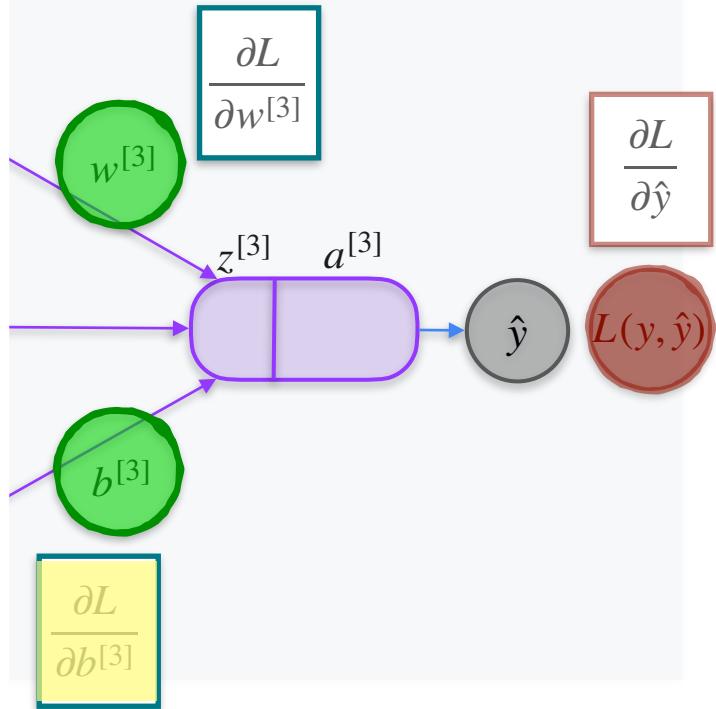
# Back Propagation Introduction



$$\frac{\partial L}{\partial w^{[3]}} = \frac{\partial z^{[3]}}{\partial w^{[3]}} \cdot \frac{\partial a^{[3]}}{\partial z^{[3]}} \cdot \frac{\partial L}{\partial a^{[3]}}$$

$$\frac{\partial L}{\partial b^{[3]}}$$

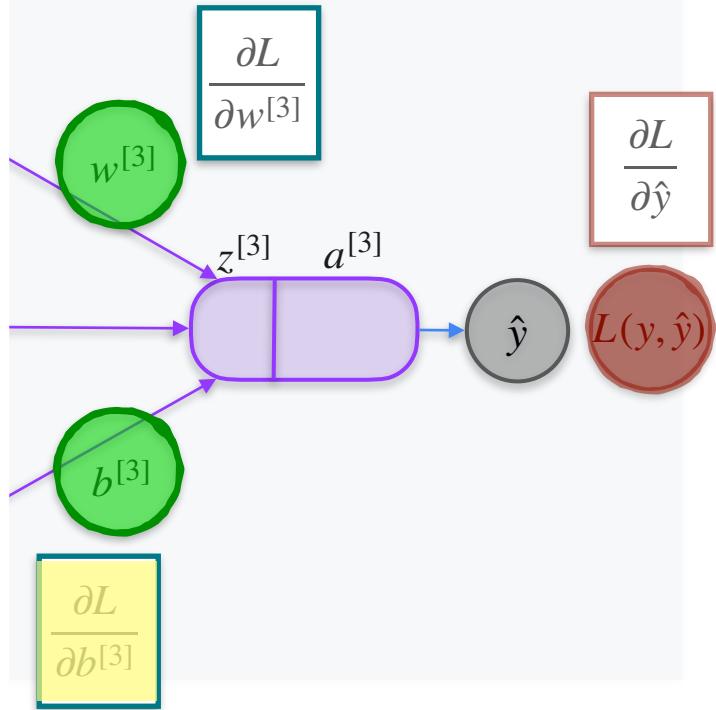
# Back Propagation Introduction



$$\frac{\partial L}{\partial w^{[3]}} = \frac{\partial z^{[3]}}{\partial w^{[3]}} \cdot \frac{\partial a^{[3]}}{\partial z^{[3]}} \cdot \frac{\partial L}{\partial a^{[3]}}$$

$$\begin{array}{c} \frac{\partial L}{\partial b^{[3]}} \\ \frac{\partial z^{[3]}}{\partial b^{[3]}} \\ \frac{\partial a^{[3]}}{\partial z^{[3]}} \\ \frac{\partial L}{\partial a^{[3]}} \end{array}$$

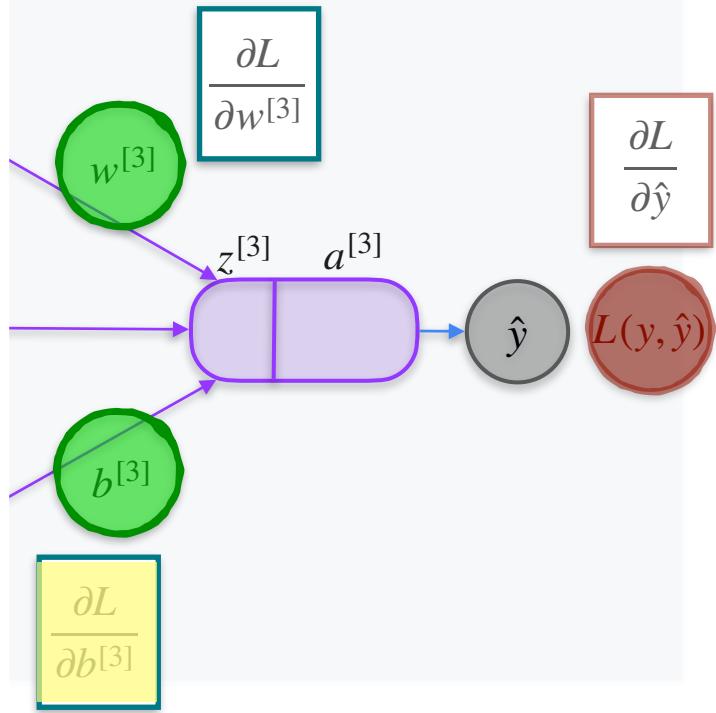
# Back Propagation Introduction



$$\frac{\partial L}{\partial w^{[3]}} = \frac{\partial z^{[3]}}{\partial w^{[3]}} \cdot \frac{\partial a^{[3]}}{\partial z^{[3]}} \cdot \frac{\partial L}{\partial a^{[3]}}$$

$$\frac{\partial L}{\partial b^{[3]}} = \frac{\partial z^{[3]}}{\partial b^{[3]}} \cdot \frac{\partial a^{[3]}}{\partial z^{[3]}} \cdot \frac{\partial L}{\partial a^{[3]}}$$

# Back Propagation Introduction

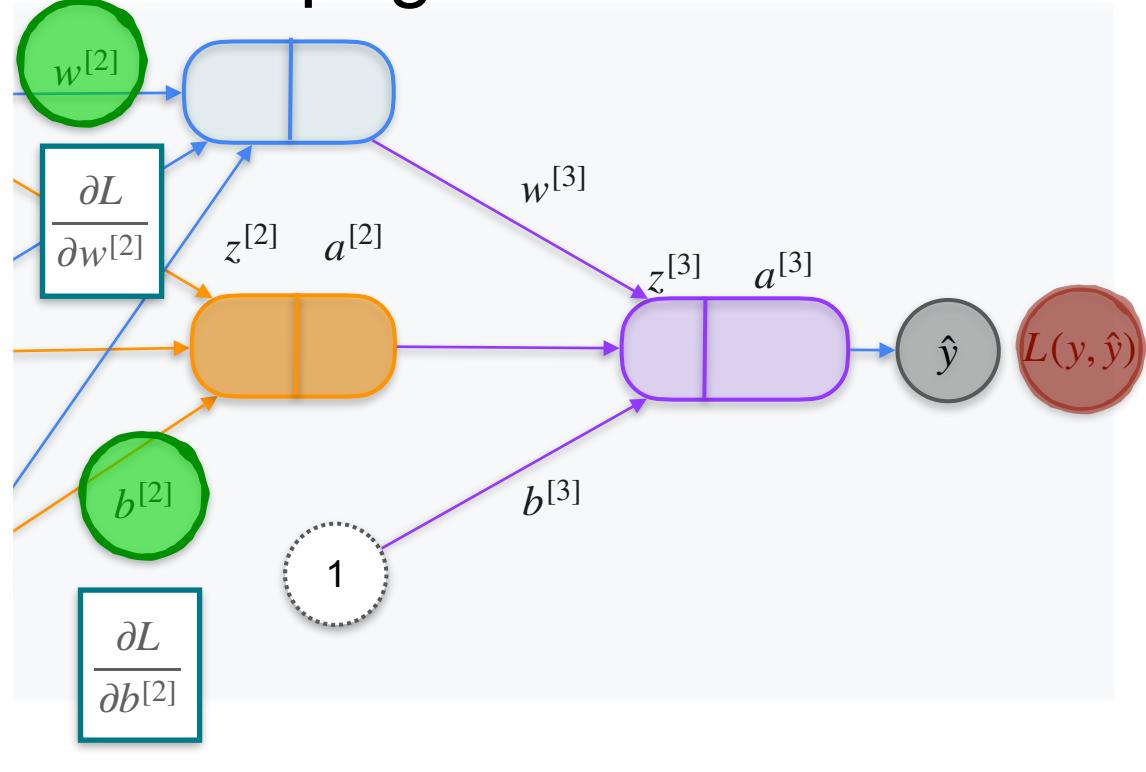


$$\frac{\partial L}{\partial w^{[3]}} = \frac{\partial z^{[3]}}{\partial w^{[3]}} \cdot \frac{\partial a^{[3]}}{\partial z^{[3]}} \cdot \frac{\partial L}{\partial a^{[3]}}$$

$$\frac{\partial L}{\partial b^{[3]}} = \frac{\partial z^{[3]}}{\partial b^{[3]}} \cdot \frac{\partial a^{[3]}}{\partial z^{[3]}} \cdot \frac{\partial L}{\partial a^{[3]}}$$

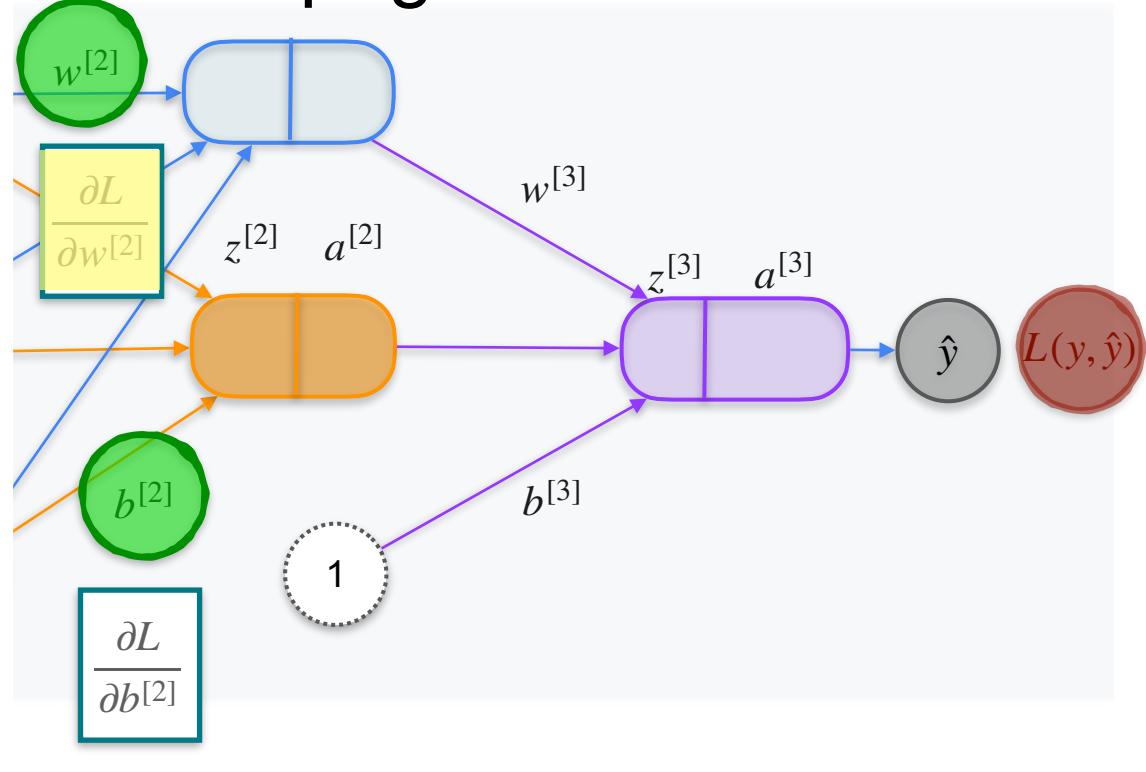
$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

# Back Propagation Introduction

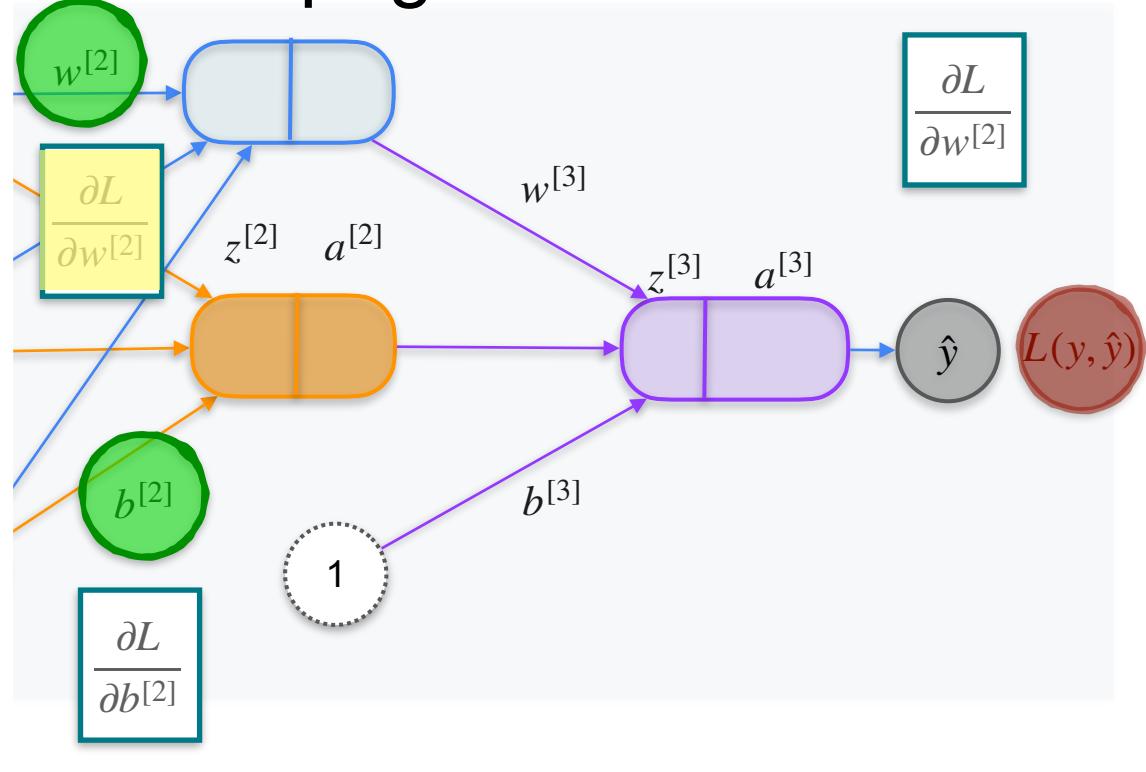


$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

# Back Propagation Introduction

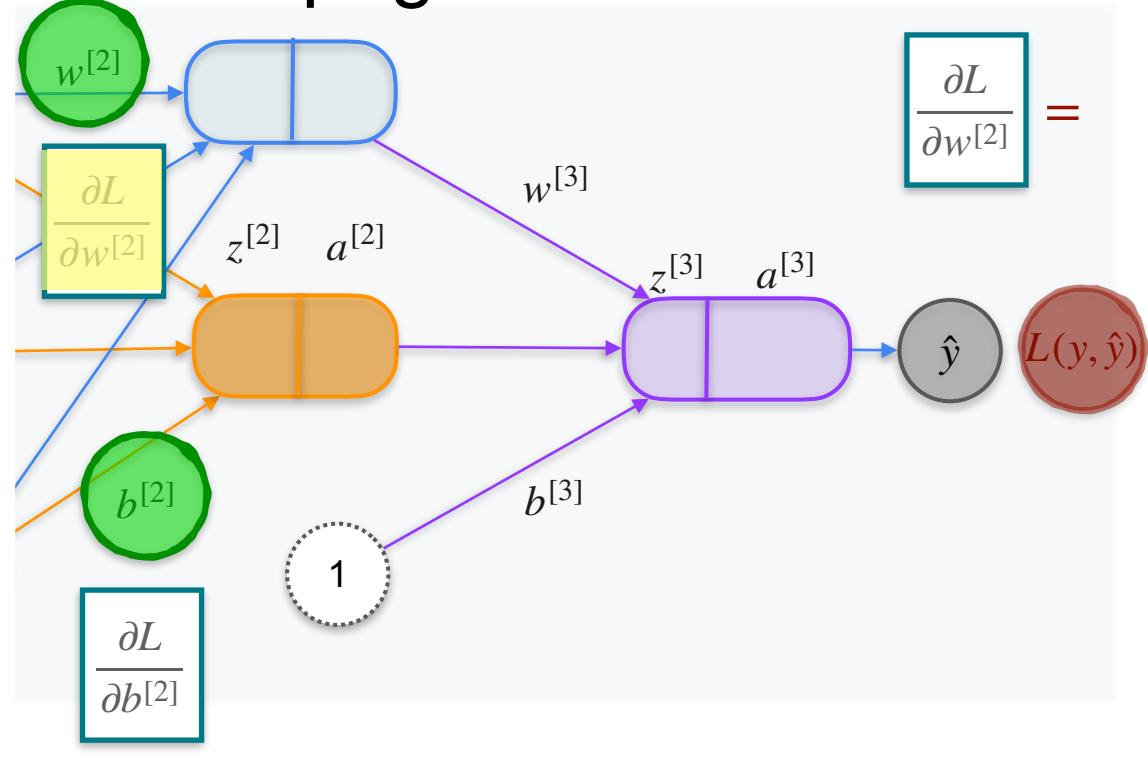


# Back Propagation Introduction



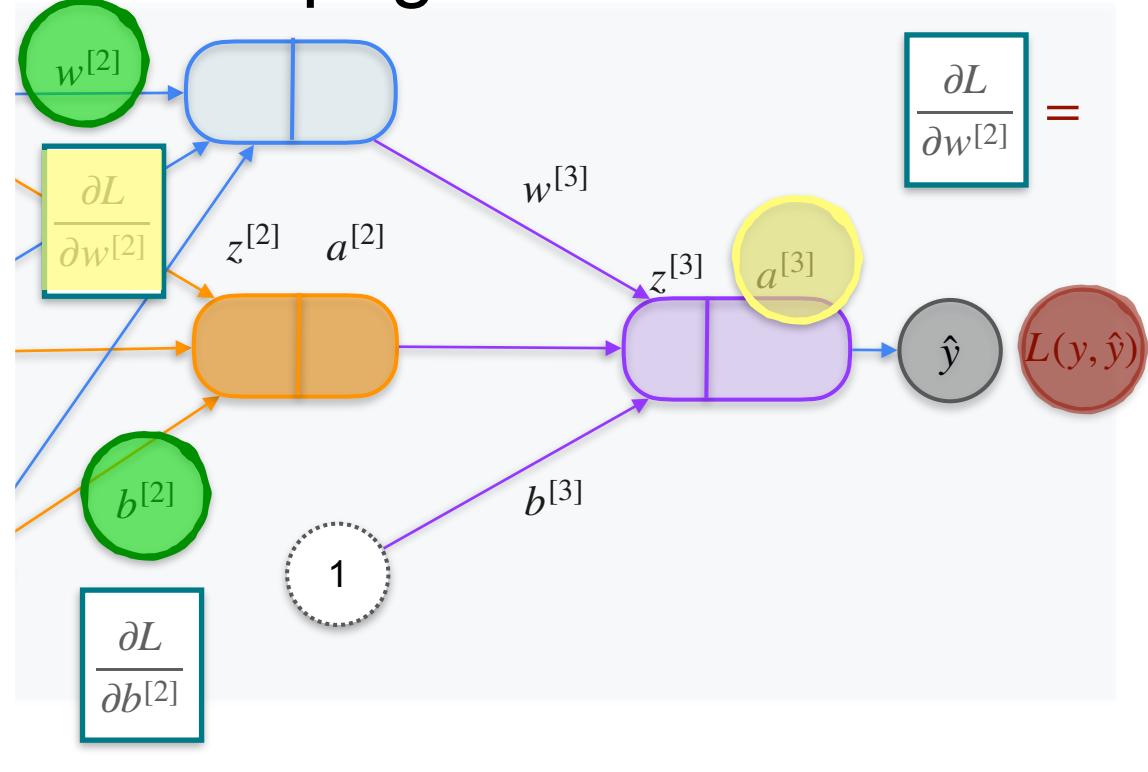
$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

# Back Propagation Introduction

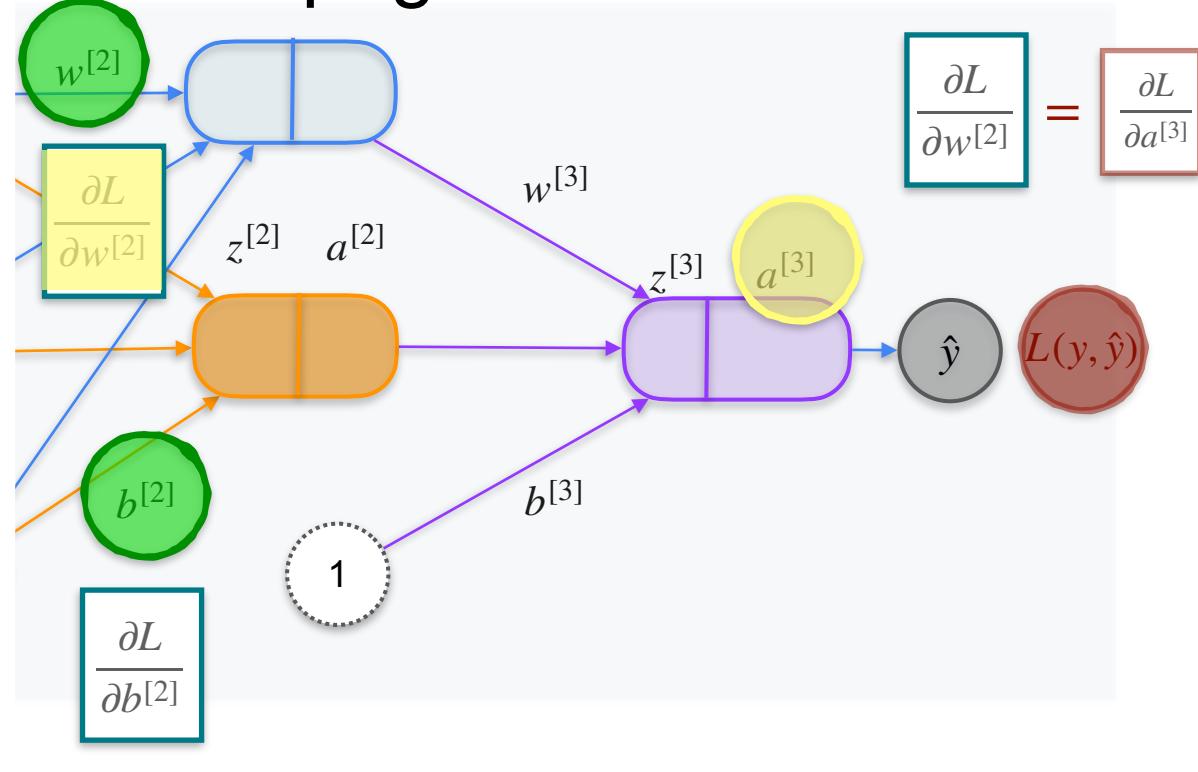


$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

# Back Propagation Introduction

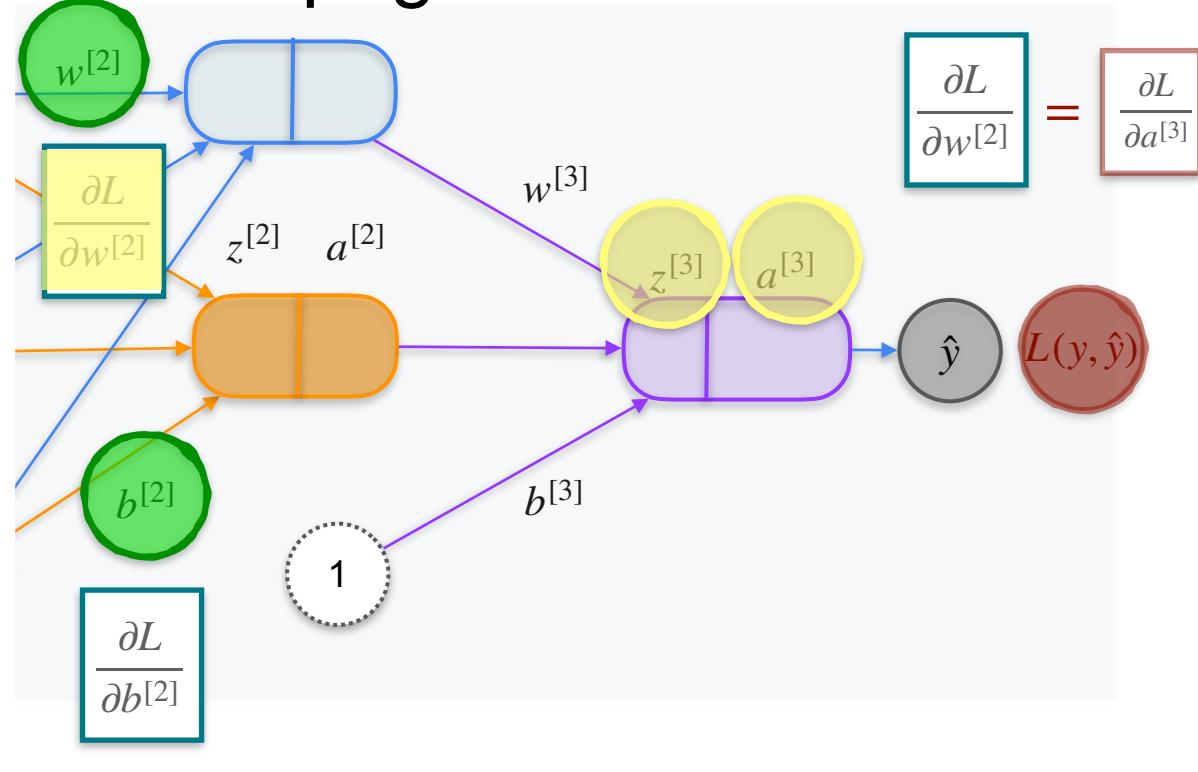


# Back Propagation Introduction

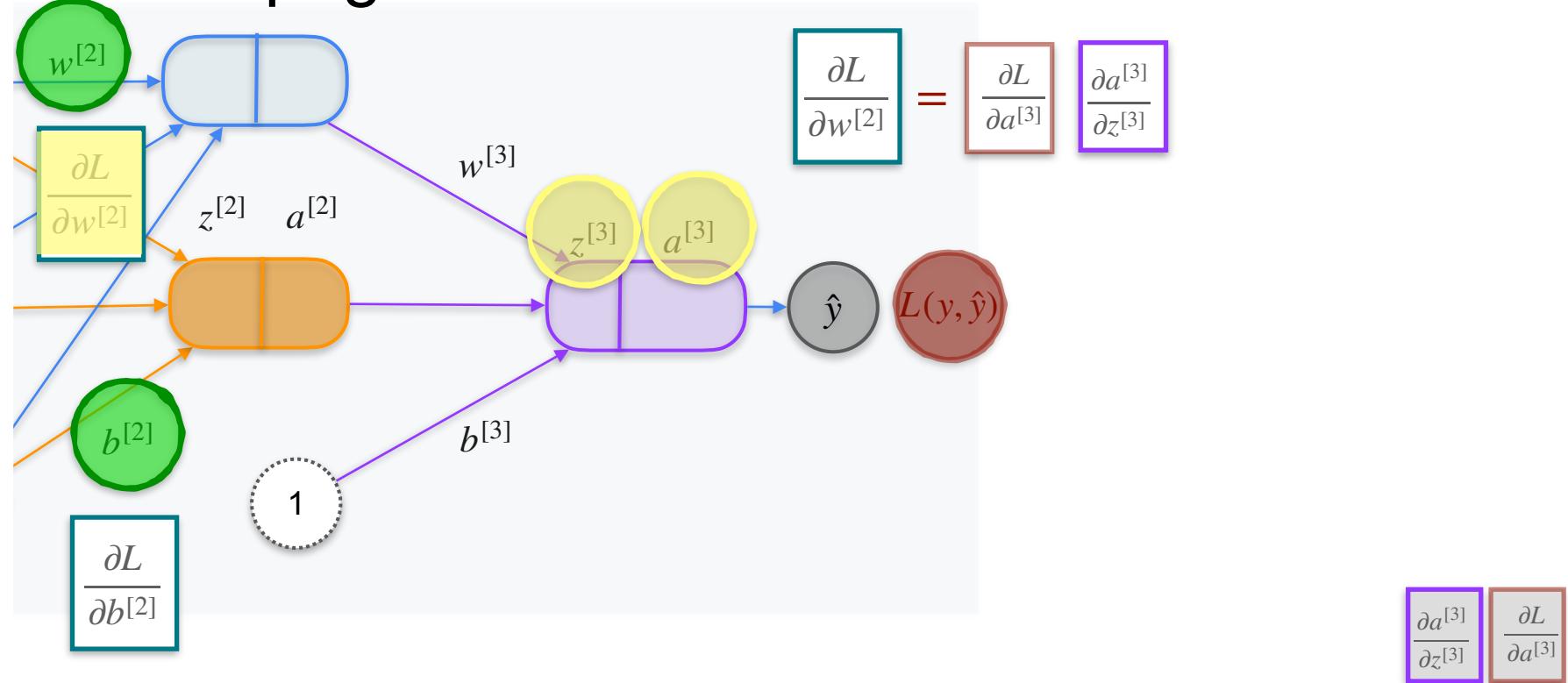


$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

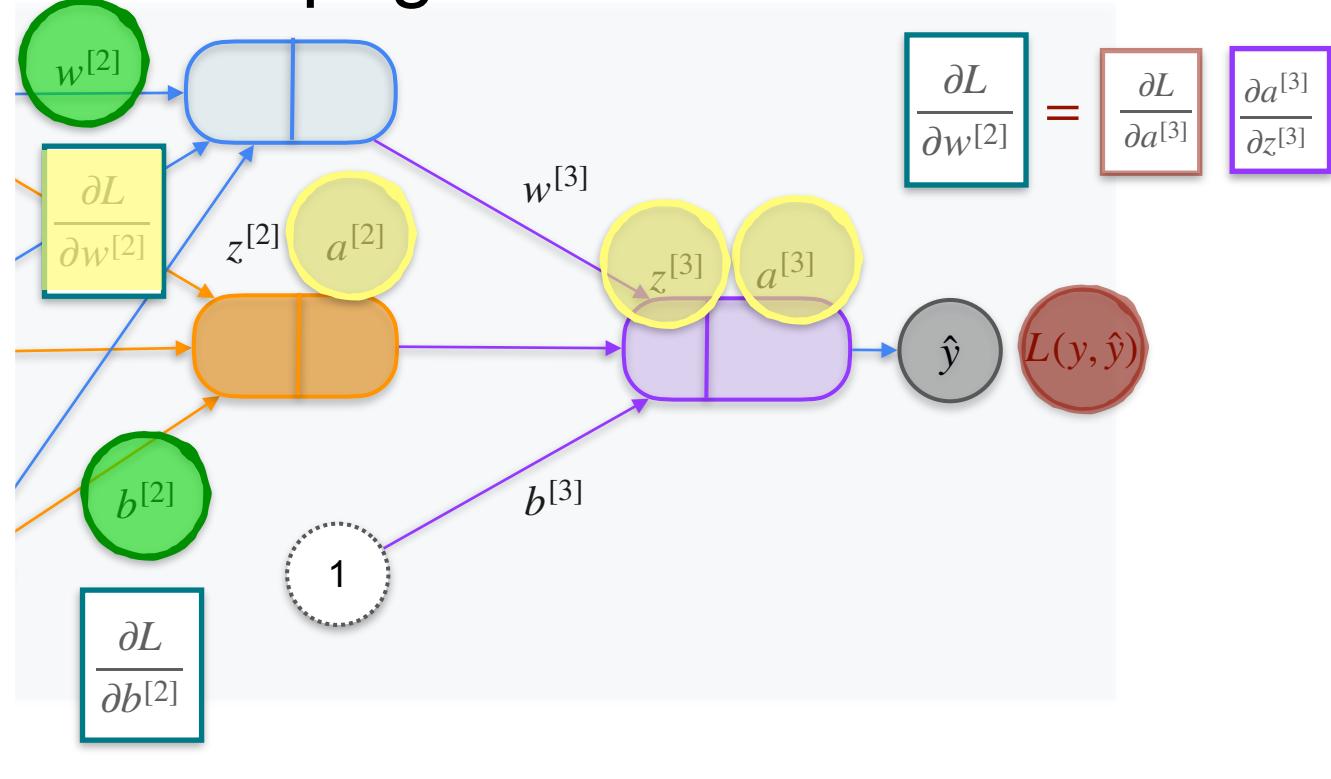
# Back Propagation Introduction



# Back Propagation Introduction

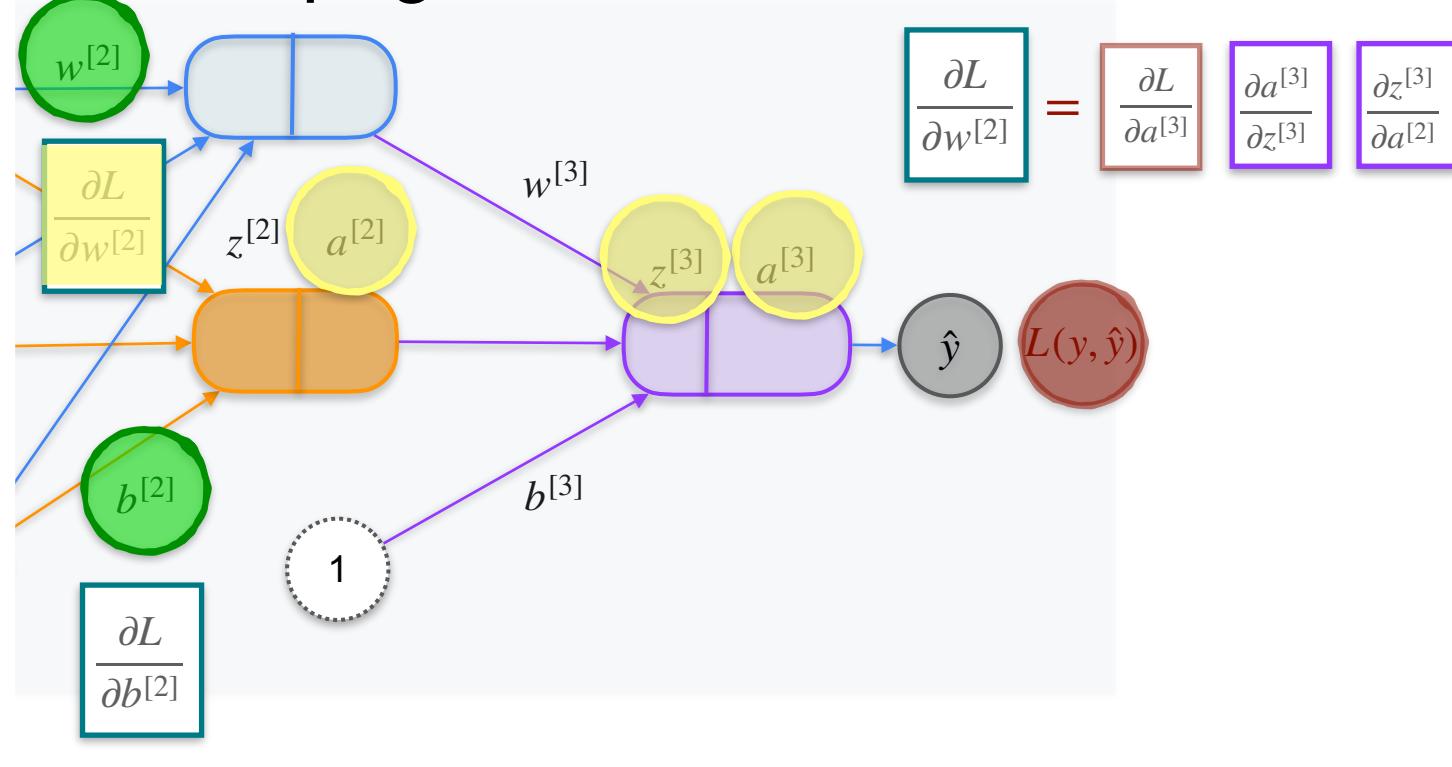


# Back Propagation Introduction

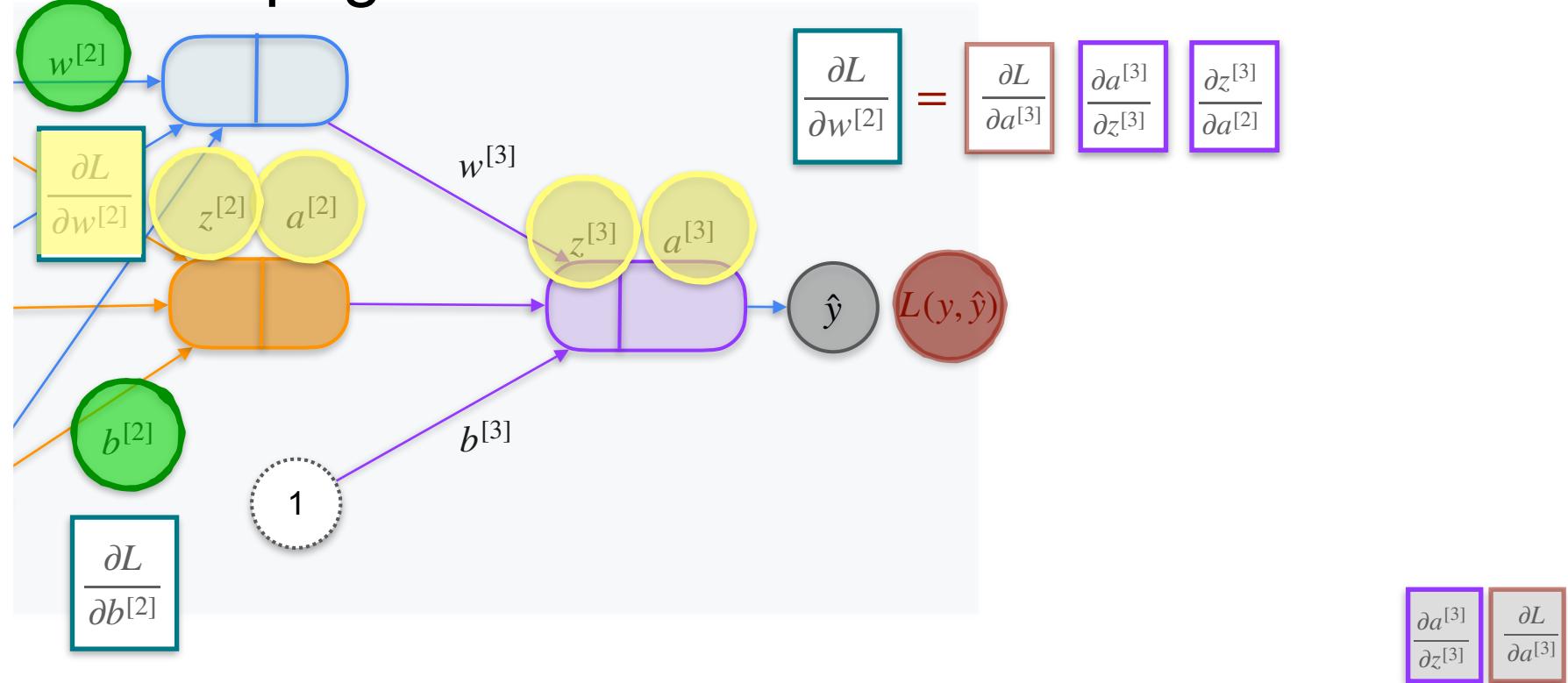


$$\frac{\partial a^{[3]}}{\partial z^{[3]}}$$
$$\frac{\partial L}{\partial a^{[3]}}$$

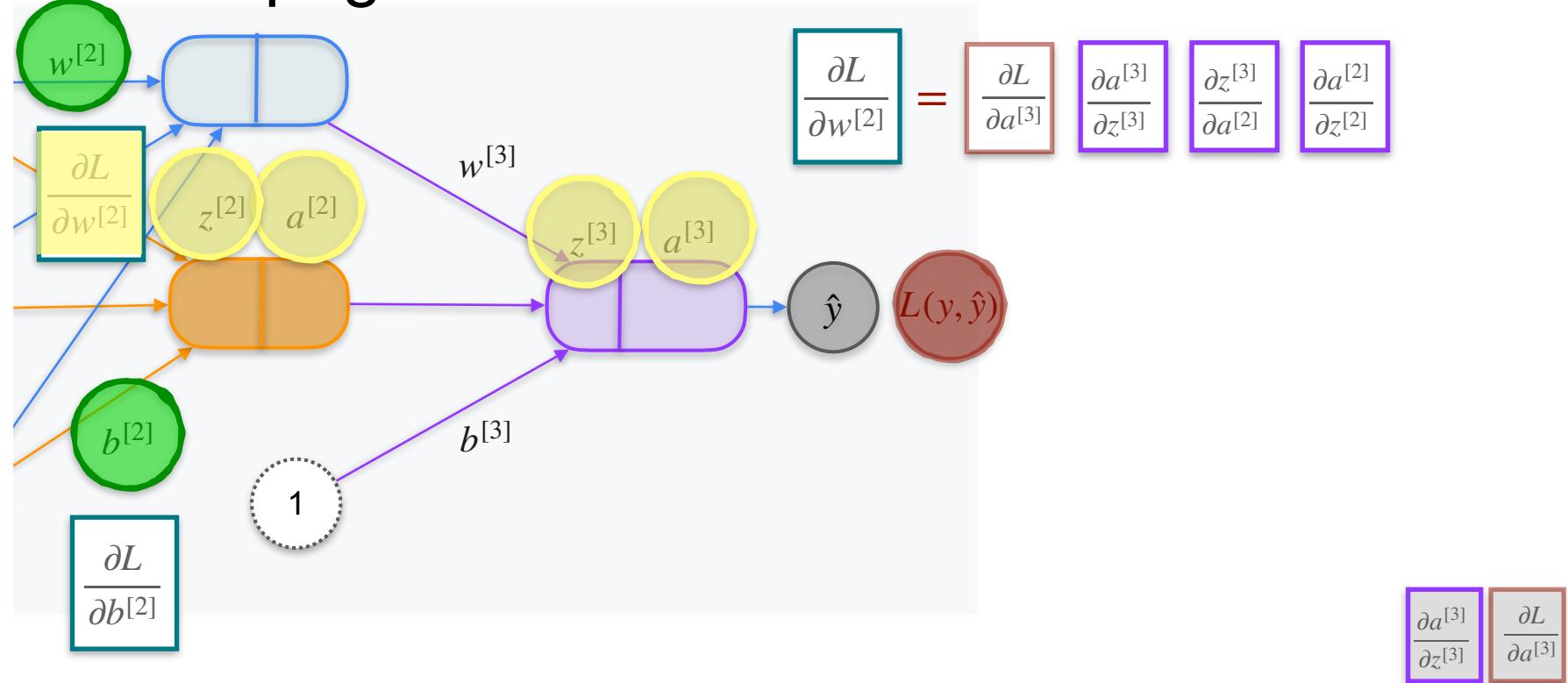
# Back Propagation Introduction



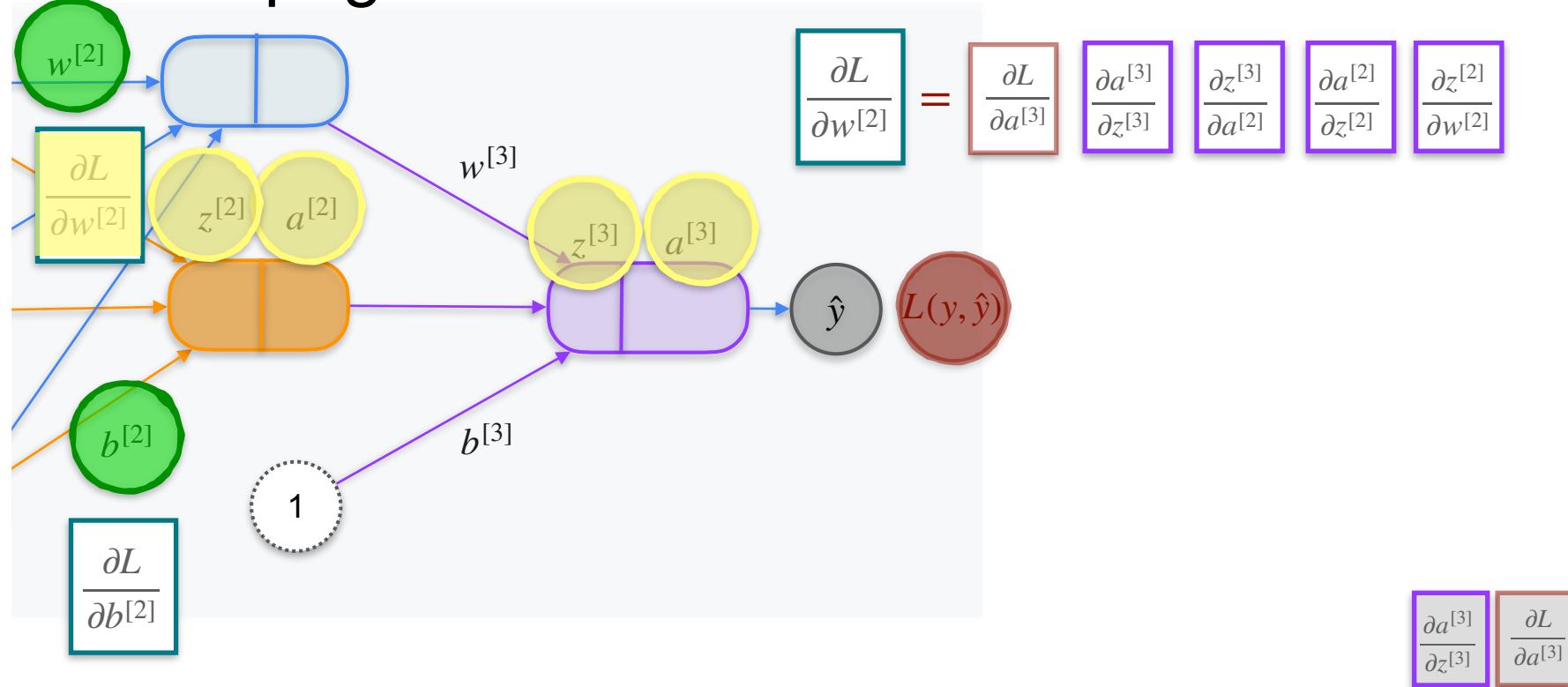
# Back Propagation Introduction



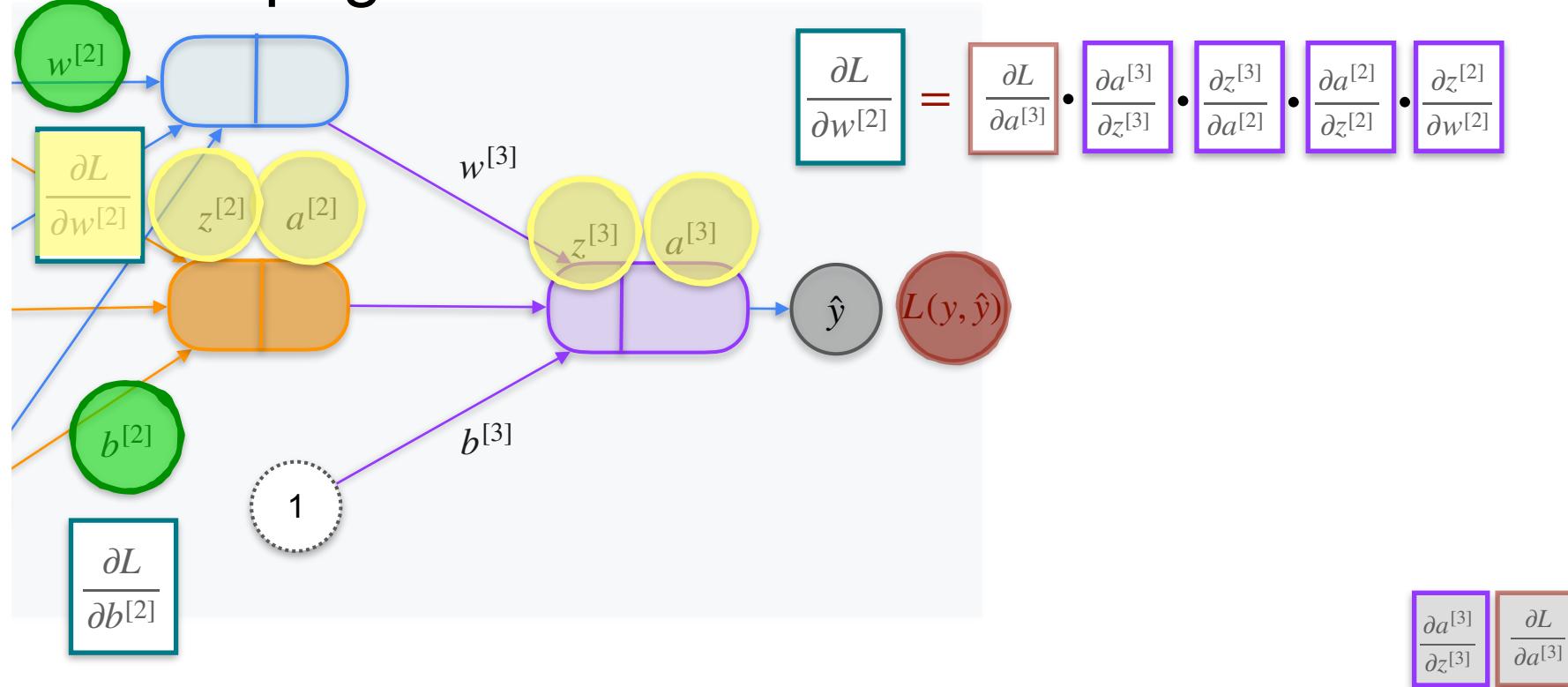
# Back Propagation Introduction



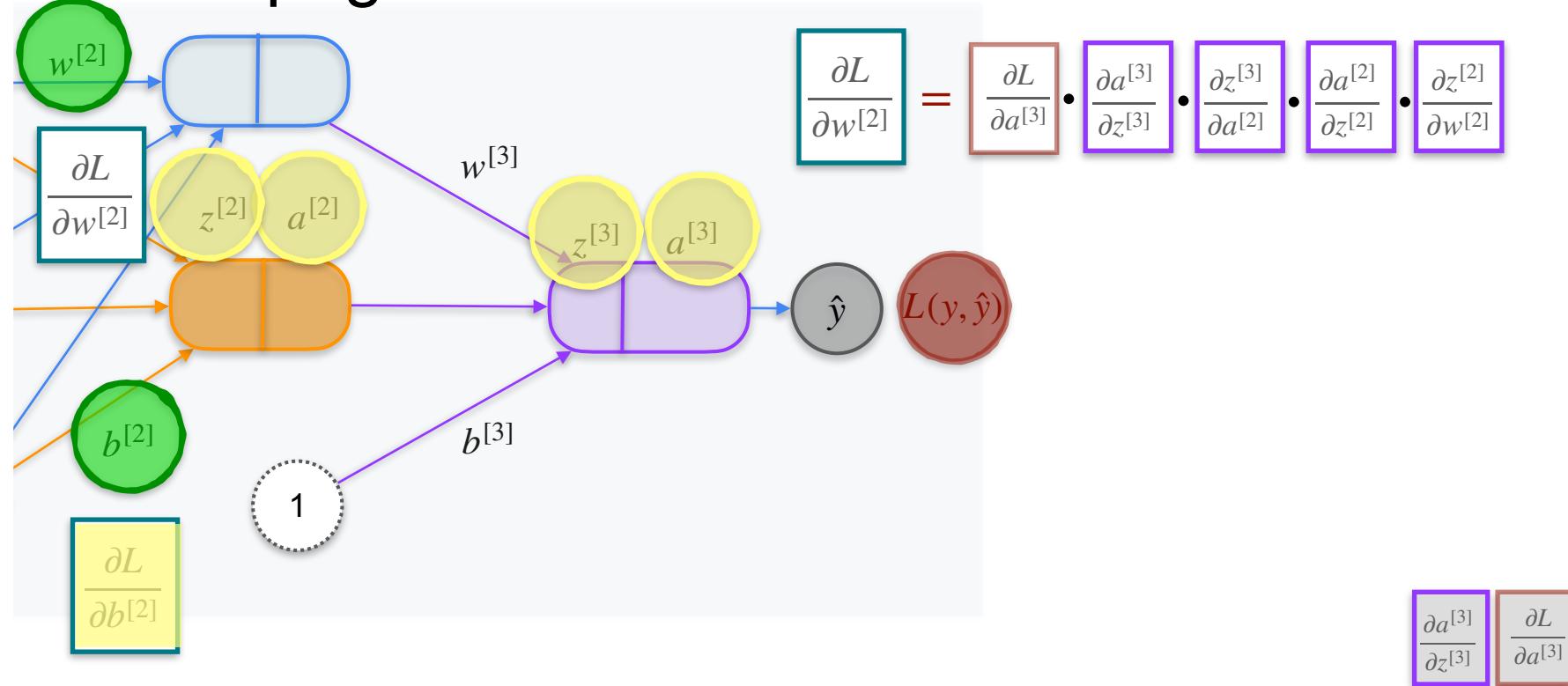
# Back Propagation Introduction



# Back Propagation Introduction

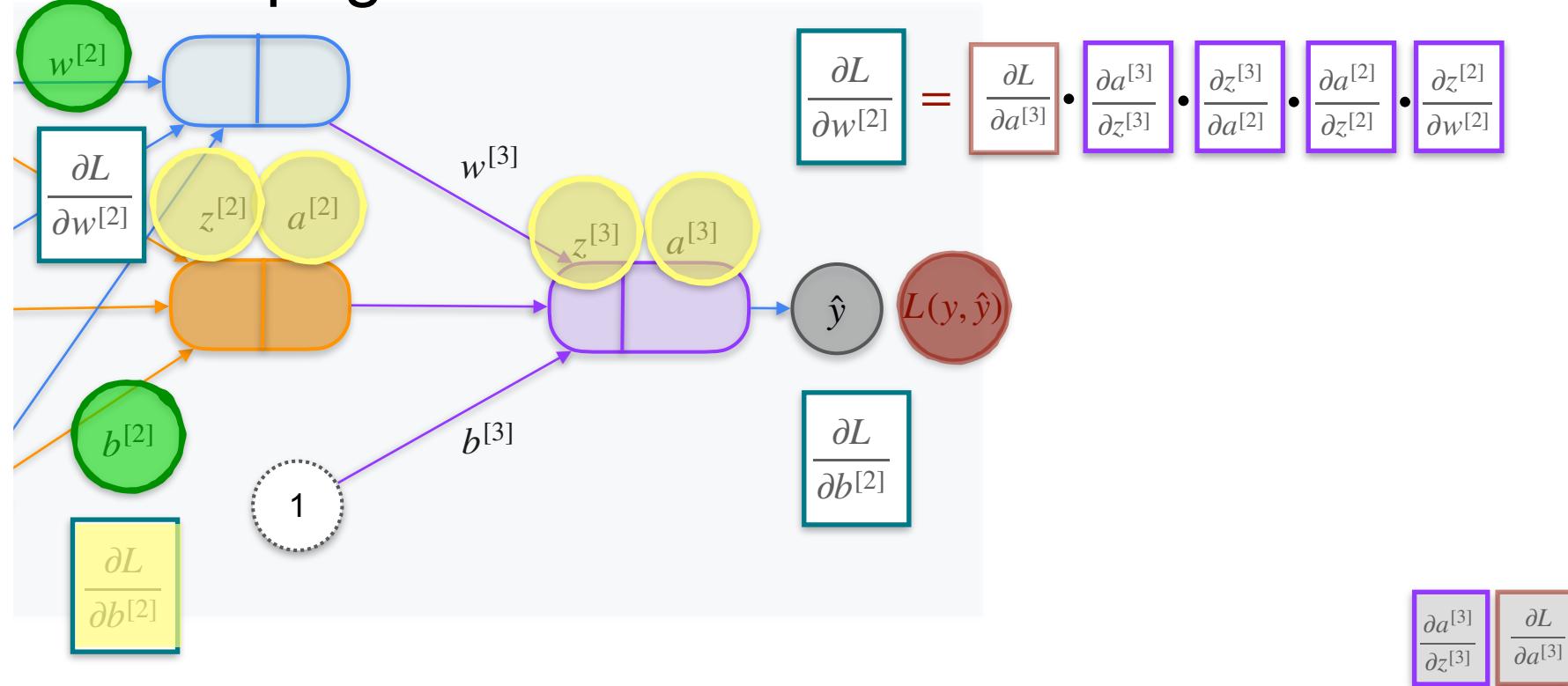


# Back Propagation Introduction



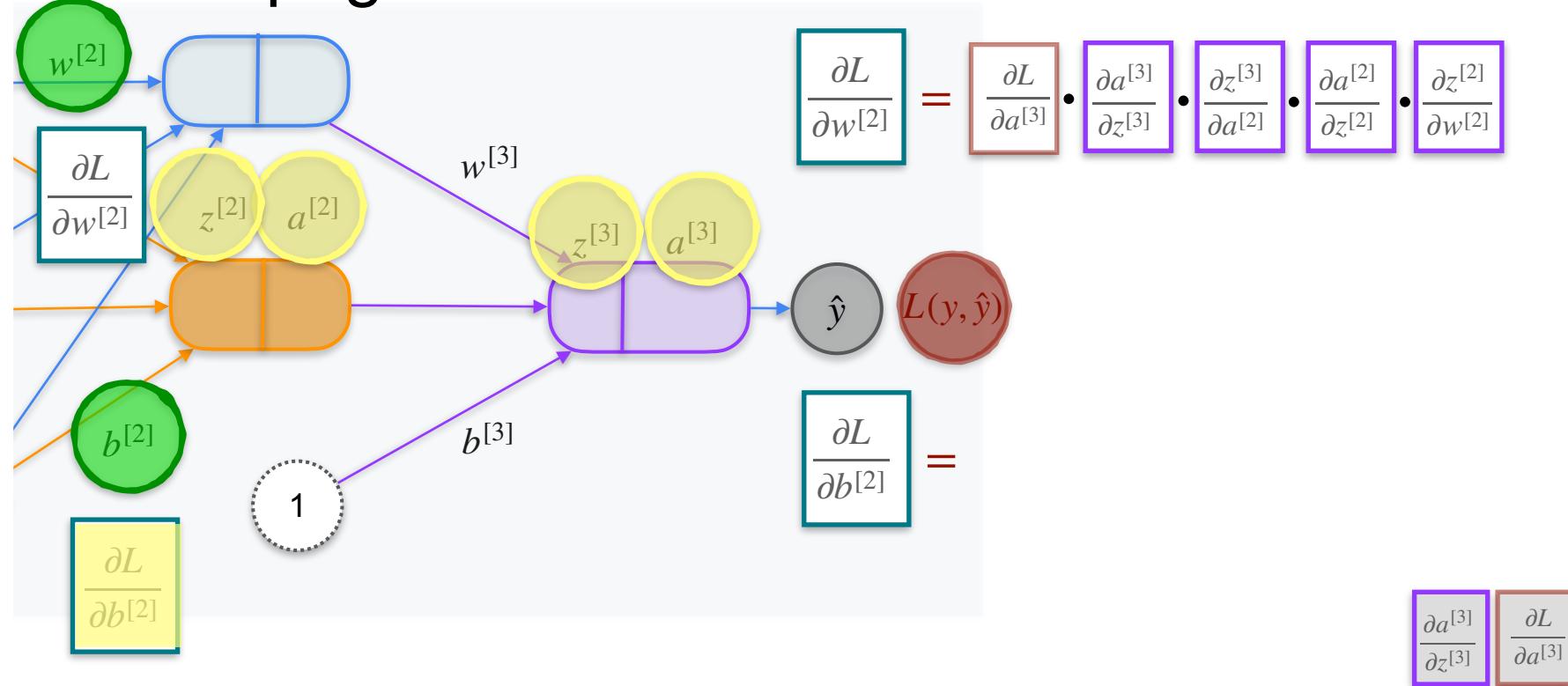
$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

# Back Propagation Introduction

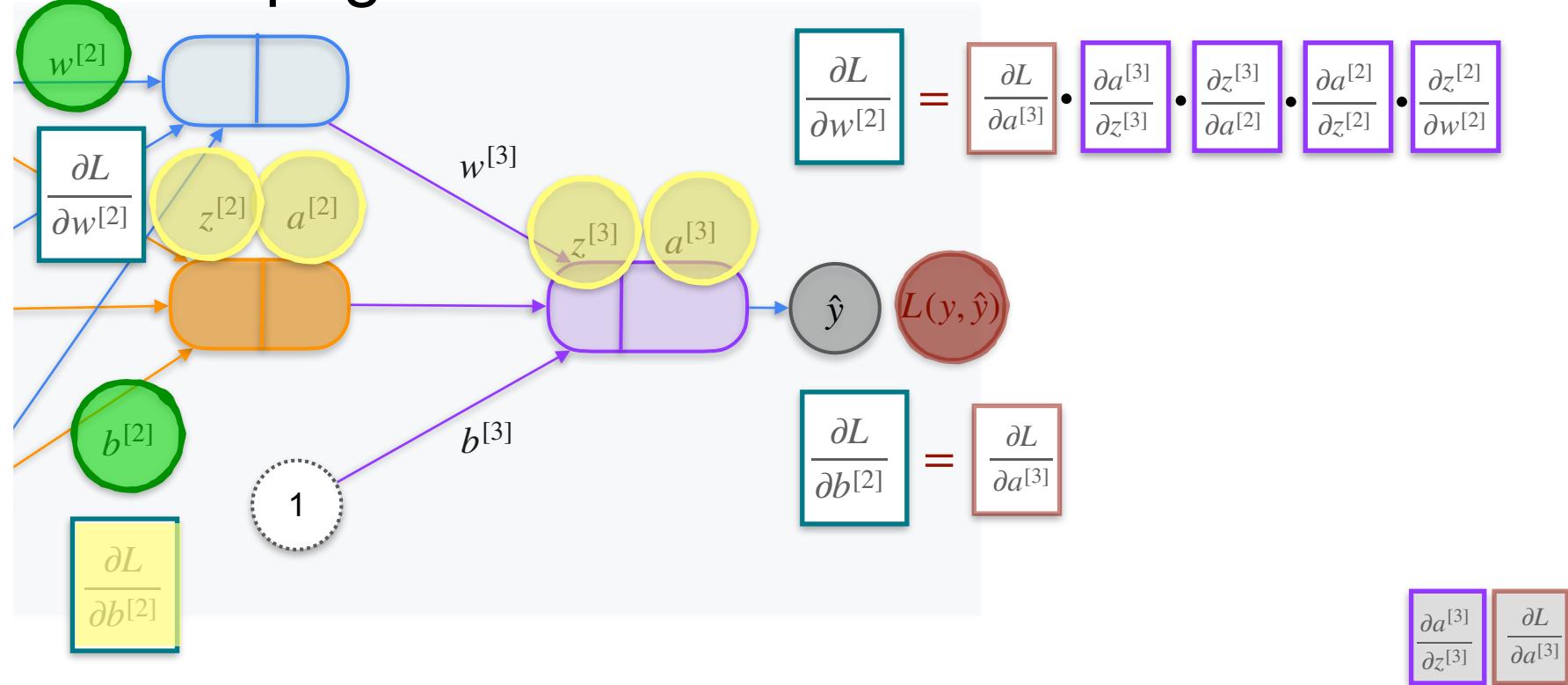


$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

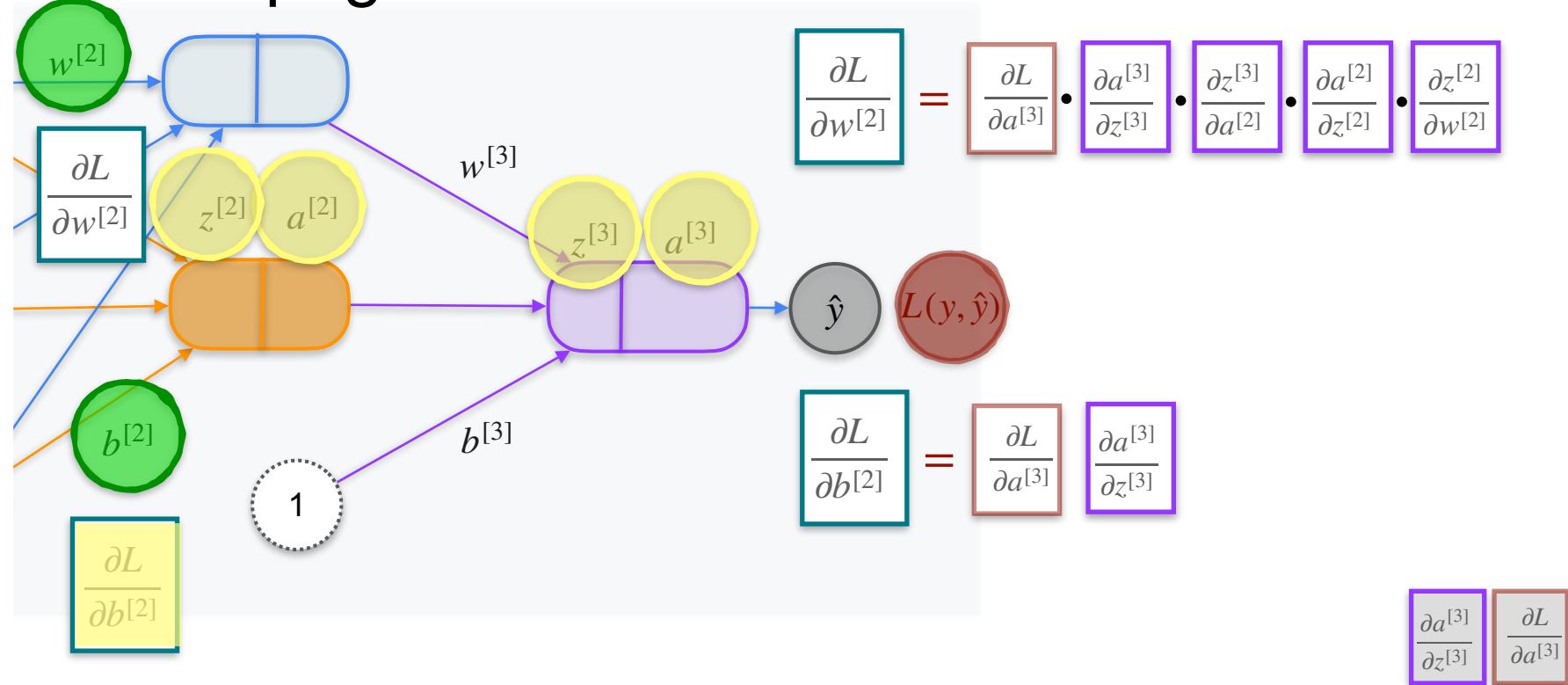
# Back Propagation Introduction



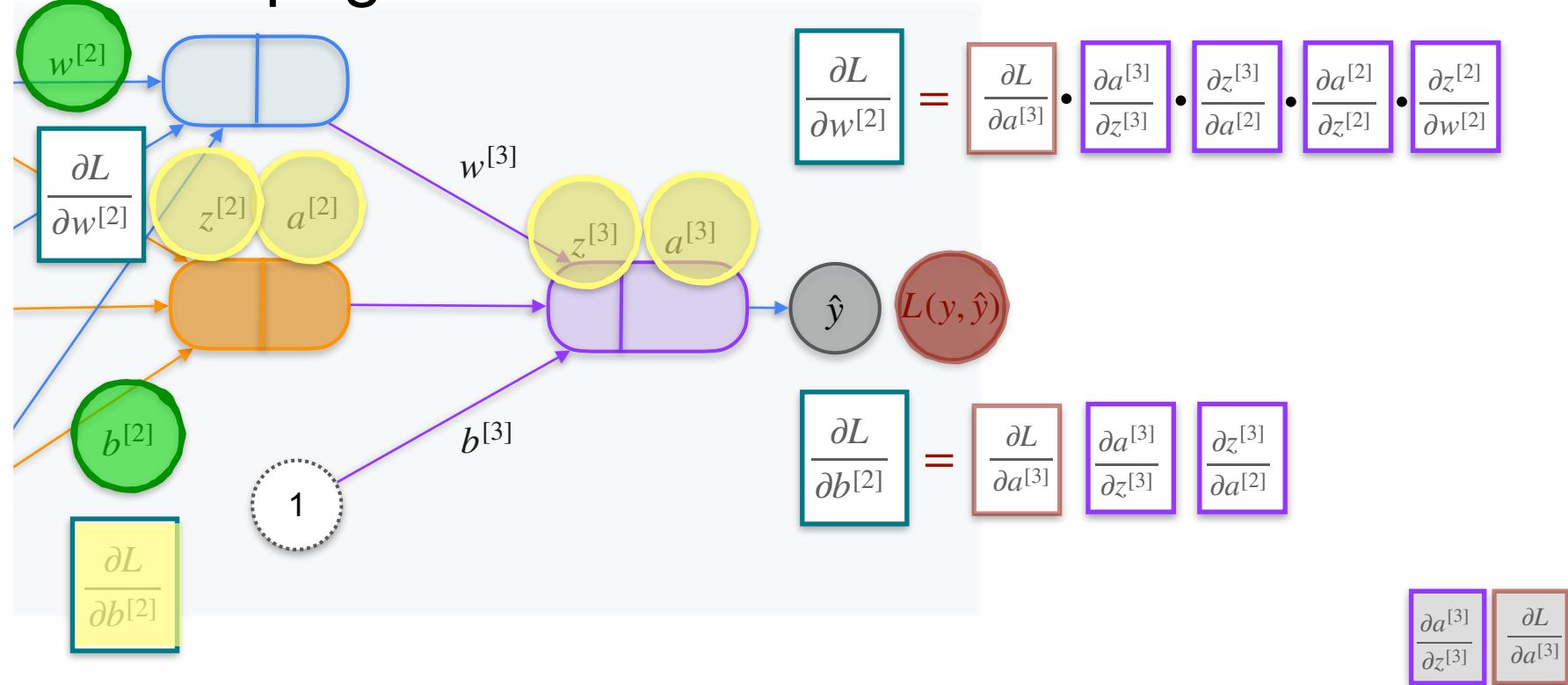
# Back Propagation Introduction



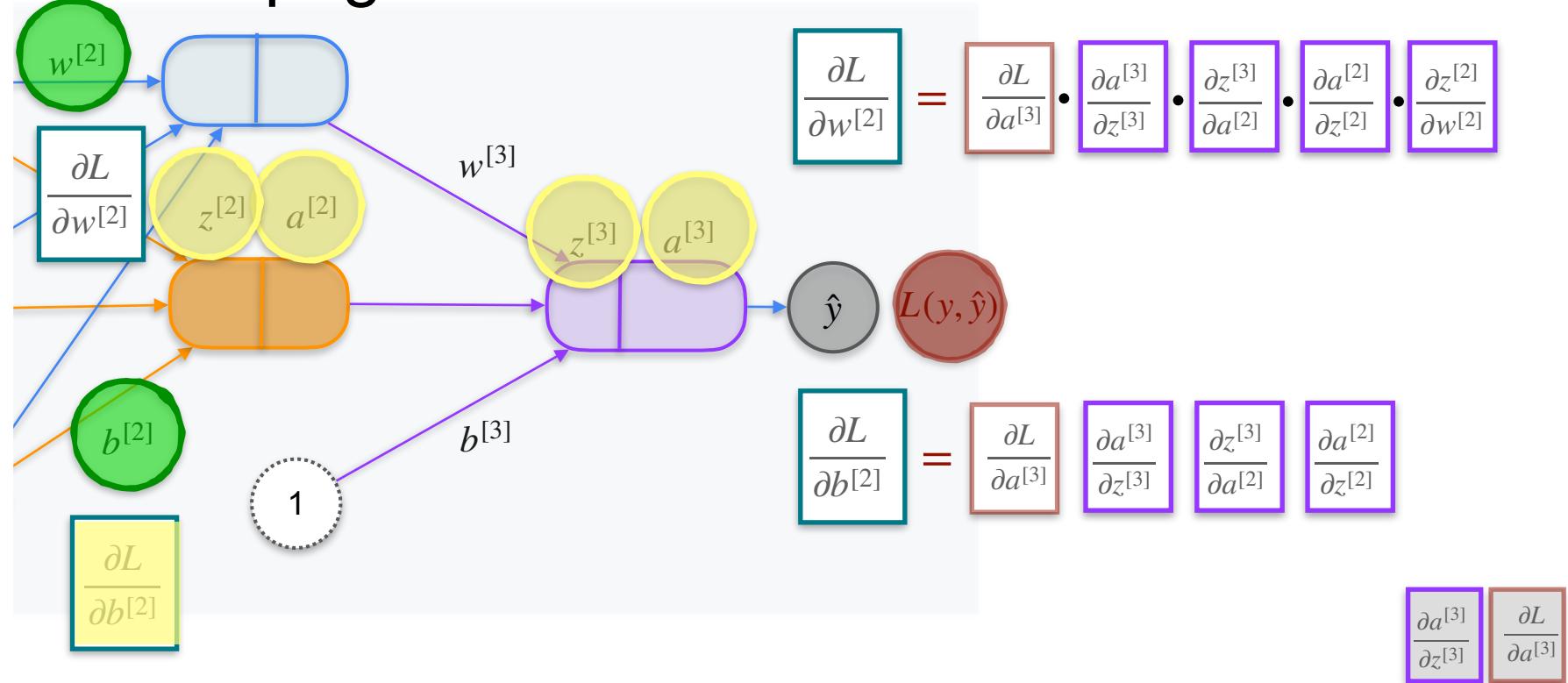
# Back Propagation Introduction



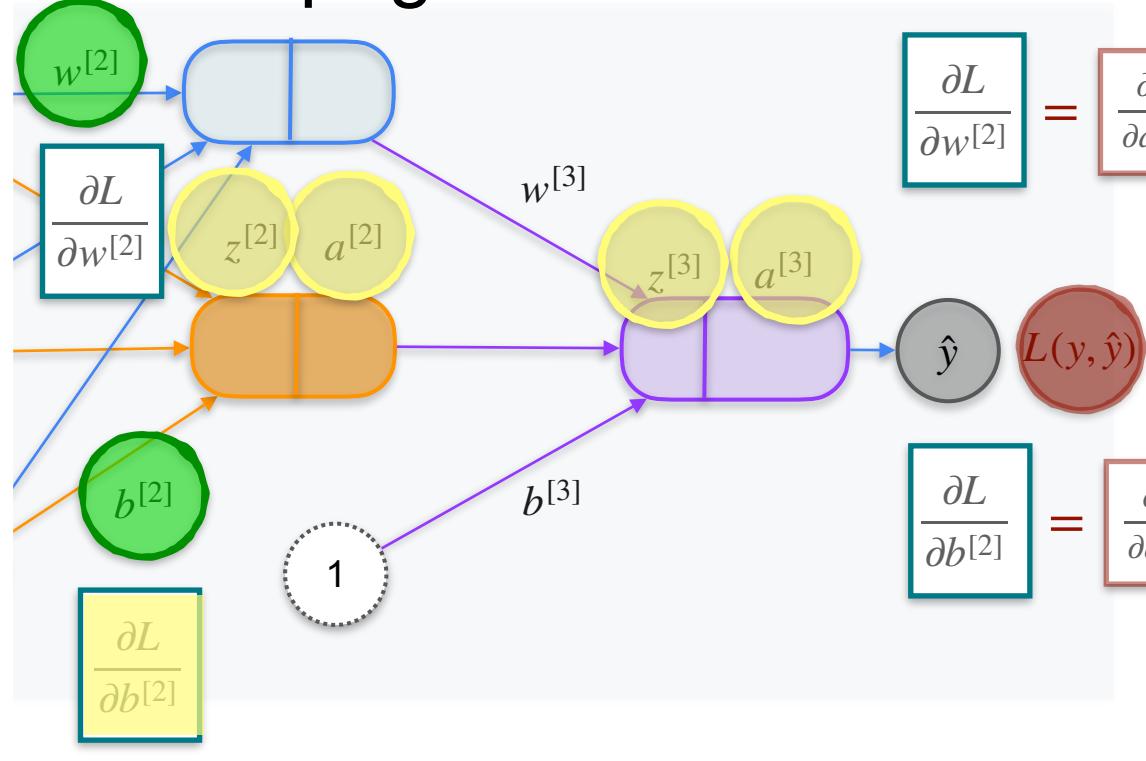
# Back Propagation Introduction



# Back Propagation Introduction



# Back Propagation Introduction

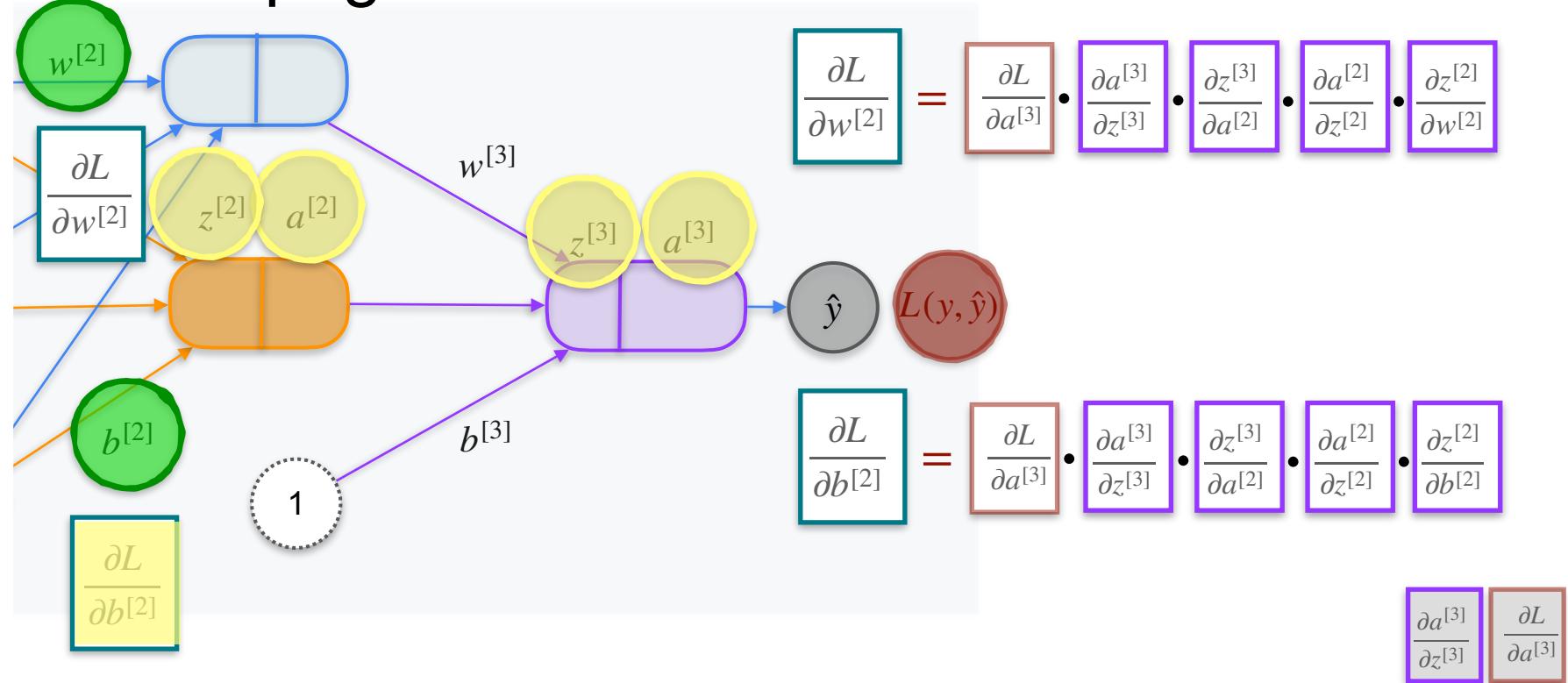


$$\frac{\partial L}{\partial w^{[2]}} = \frac{\partial L}{\partial a^{[3]}} \cdot \frac{\partial a^{[3]}}{\partial z^{[3]}} \cdot \frac{\partial z^{[3]}}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} \cdot \frac{\partial z^{[2]}}{\partial w^{[2]}}$$

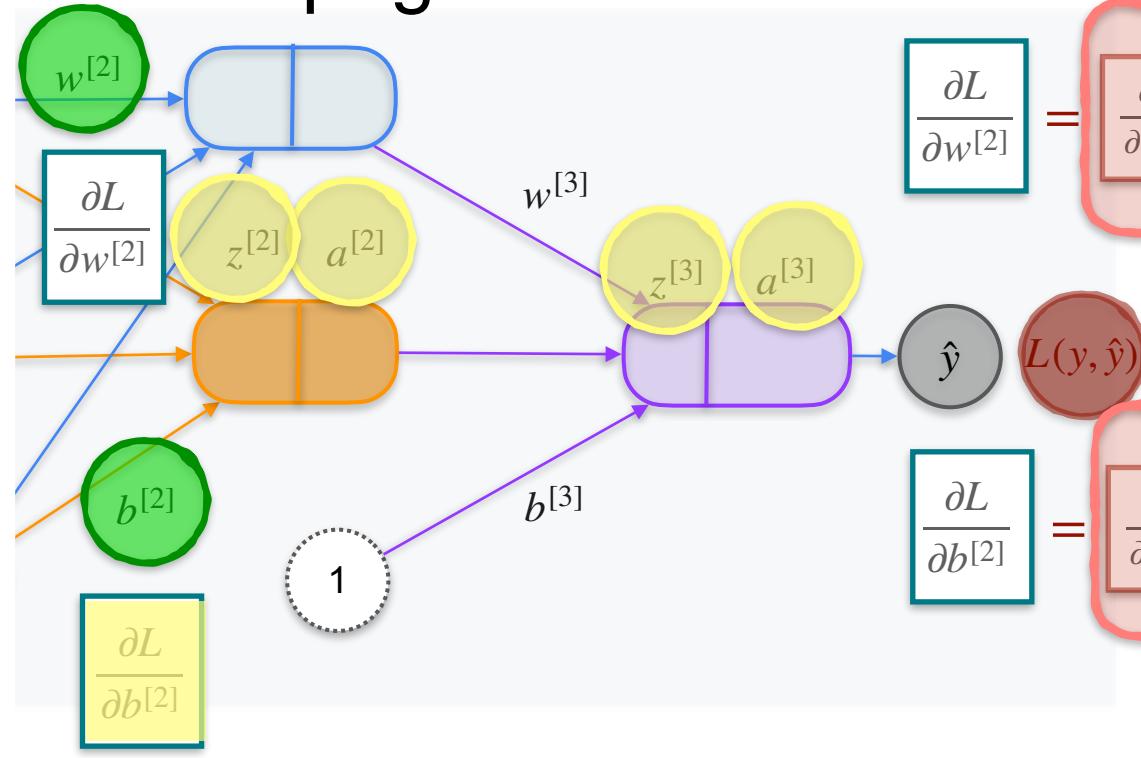
$$\frac{\partial L}{\partial b^{[2]}} = \frac{\partial L}{\partial a^{[3]}} \frac{\partial a^{[3]}}{\partial z^{[3]}} \frac{\partial z^{[3]}}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial b^{[2]}}$$

$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

# Back Propagation Introduction



# Back Propagation Introduction

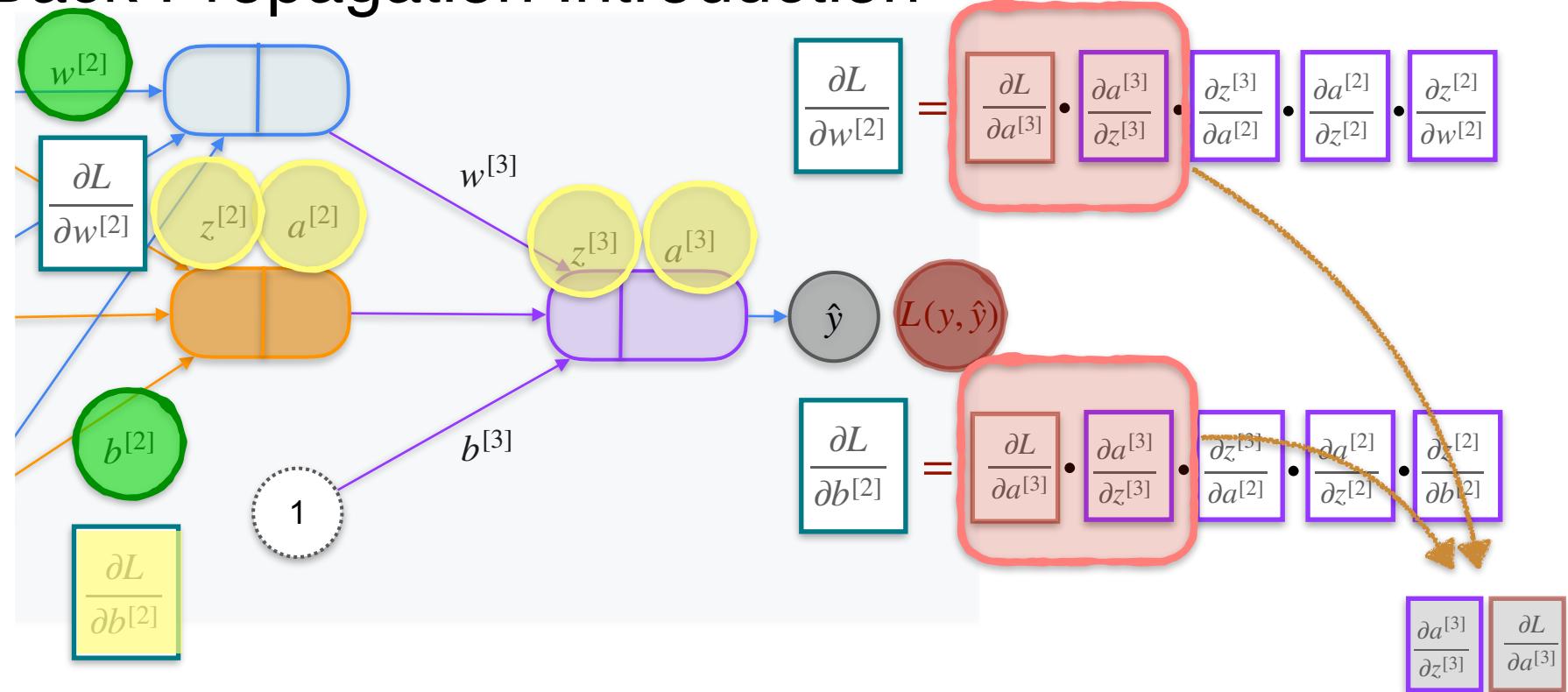


$$\frac{\partial L}{\partial w^{[2]}} = \boxed{\frac{\partial L}{\partial a^{[3]}}} \cdot \boxed{\frac{\partial a^{[3]}}{\partial z^{[3]}}} \cdot \boxed{\frac{\partial z^{[3]}}{\partial a^{[2]}}} \cdot \boxed{\frac{\partial a^{[2]}}{\partial z^{[2]}}} \cdot \boxed{\frac{\partial z^{[2]}}{\partial w^{[2]}}}$$

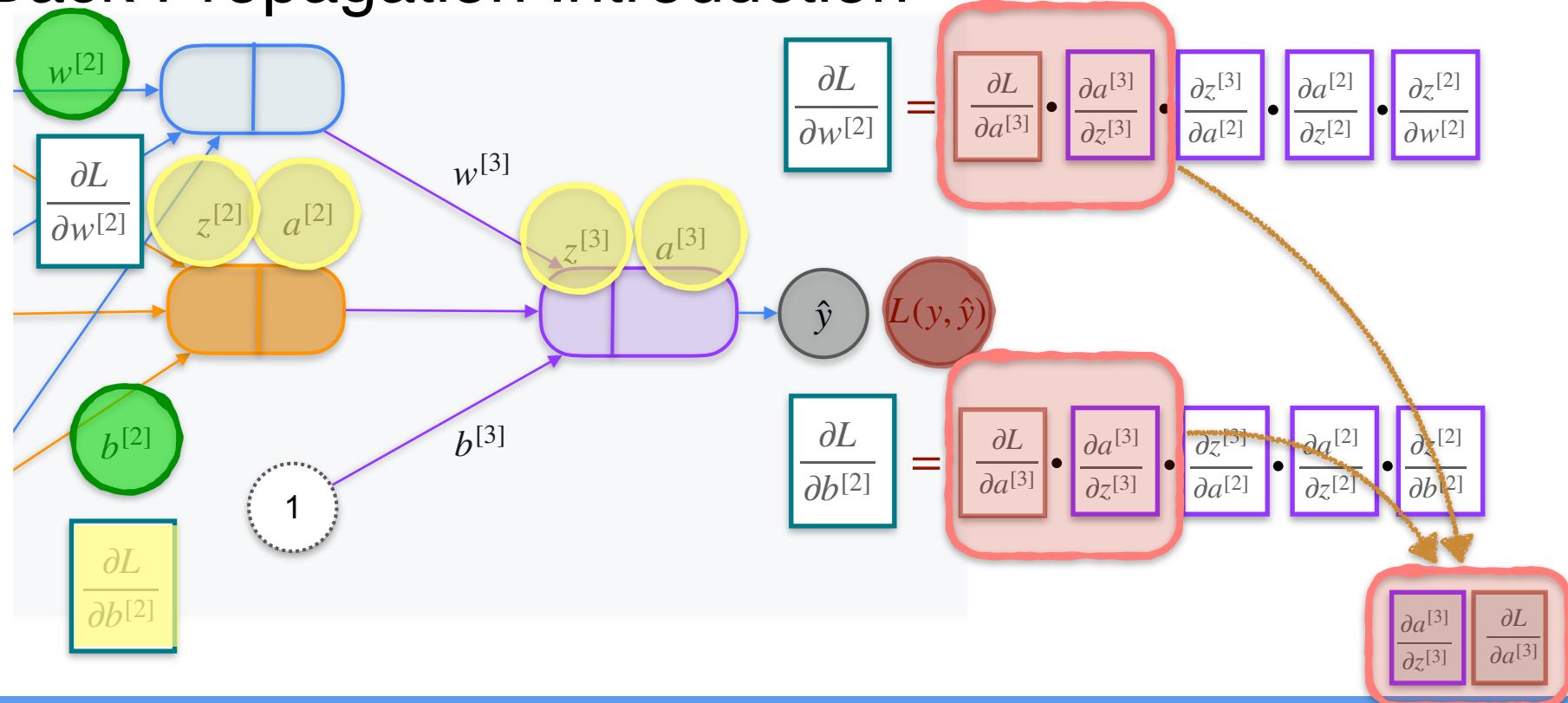
$$\frac{\partial L}{\partial b^{[2]}} = \boxed{\frac{\partial L}{\partial a^{[3]}}} \cdot \boxed{\frac{\partial a^{[3]}}{\partial z^{[3]}}} \cdot \boxed{\frac{\partial z^{[3]}}{\partial a^{[2]}}} \cdot \boxed{\frac{\partial a^{[2]}}{\partial z^{[2]}}} \cdot \boxed{\frac{\partial z^{[2]}}{\partial b^{[2]}}}$$

$$\frac{\partial a^{[3]}}{\partial z^{[3]}} \quad \frac{\partial L}{\partial a^{[3]}}$$

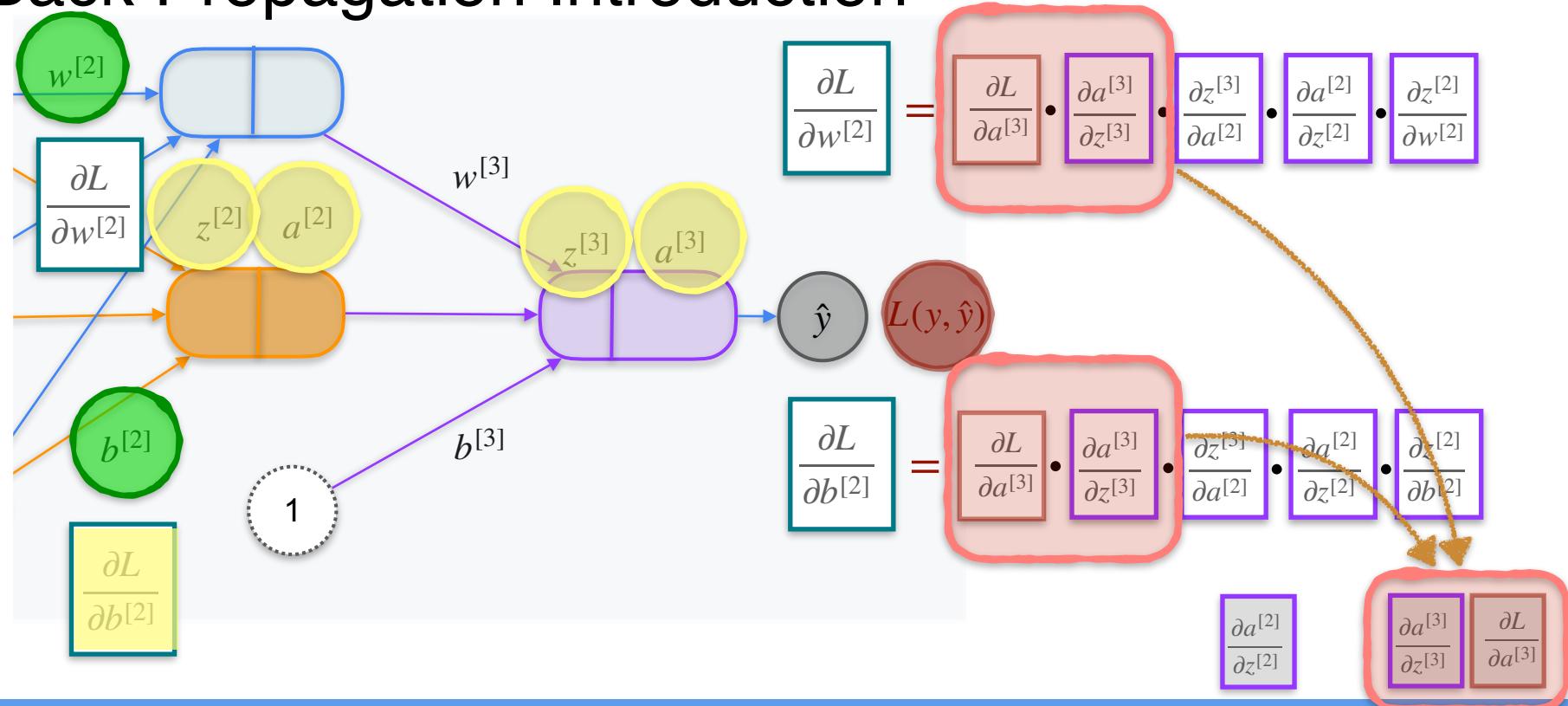
# Back Propagation Introduction



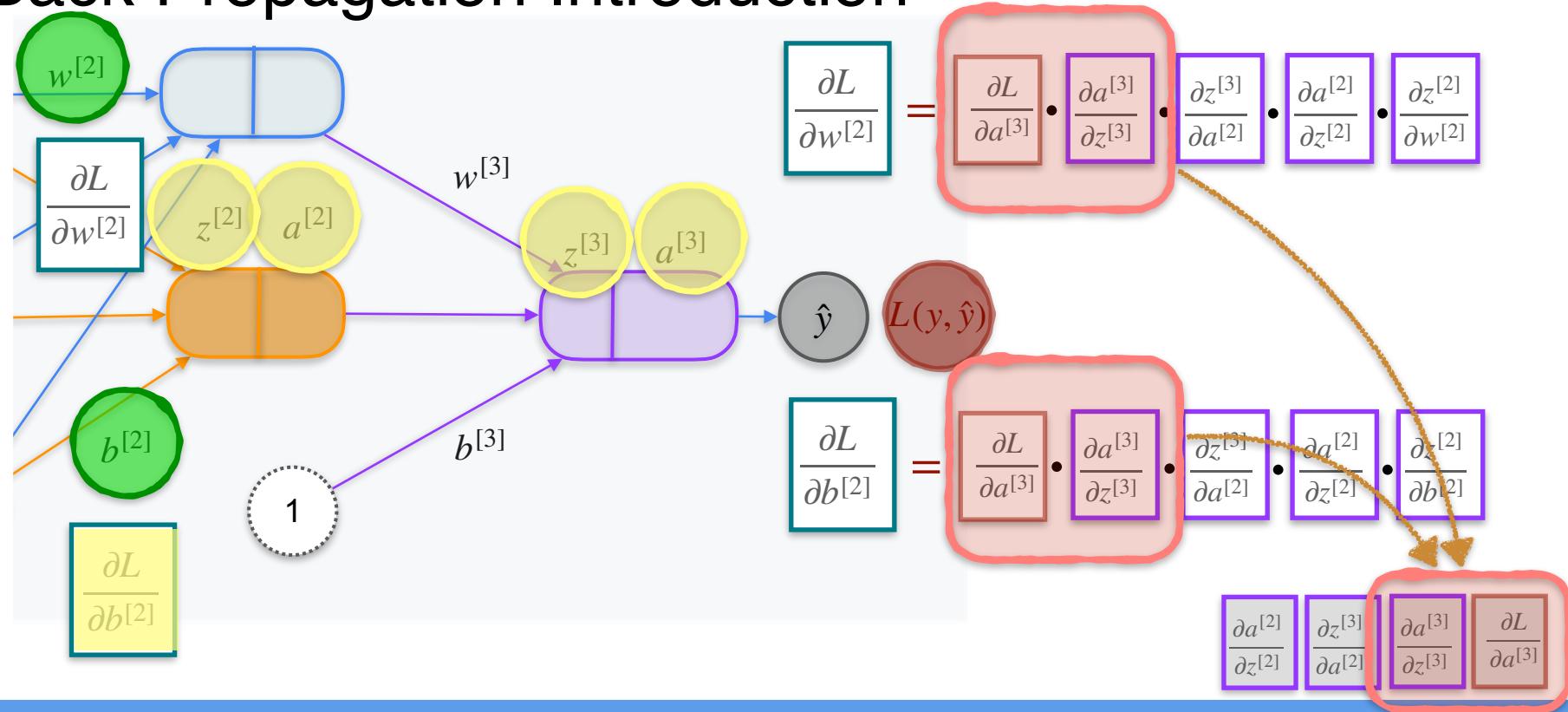
# Back Propagation Introduction



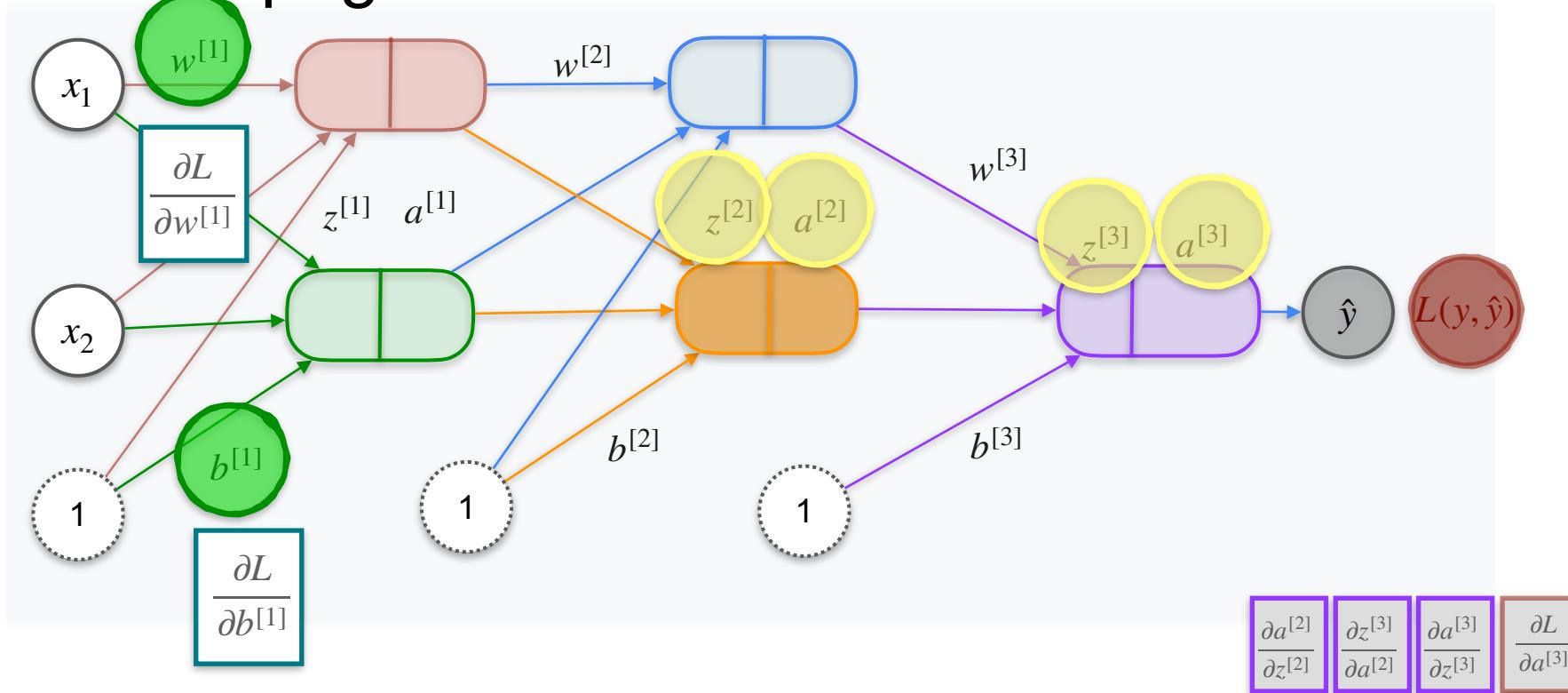
# Back Propagation Introduction



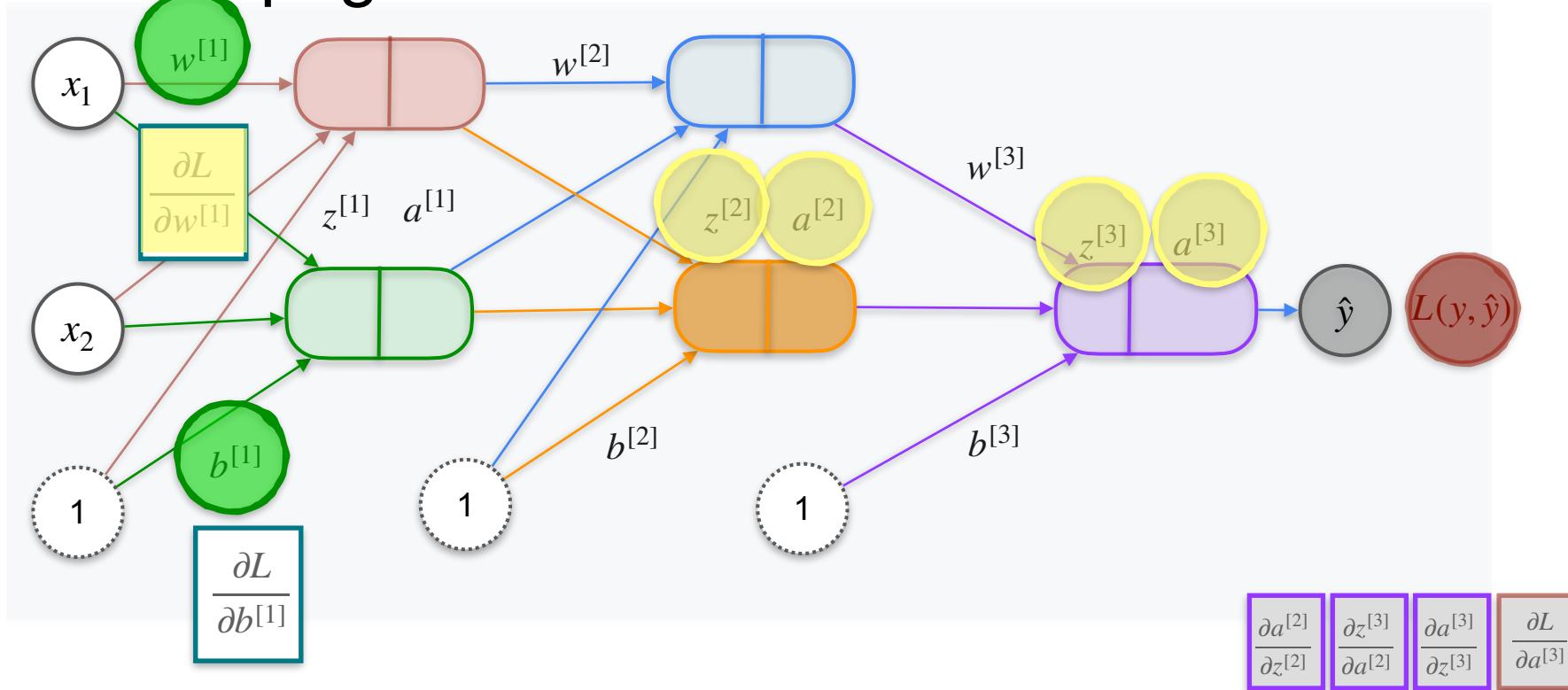
# Back Propagation Introduction



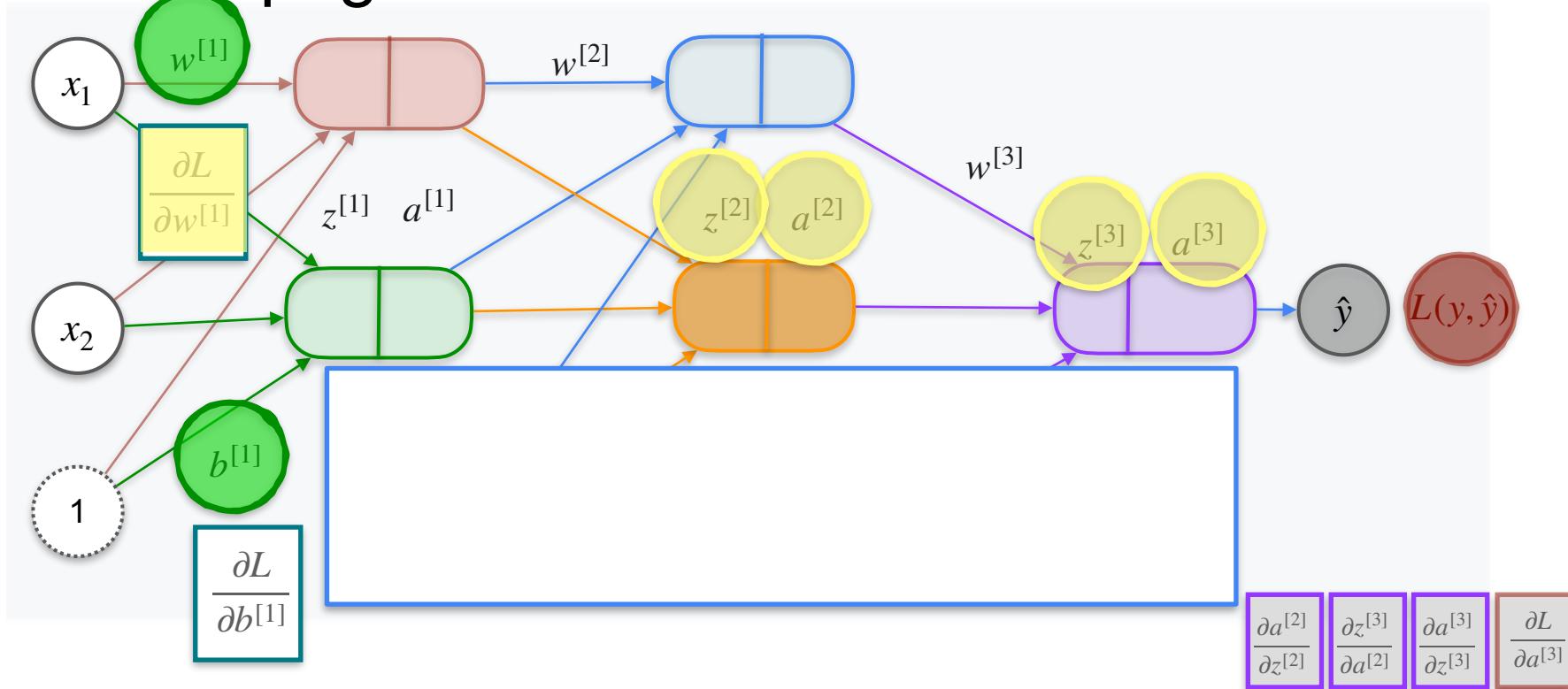
# Back Propagation Introduction



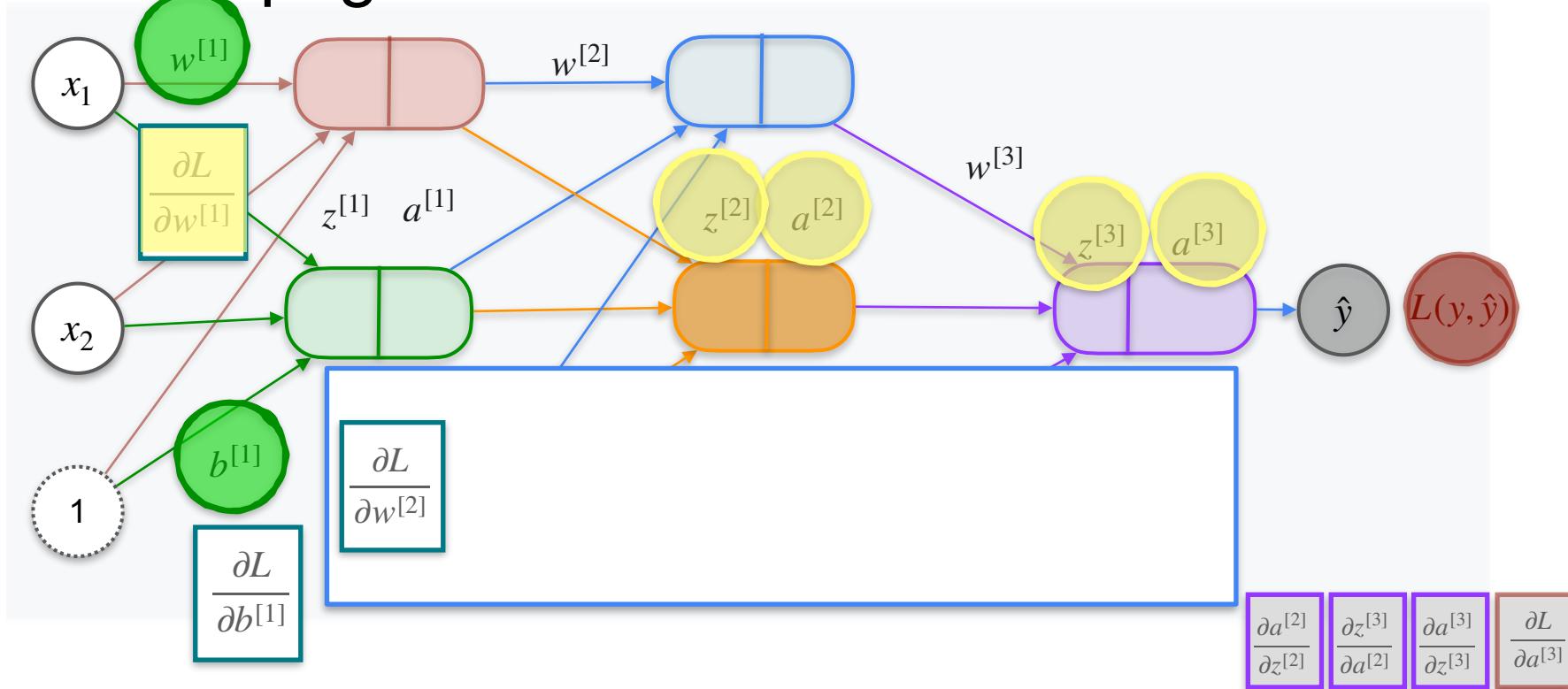
# Back Propagation Introduction



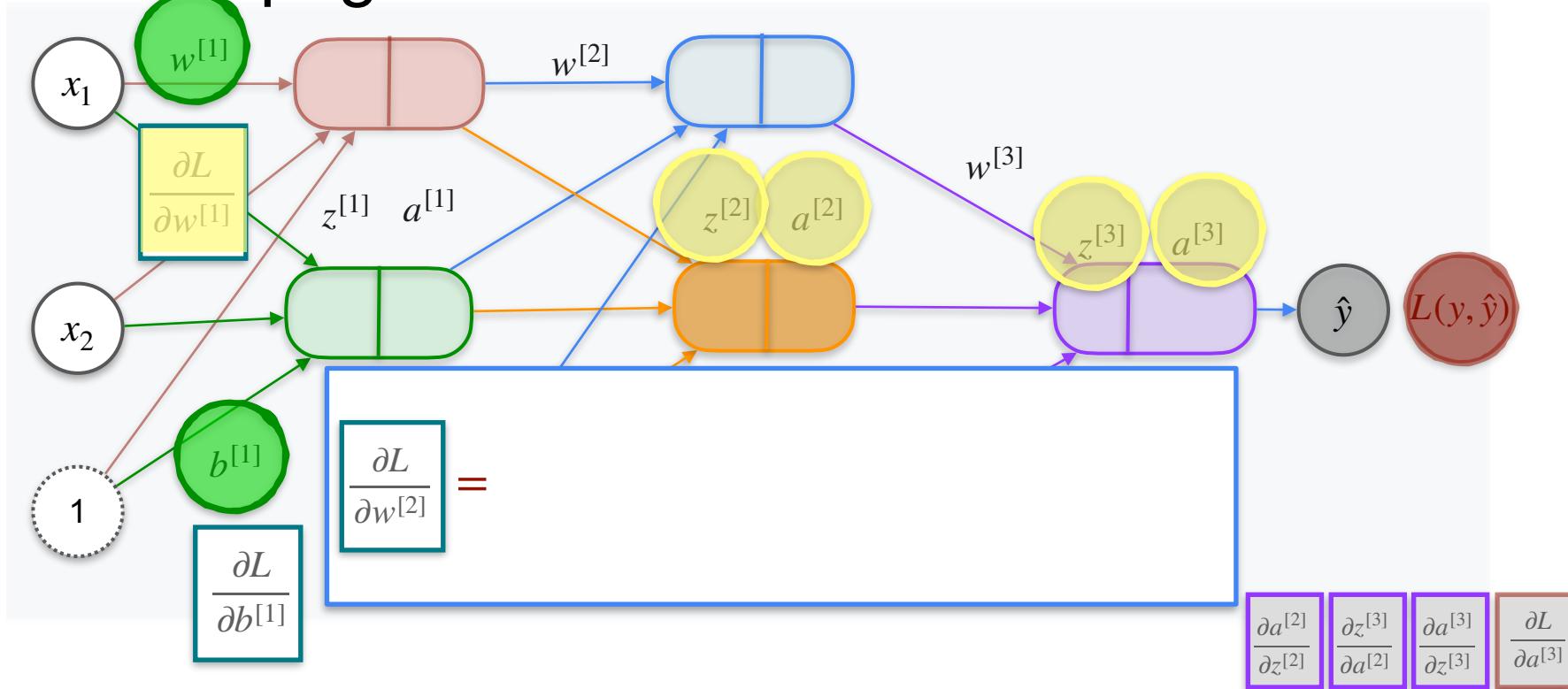
# Back Propagation Introduction



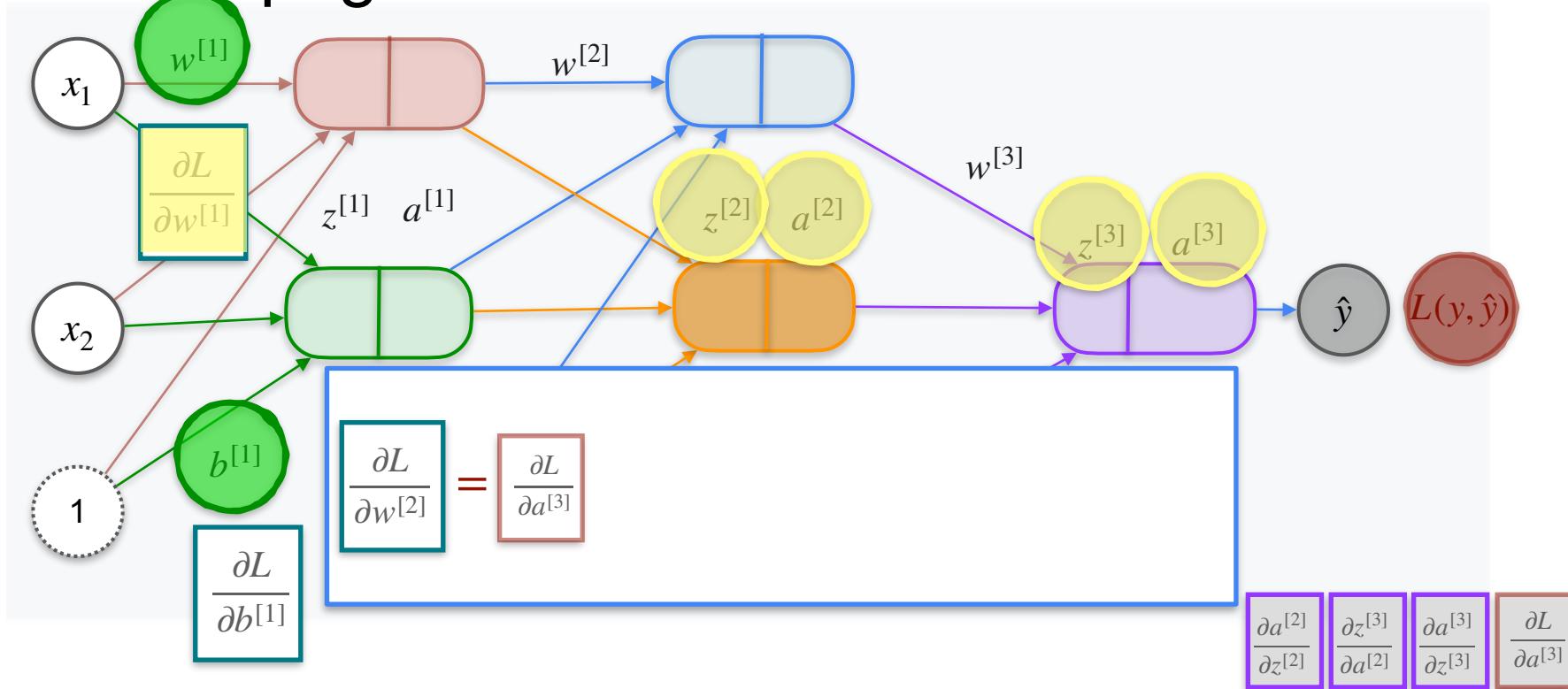
# Back Propagation Introduction



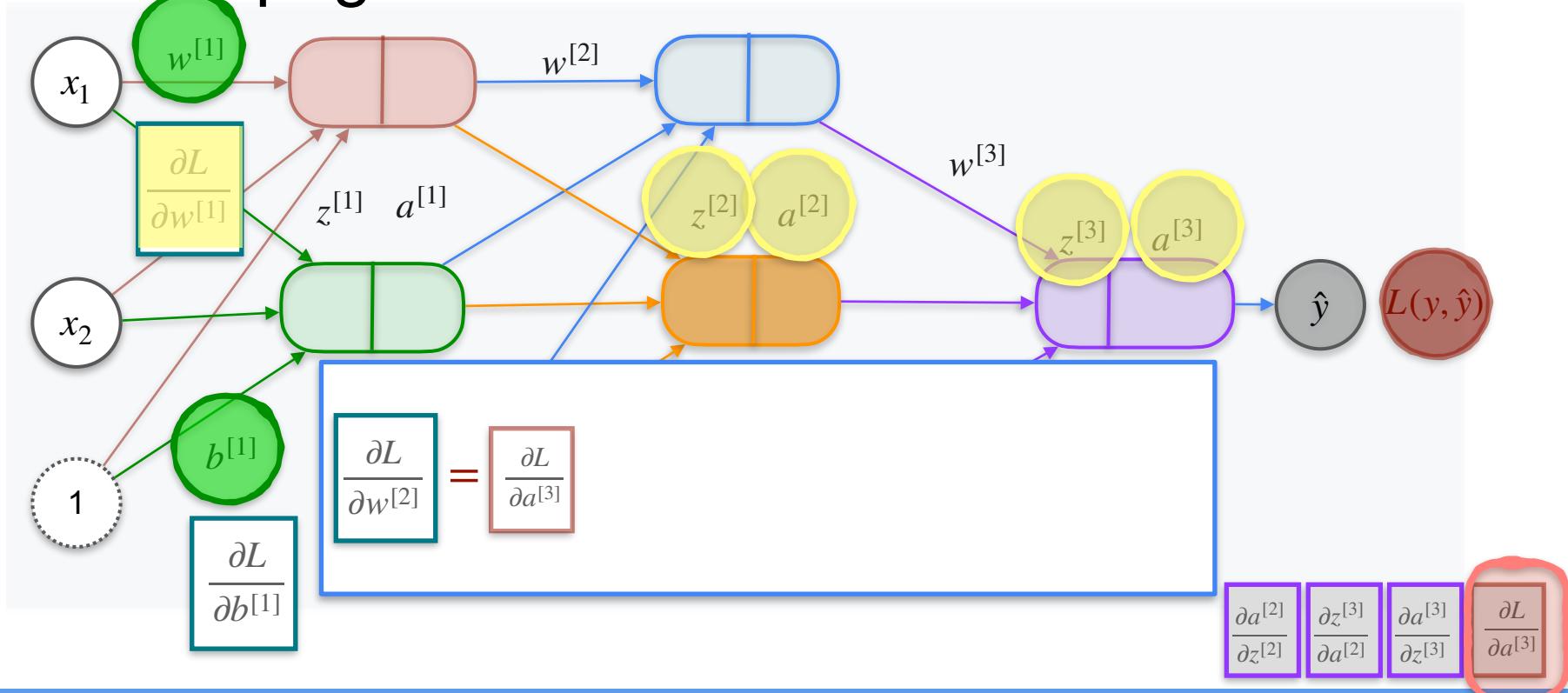
# Back Propagation Introduction



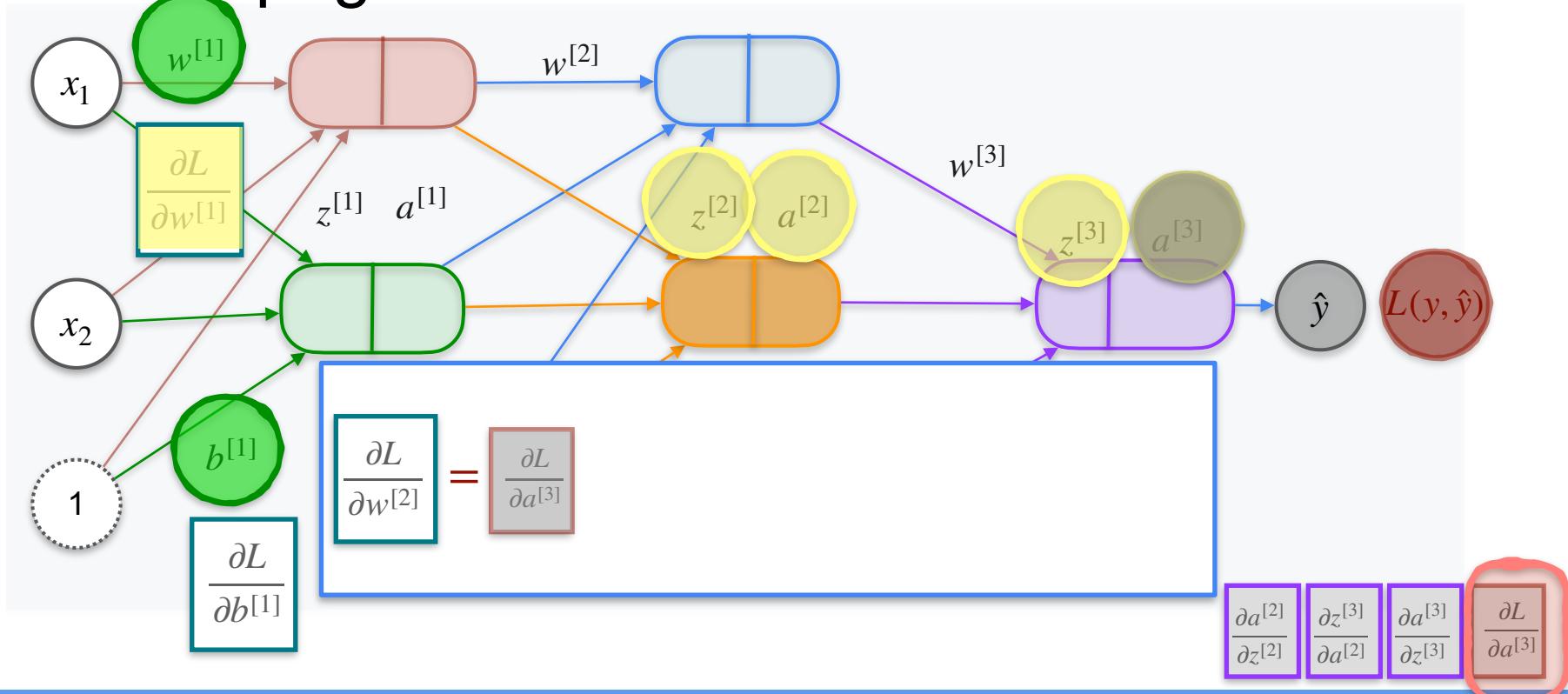
# Back Propagation Introduction



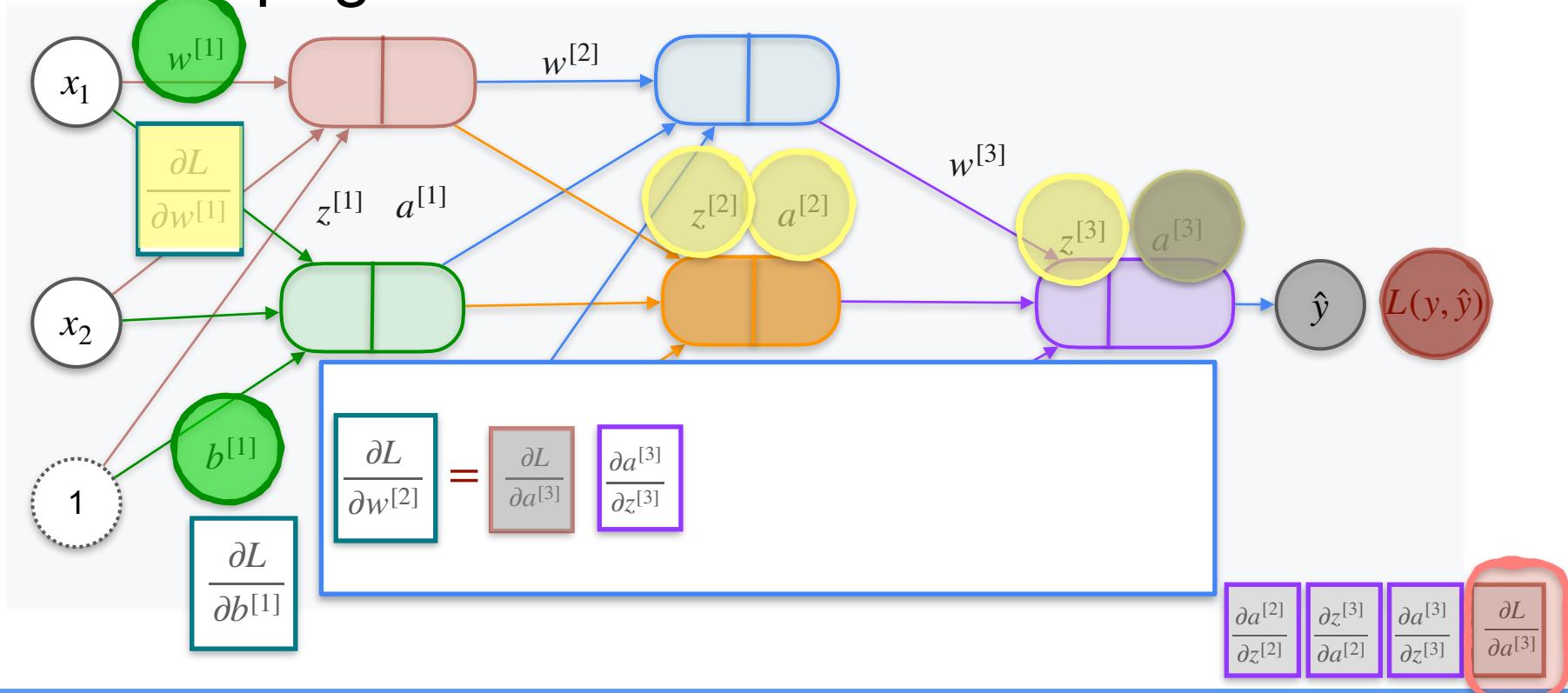
# Back Propagation Introduction



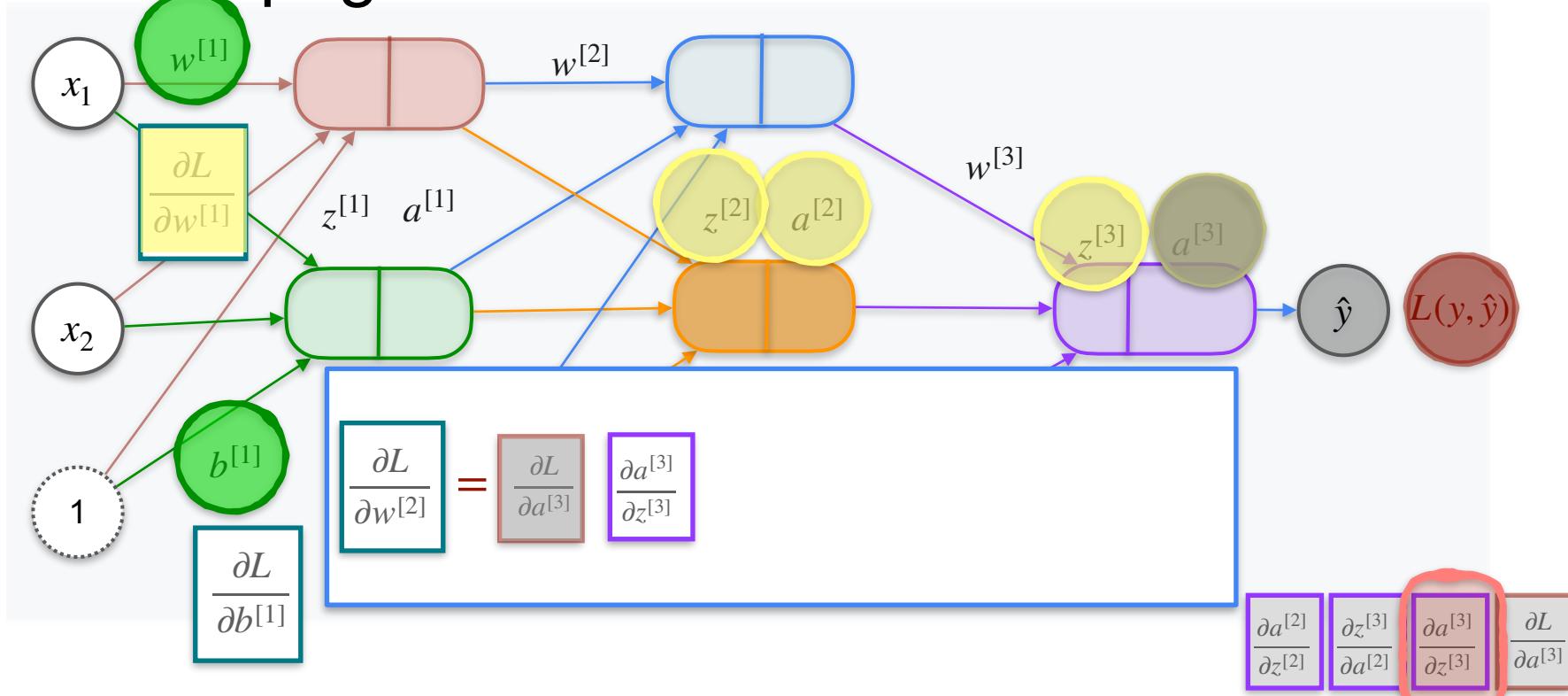
# Back Propagation Introduction



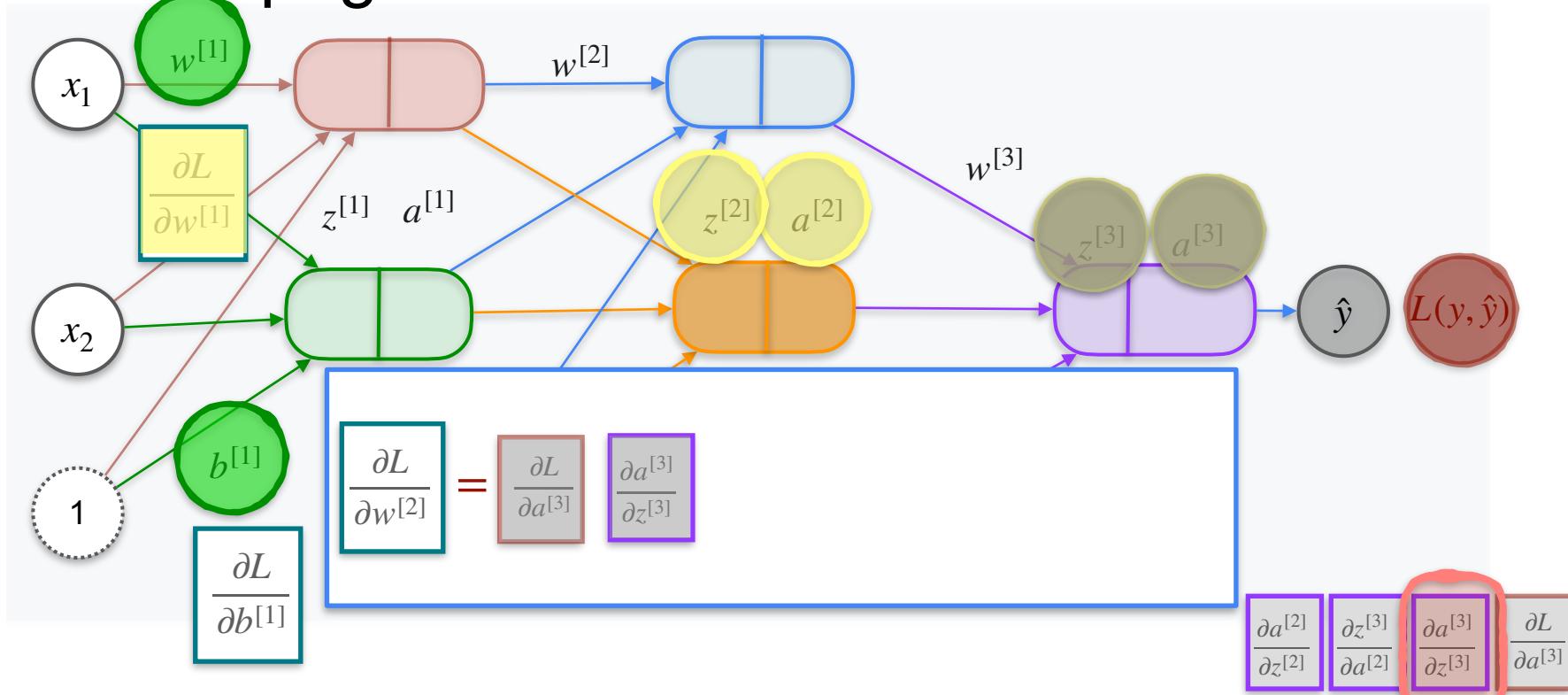
# Back Propagation Introduction



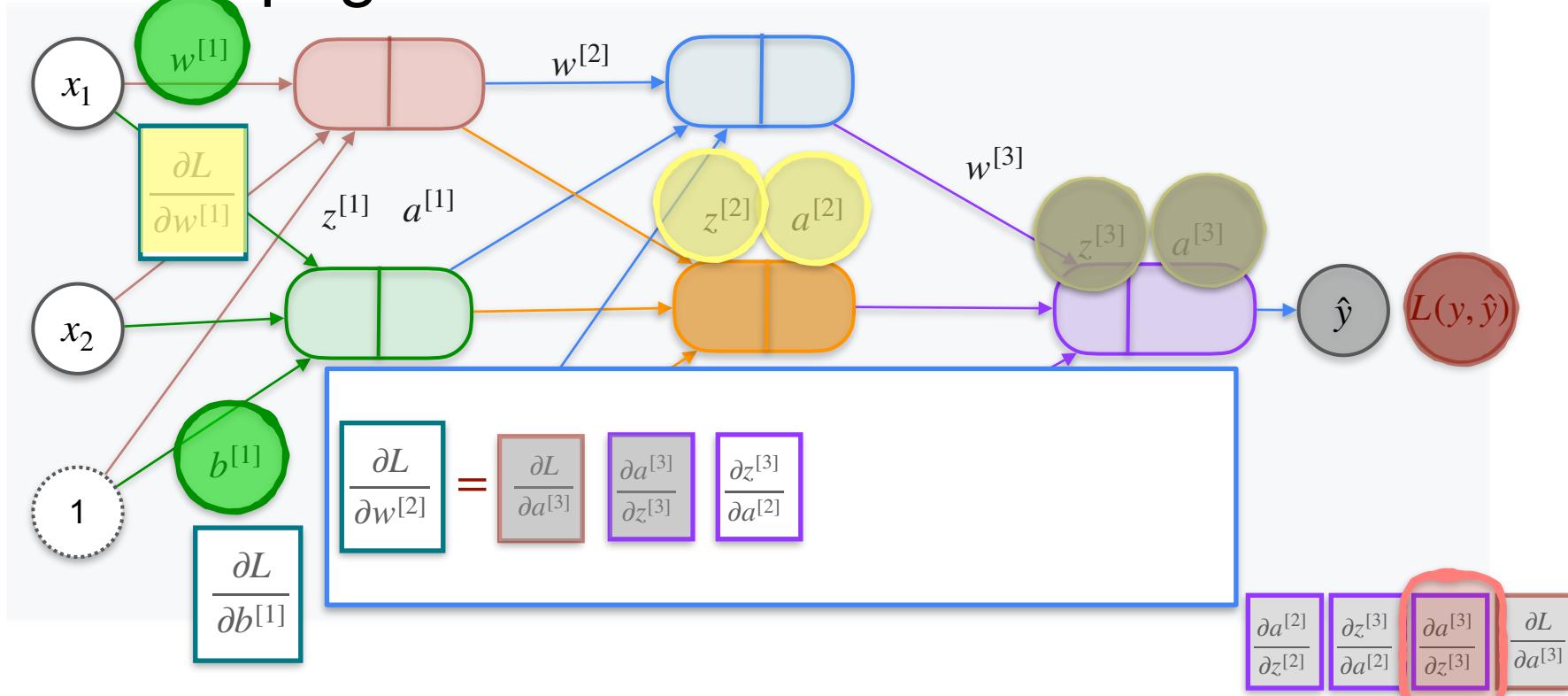
# Back Propagation Introduction



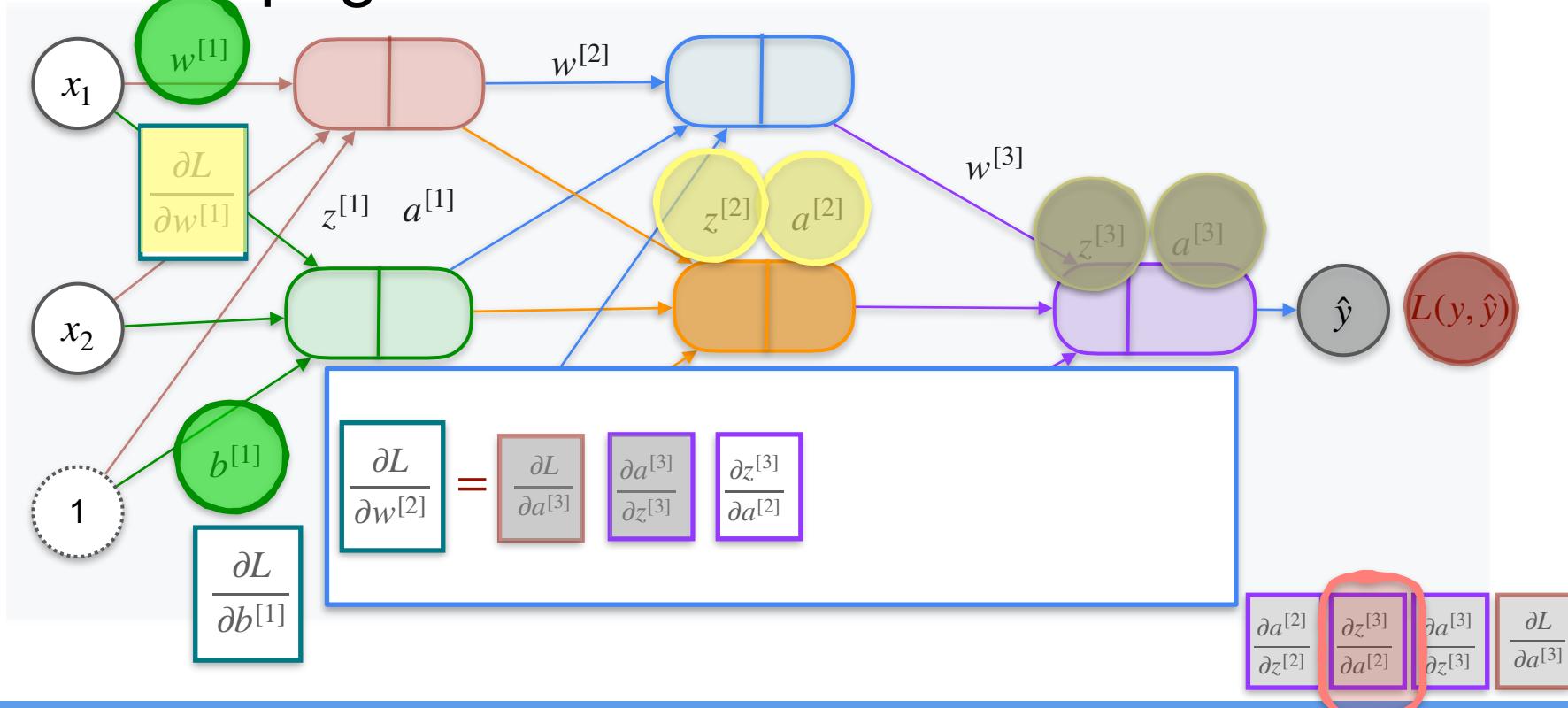
# Back Propagation Introduction



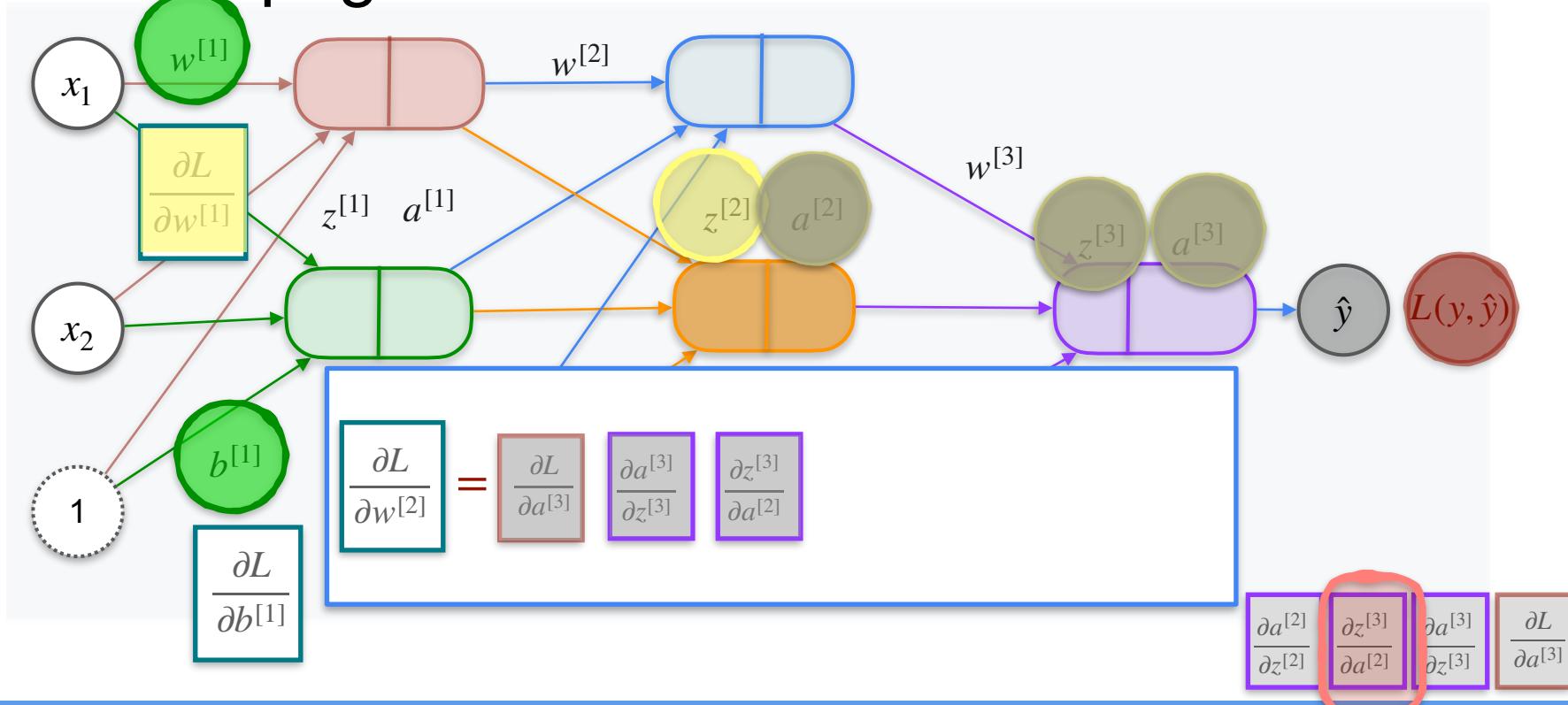
# Back Propagation Introduction



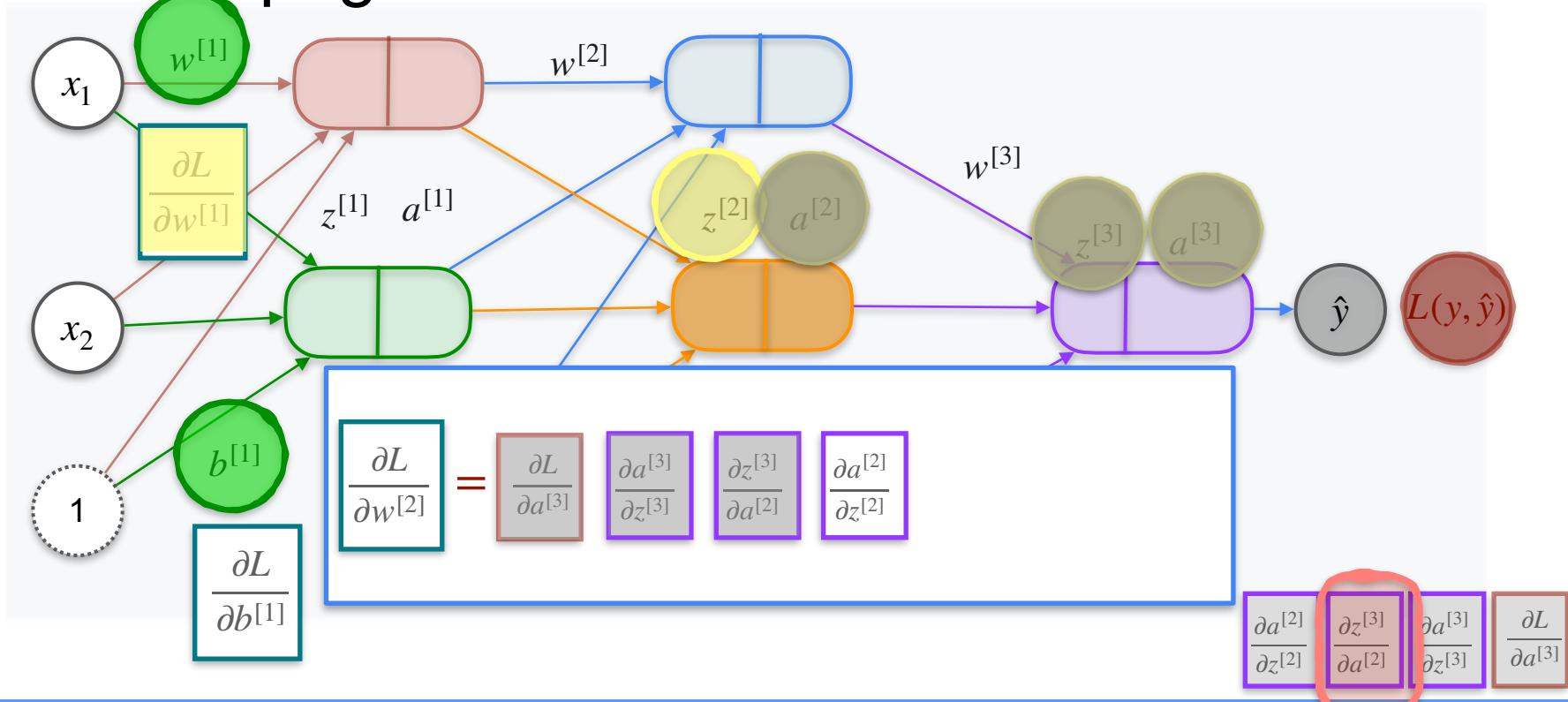
# Back Propagation Introduction



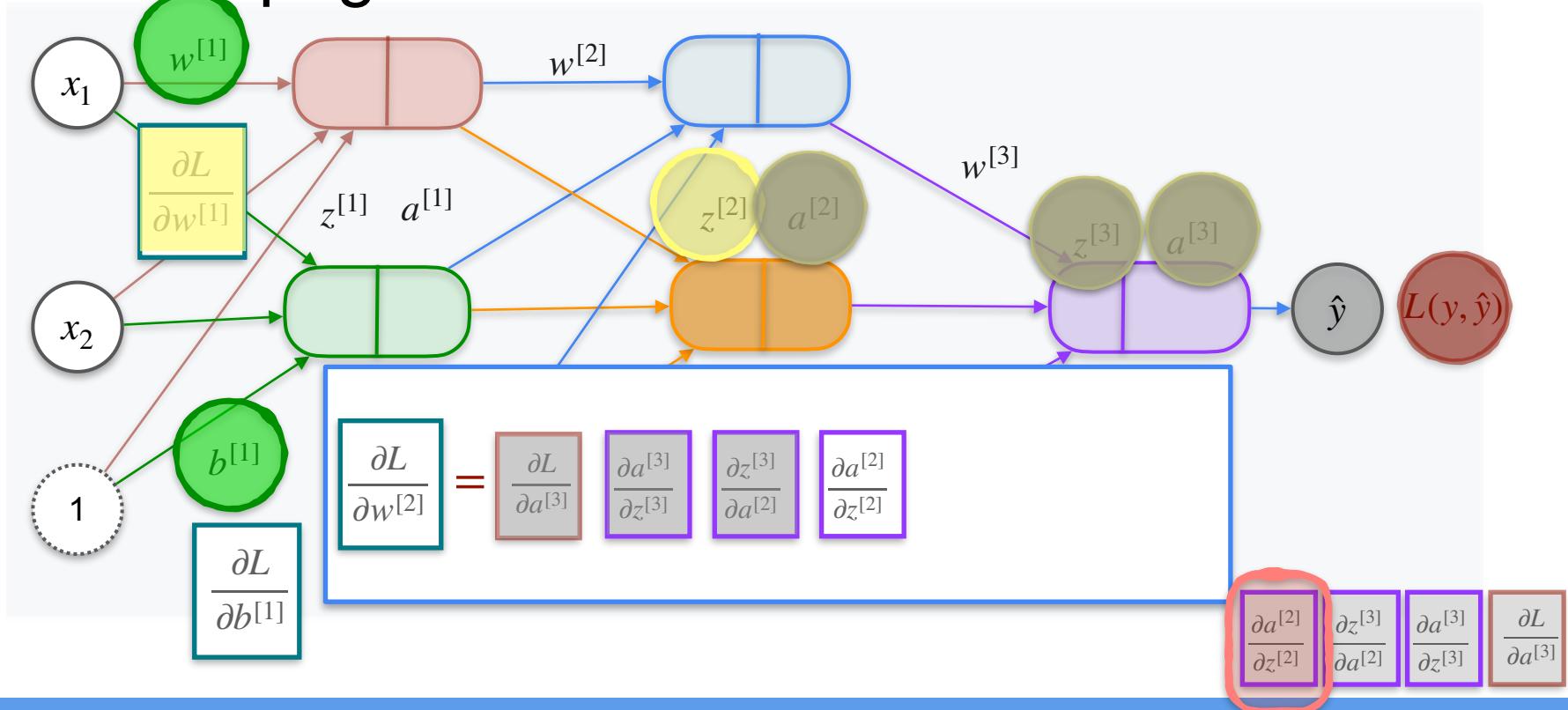
# Back Propagation Introduction



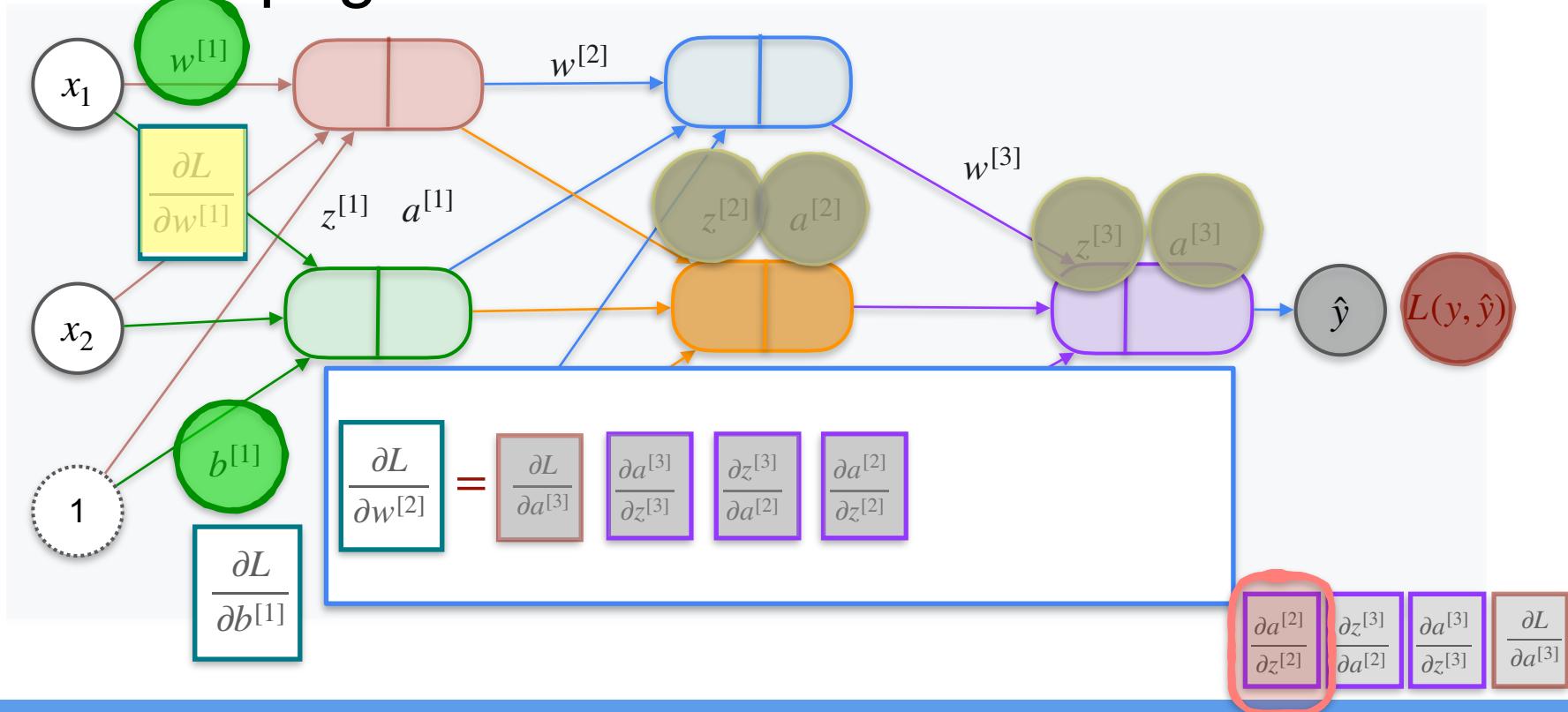
# Back Propagation Introduction



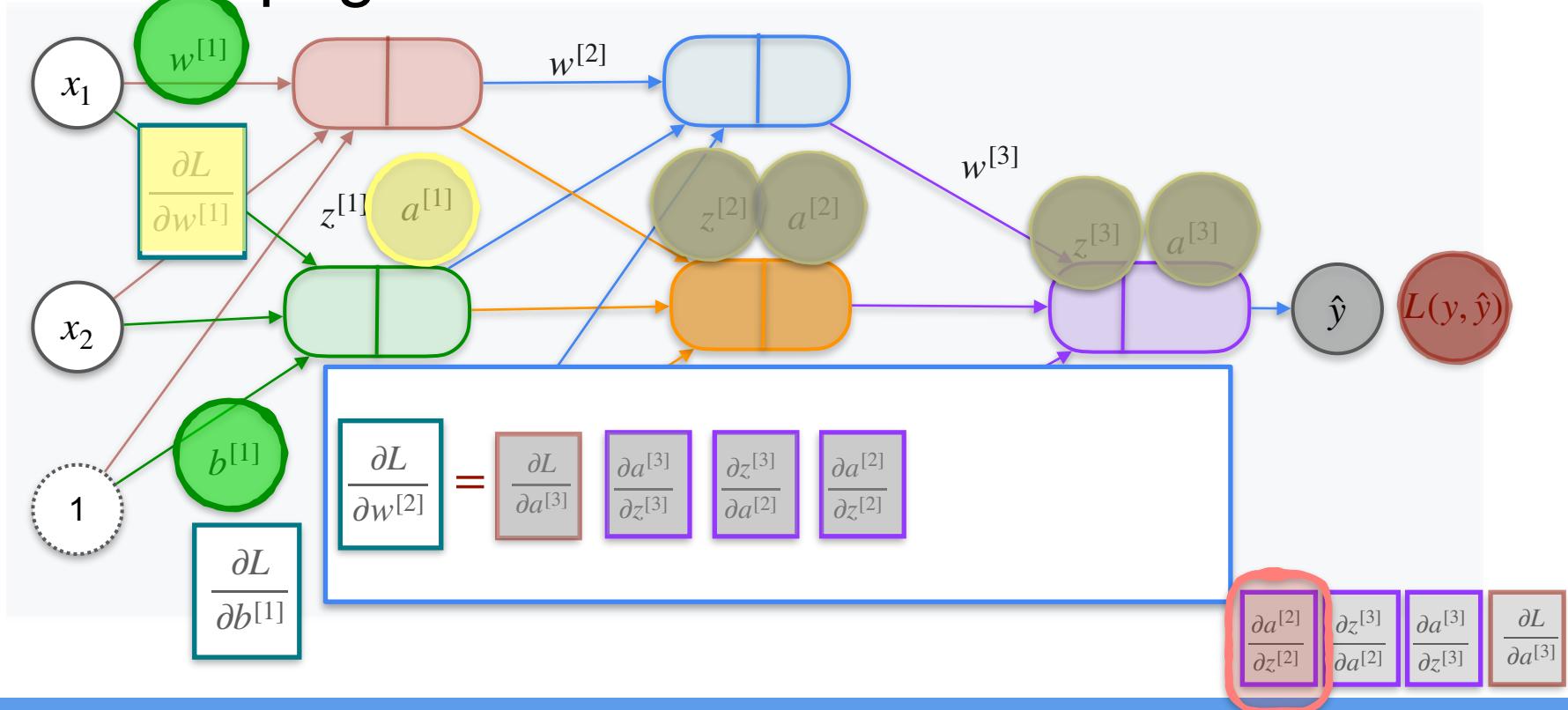
# Back Propagation Introduction



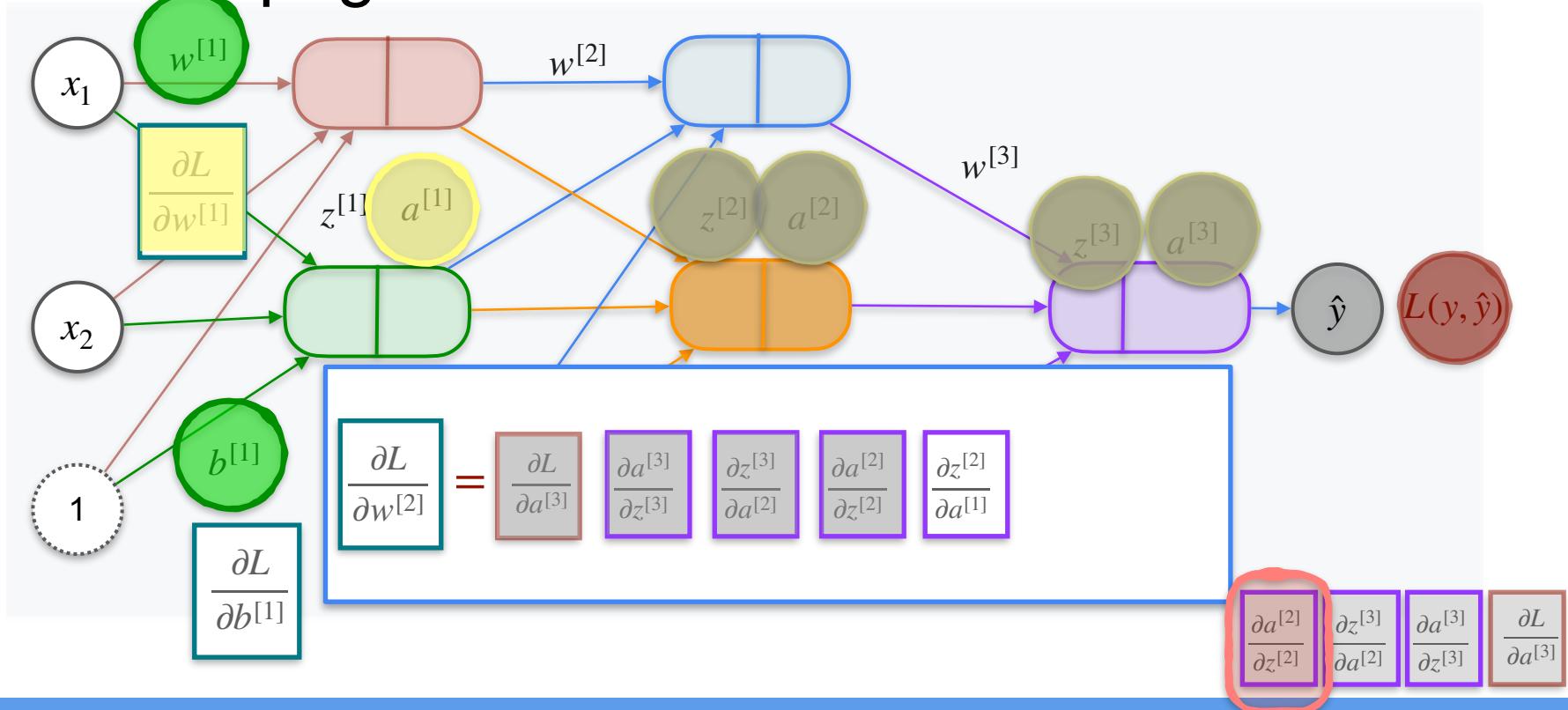
# Back Propagation Introduction



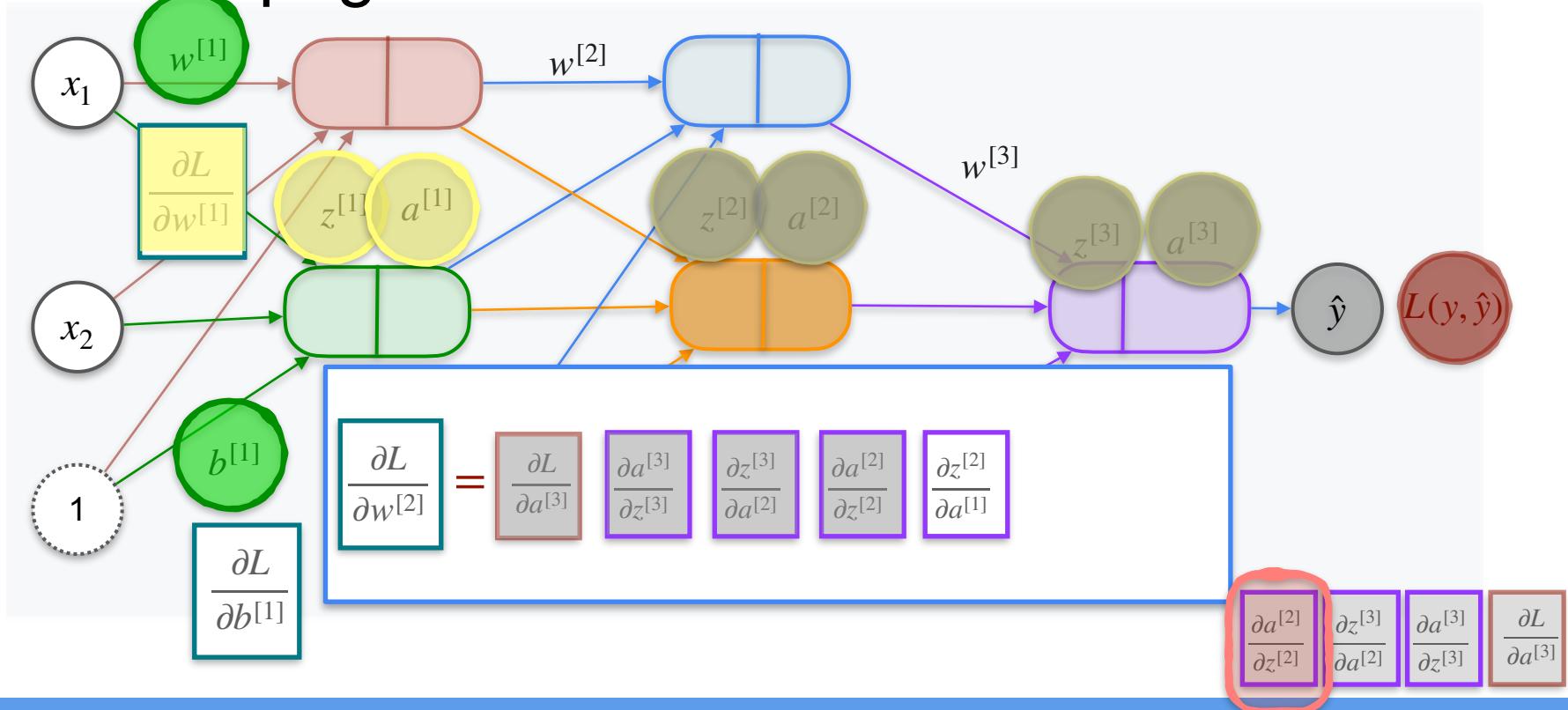
# Back Propagation Introduction



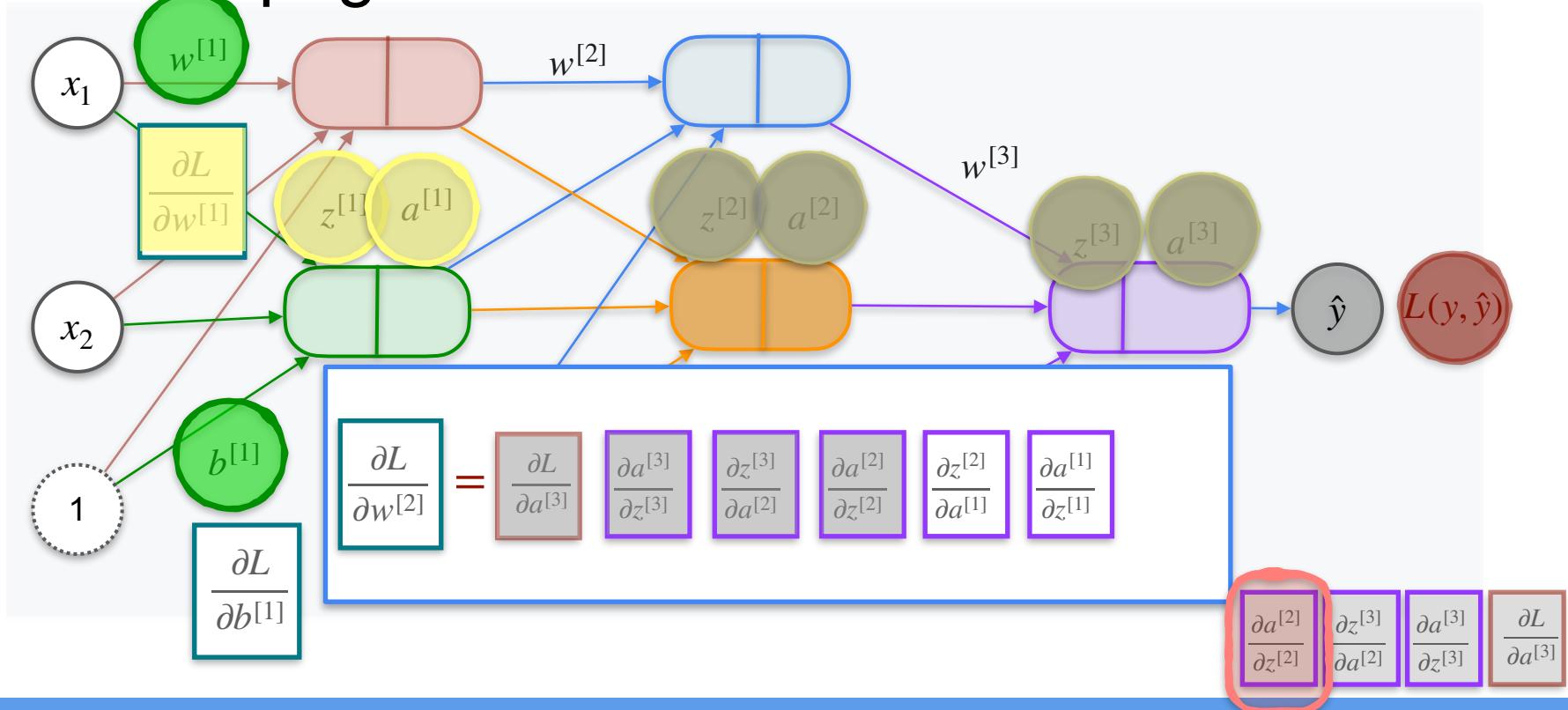
# Back Propagation Introduction



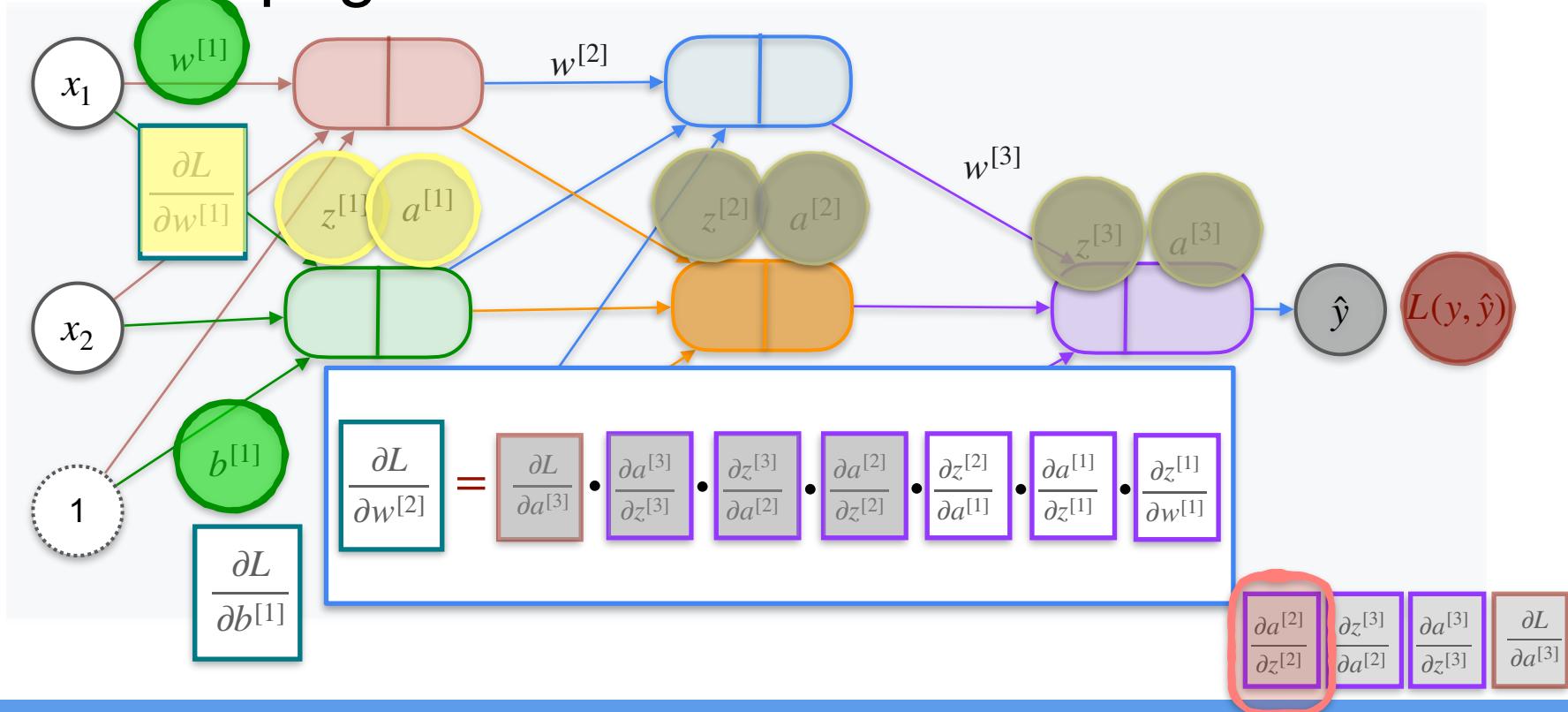
# Back Propagation Introduction



# Back Propagation Introduction



# Back Propagation Introduction





DeepLearning.AI

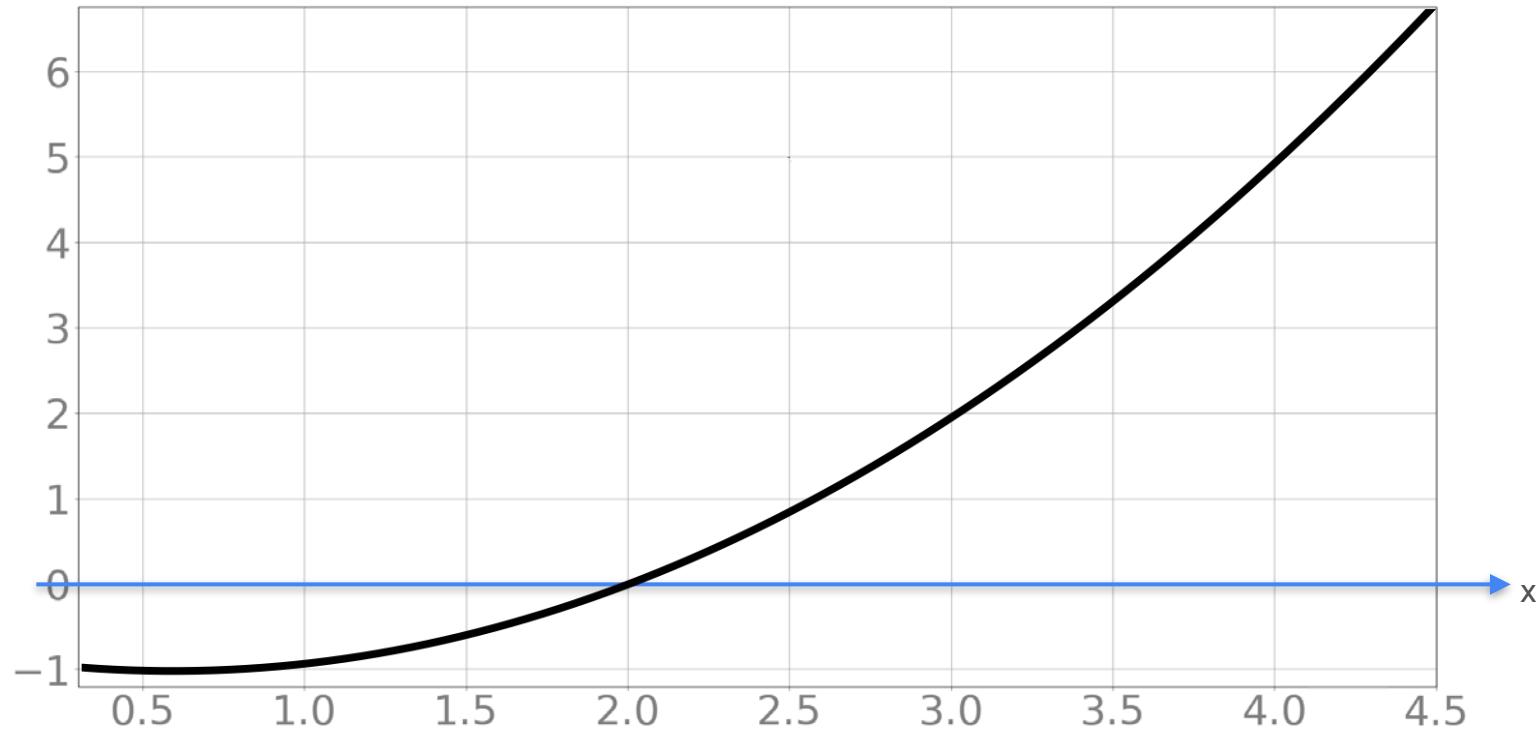
# Optimization in Neural Networks and Newton's Method

---

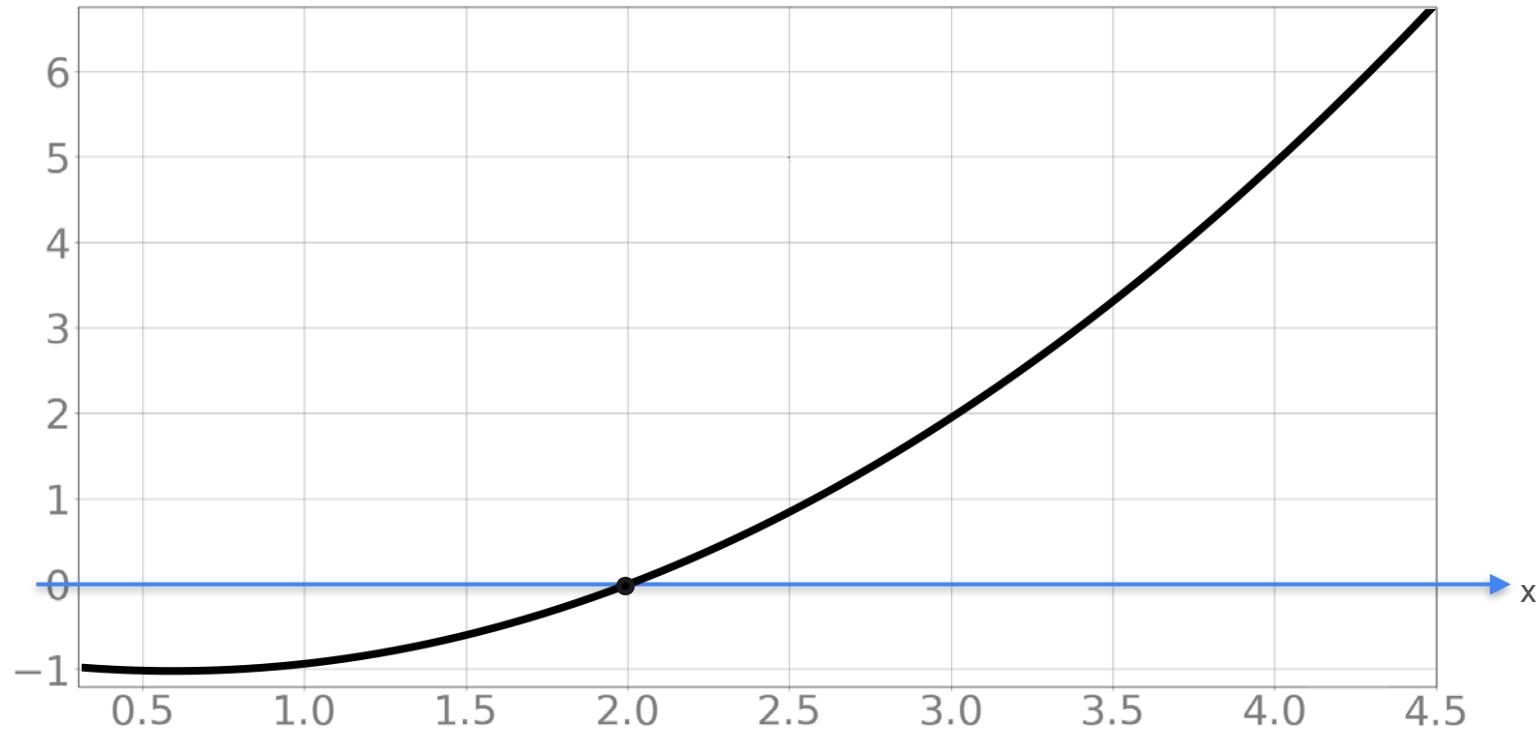
## Newton's method

# Newton's Method

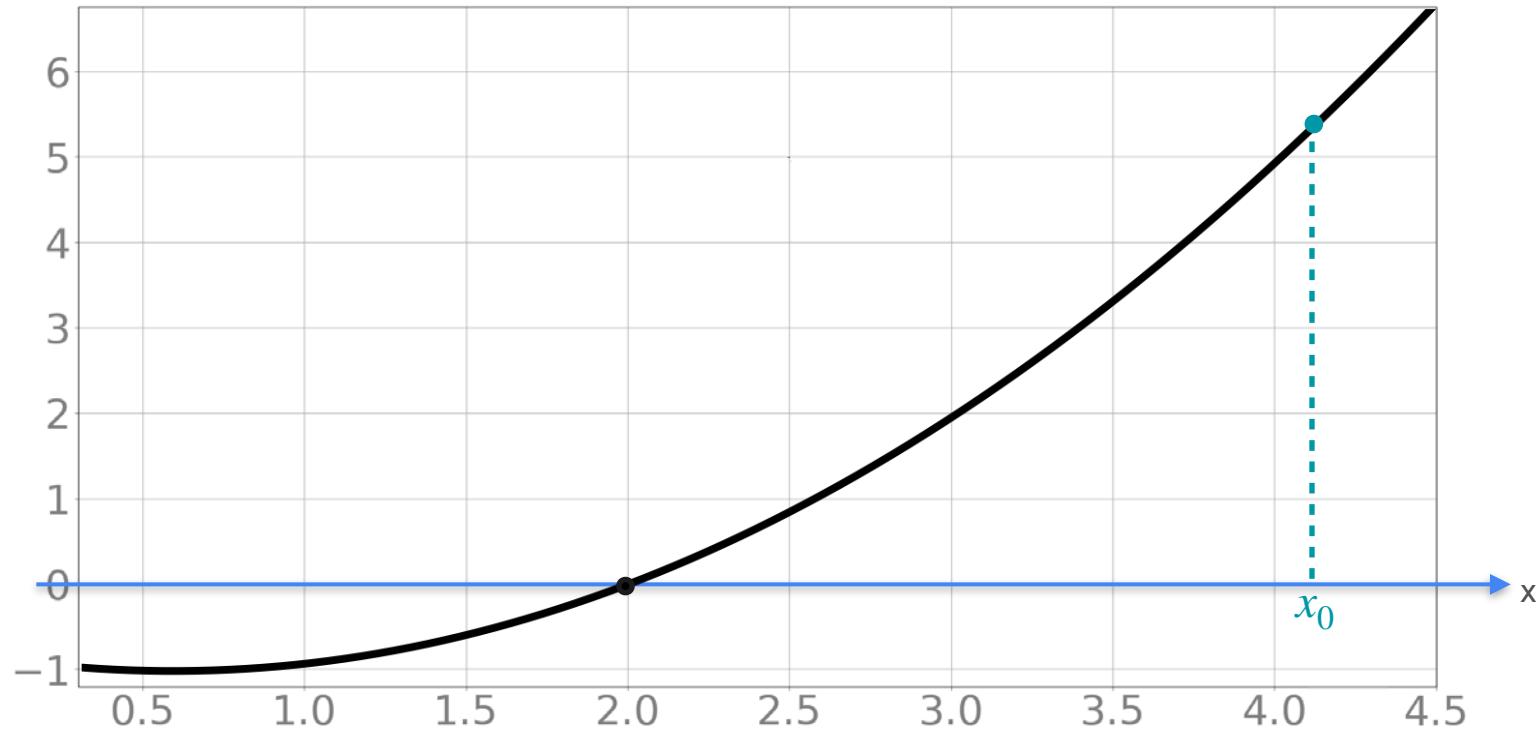
# Newton's Method



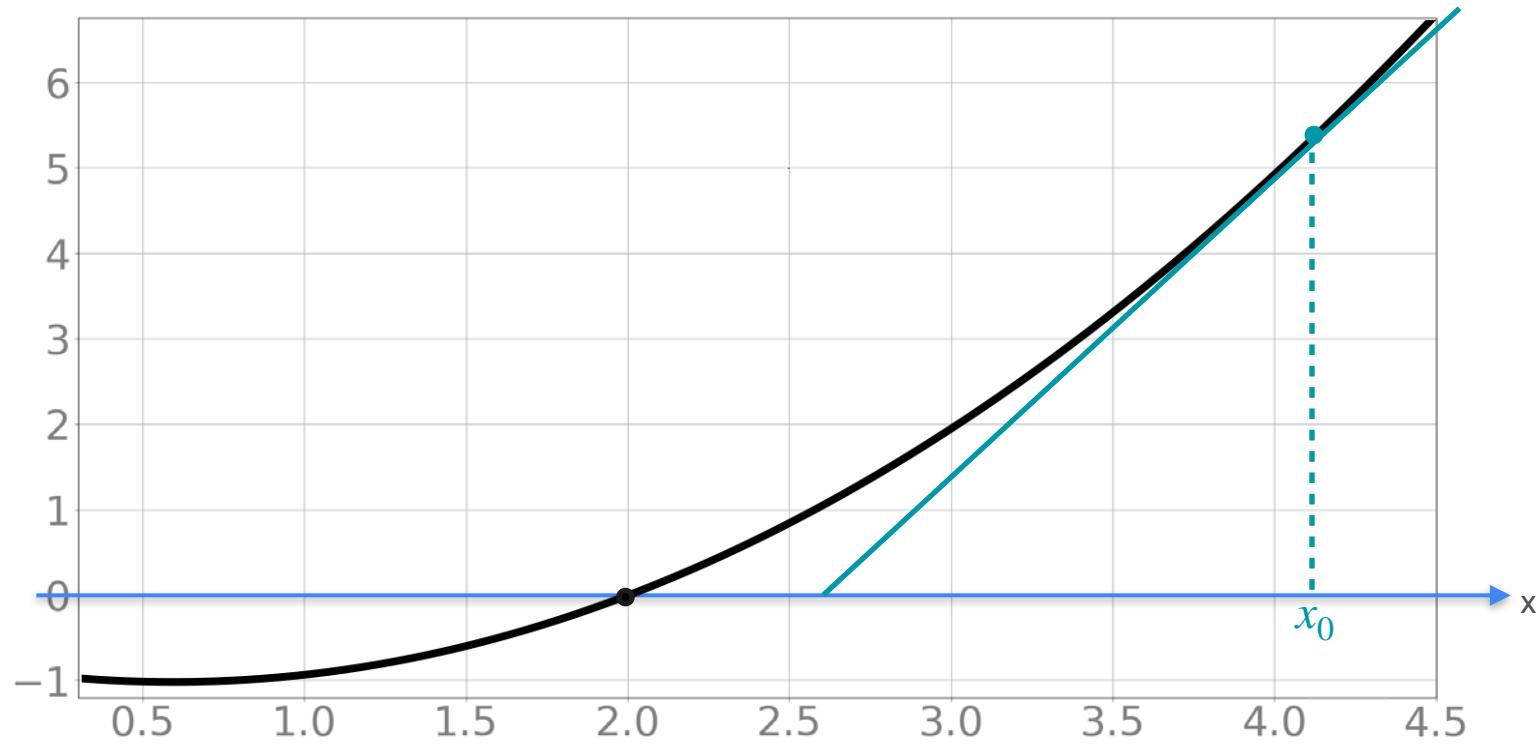
# Newton's Method



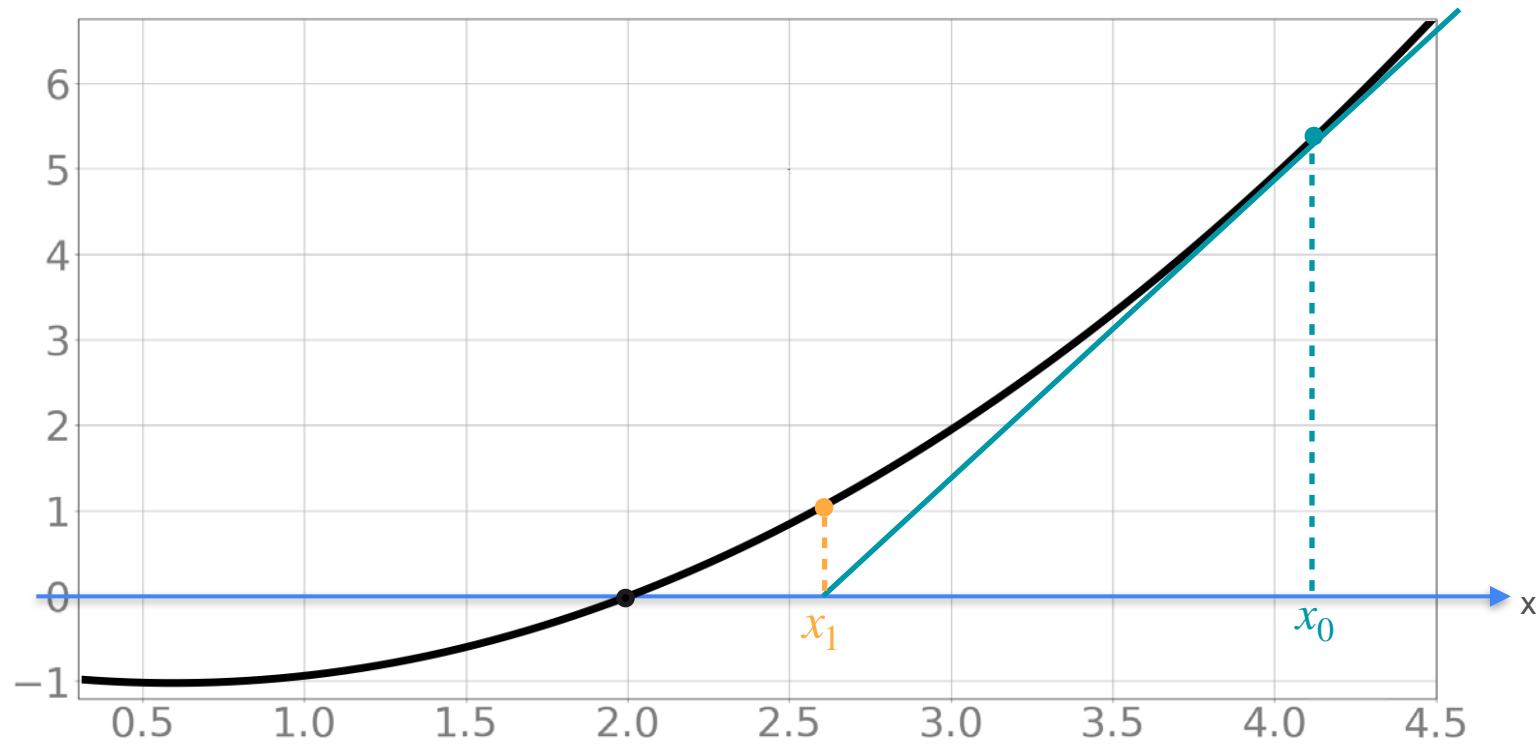
# Newton's Method



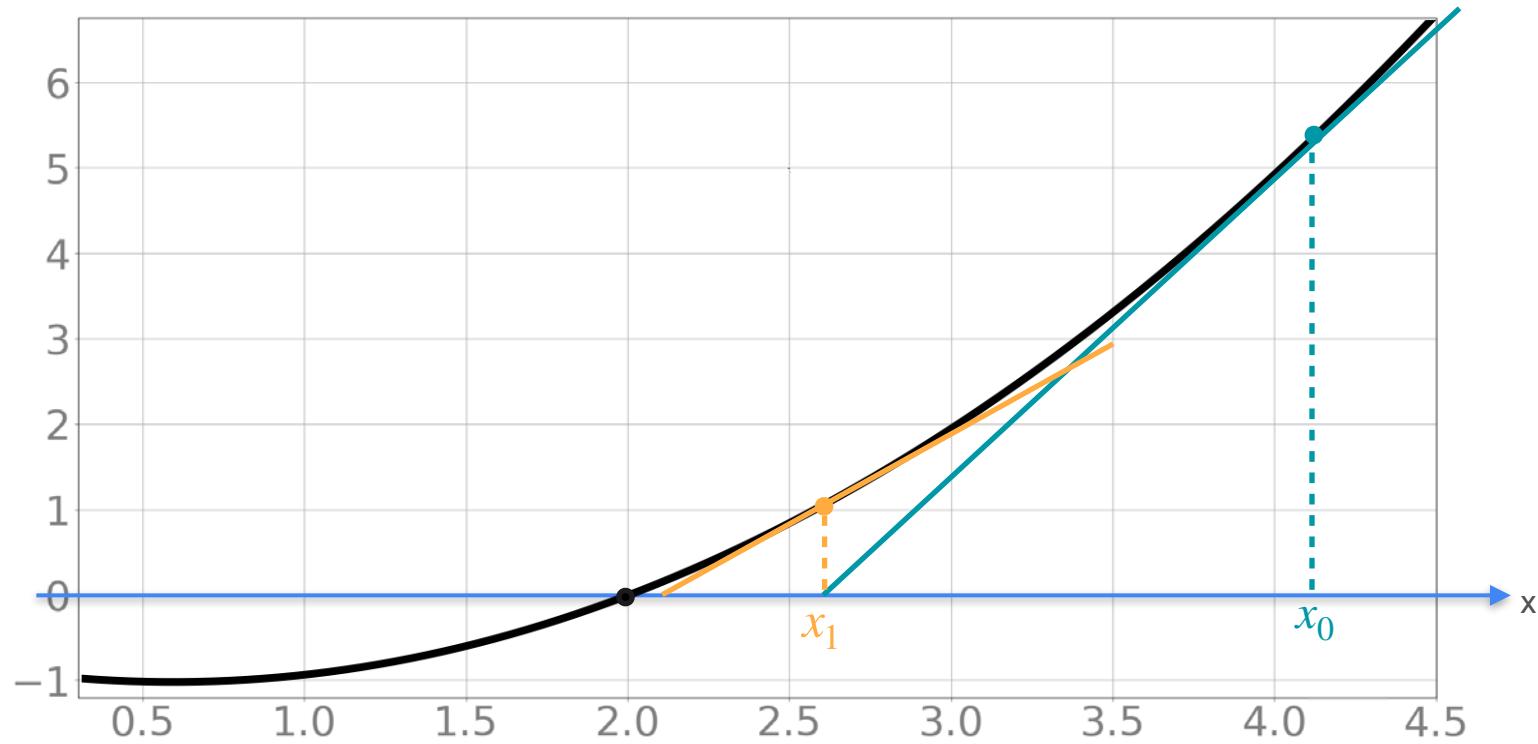
# Newton's Method



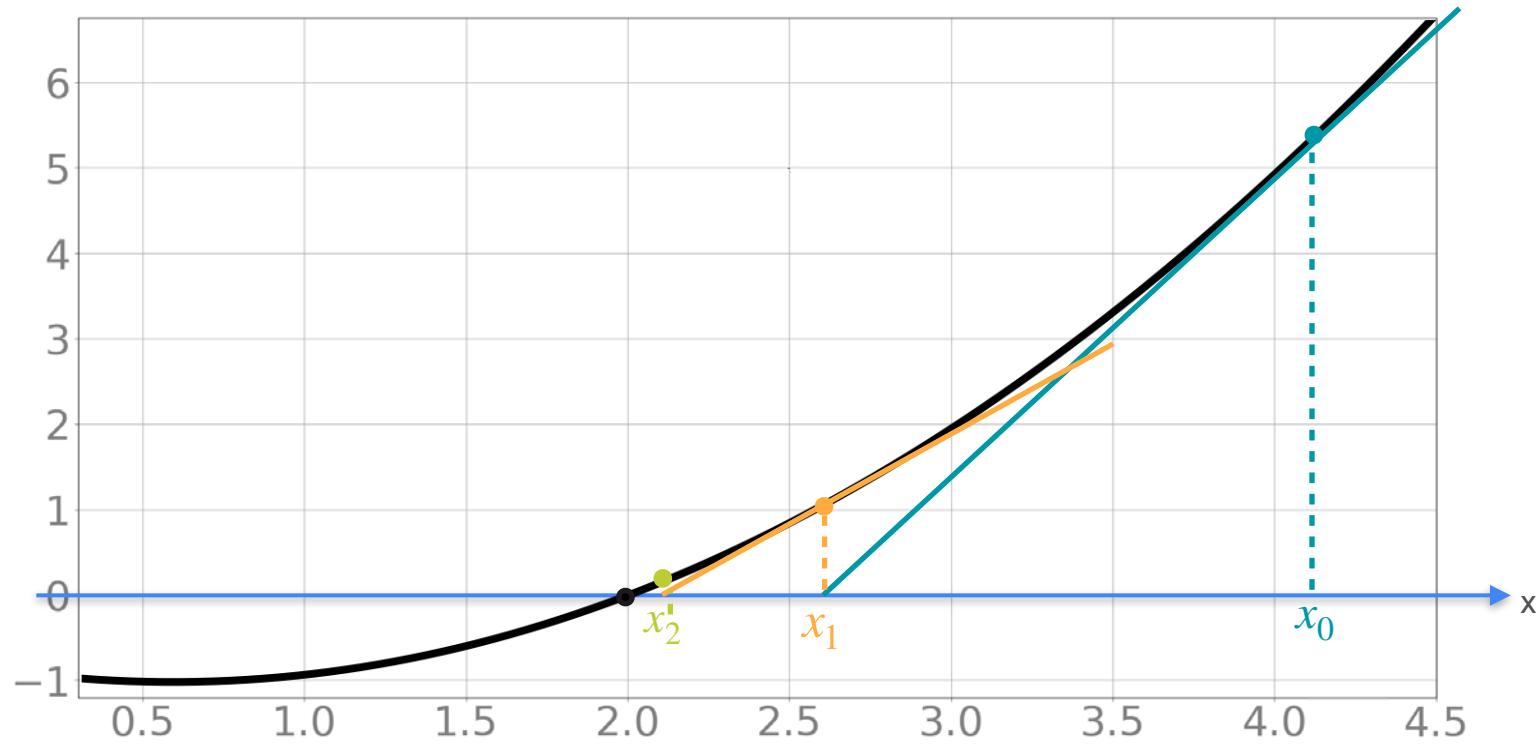
# Newton's Method



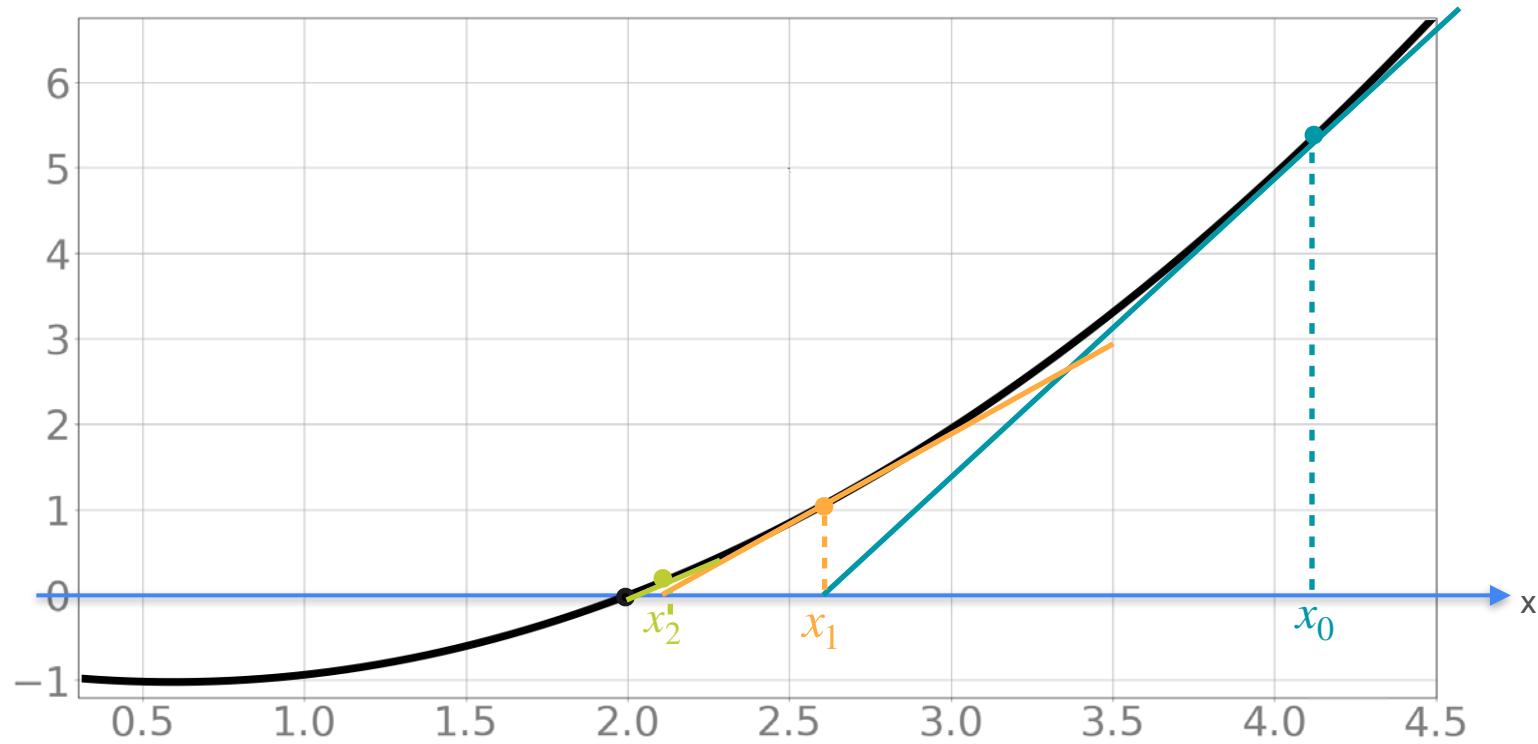
# Newton's Method



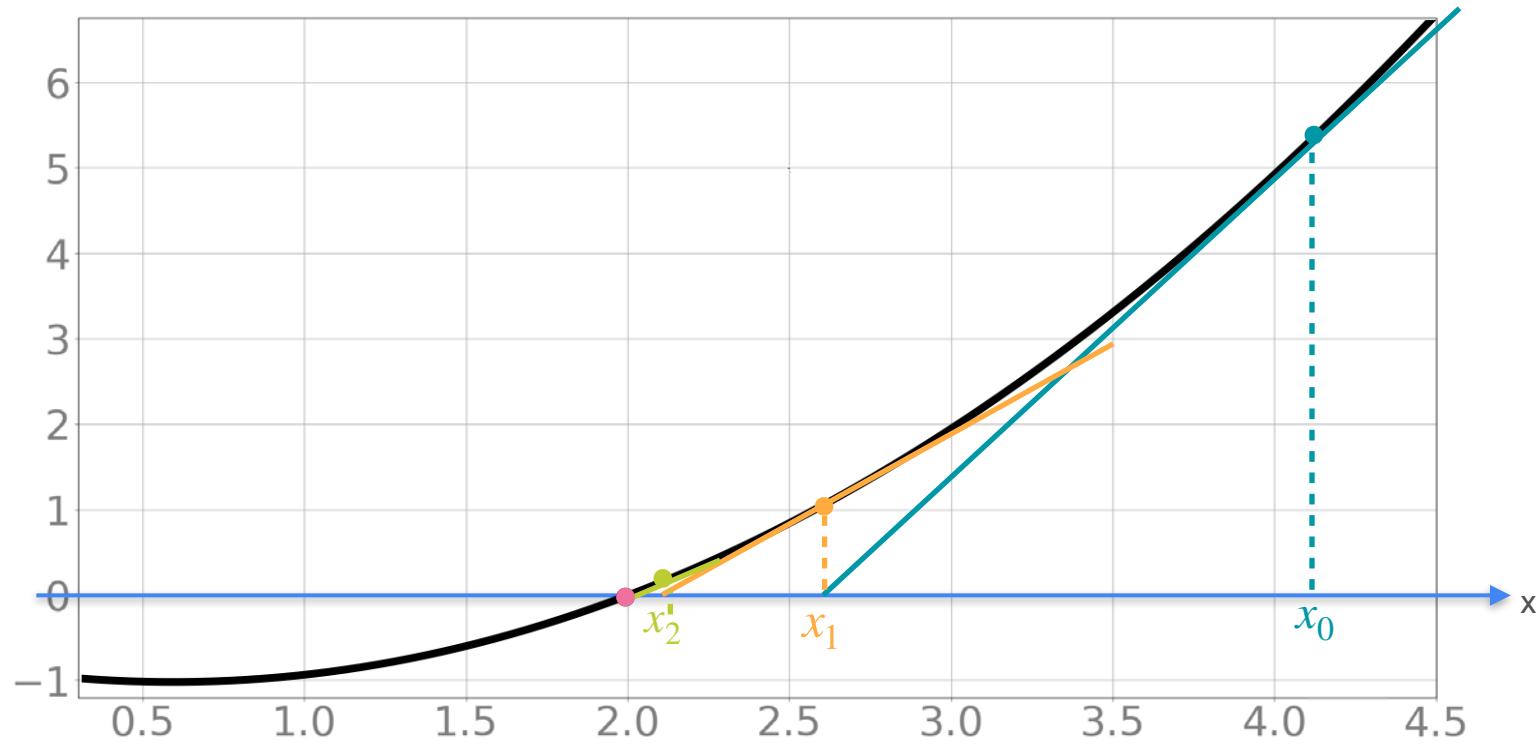
# Newton's Method



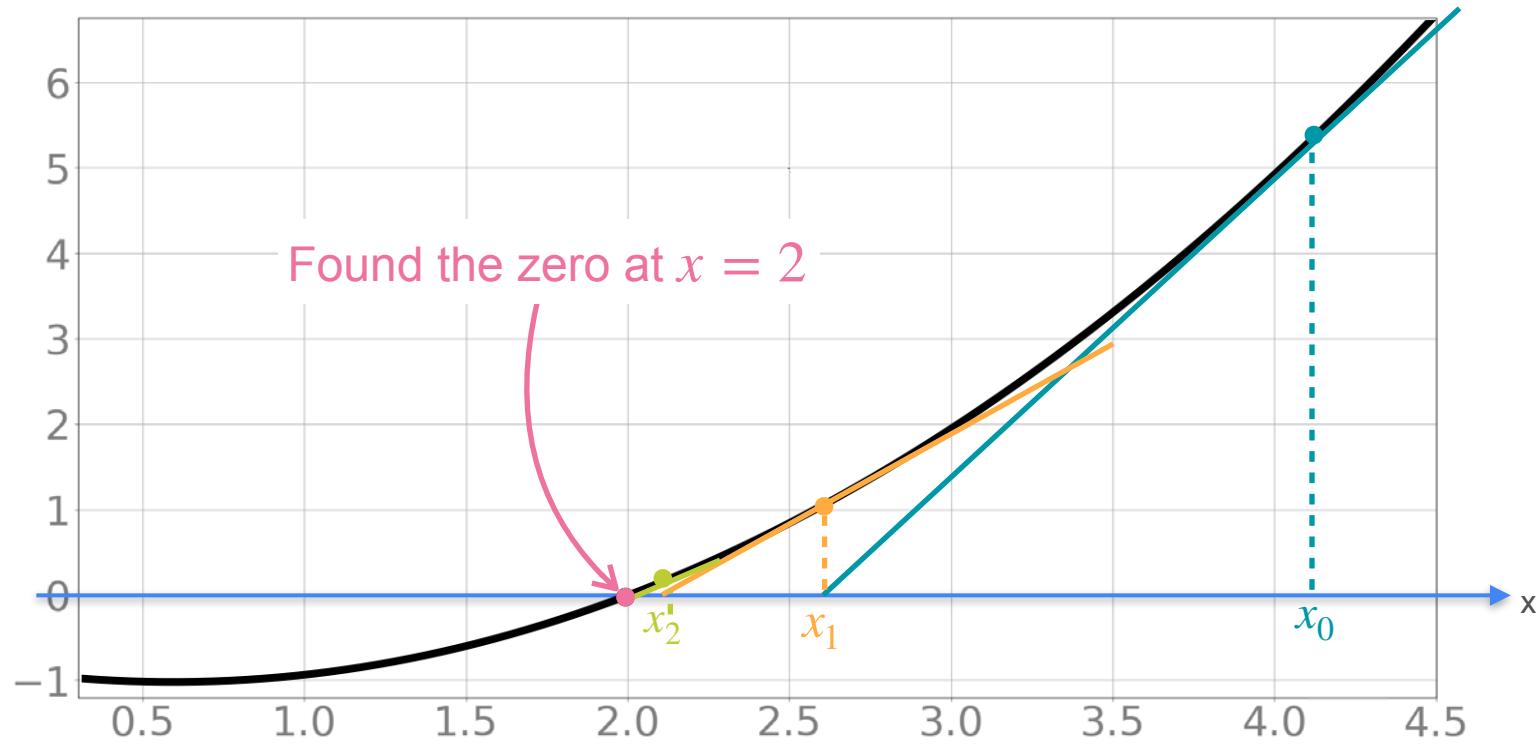
# Newton's Method



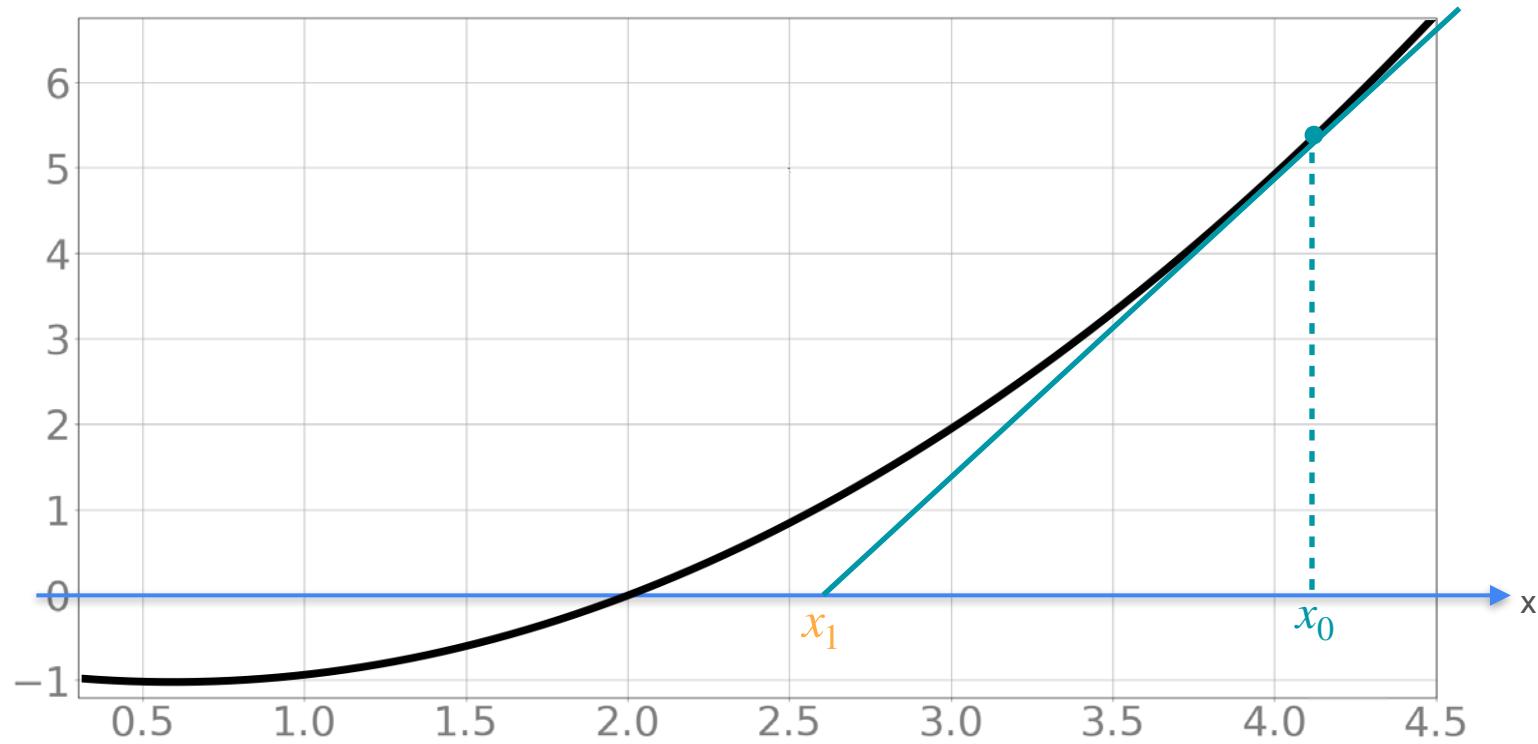
# Newton's Method



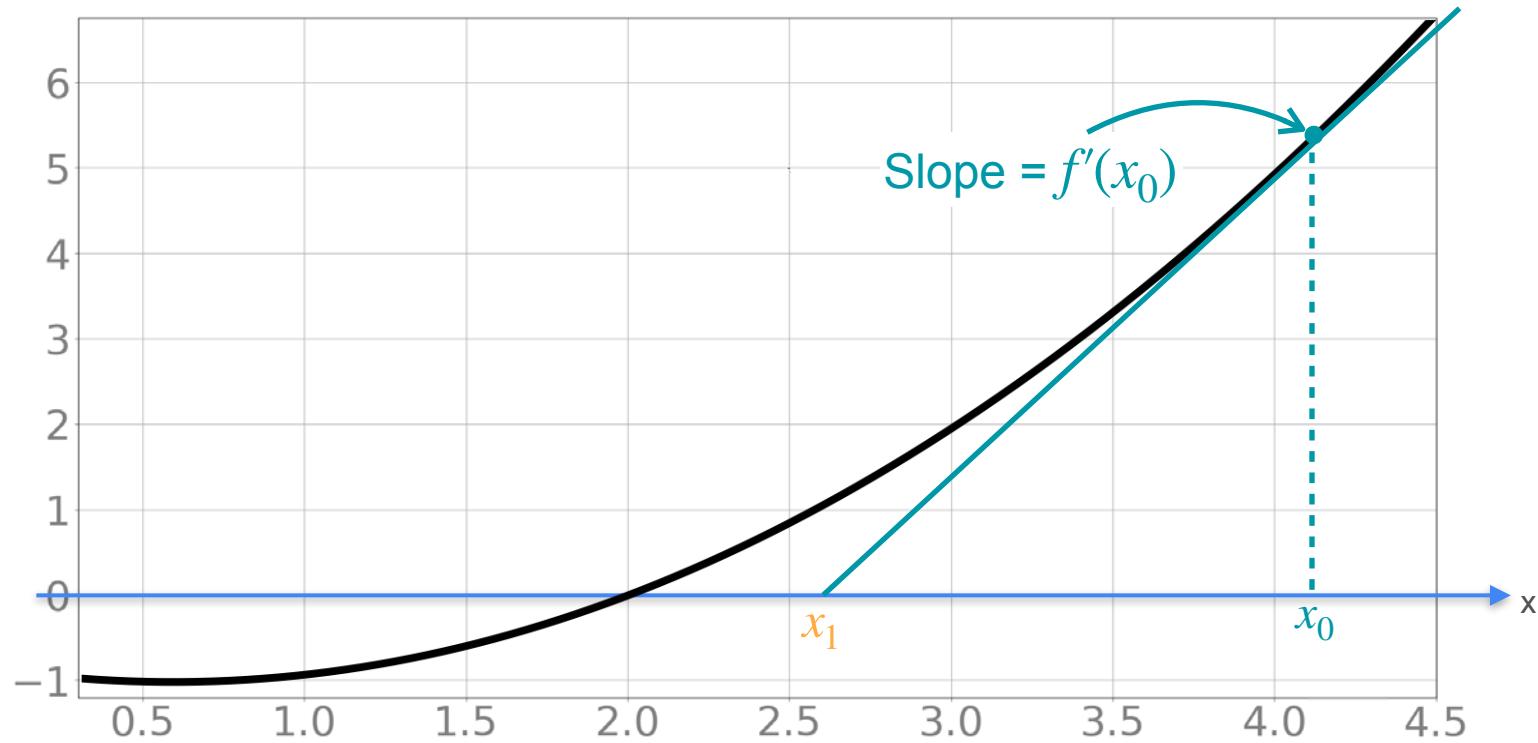
# Newton's Method



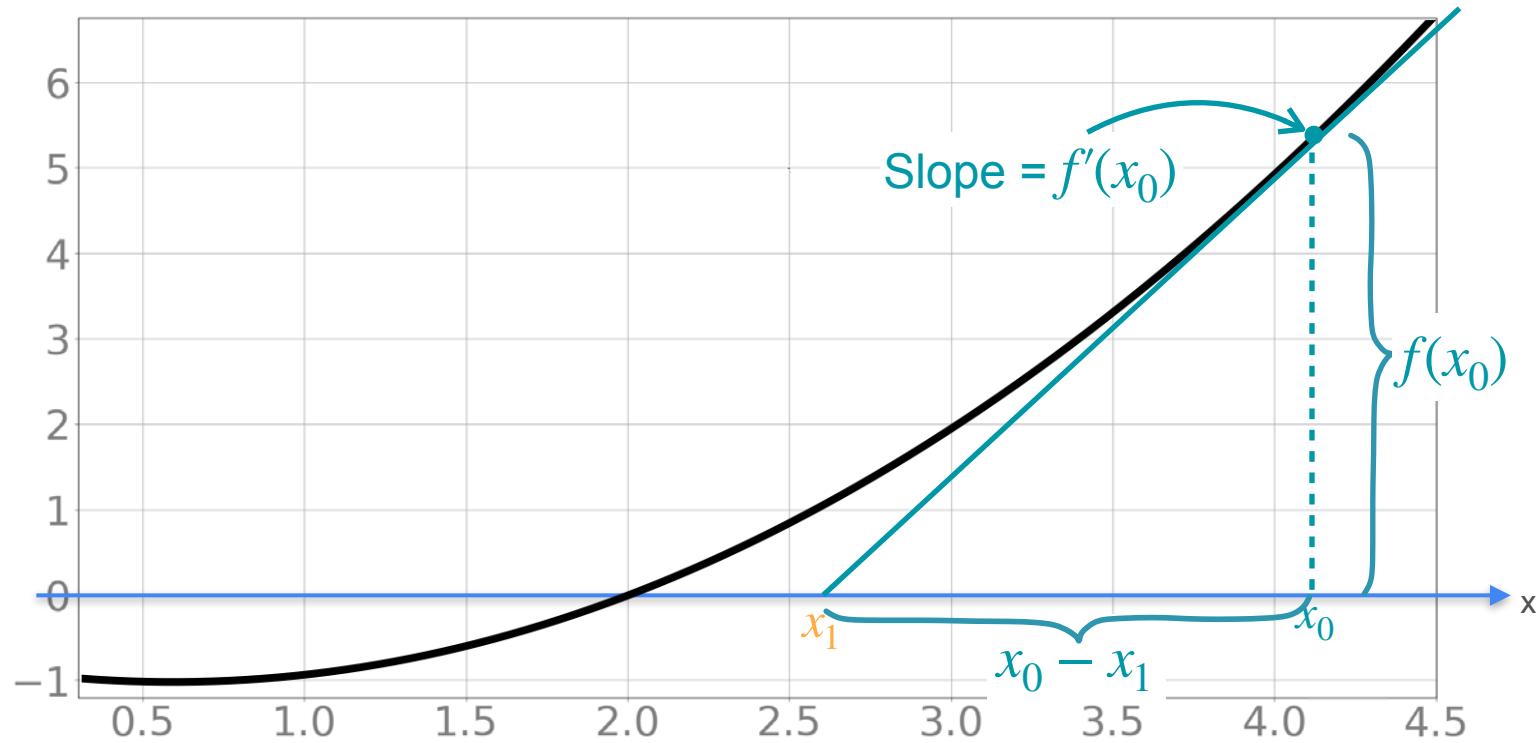
# Update Approximation



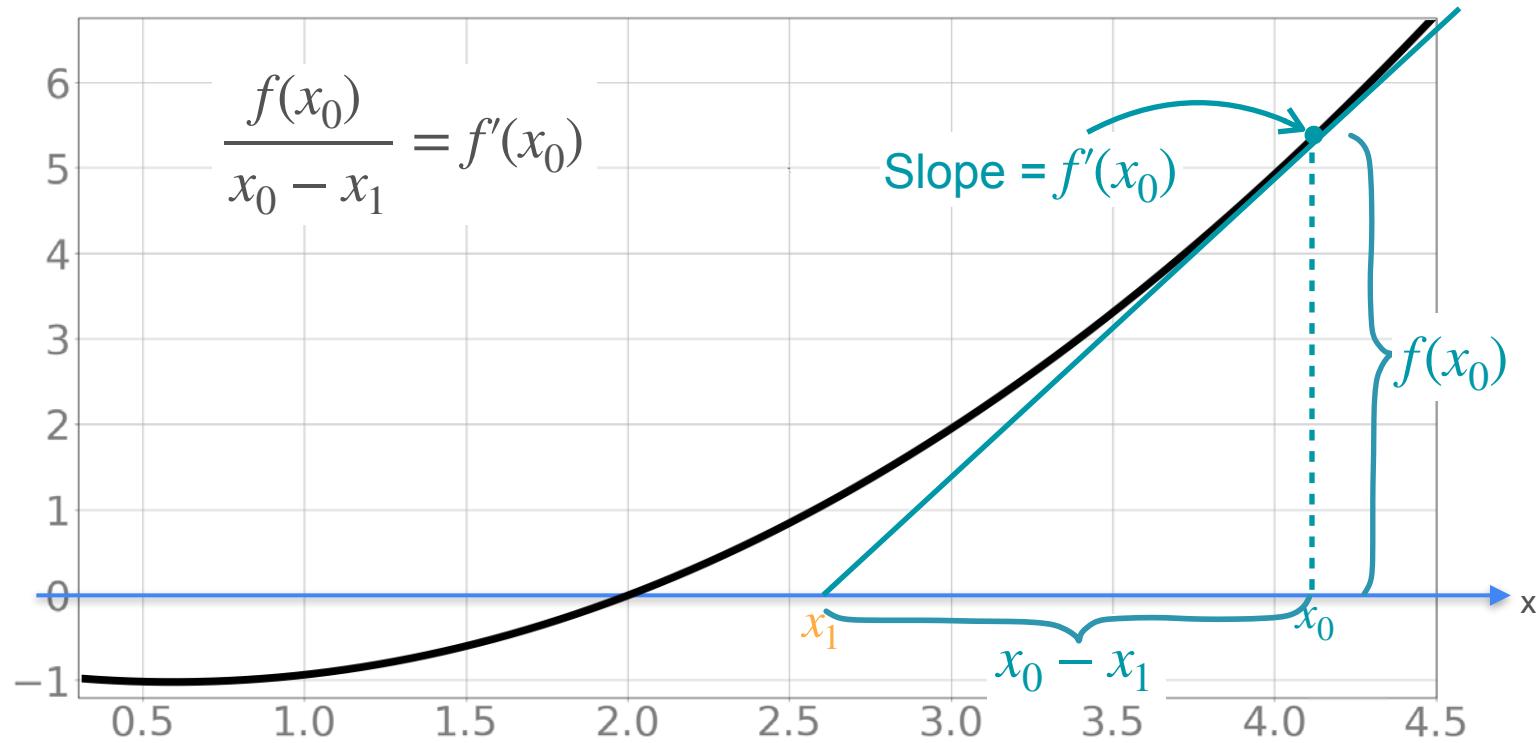
# Update Approximation



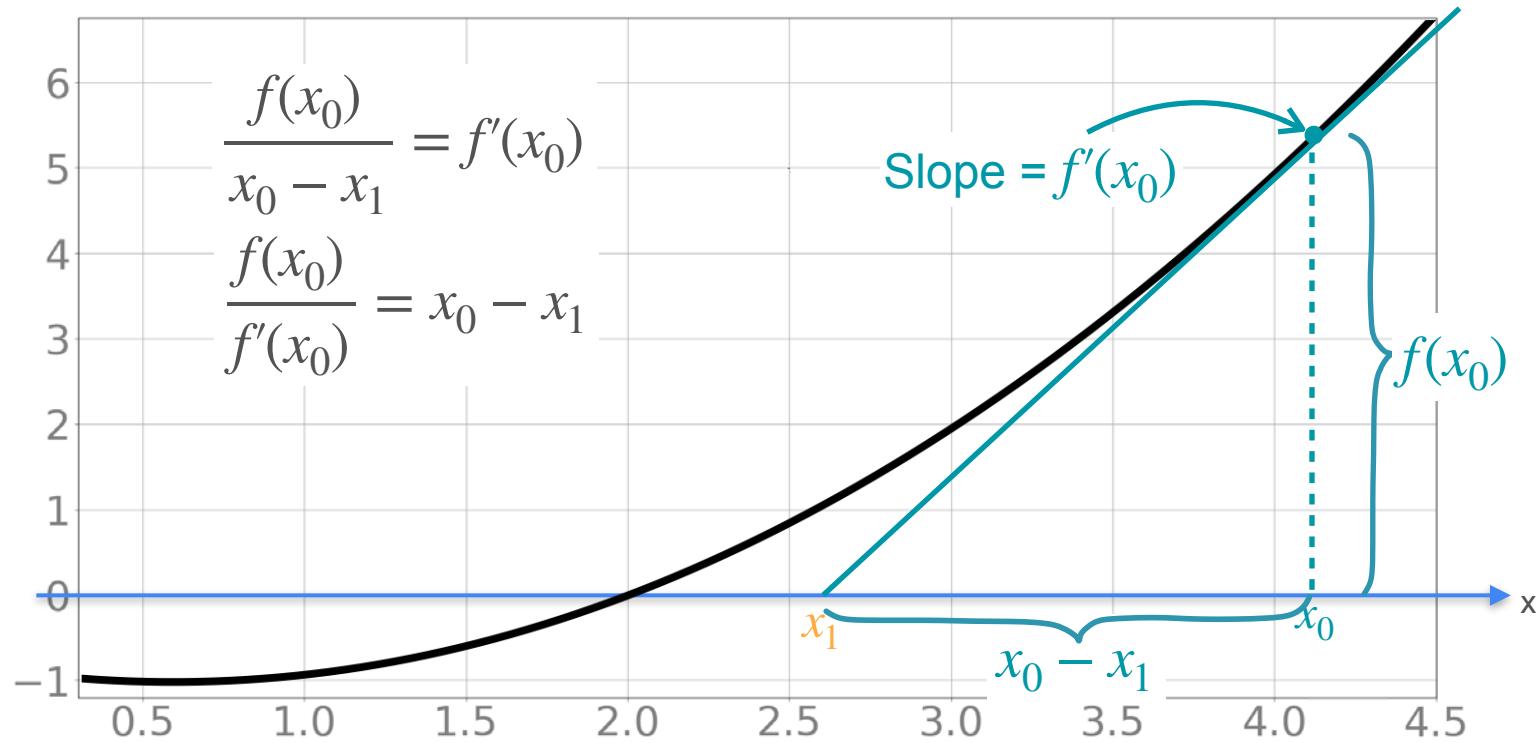
# Update Approximation



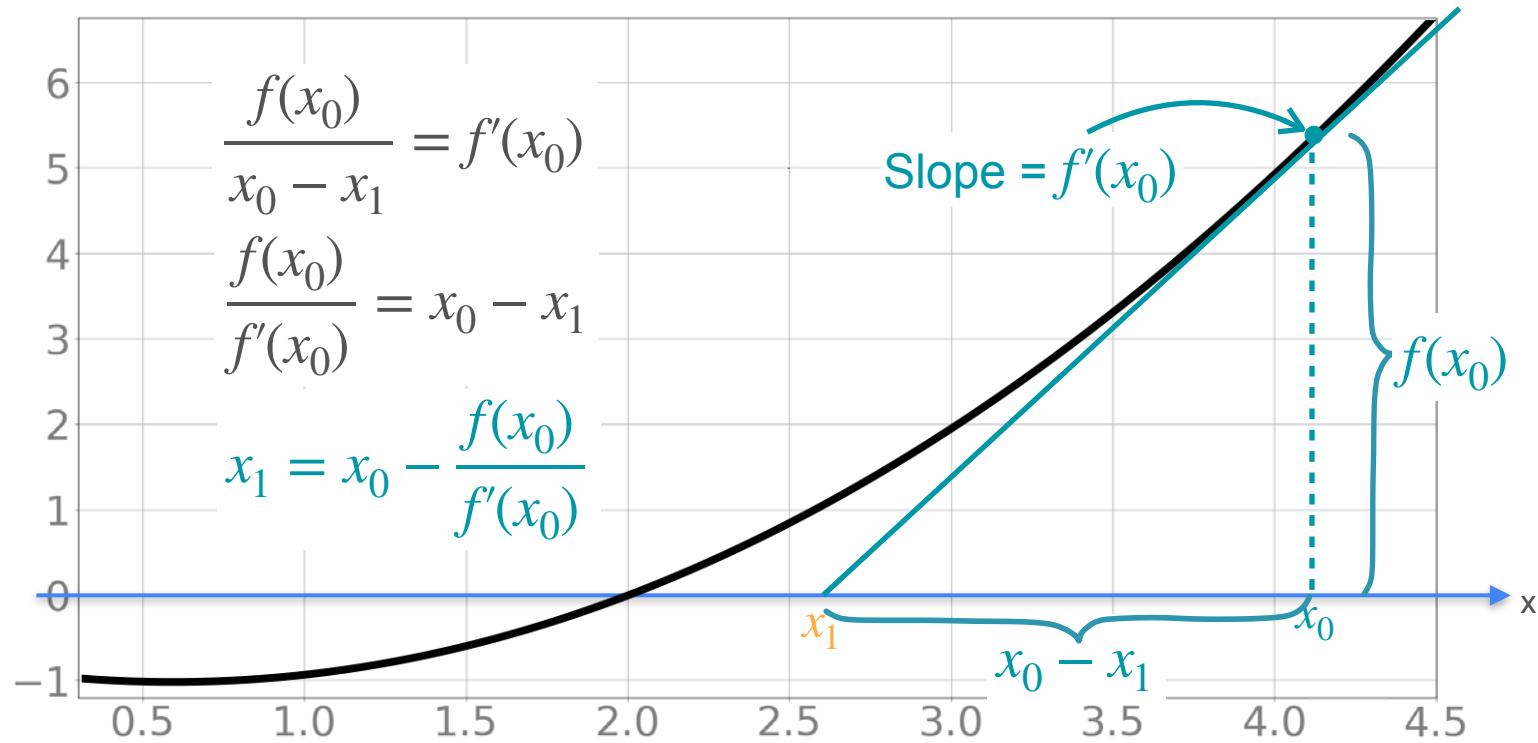
# Update Approximation



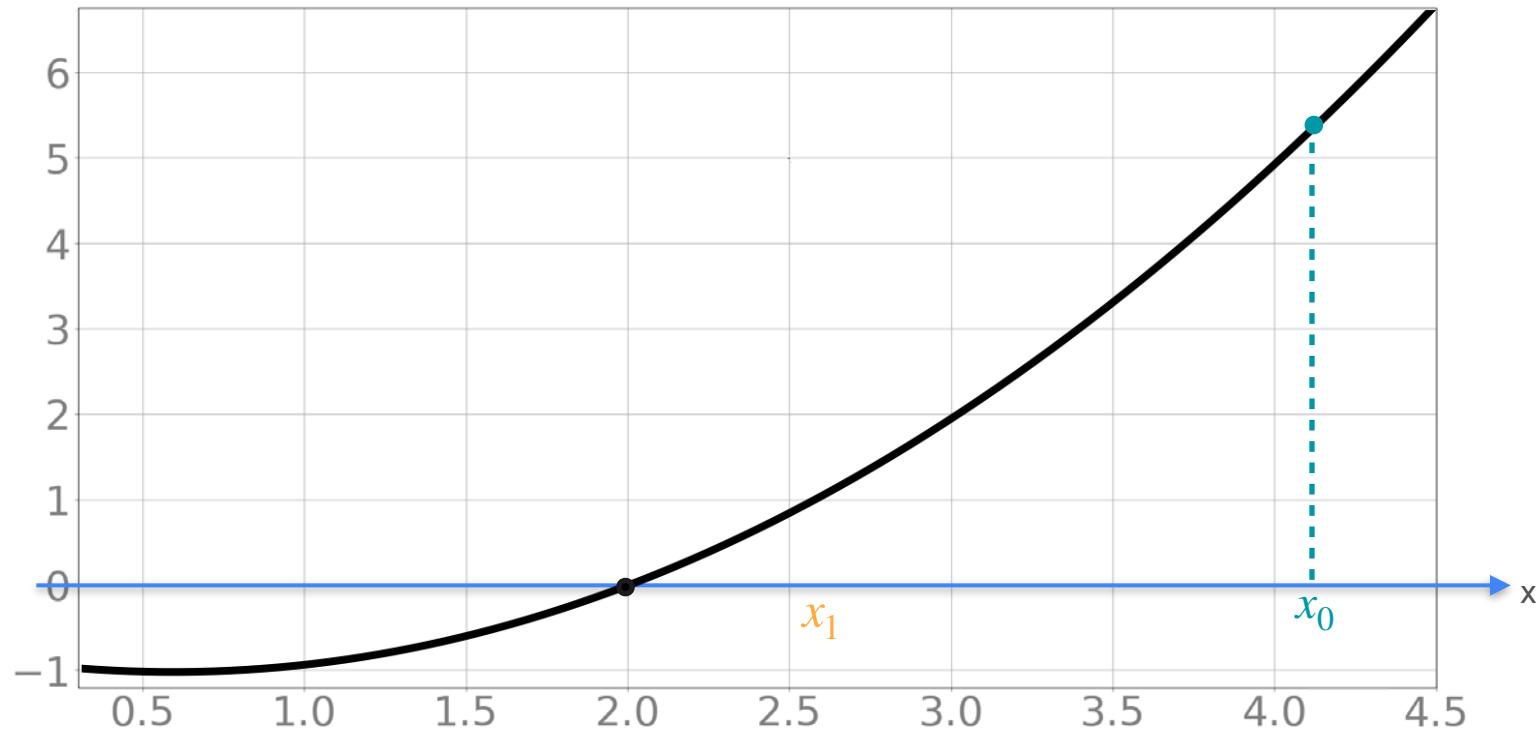
# Update Approximation



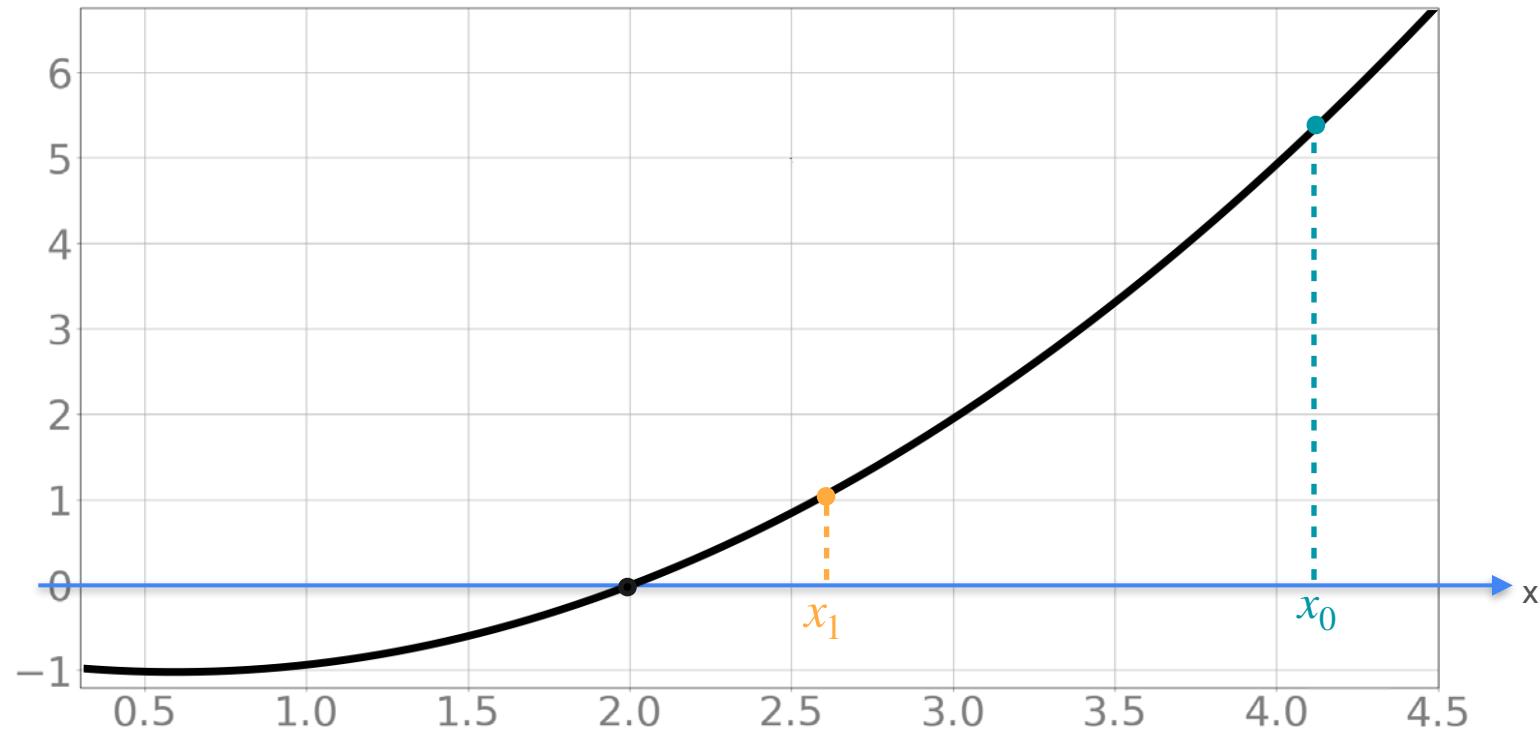
# Update Approximation



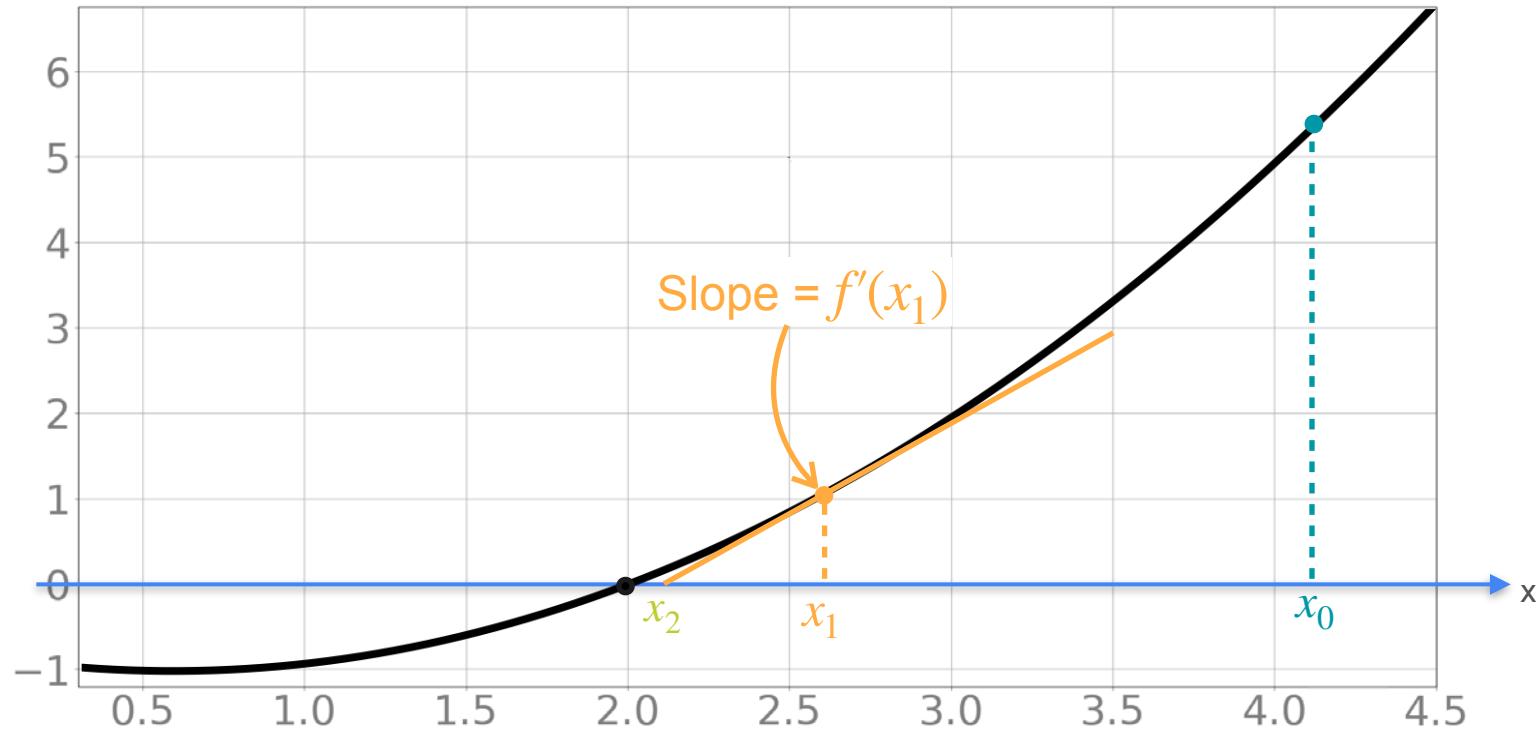
# Update Approximation



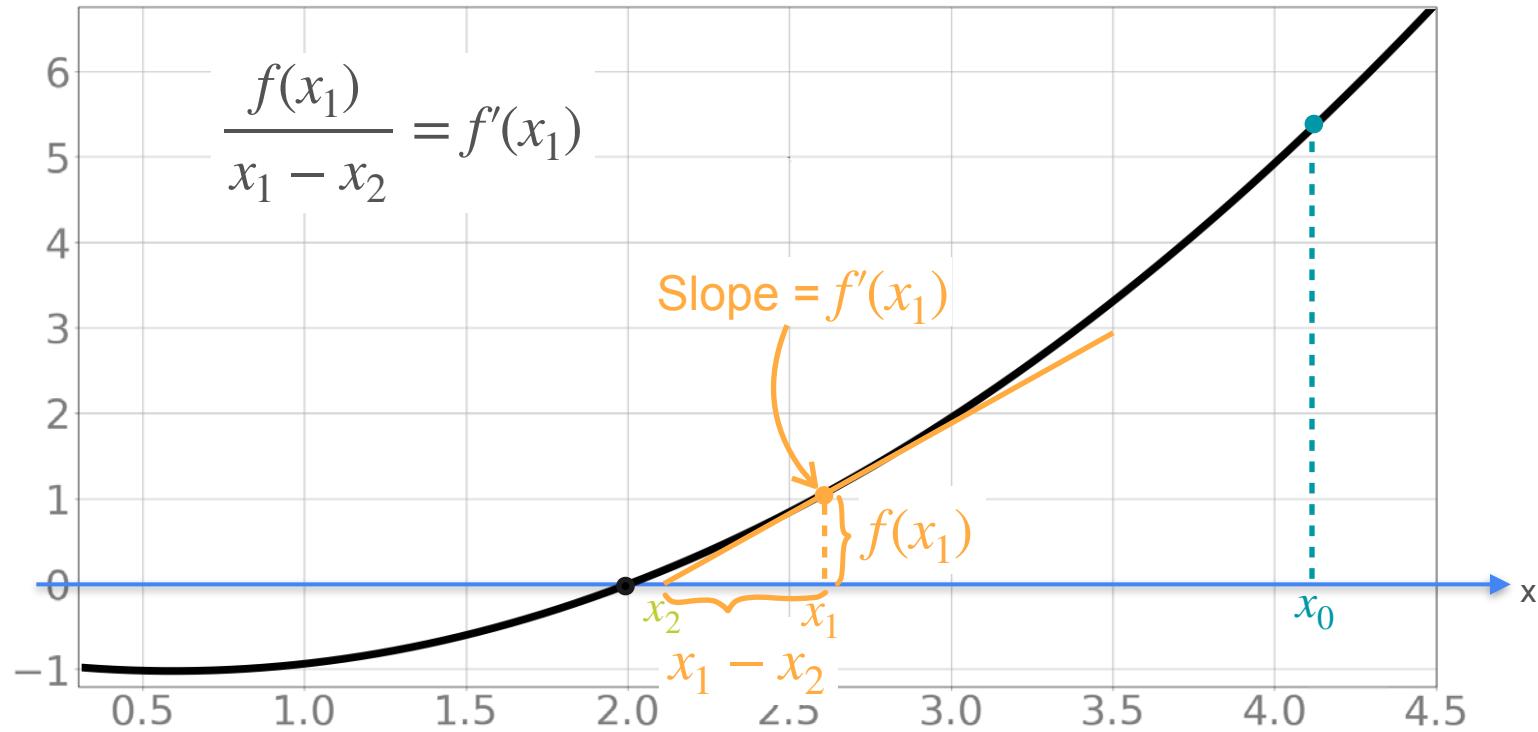
# Update Approximation



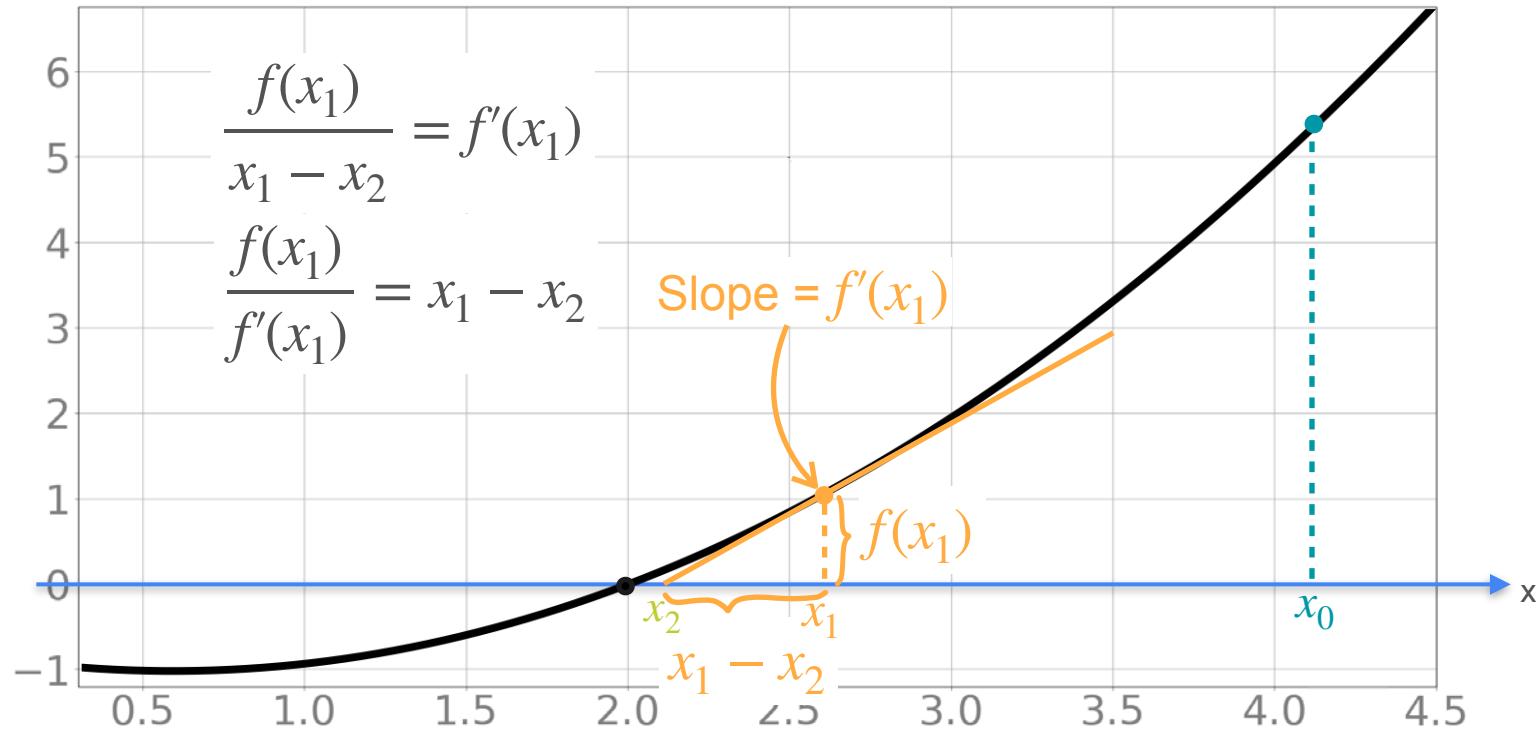
# Update Approximation



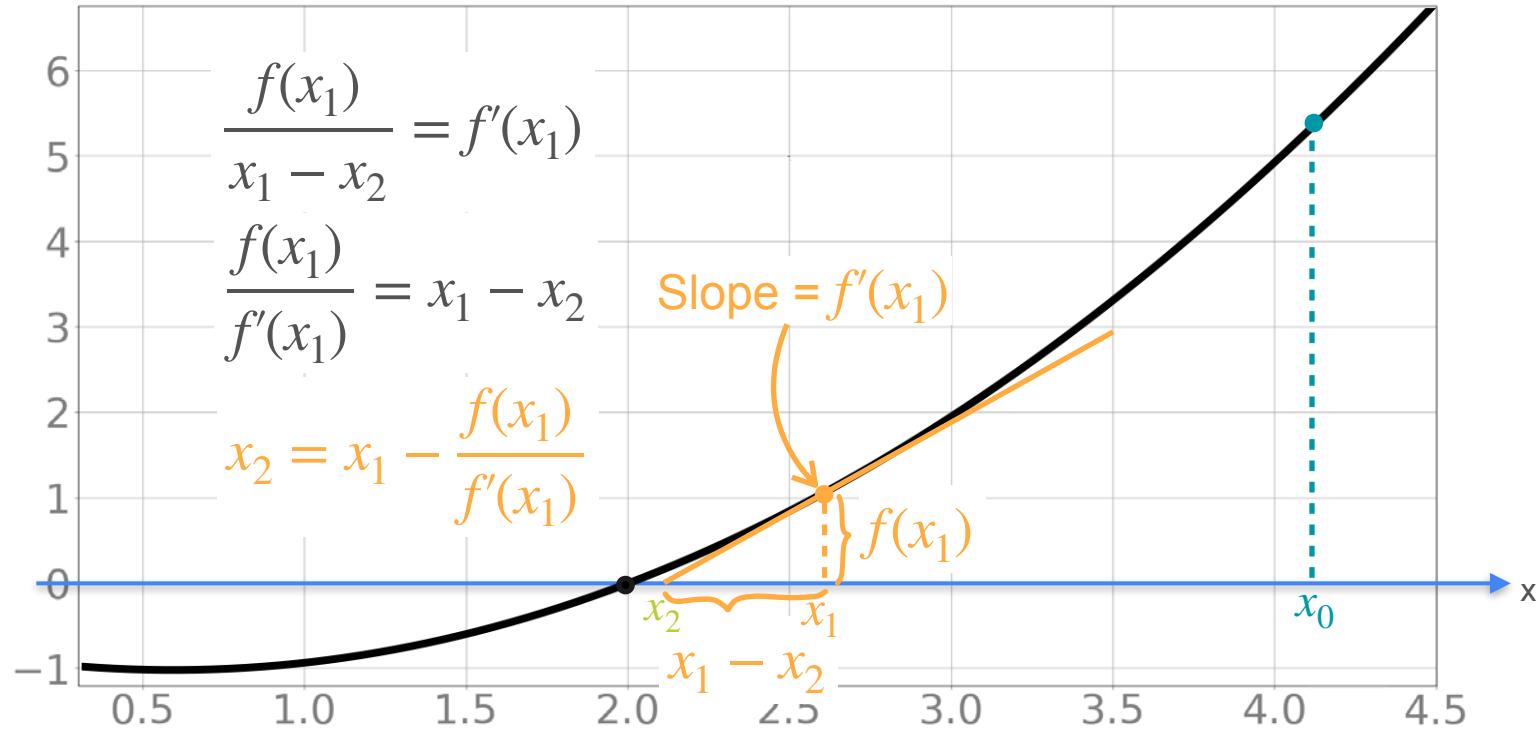
# Update Approximation



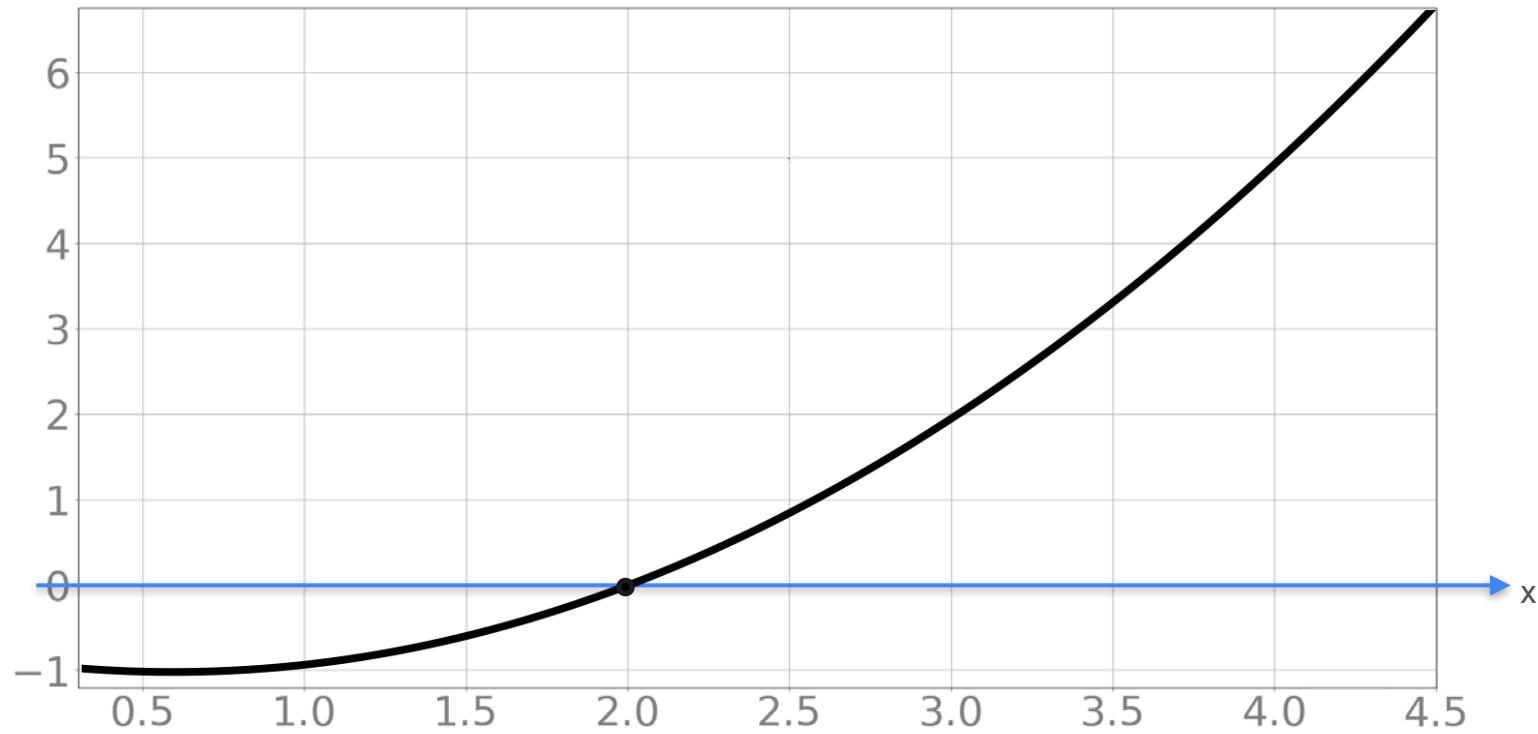
# Update Approximation



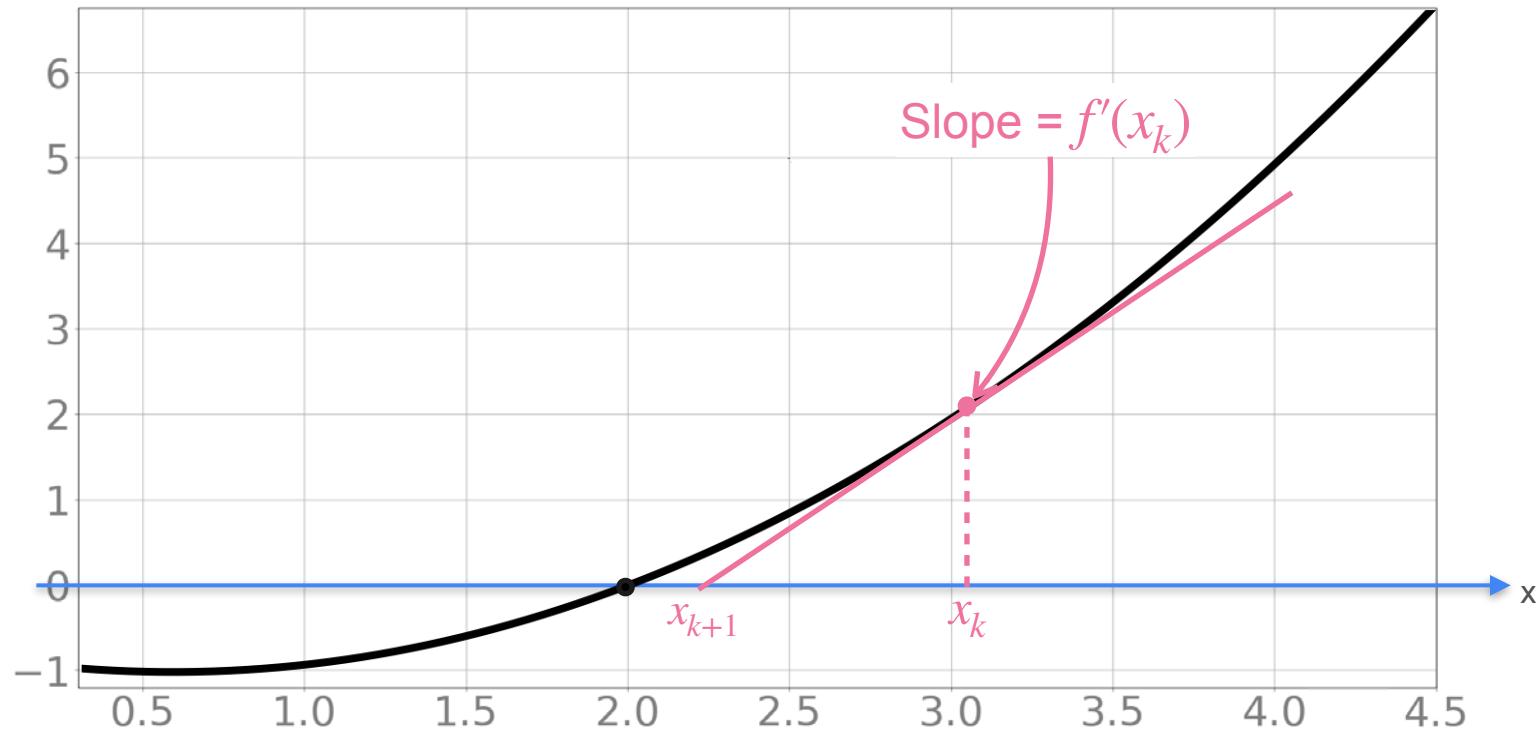
# Update Approximation



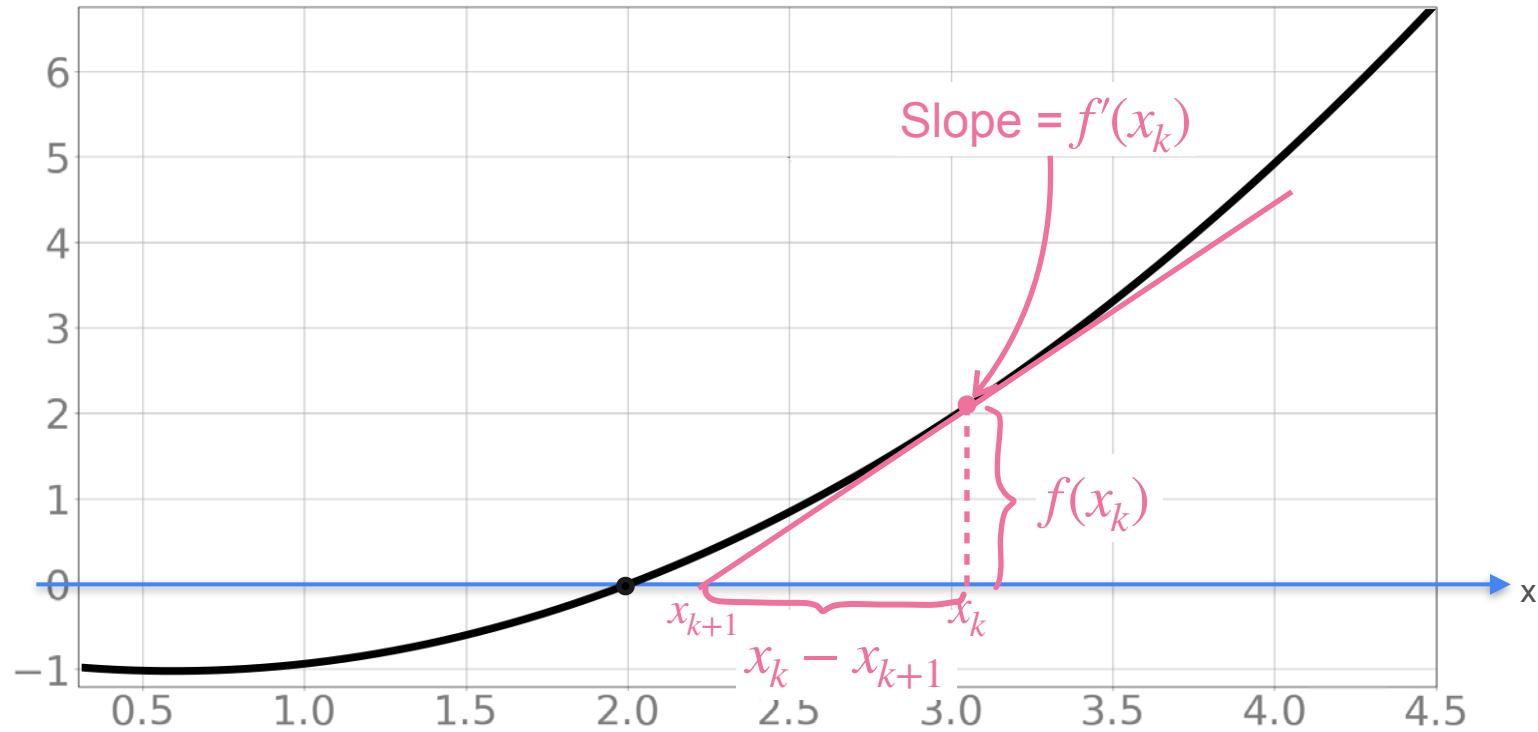
# Update Approximation



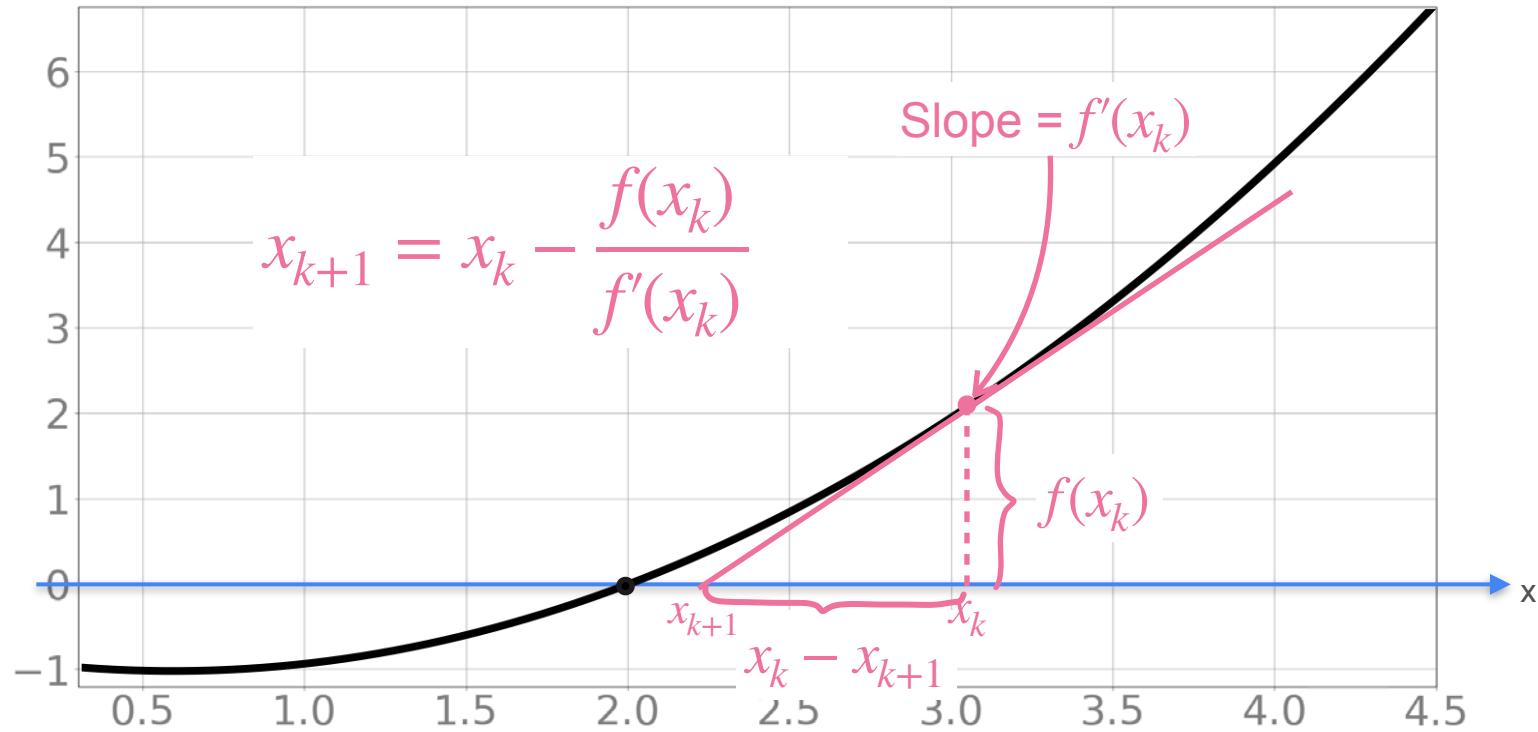
# Update Approximation



# Update Approximation



# Update Approximation



# Newton's Method for Optimization

# Newton's Method for Optimization



# Newton's Method for Optimization

Newton's method

Goal: find a zero of  $f(x)$



# Newton's Method for Optimization

Newton's method

Goal: find a zero of  $f(x)$



NM for Optimization

Goal: minimize  $g(x)$  find zeros of  $g'(x)$

# Newton's Method for Optimization

Newton's method

Goal: find a zero of  $f(x)$



NM for Optimization

Goal: minimize  $g(x)$  → find zeros of  $g'(x)$

$$f(x) \mapsto g'(x)$$

$$f'(x) \mapsto (g'(x))'$$

# Newton's Method for Optimization

Newton's method

Goal: find a zero of  $f(x)$

1) Start with some  $x_0$



NM for Optimization

Goal: minimize  $g(x)$  find zeros of  $g'(x)$

$$f(x) \mapsto g'(x) \qquad f'(x) \mapsto (g'(x))'$$

1) Start with some  $x_0$

# Newton's Method for Optimization

## Newton's method

Goal: find a zero of  $f(x)$

1) Start with some  $x_0$

2) Update:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$



## NM for Optimization

Goal: minimize  $g(x)$  find zeros of  $g'(x)$

$$f(x) \mapsto g'(x)$$

$$f'(x) \mapsto (g'(x))'$$

1) Start with some  $x_0$

2) Update:

# Newton's Method for Optimization

## Newton's method

Goal: find a zero of  $f(x)$

1) Start with some  $x_0$

2) Update:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$



## NM for Optimization

Goal: minimize  $g(x)$  find zeros of  $g'(x)$

$$f(x) \mapsto g'(x)$$

$$f'(x) \mapsto (g'(x))'$$

1) Start with some  $x_0$

2) Update:

$$x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'}$$

# Newton's Method for Optimization

Newton's method

Goal: find a zero of  $f(x)$

1) Start with some  $x_0$

2) Update:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

3) Repeat 2) until you find the root.



NM for Optimization

Goal: minimize  $g(x)$  find zeros of  $g'(x)$

$$f(x) \mapsto g'(x)$$

$$f'(x) \mapsto (g'(x))'$$

1) Start with some  $x_0$

2) Update:

$$x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'}$$

3) Repeat 2) until you find the root.



DeepLearning.AI

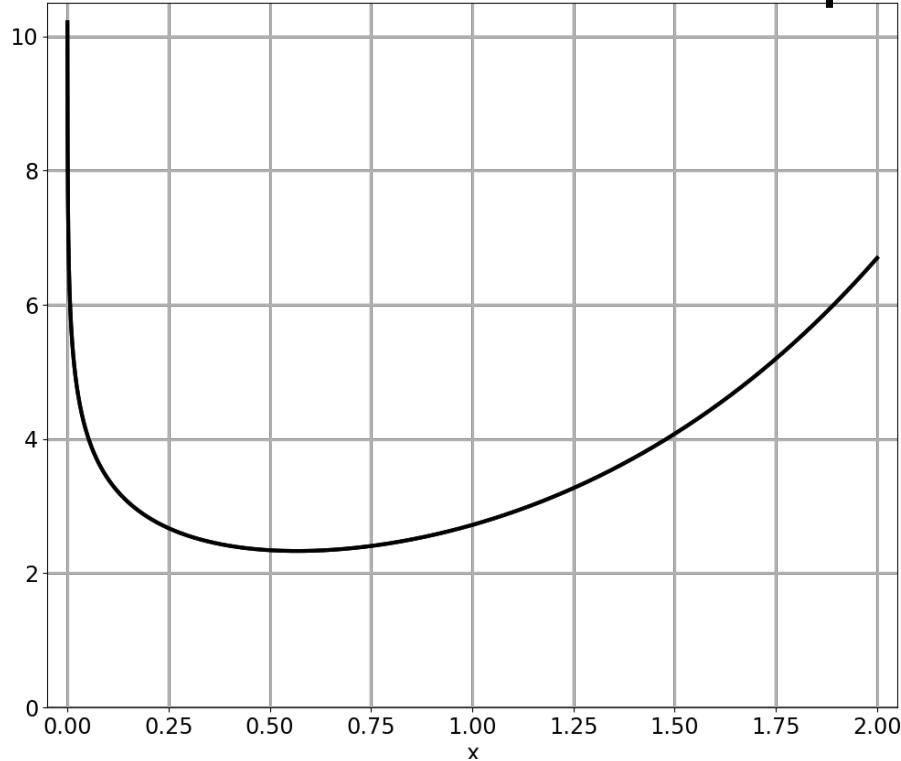
# Optimization in Neural Networks and Newton's Method

---

**Newton's method:  
An example**

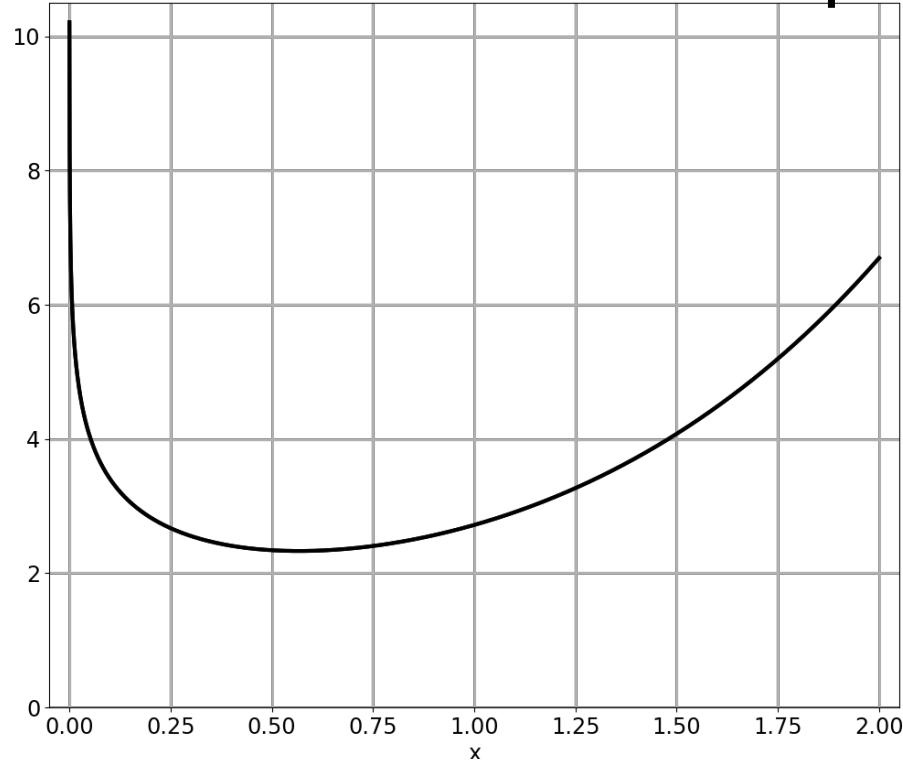
# Newton's Method for Optimization

# Newton's Method for Optimization



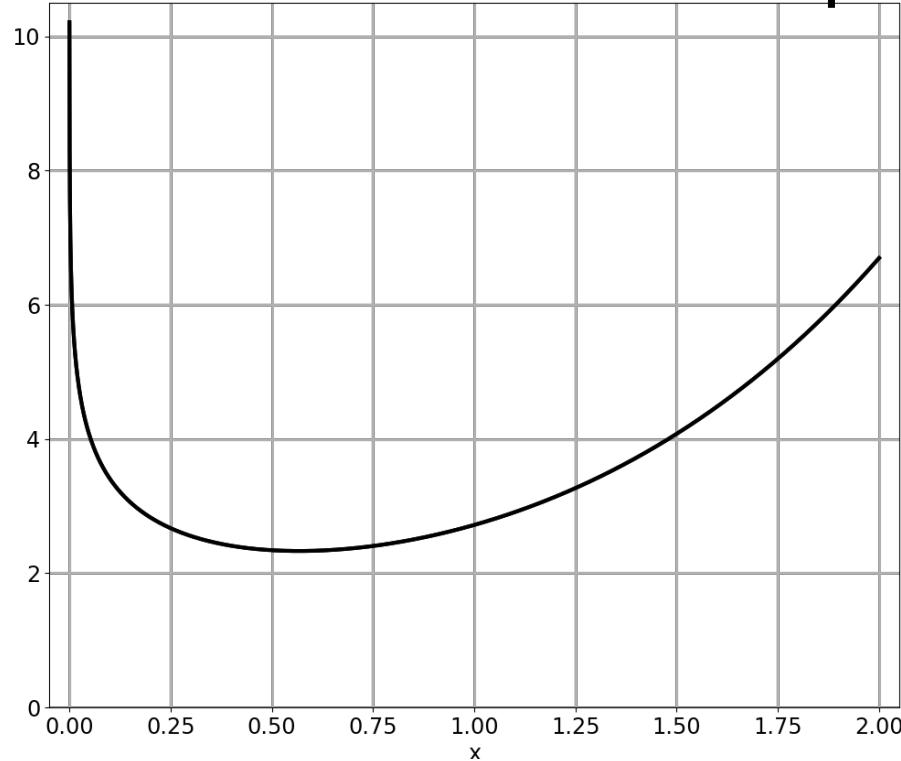
$$g(x) = e^x - \log(x)$$

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

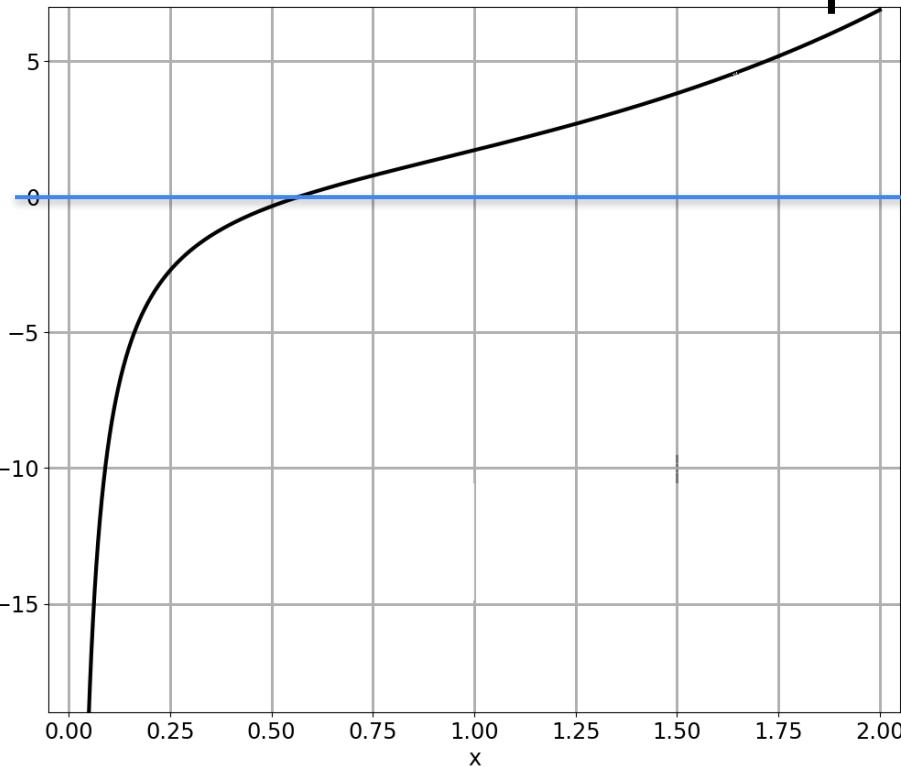
# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

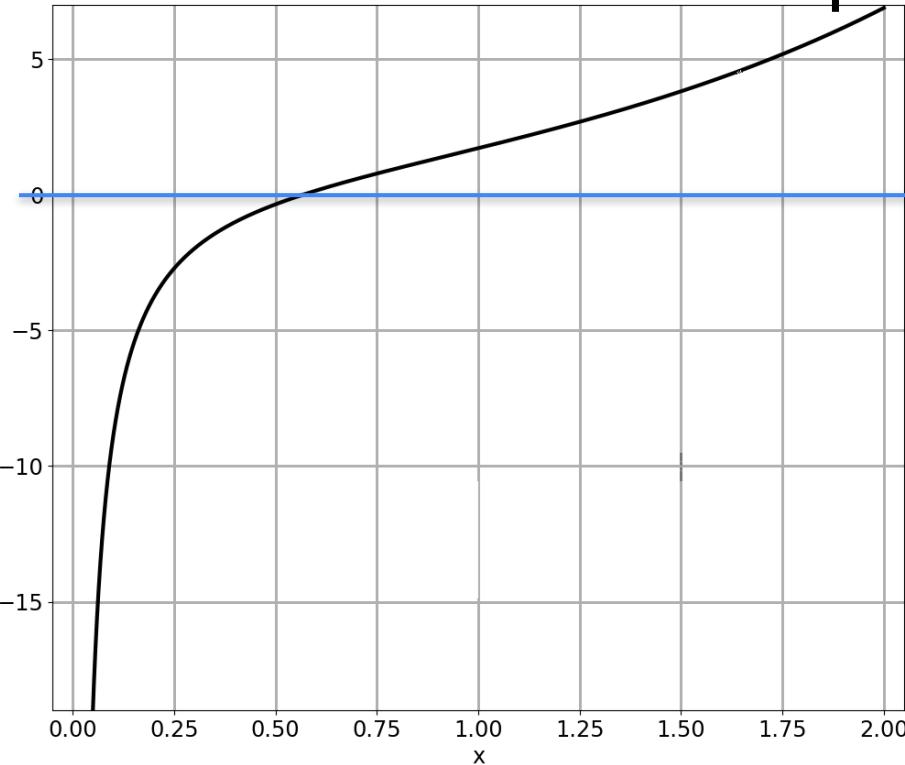
# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

# Newton's Method for Optimization

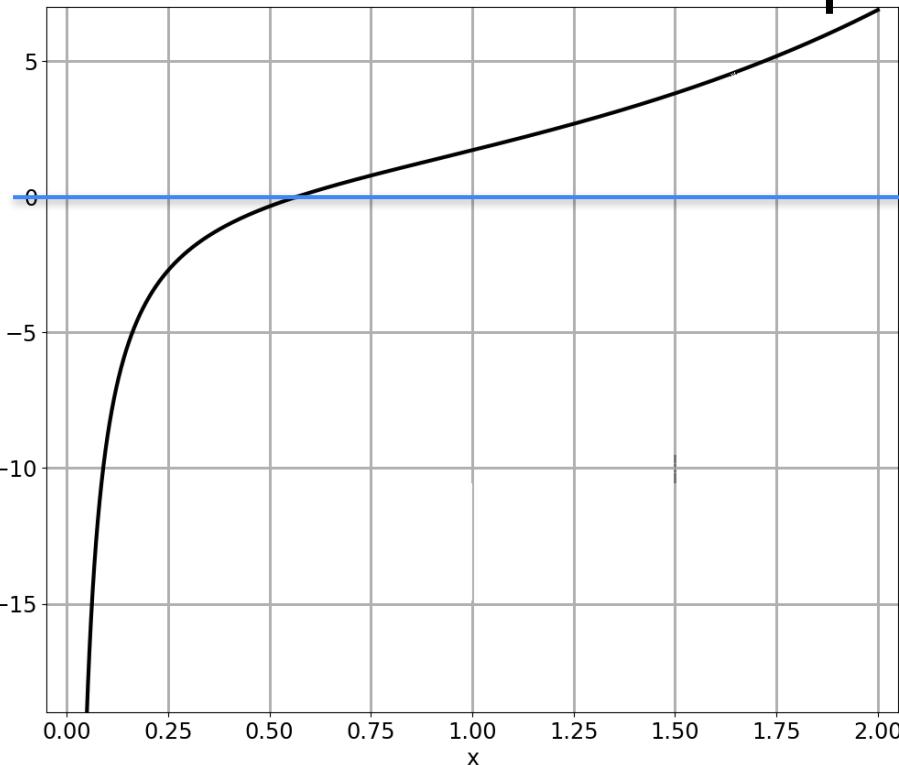


$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

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

# Newton's Method for Optimization

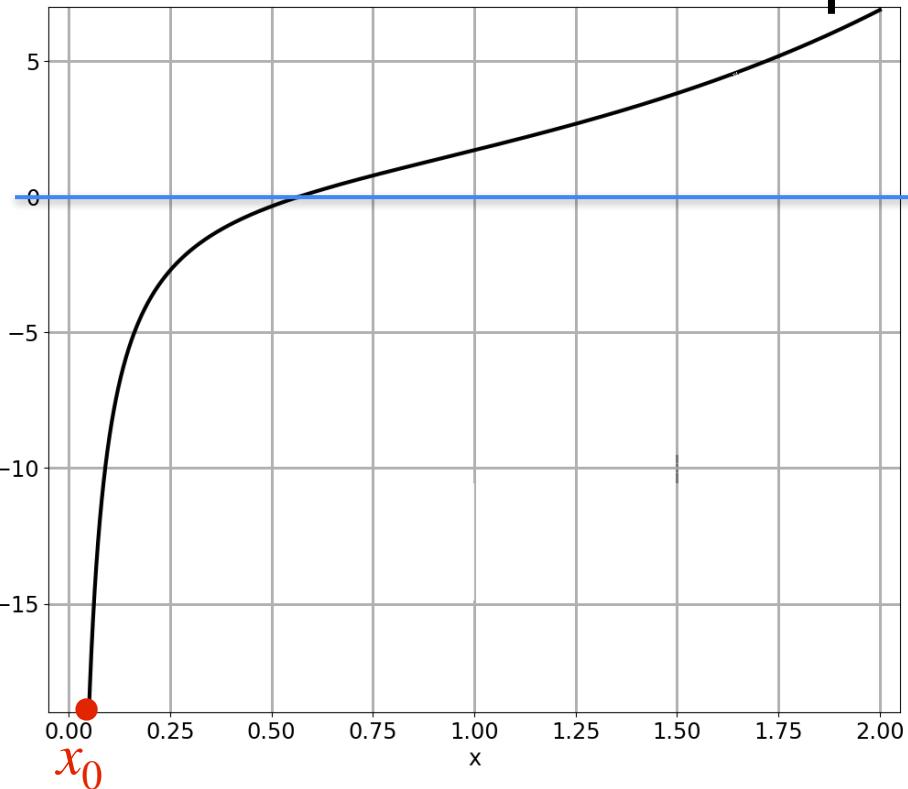


$$g(x) = e^x - \log(x) \quad \underbrace{g'(x) = e^x - 1/x}_{f(x)}$$

Minimum:  $x^* = 0.5671$

$$(g'(x))' = e^x + \frac{1}{x^2} \quad \underbrace{\phantom{(g'(x))'} f'(x)}_{\frac{1}{x^2}}$$

# Newton's Method for Optimization

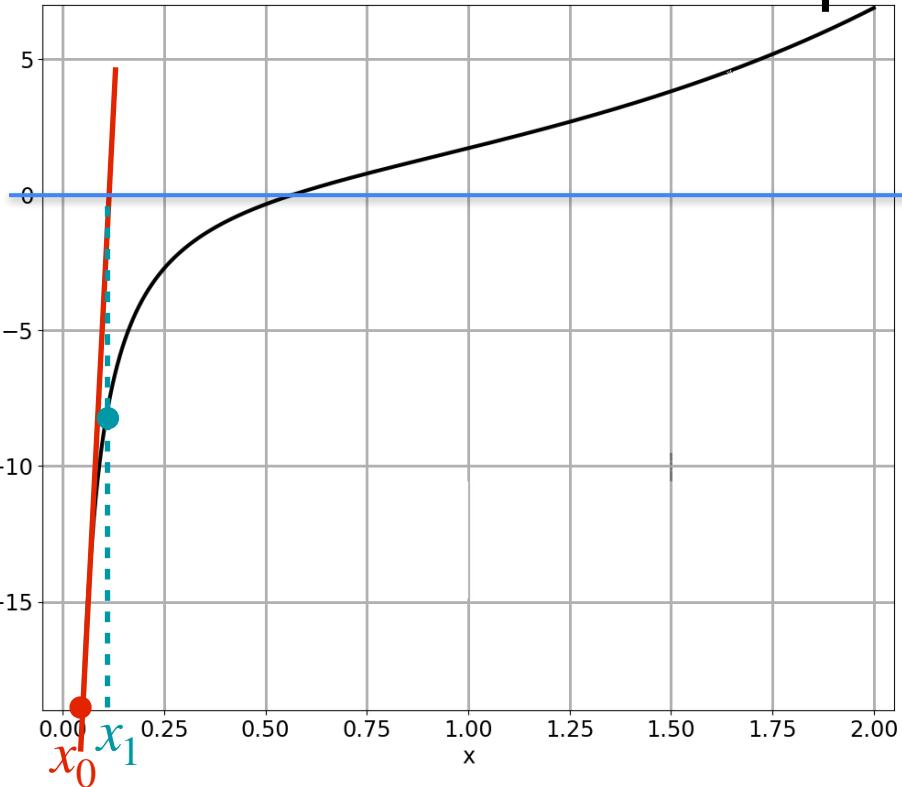


$$g(x) = e^x - \log(x) \quad \overbrace{g'(x) = e^x - 1/x}^{f(x)}$$

Minimum:  $x^* = 0.5671$

$$(g'(x))' = e^x + \frac{1}{x^2}$$
$$x_0 = 0.05 \quad \overbrace{f'(x)}^{g'(x)}$$

# Newton's Method for Optimization

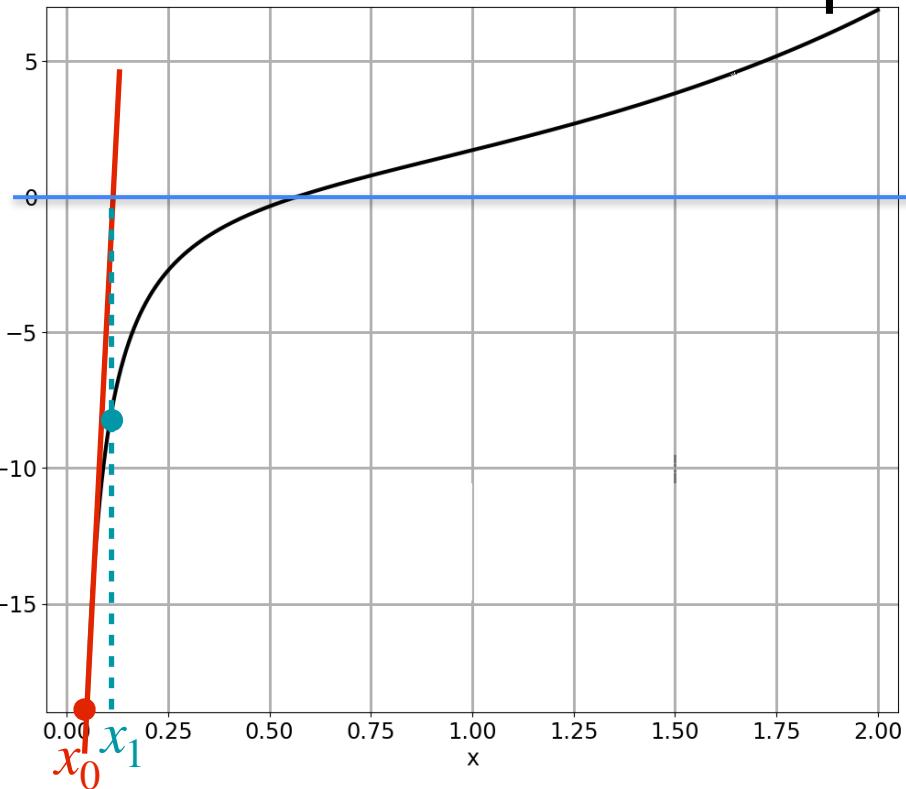


$$g(x) = e^x - \log(x)$$
$$g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

$$(g'(x))' = e^x + \frac{1}{x^2}$$
$$f'(x)$$
$$x_0 = 0.05$$

# Newton's Method for Optimization

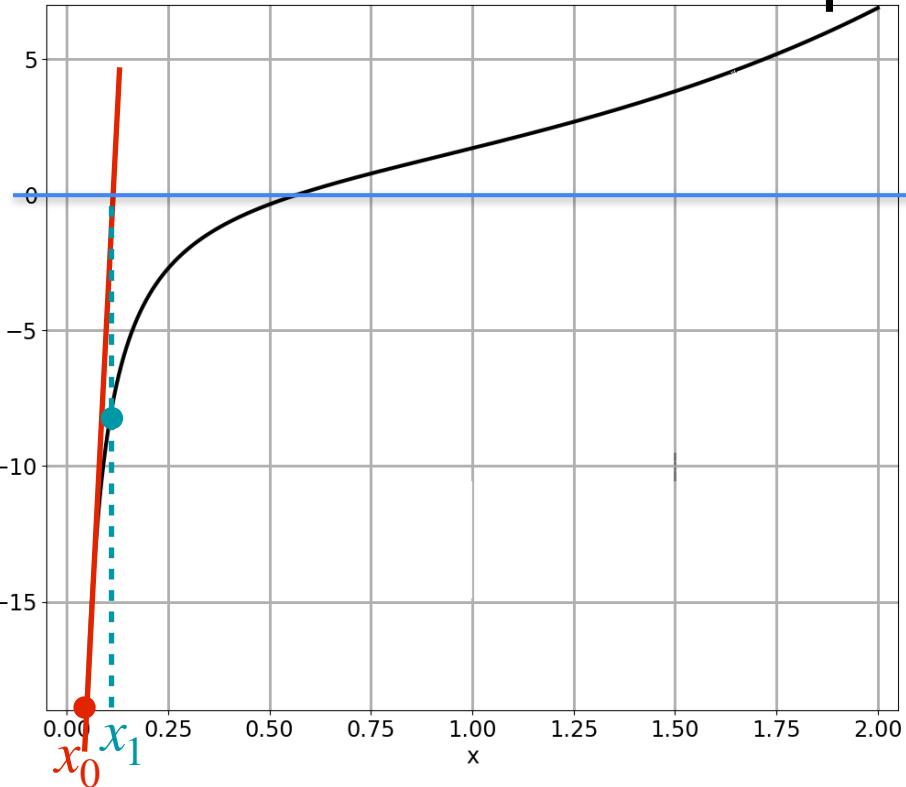


$$g(x) = e^x - \log(x) \quad \overbrace{g'(x)}^{f(x)} = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

$$(g'(x))' = e^x + \frac{1}{x^2}$$
$$x_0 = 0.05$$
$$x_1 = x_0 - \frac{g'(x_0)}{(g'(x_0))'}$$
$$= 0.05 - \frac{\left(e^{0.05} - \frac{1}{0.05}\right)}{\left(e^{0.05} + \frac{1}{0.05^2}\right)}$$

# Newton's Method for Optimization

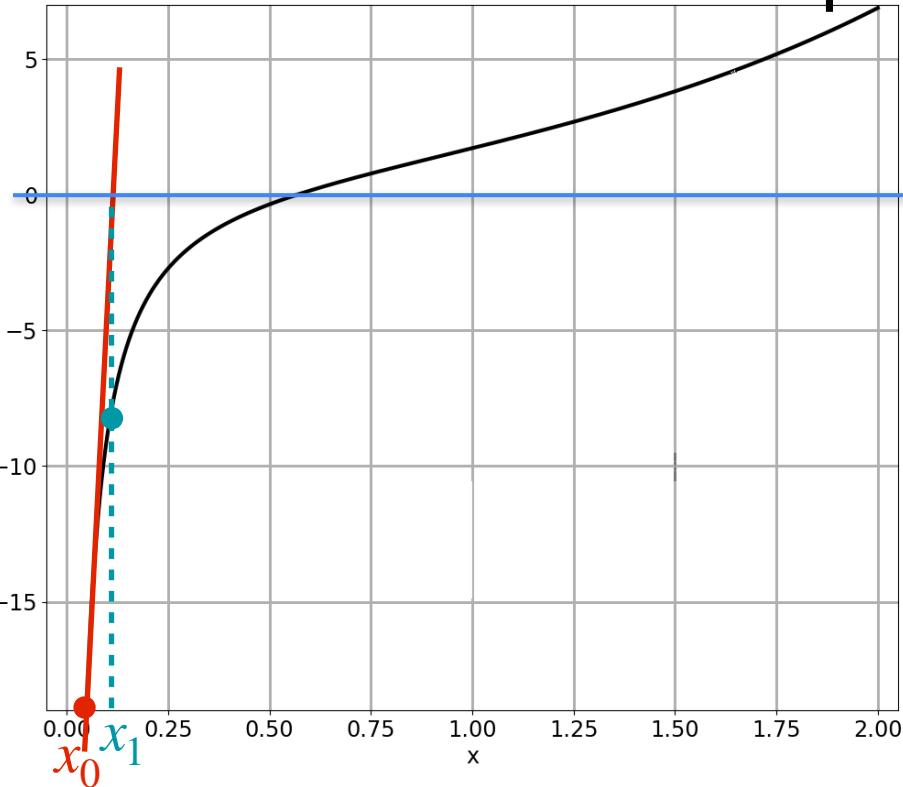


$$g(x) = e^x - \log(x) \quad \overbrace{g'(x)}^{f'(x)} = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

$$(g'(x))' = e^x + \frac{1}{x^2}$$
$$x_0 = 0.05$$
$$x_1 = x_0 - \frac{g'(x_0)}{(g'(x_0))'} \quad \overbrace{f'(x)}^{(g'(x))'}$$
$$= 0.05 - \frac{\left(e^{0.05} - \frac{1}{0.05}\right)}{\left(e^{0.05} + \frac{1}{0.05^2}\right)} = 0.097$$

# Newton's Method for Optimization

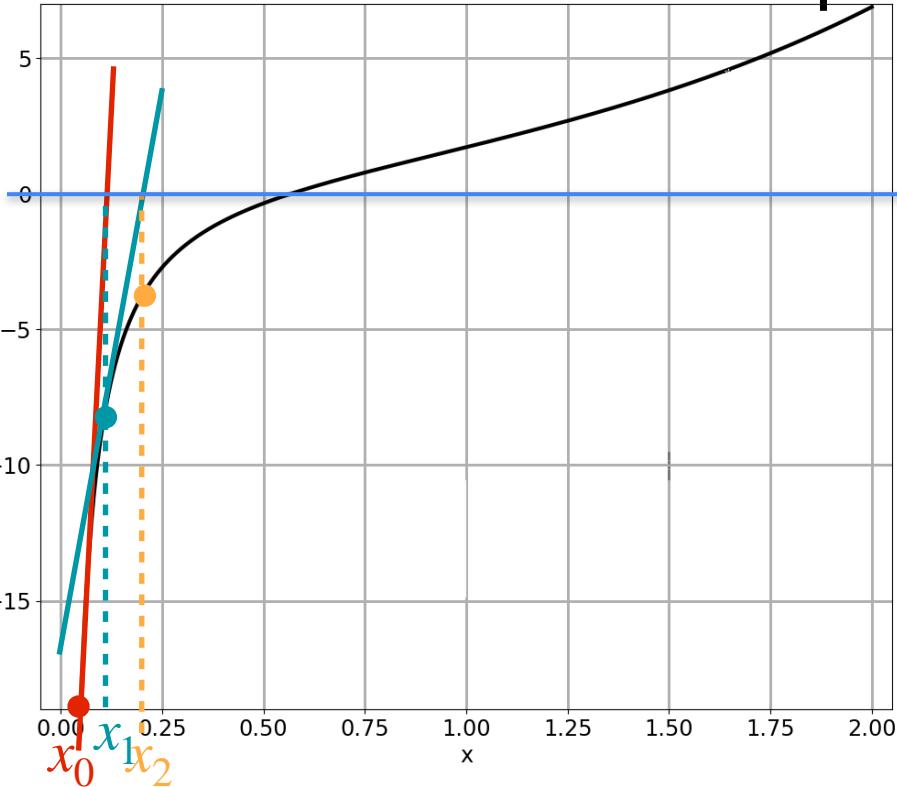


$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

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

# Newton's Method for Optimization



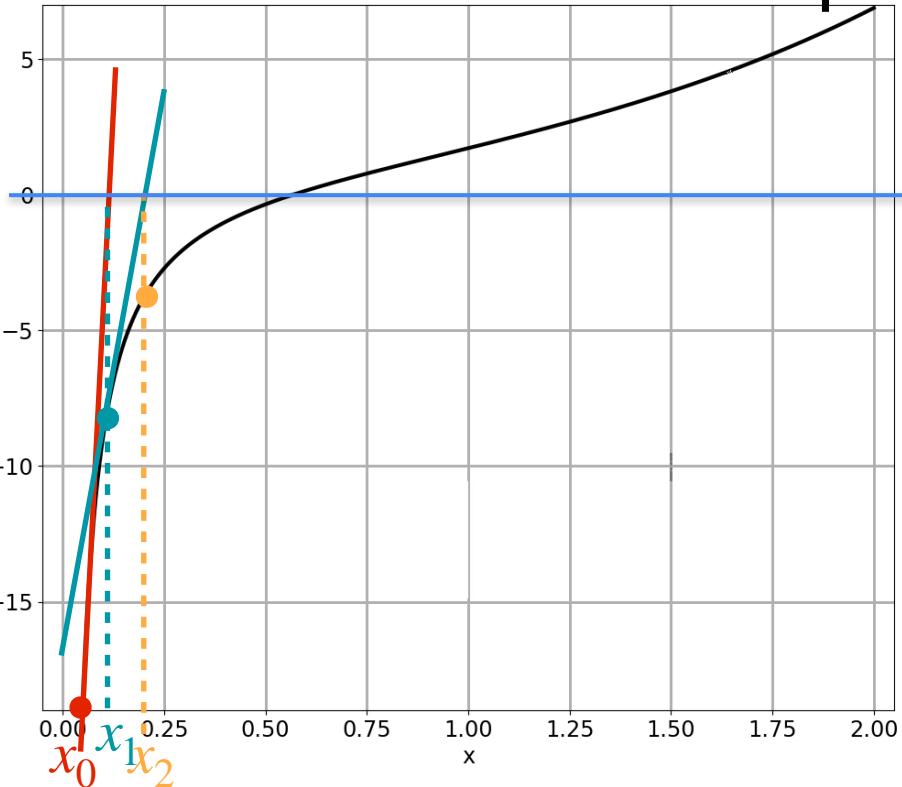
$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

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

$$x_1 = 0.097$$

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

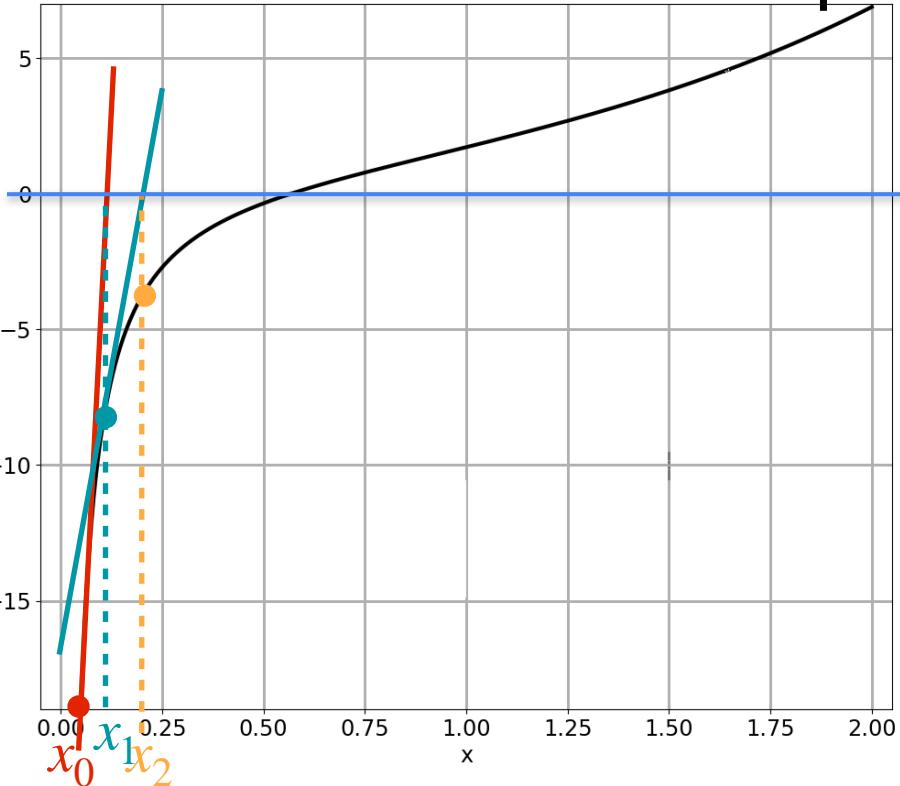
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_1 = 0.097$$

$$x_2 = x_1 - \frac{g'(x_1)}{(g'(x_1))'}$$

$$= 0.097 - \frac{\left(e^{0.097} - \frac{1}{0.097}\right)}{\left(e^{0.097} + \frac{1}{0.097^2}\right)} = 0.183$$

# Newton's Method for Optimization

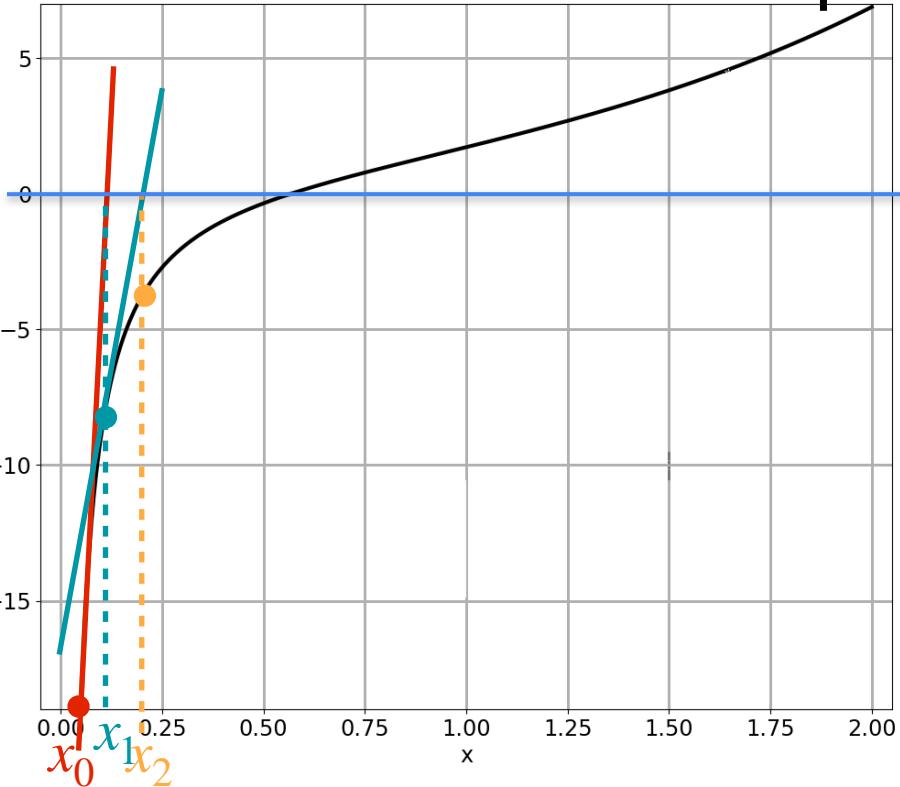


$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

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

# Newton's Method for Optimization



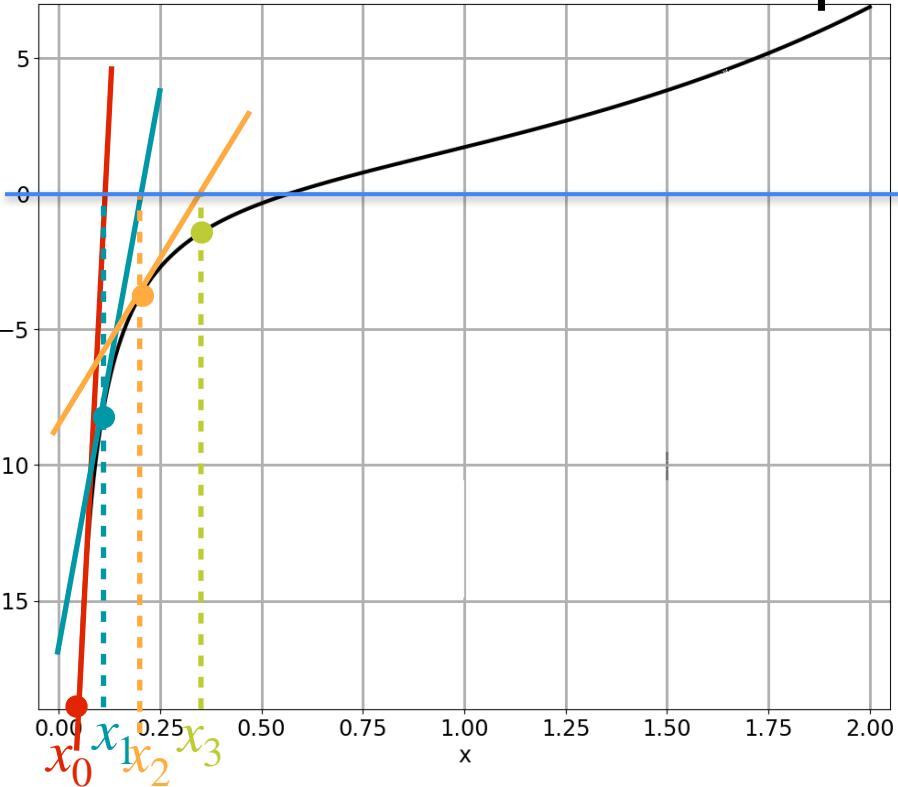
$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

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

$$x_2 = 0.183$$

# Newton's Method for Optimization



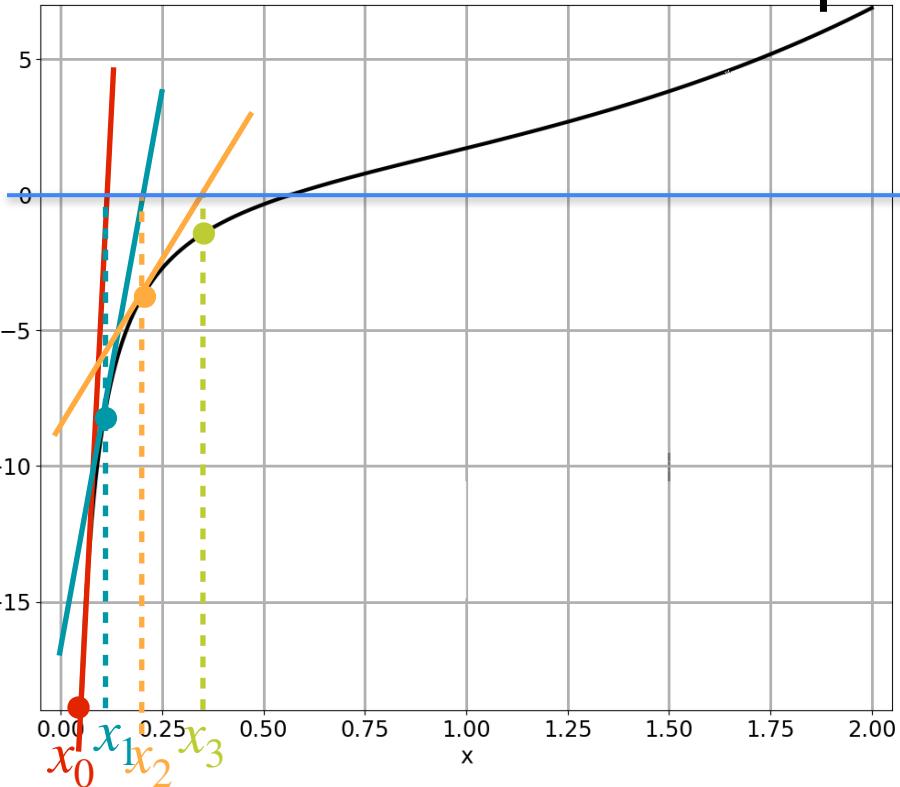
$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

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

$$x_2 = 0.183$$

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

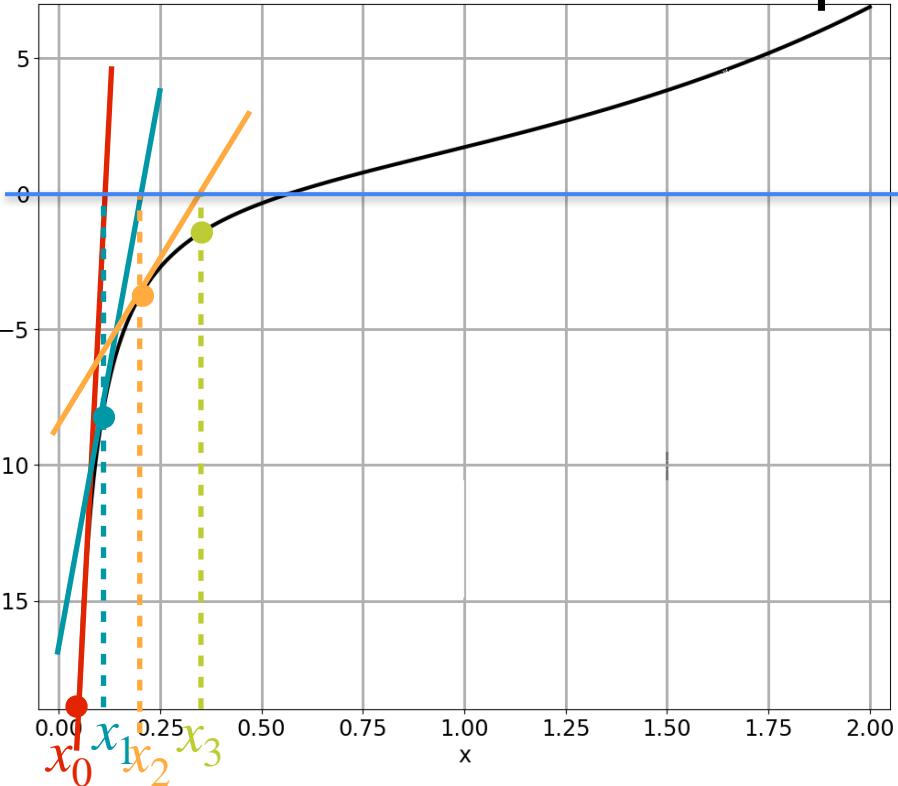
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_2 = 0.183$$

$$x_3 = x_2 - \frac{g'(x_2)}{(g'(x_2))'}$$

$$= 0.183 - \frac{\left(e^{0.183} - \frac{1}{0.183}\right)}{\left(e^{0.183} + \frac{1}{0.183^2}\right)}$$

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

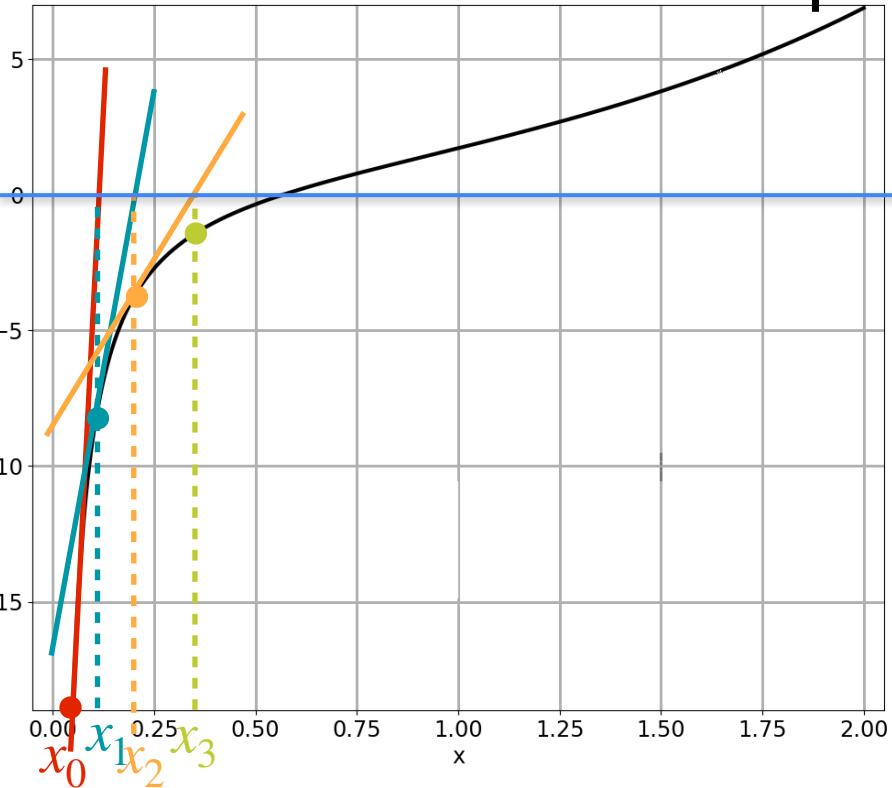
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_2 = 0.183$$

$$x_3 = x_2 - \frac{g'(x_2)}{(g'(x_2))'}$$

$$= 0.183 - \frac{\left(e^{0.183} - \frac{1}{0.183}\right)}{\left(e^{0.183} + \frac{1}{0.183^2}\right)} = 0.320$$

# Newton's Method for Optimization



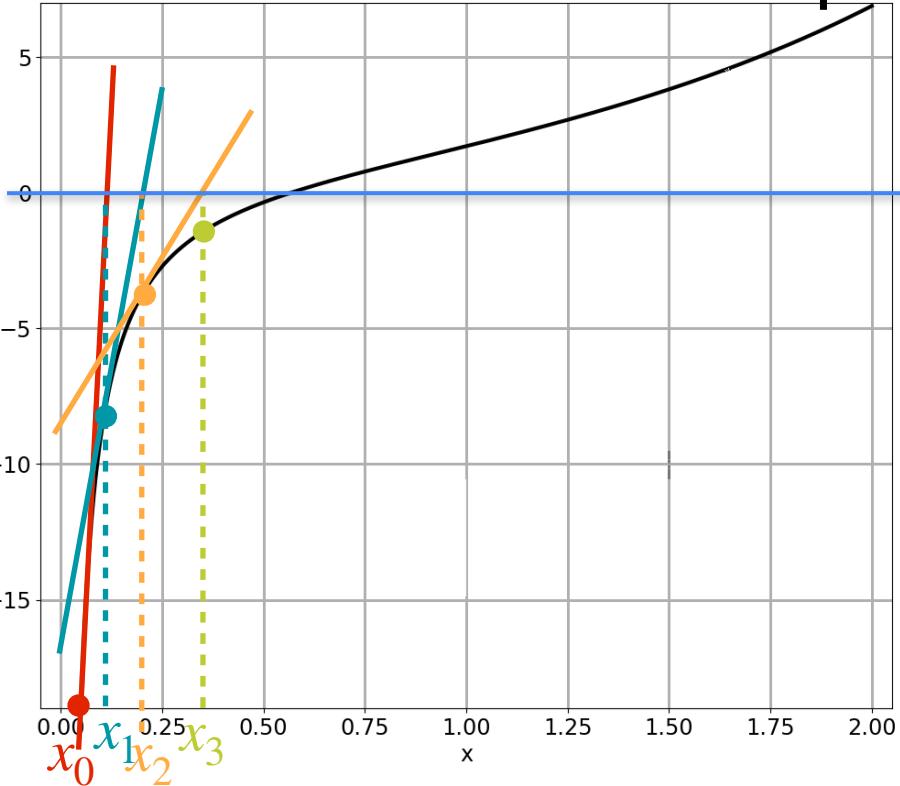
$$g(x) = e^x - \log(x)$$

$$g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

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

# Newton's Method for Optimization



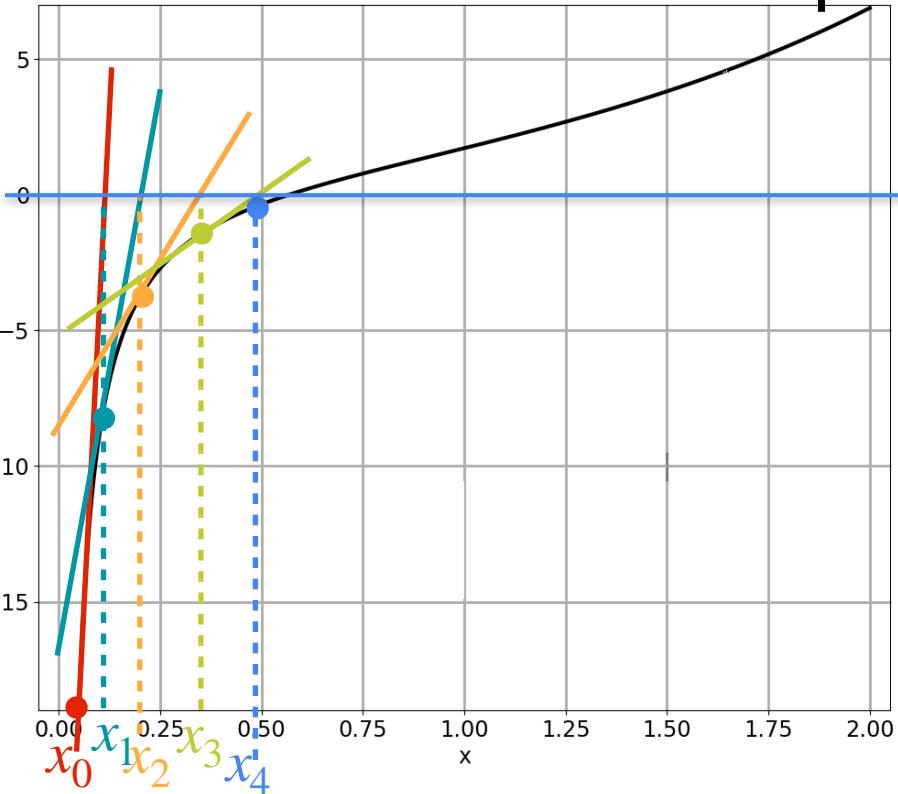
$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

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

$$x_3 = 0.320$$

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

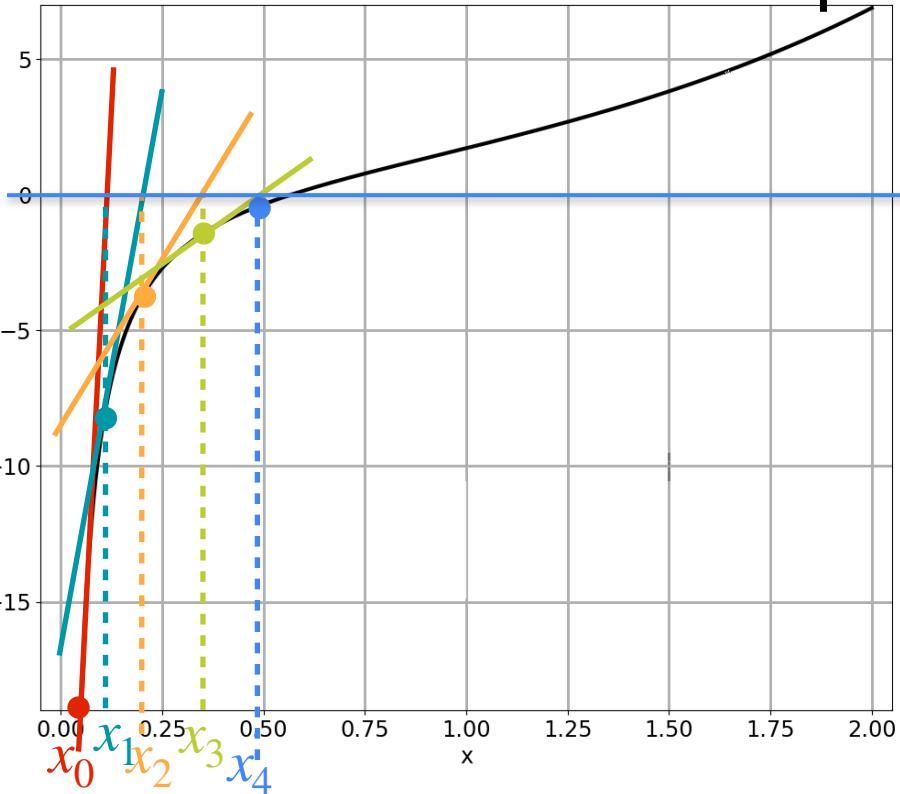
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_3 = 0.320$$

$$x_4 = x_3 - \frac{g'(x_3)}{(g'(x_3))'}$$

$$= 0.320 - \frac{\left(e^{0.320} - \frac{1}{0.320}\right)}{\left(e^{0.320} + \frac{1}{0.320^2}\right)}$$

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

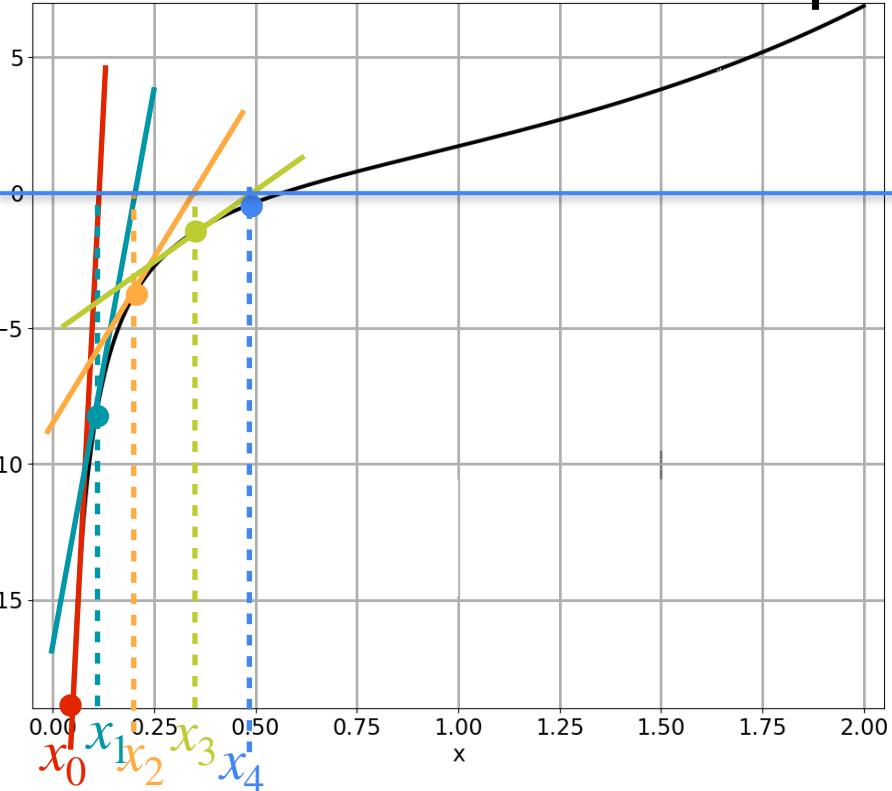
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_3 = 0.320$$

$$x_4 = x_3 - \frac{g'(x_3)}{(g'(x_3))'}$$

$$= 0.320 - \frac{\left(e^{0.320} - \frac{1}{0.320}\right)}{\left(e^{0.320} + \frac{1}{0.320^2}\right)} = 0.477$$

# Newton's Method for Optimization

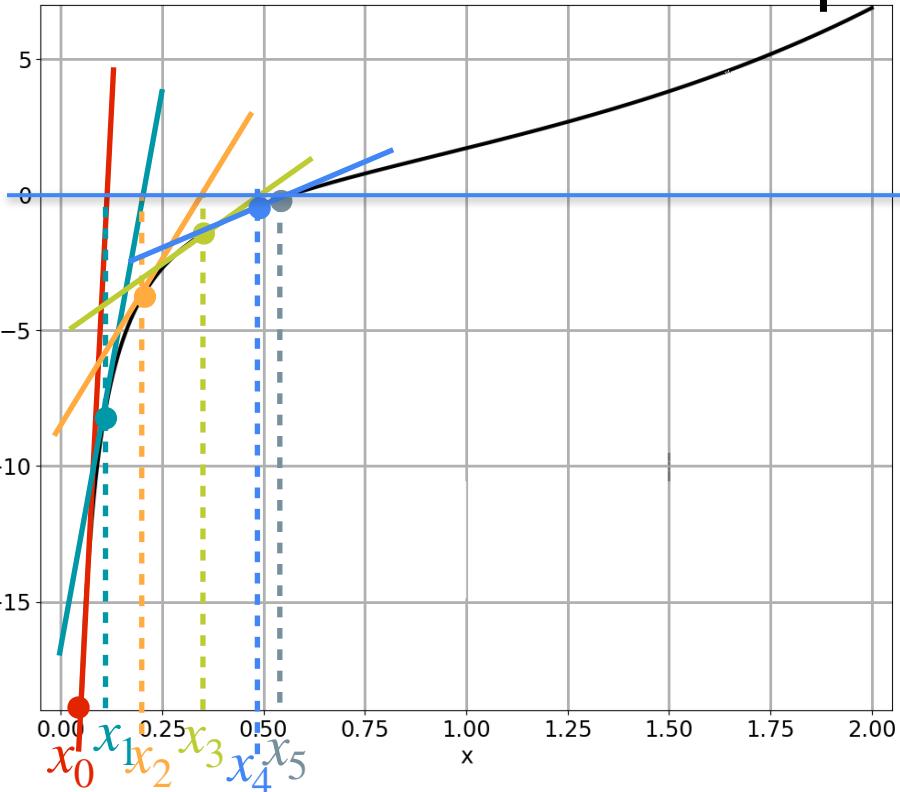


$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.5671$

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

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.5671$$

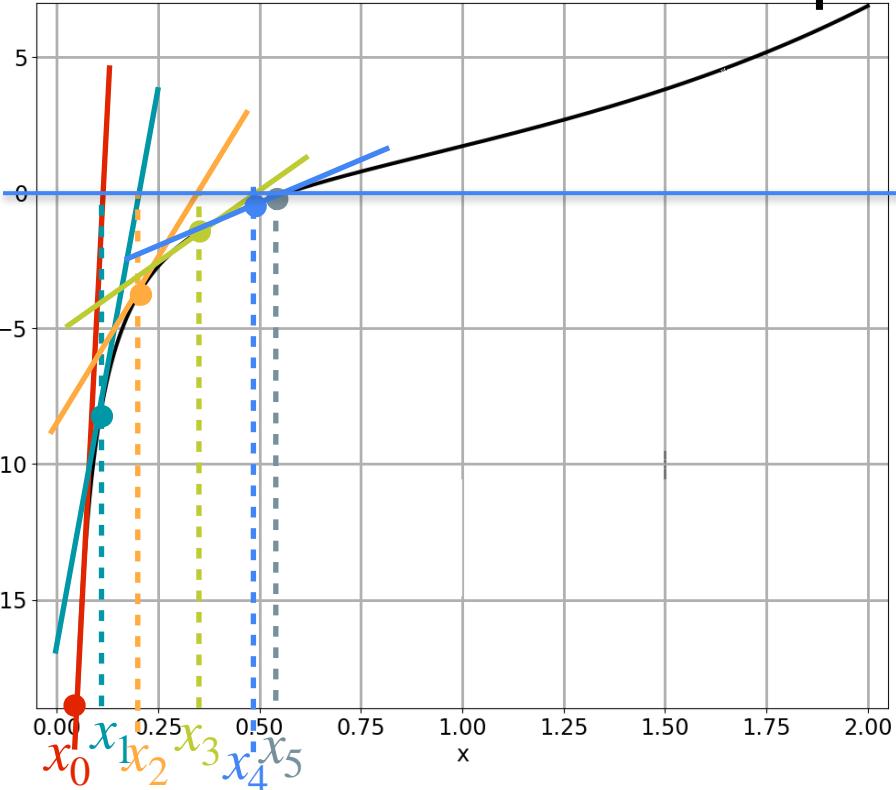
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_4 = 0.477$$

$$x_5 = x_4 - \frac{g'(x_4)}{(g'(x_4))'}$$

$$= 0.447 - \frac{\left(e^{0.447} - \frac{1}{0.447}\right)}{\left(e^{0.447} + \frac{1}{0.447^2}\right)} = 0.558$$

# Newton's Method for Optimization

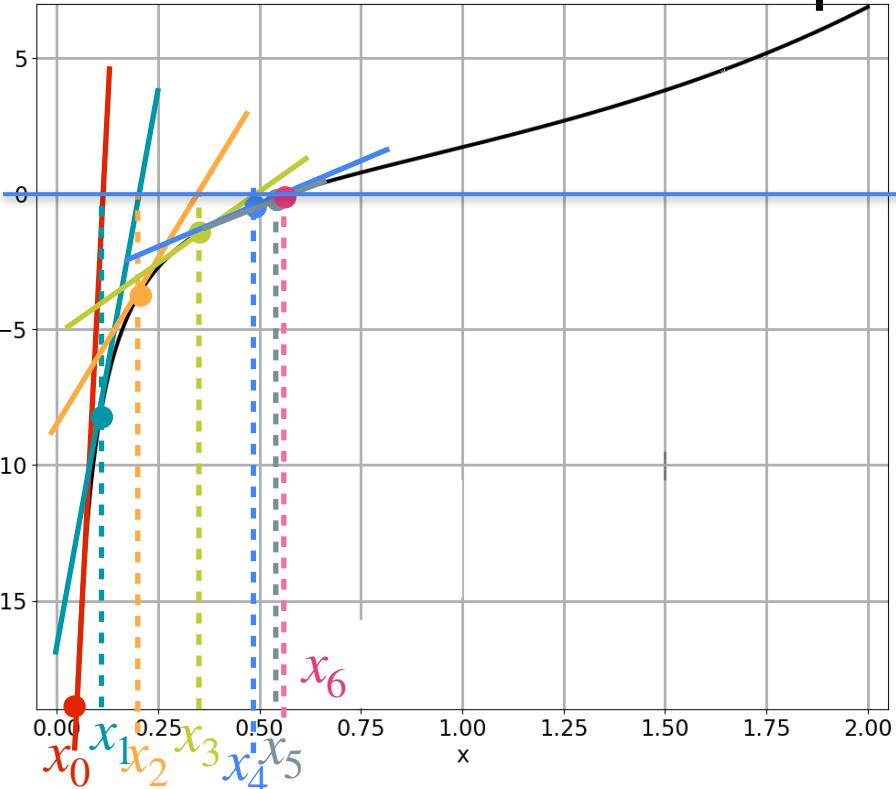


$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.567$

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

# Newton's Method for Optimization

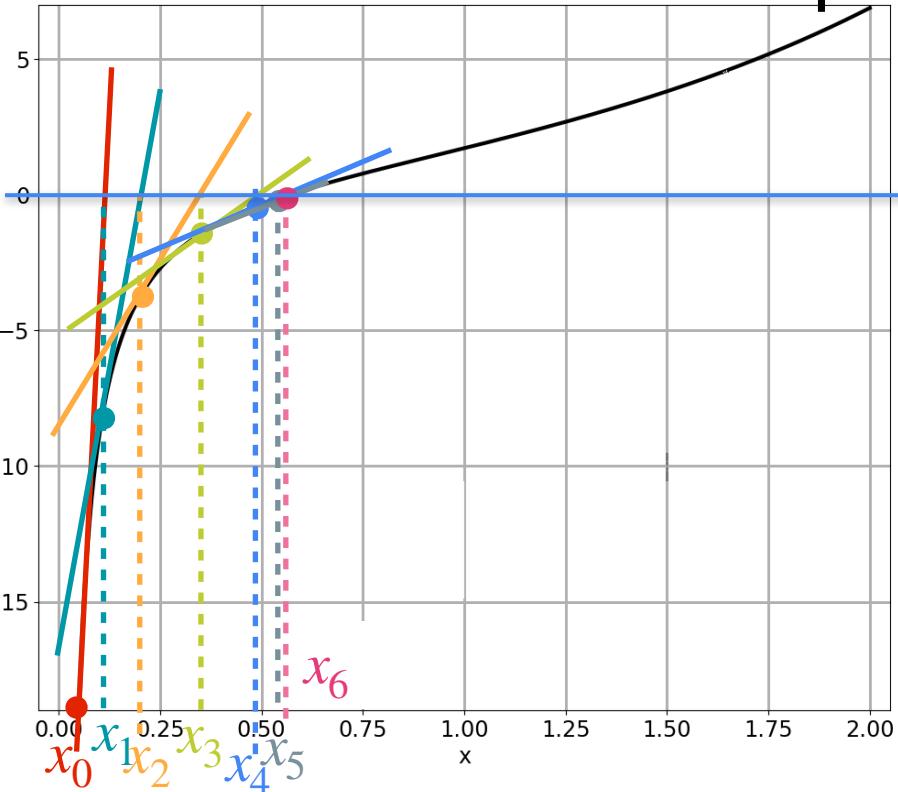


$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.567$

$$(g'(x))' = e^x + \frac{1}{x^2}$$
$$x_5 = 0.558$$

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

$$\text{Minimum: } x^* = 0.567$$

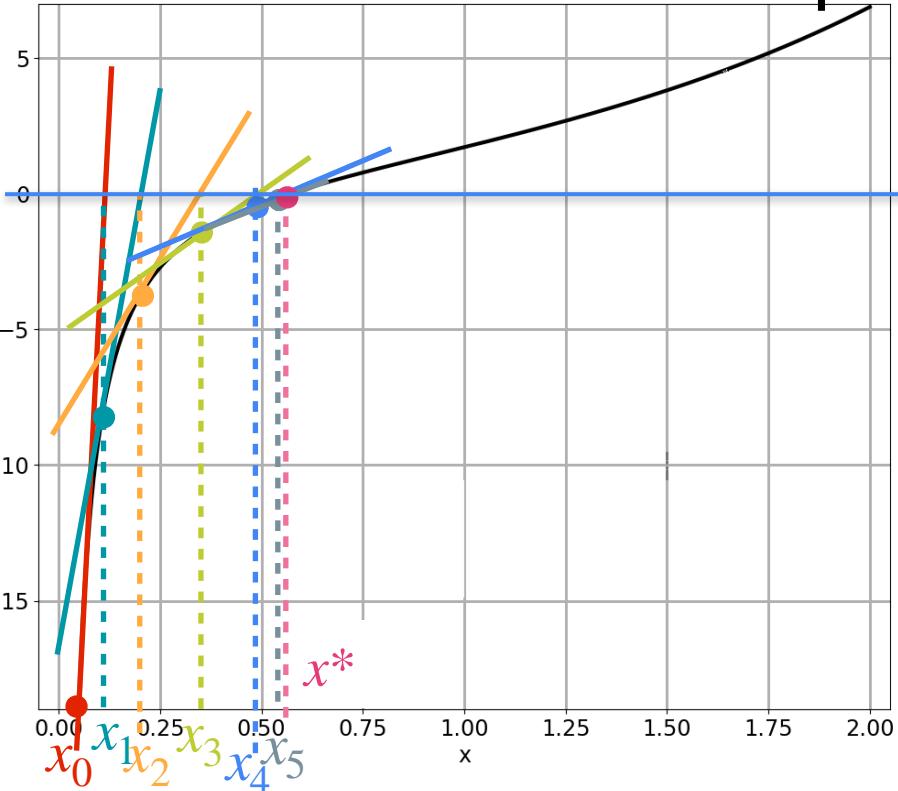
$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_5 = 0.558$$

$$x_6 = x_5 - \frac{g'(x_5)}{(g'(x_5))'}$$

$$= 0.558 - \frac{\left(e^{0.558} - \frac{1}{0.558}\right)}{\left(e^{0.558} + \frac{1}{0.558^2}\right)}$$

# Newton's Method for Optimization



$$g(x) = e^x - \log(x) \quad g'(x) = e^x - 1/x$$

Minimum:  $x^* = 0.567$

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

$$x_5 = 0.558$$

$$x^* = x_5 - \frac{g'(x_5)}{(g'(x_5))'}$$

$$= 0.558 - \frac{\left(e^{0.558} - \frac{1}{0.558}\right)}{\left(e^{0.558} + \frac{1}{0.558^2}\right)} = 0.567$$



DeepLearning.AI

# Optimization in Neural Networks and Newton's Method

---

## The second derivative

# Second Derivative

# Second Derivative

Newton's method:

# Second Derivative

Newton's method:  $x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'}$

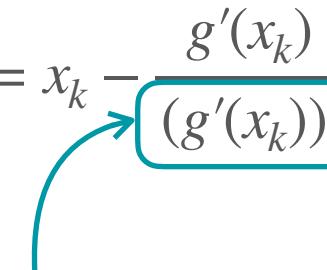
# Second Derivative

Newton's method:  $x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'} ??$

# Second Derivative

Newton's method:  $x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'}$  ??

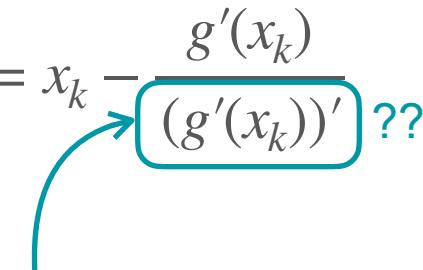
Second derivative



# Second Derivative

Newton's method:  $x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'}$  ??

Second derivative



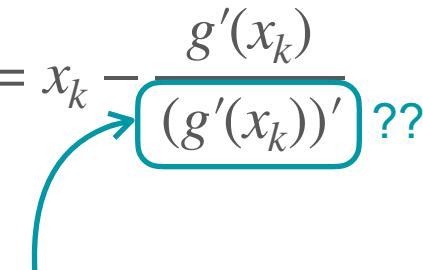
Leibniz notation:

$$\frac{d^2f(x)}{dx^2} = \frac{d}{dx} \left( \frac{df(x)}{dx} \right)$$

# Second Derivative

Newton's method:  $x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'}$  ??

Second derivative



Leibniz notation:

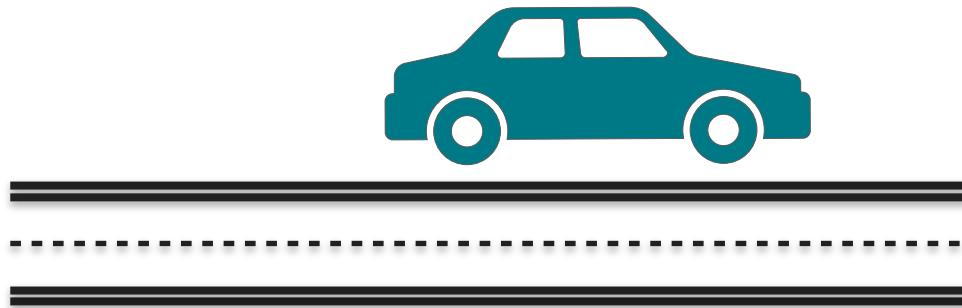
$$\frac{d^2f(x)}{dx^2} = \frac{d}{dx} \left( \frac{df(x)}{dx} \right)$$

Lagrange notation:

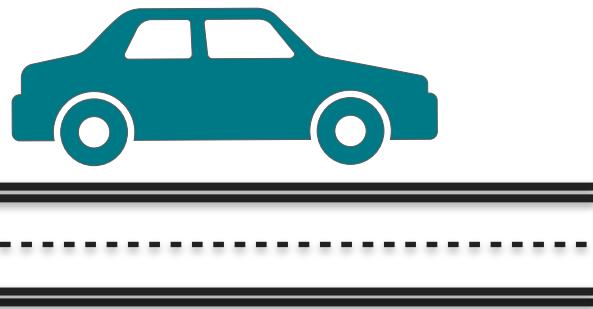
$$f''(x)$$

# Understanding Second Derivative

# Understanding Second Derivative

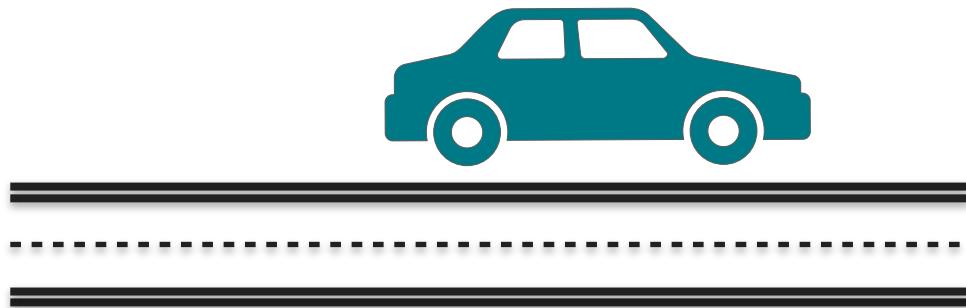


# Understanding Second Derivative



$x$  Distance

# Understanding Second Derivative

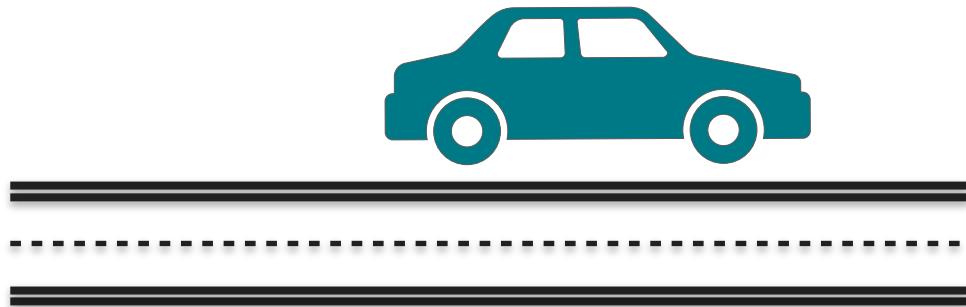


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

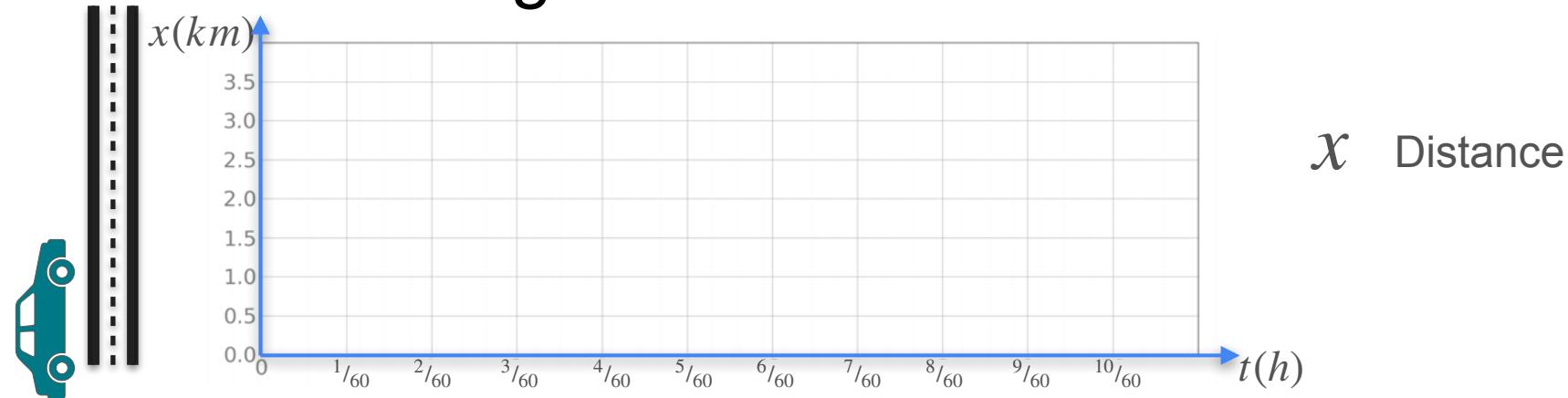
# Understanding Second Derivative



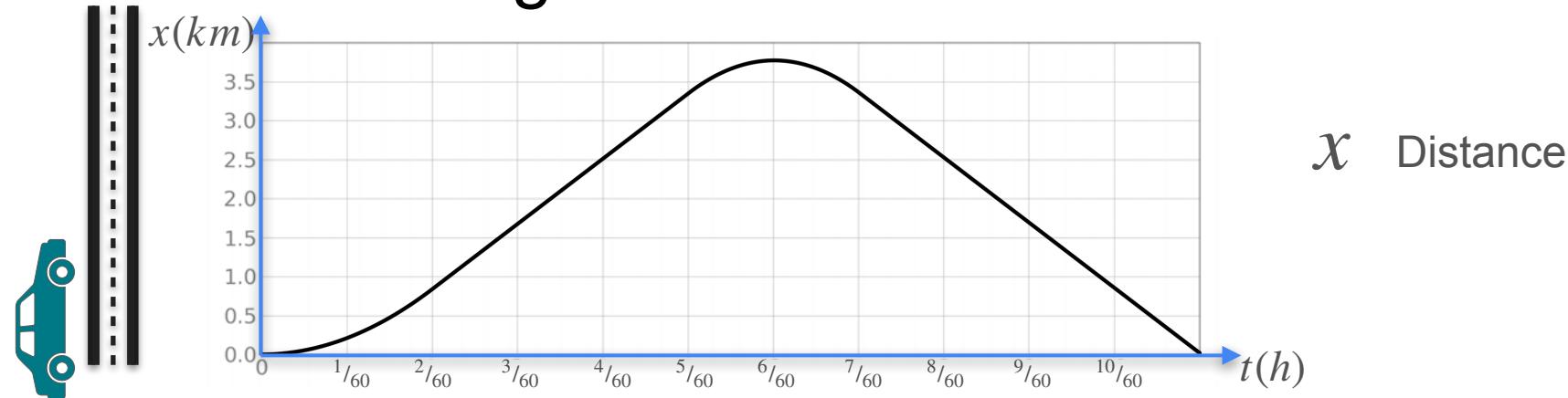
$x$	Distance	
$v$	Velocity	$\frac{dx}{dt}$
$a$	Acceleration	$\frac{dv}{dt} = \frac{d^2x}{dt^2}$

# Understanding Second Derivative

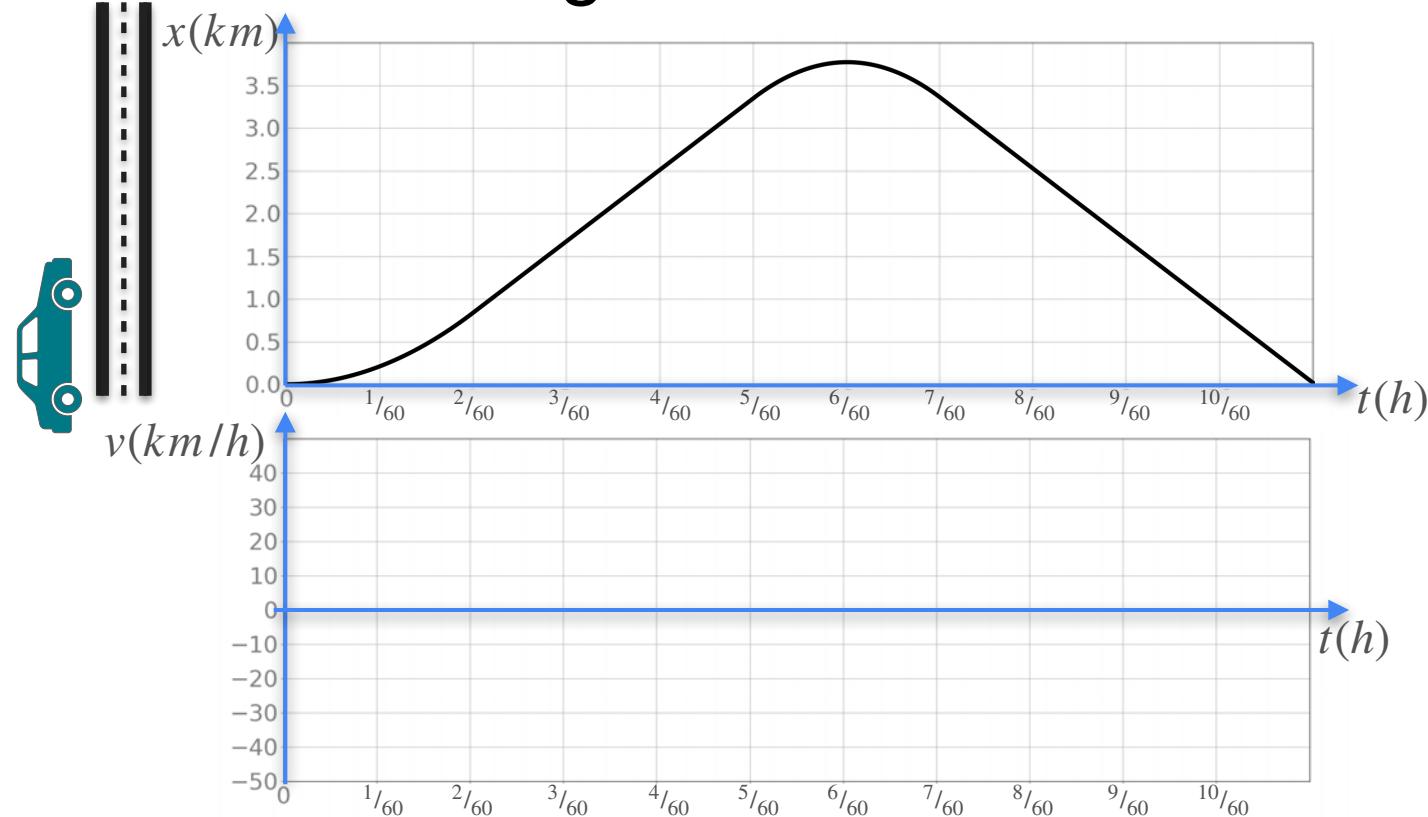
# Understanding Second Derivative



# Understanding Second Derivative



# Understanding Second Derivative

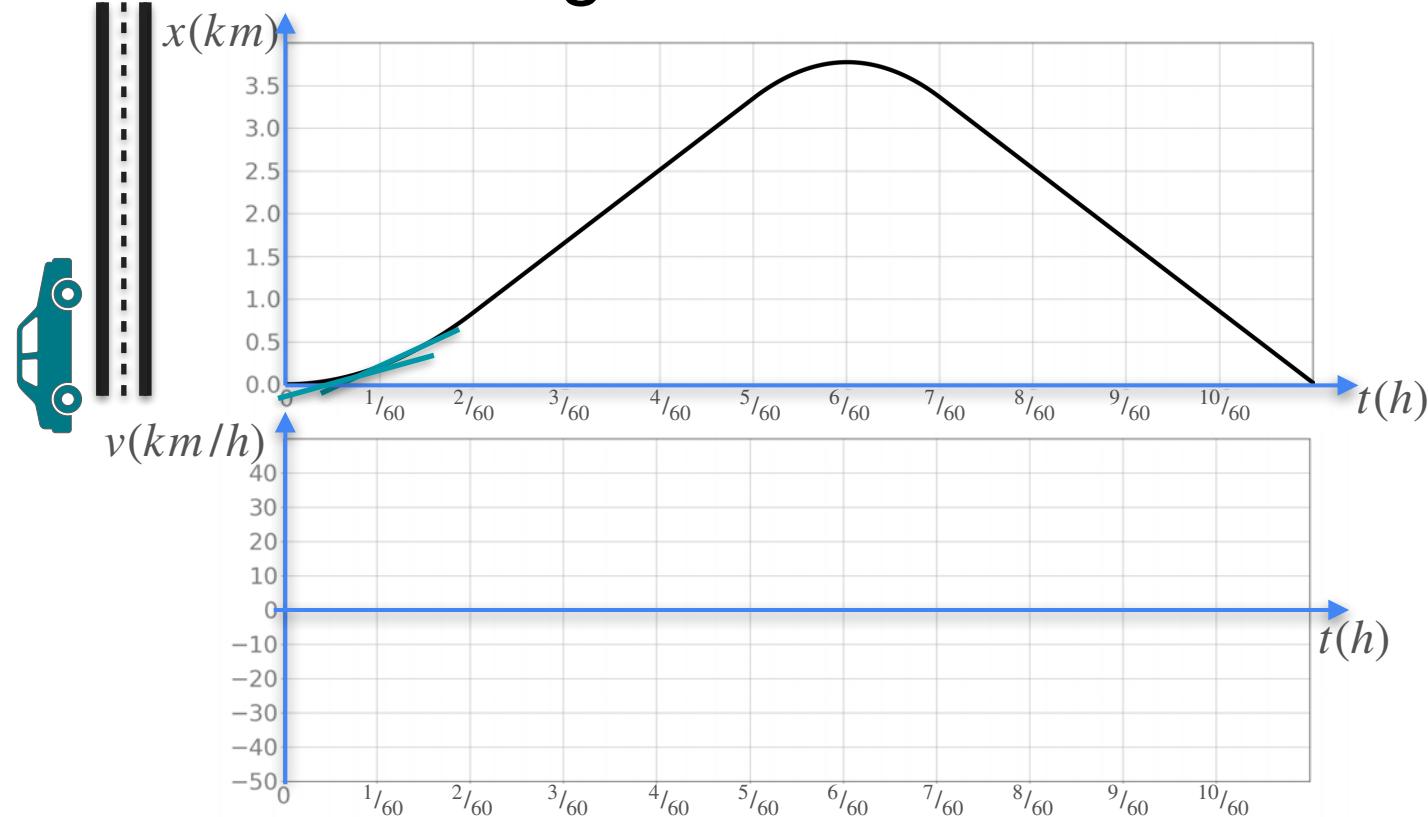


$\mathcal{X}$  Distance

$\mathcal{V}$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

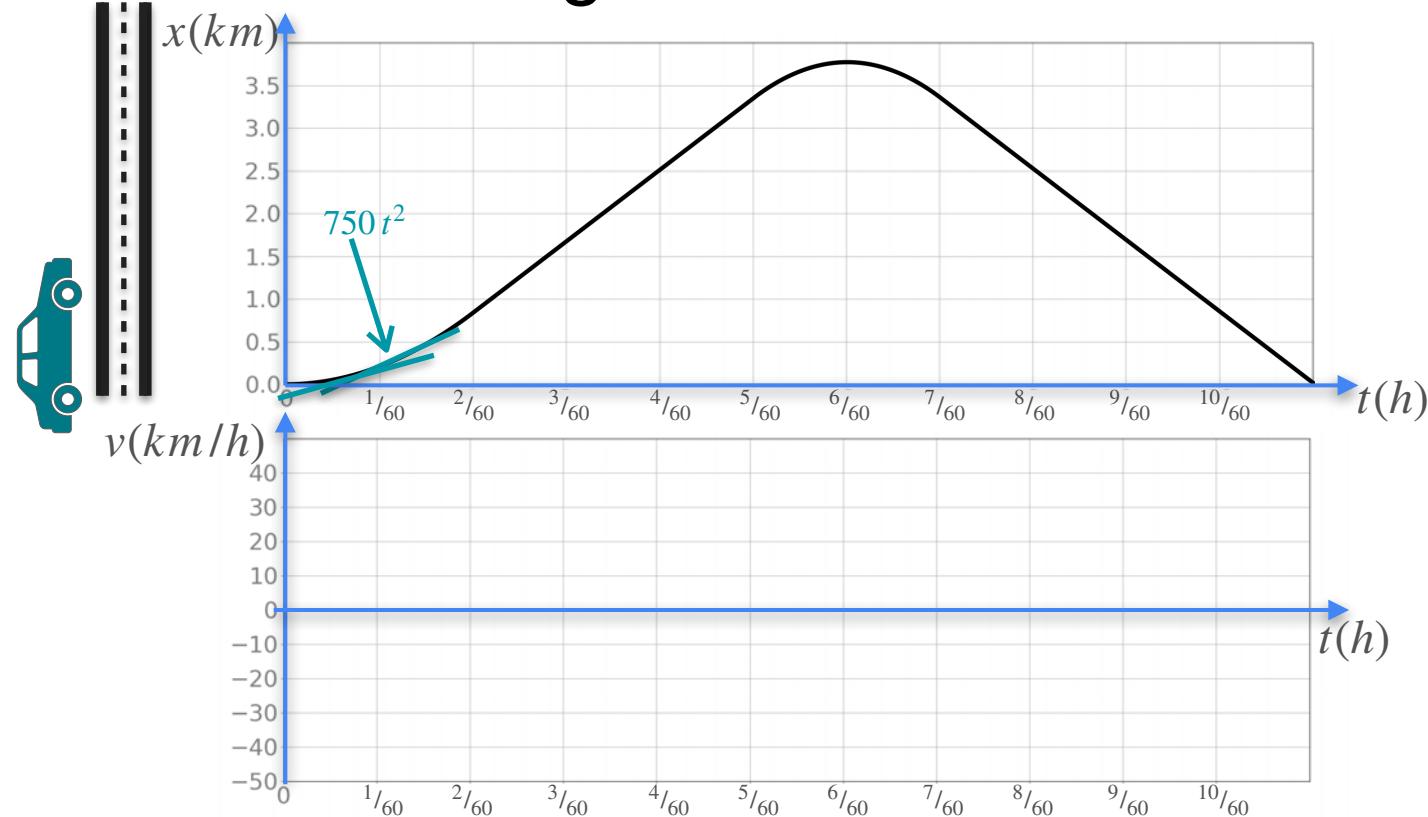


$\mathcal{X}$  Distance

$\mathcal{V}$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

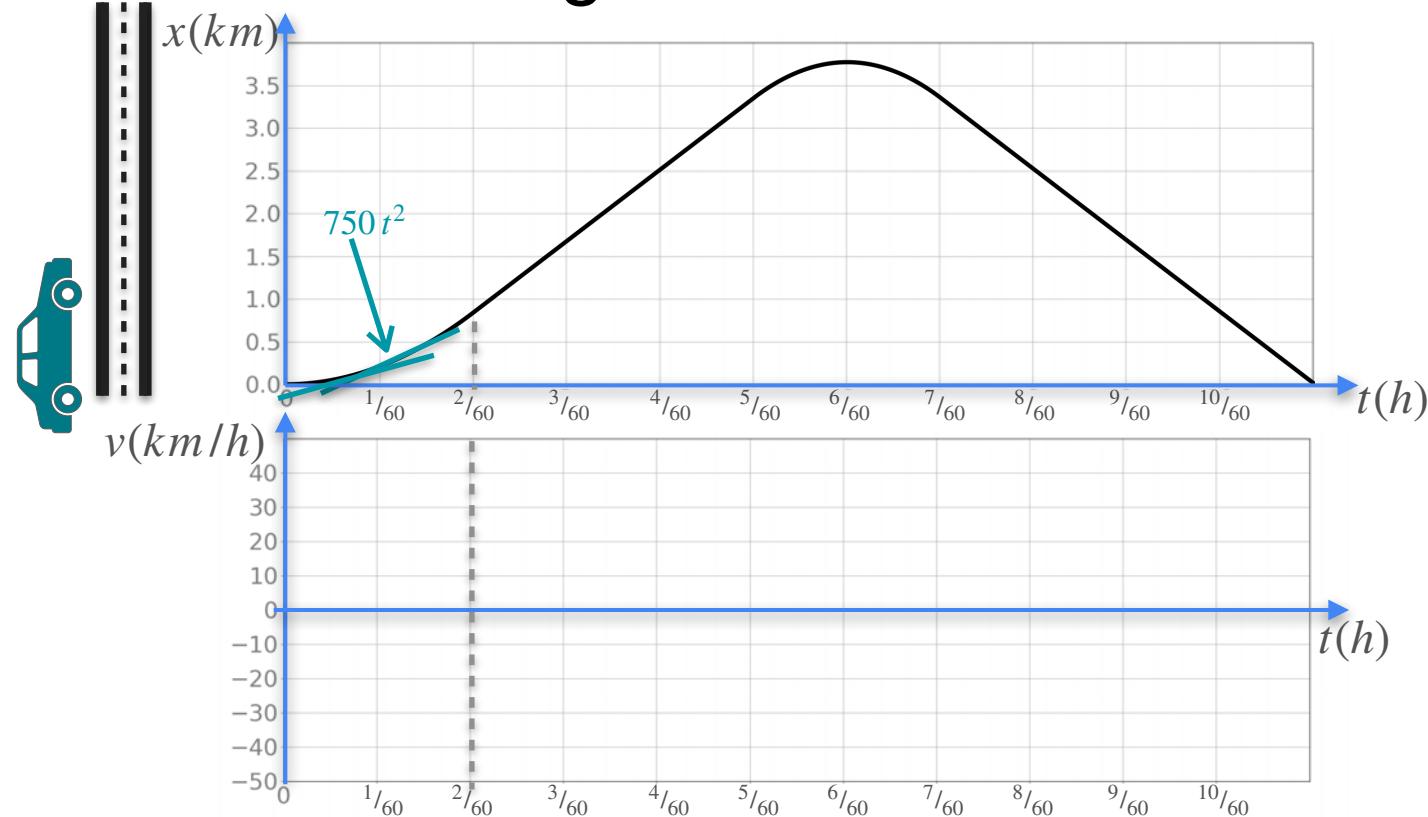


$\mathcal{X}$  Distance

$\mathcal{V}$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

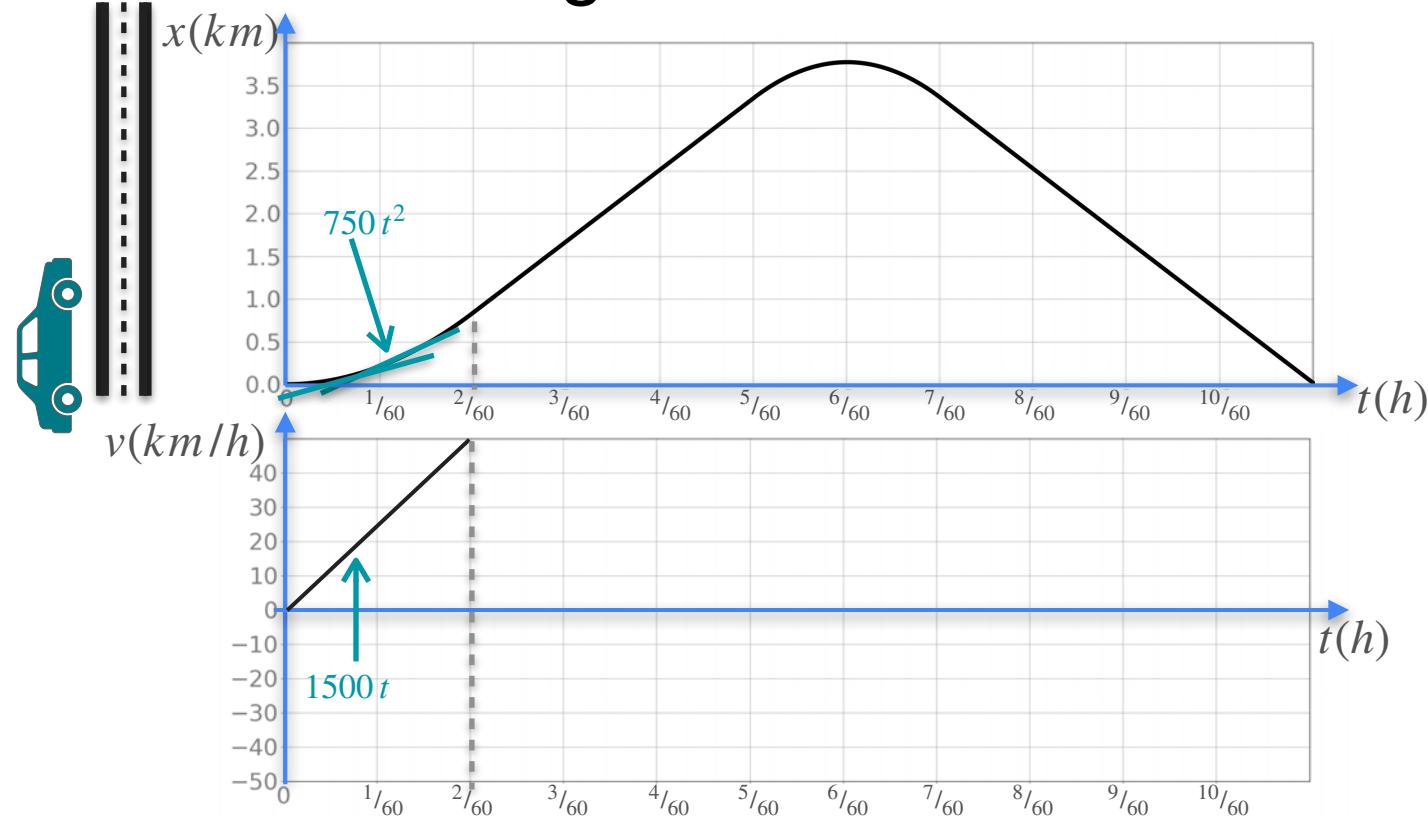


$\mathcal{X}$  Distance

$\mathcal{V}$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

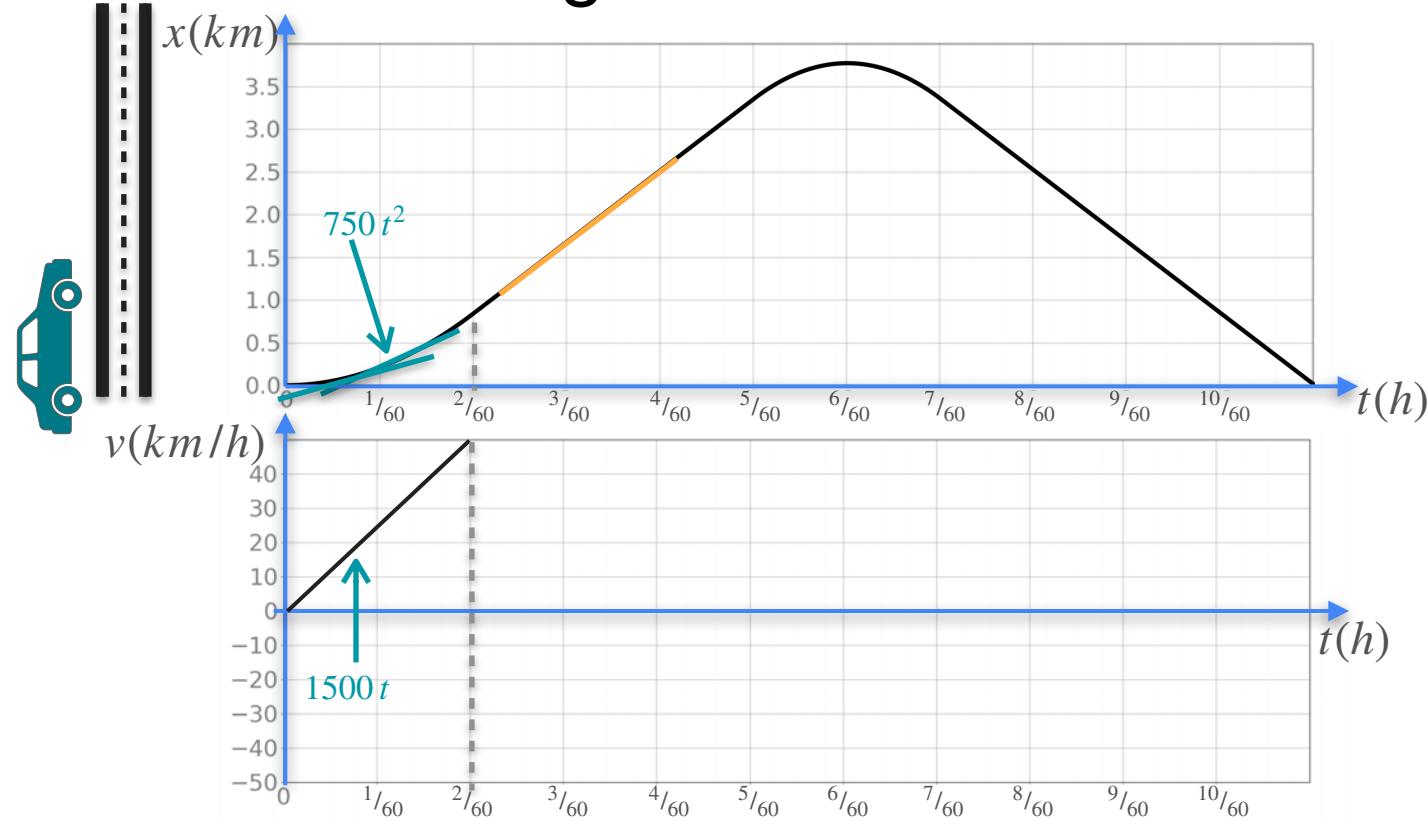


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

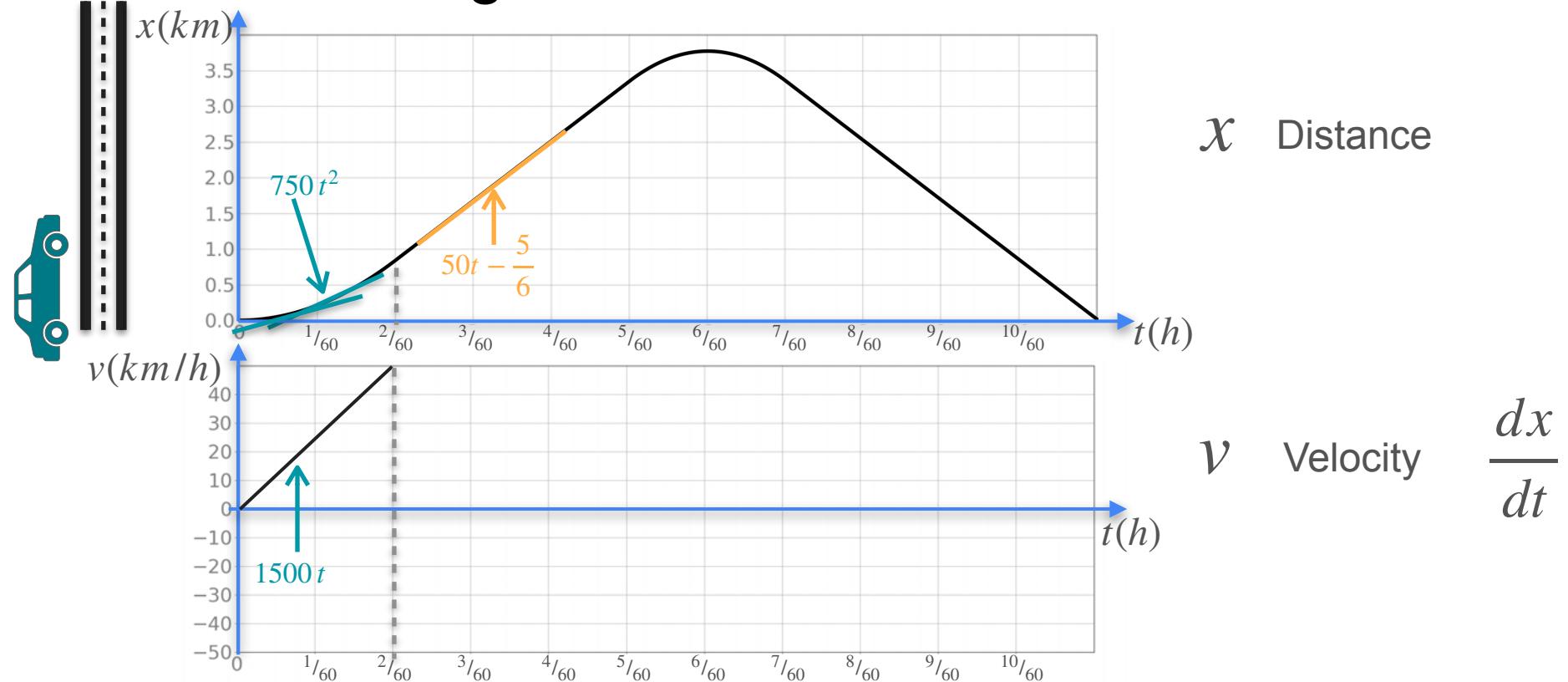


$x$  Distance

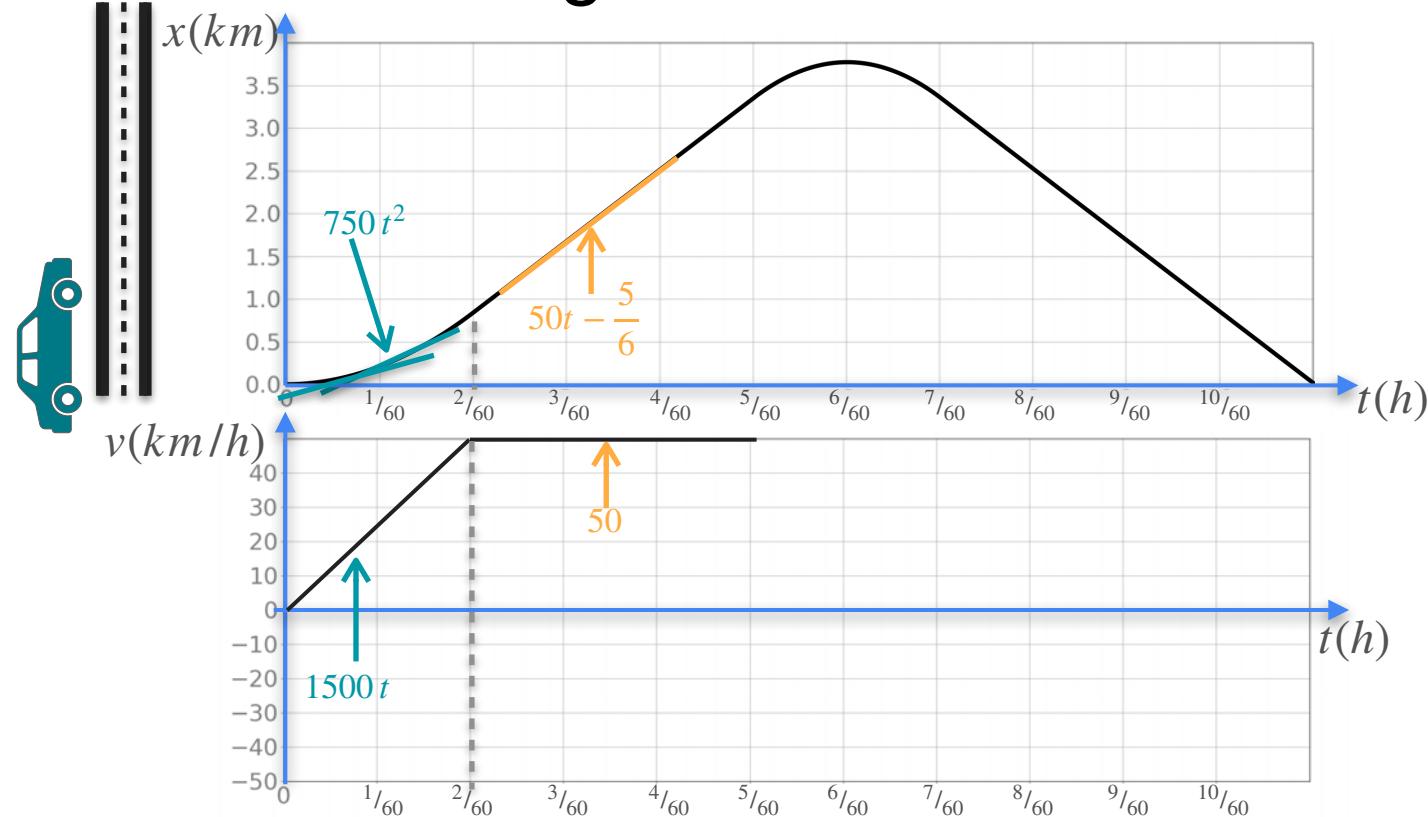
$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative



# Understanding Second Derivative

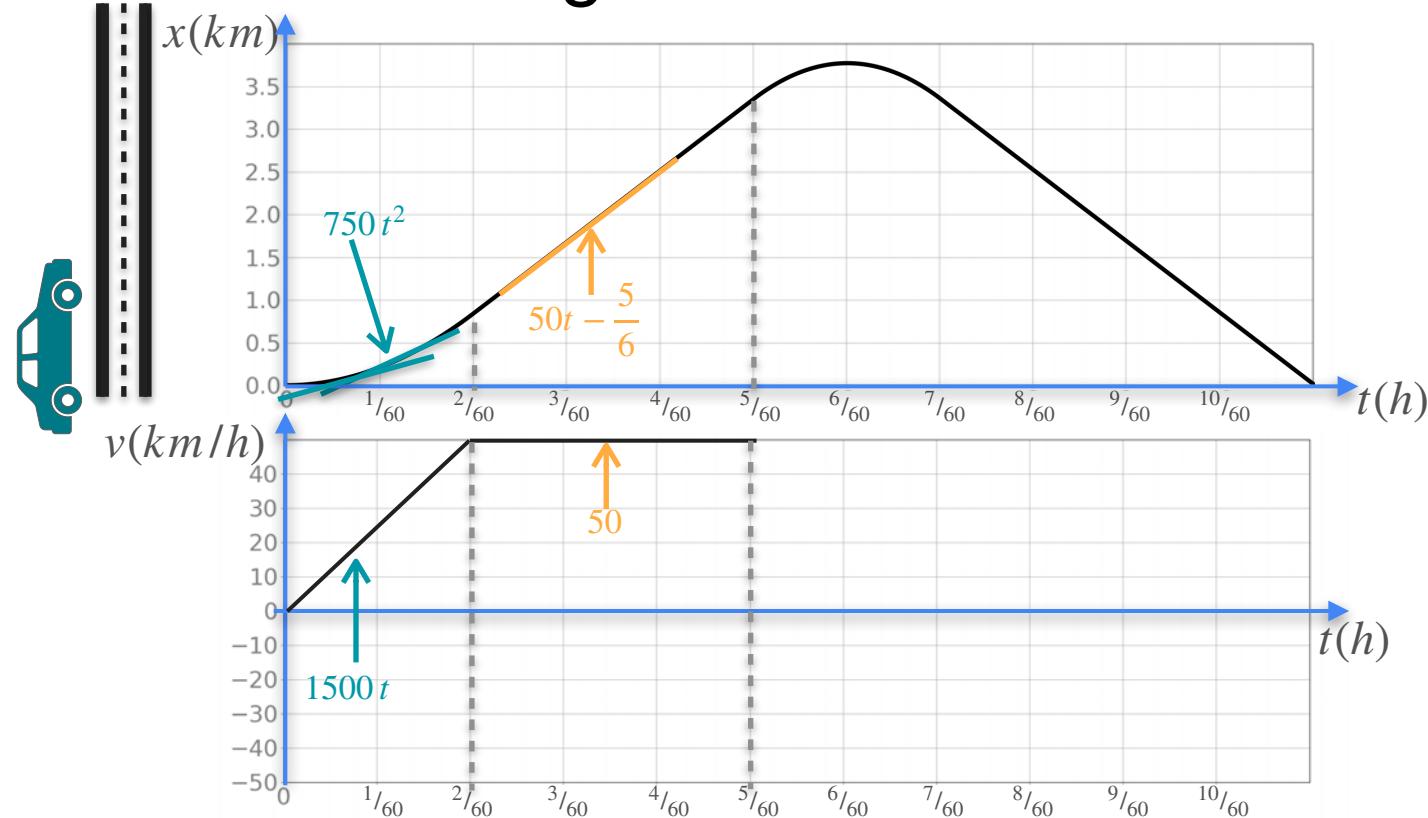


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

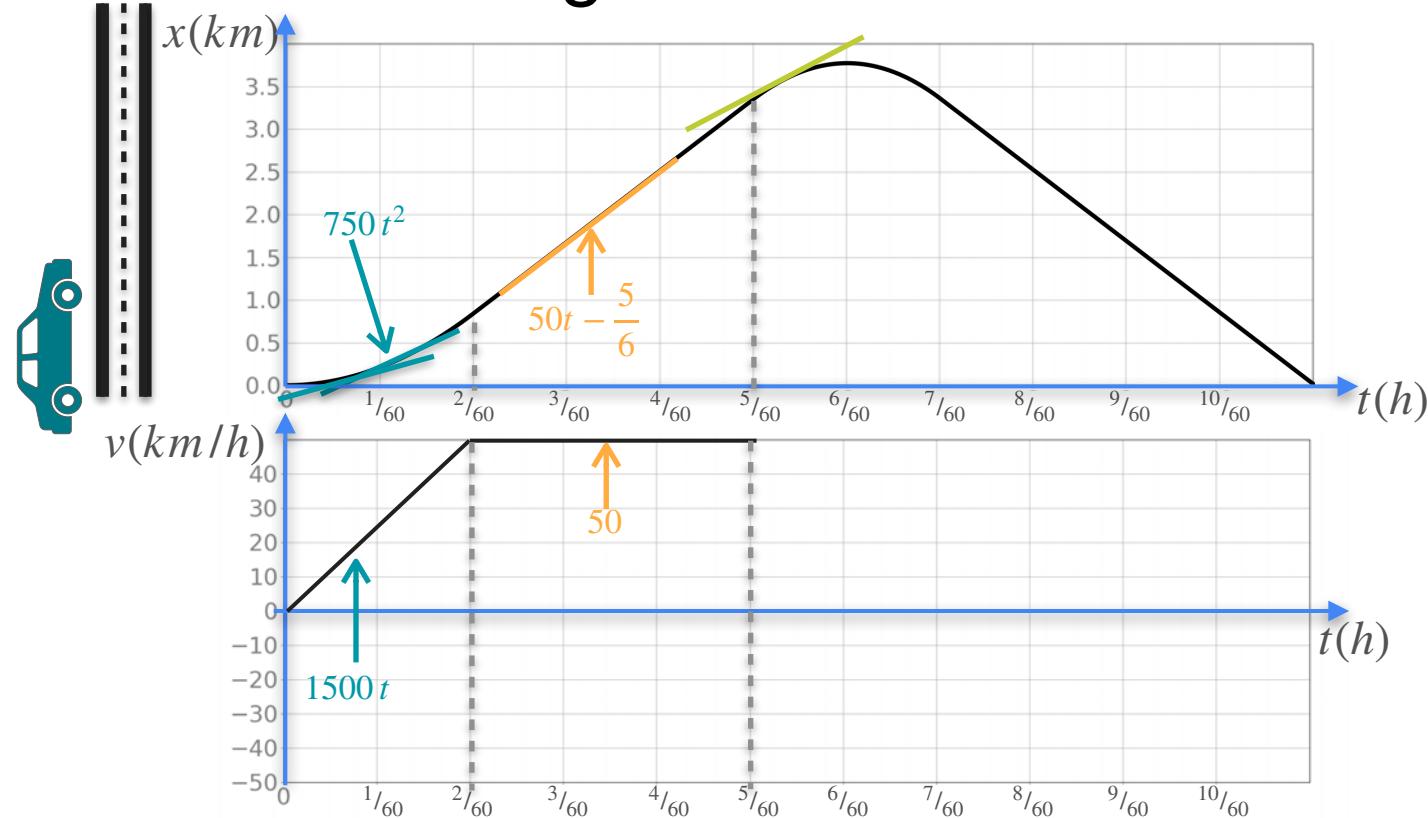


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

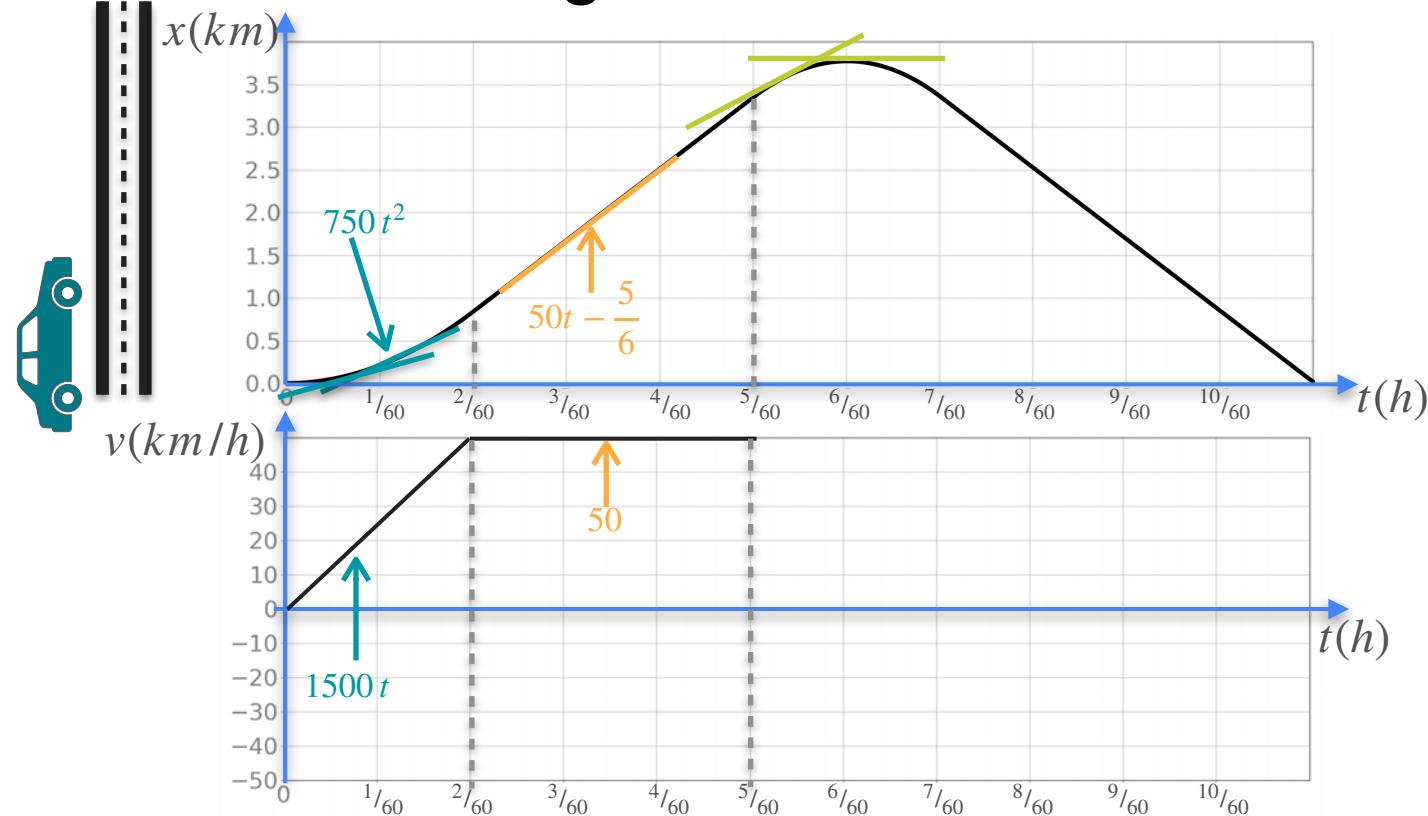


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

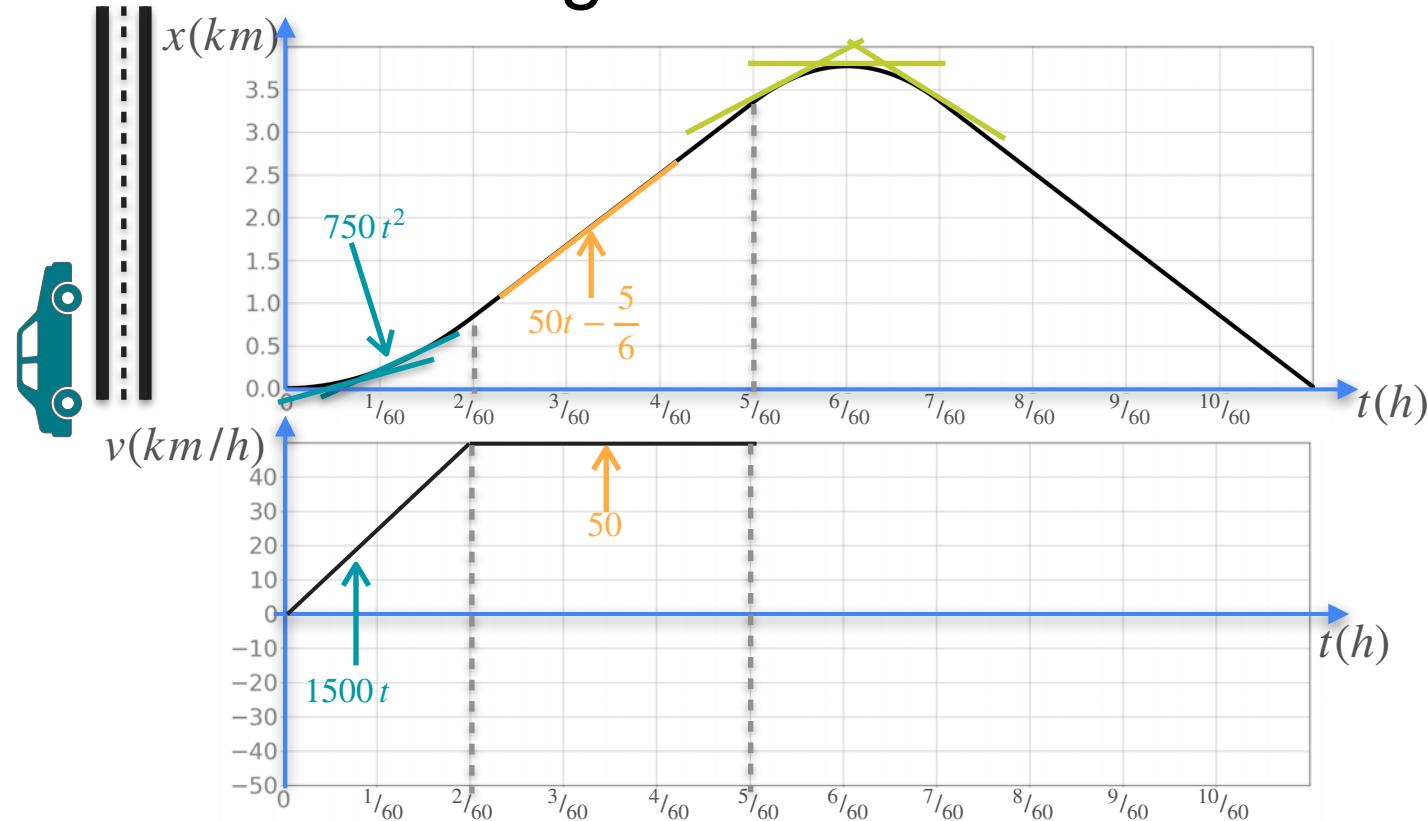


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

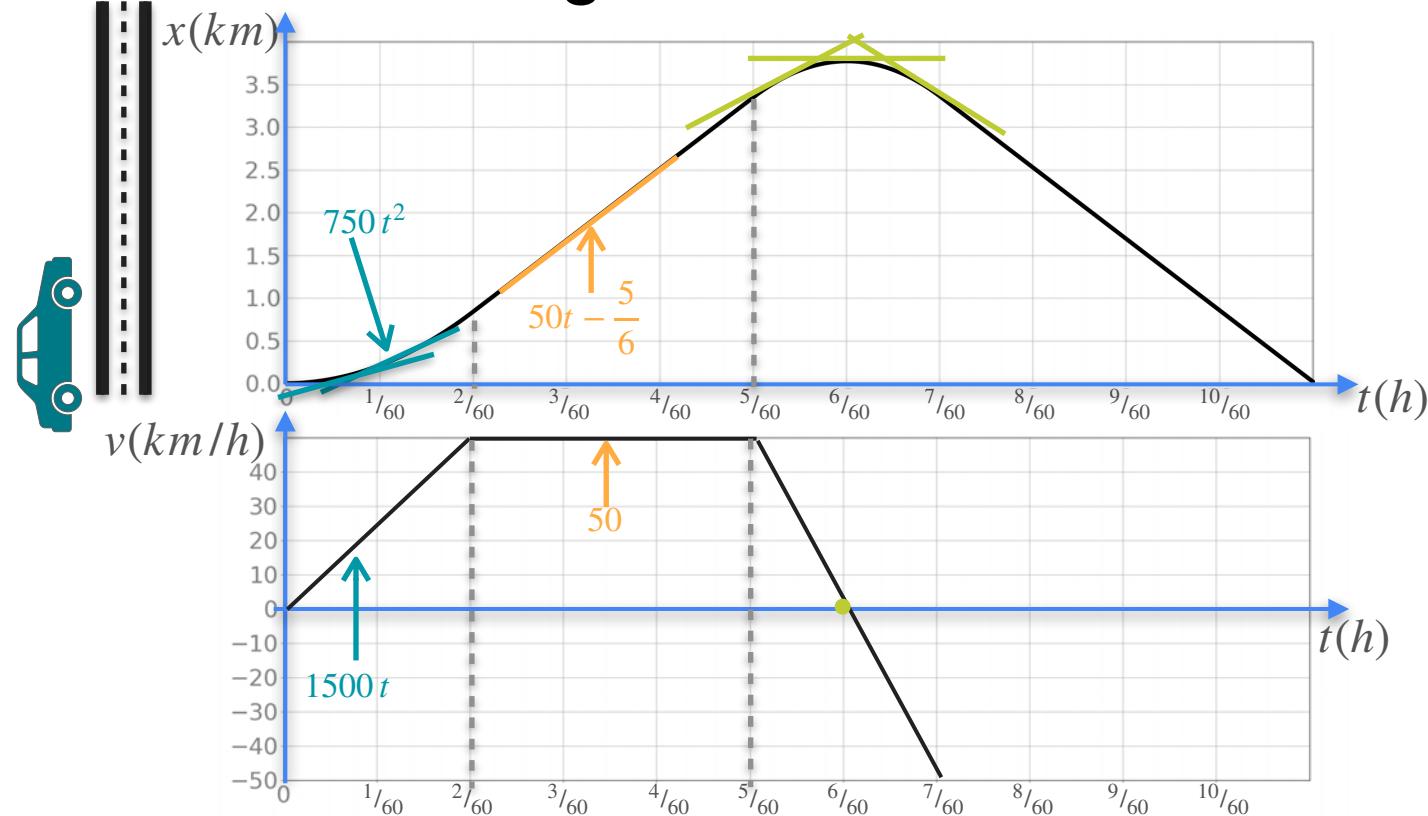


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

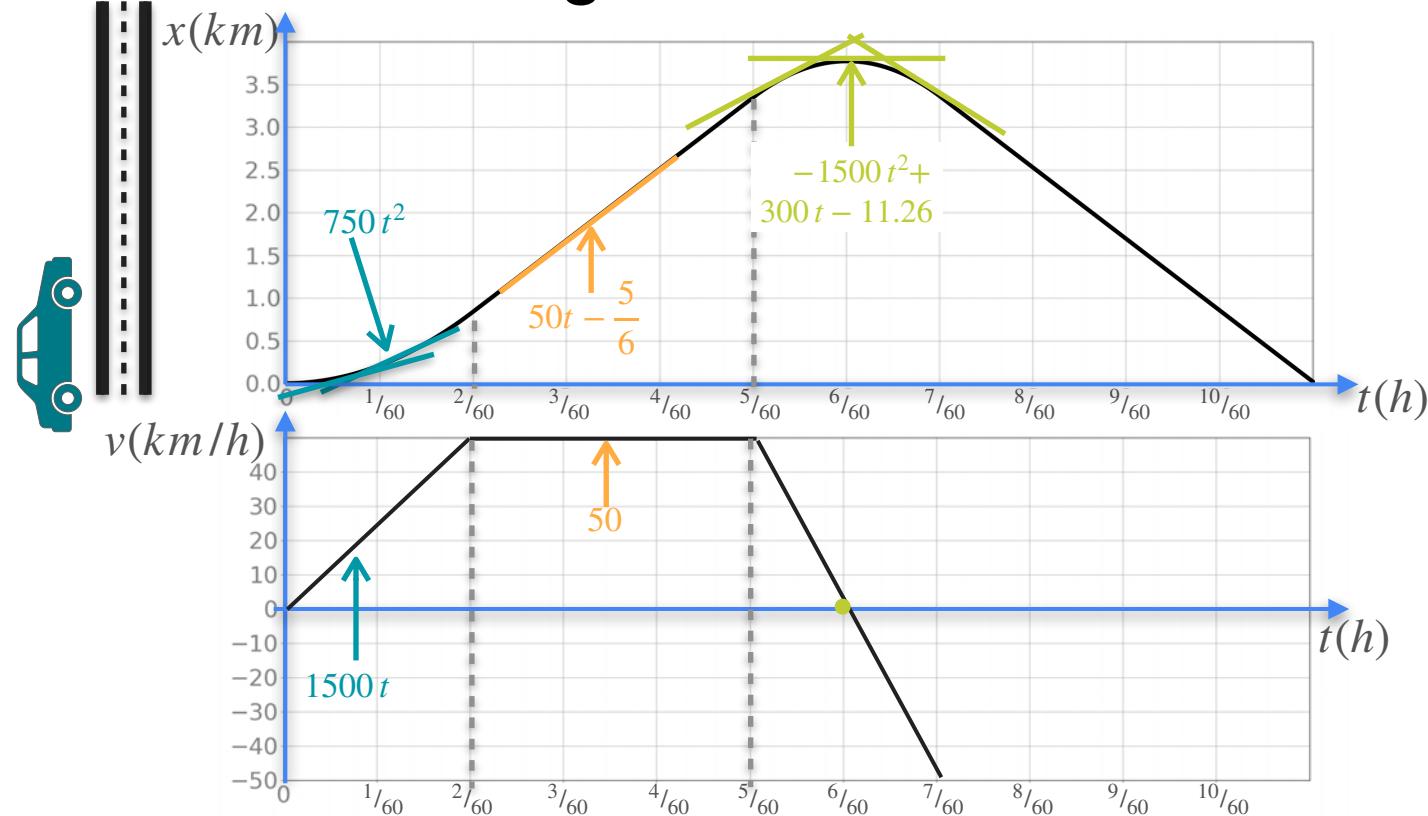


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

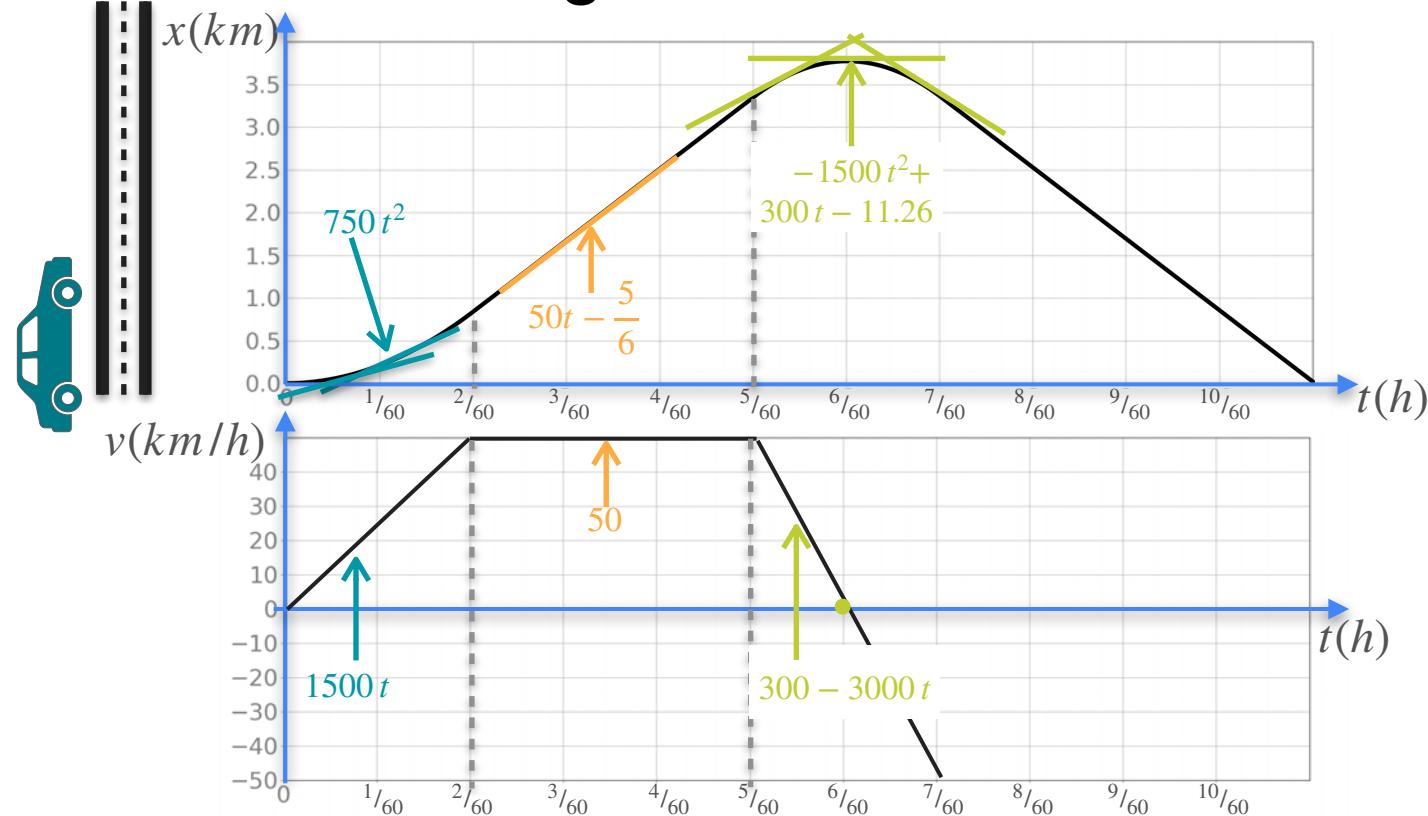


$\mathcal{X}$  Distance

$\mathcal{V}$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

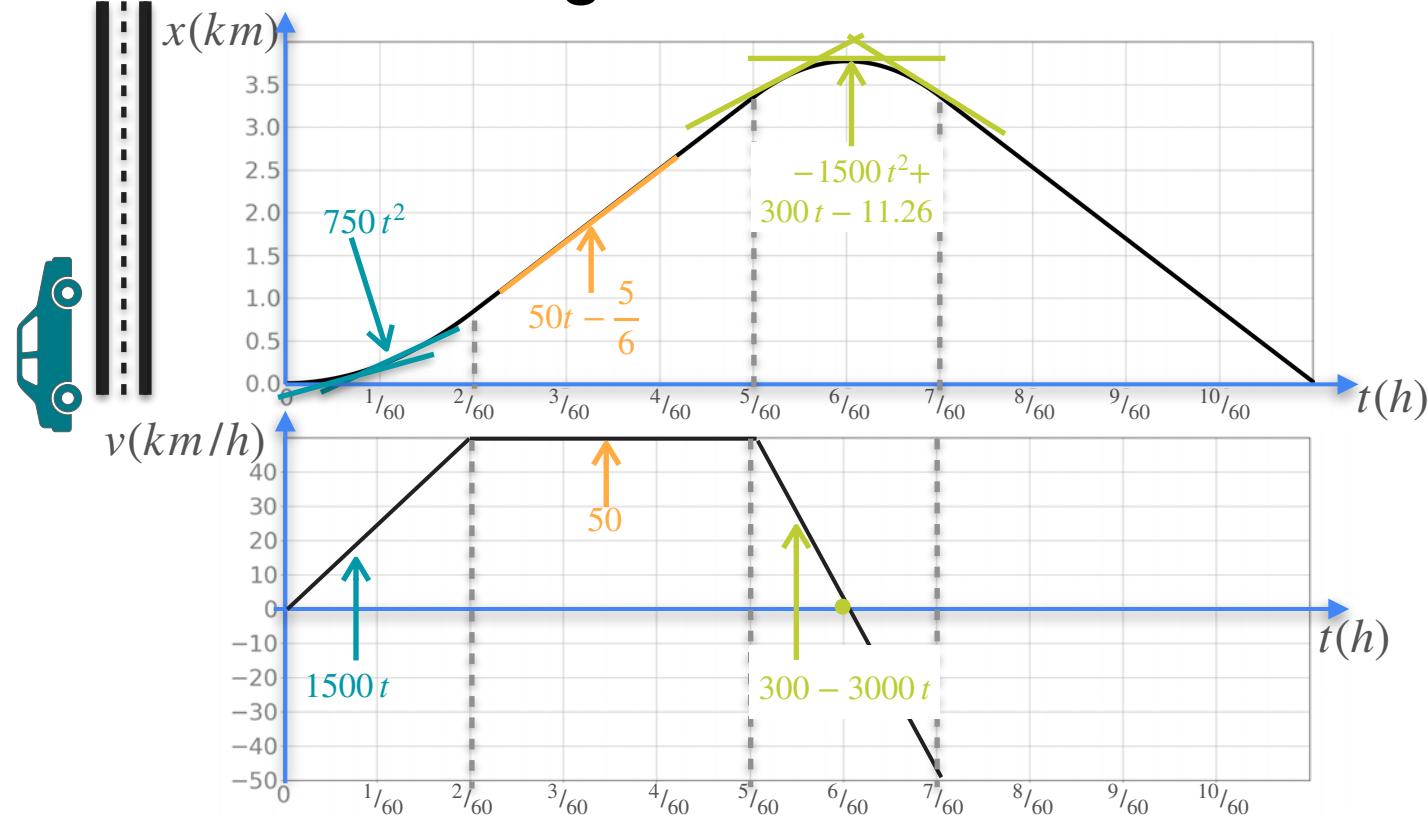


$\mathcal{X}$  Distance

$\mathcal{V}$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

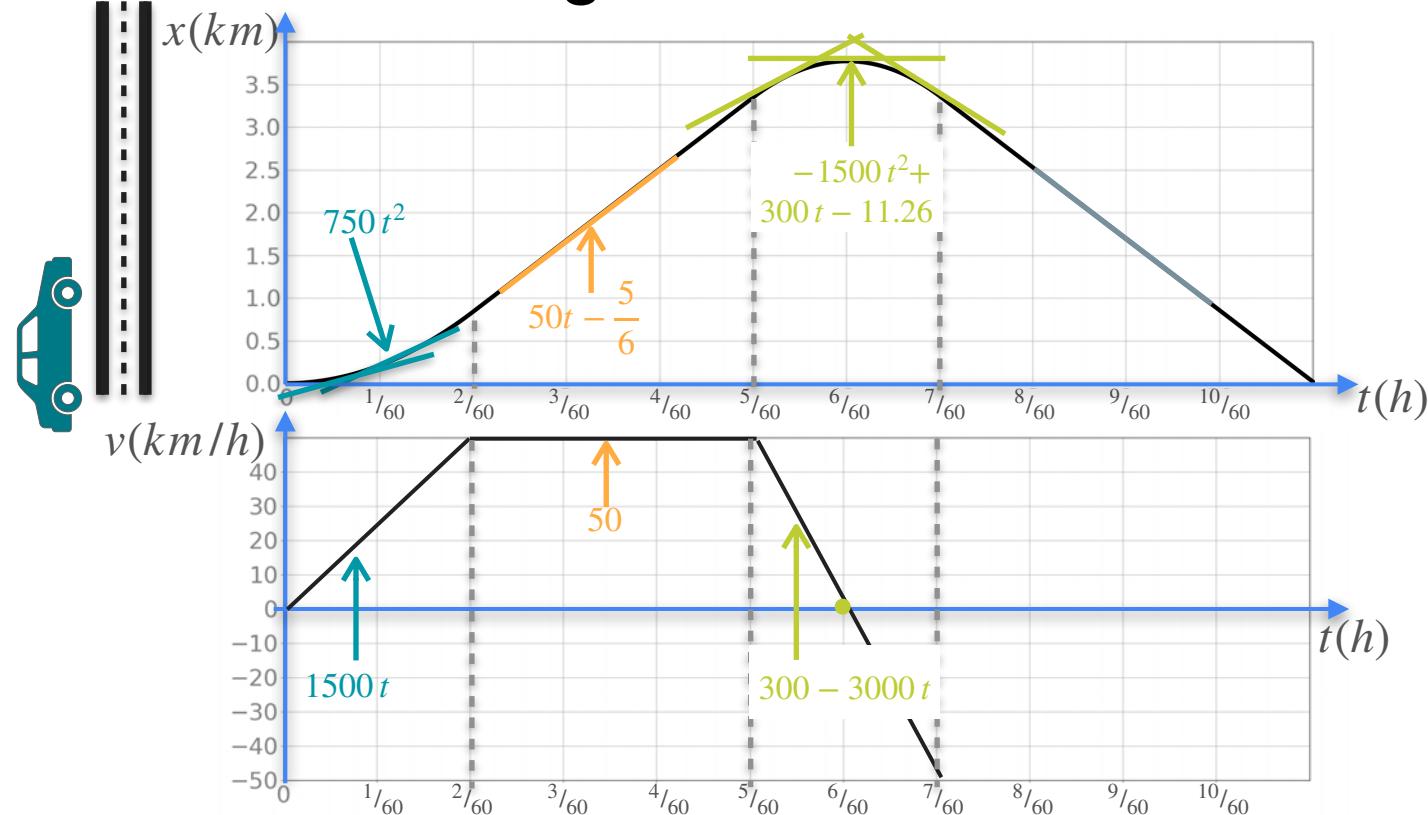


$\mathcal{X}$  Distance

$\mathcal{V}$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

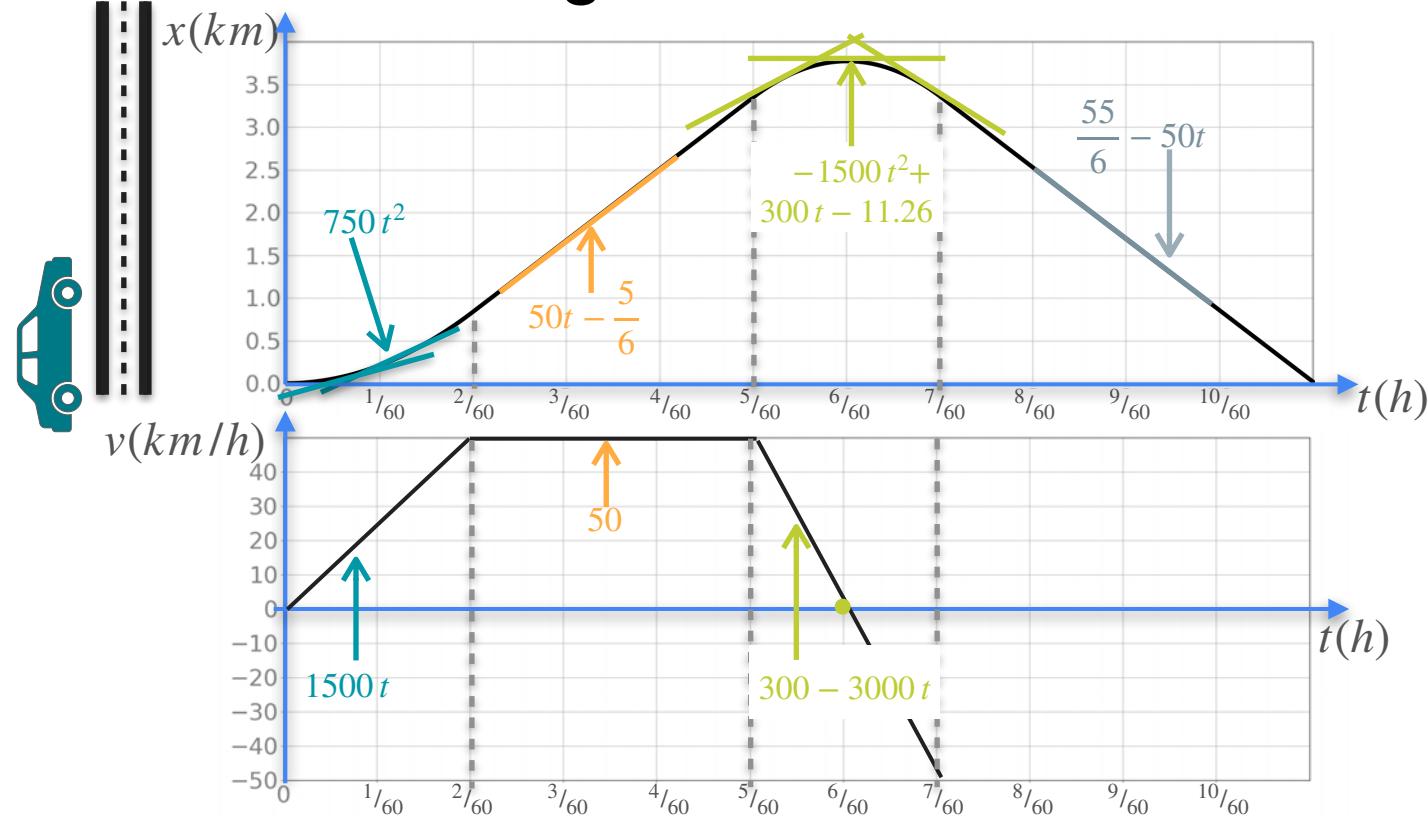


$\mathcal{X}$  Distance

$\mathcal{V}$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

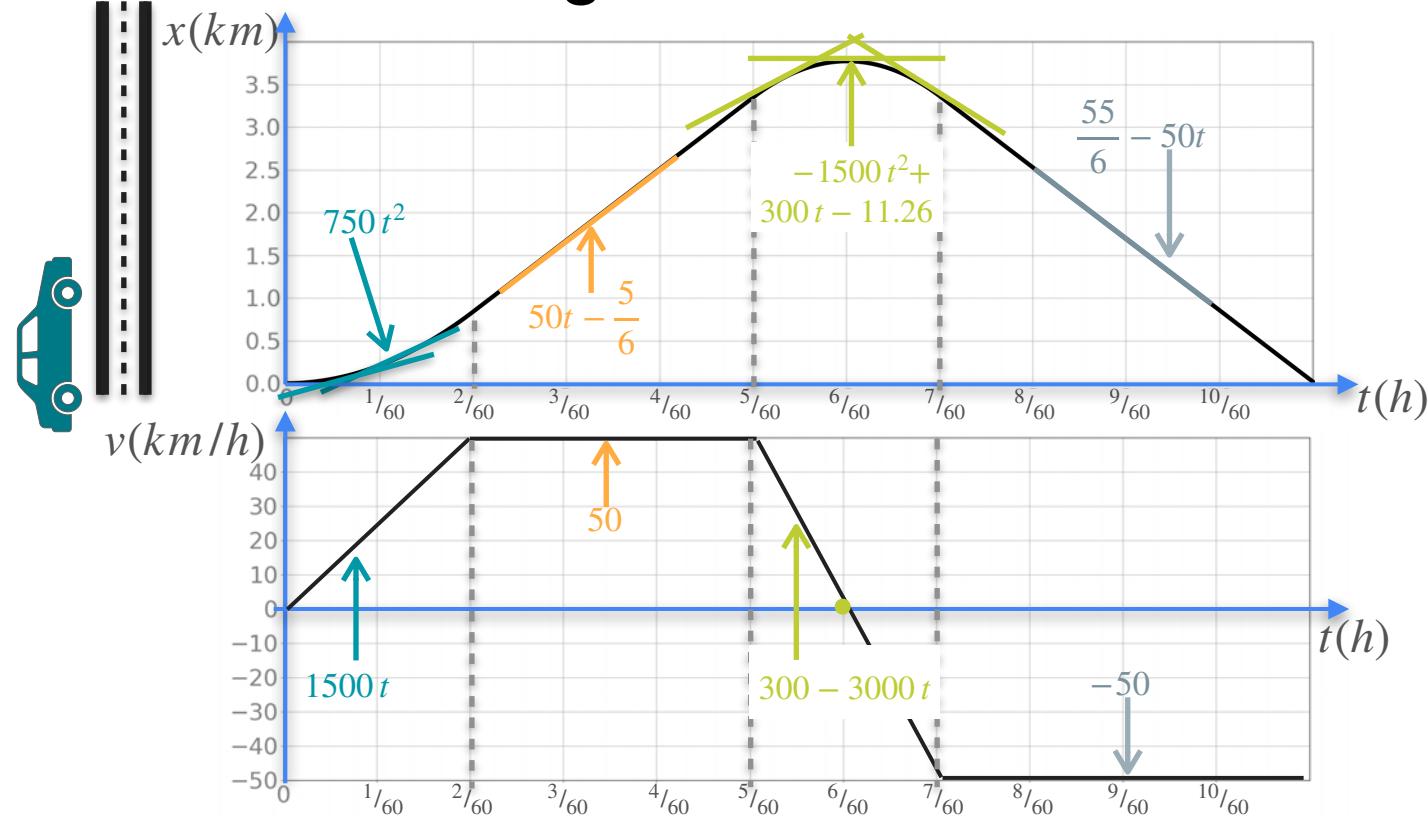


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative

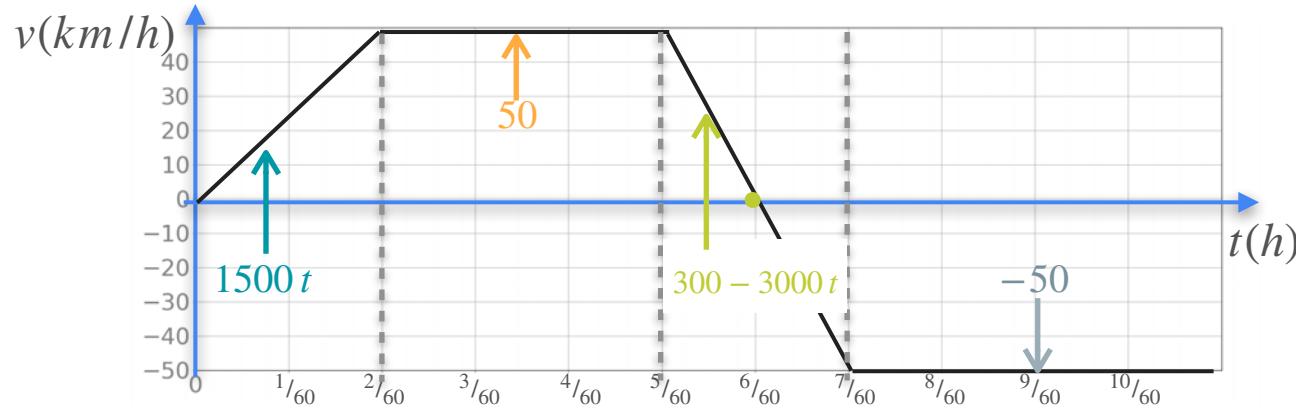


$x$  Distance

$v$  Velocity

$$\frac{dx}{dt}$$

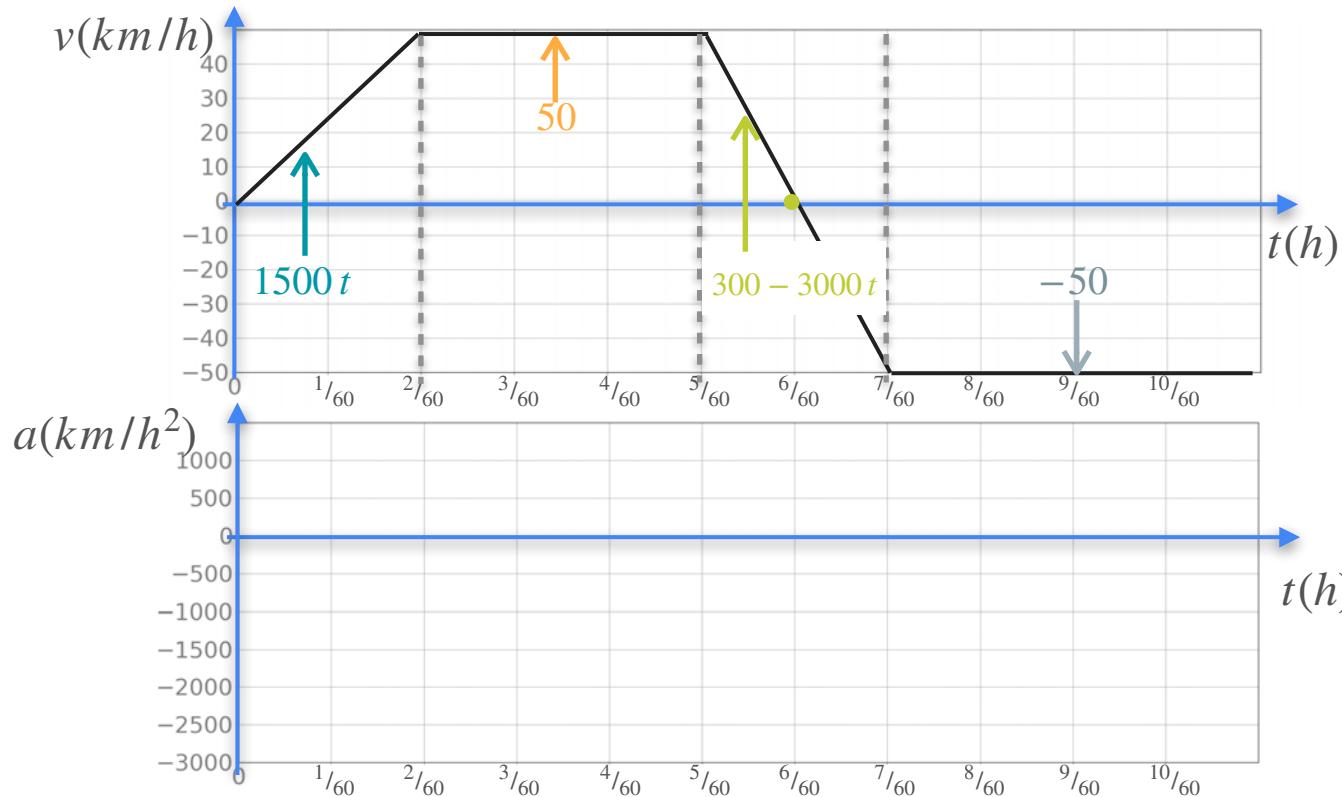
# Understanding Second Derivative



$v$  Velocity

$$\frac{dx}{dt}$$

# Understanding Second Derivative



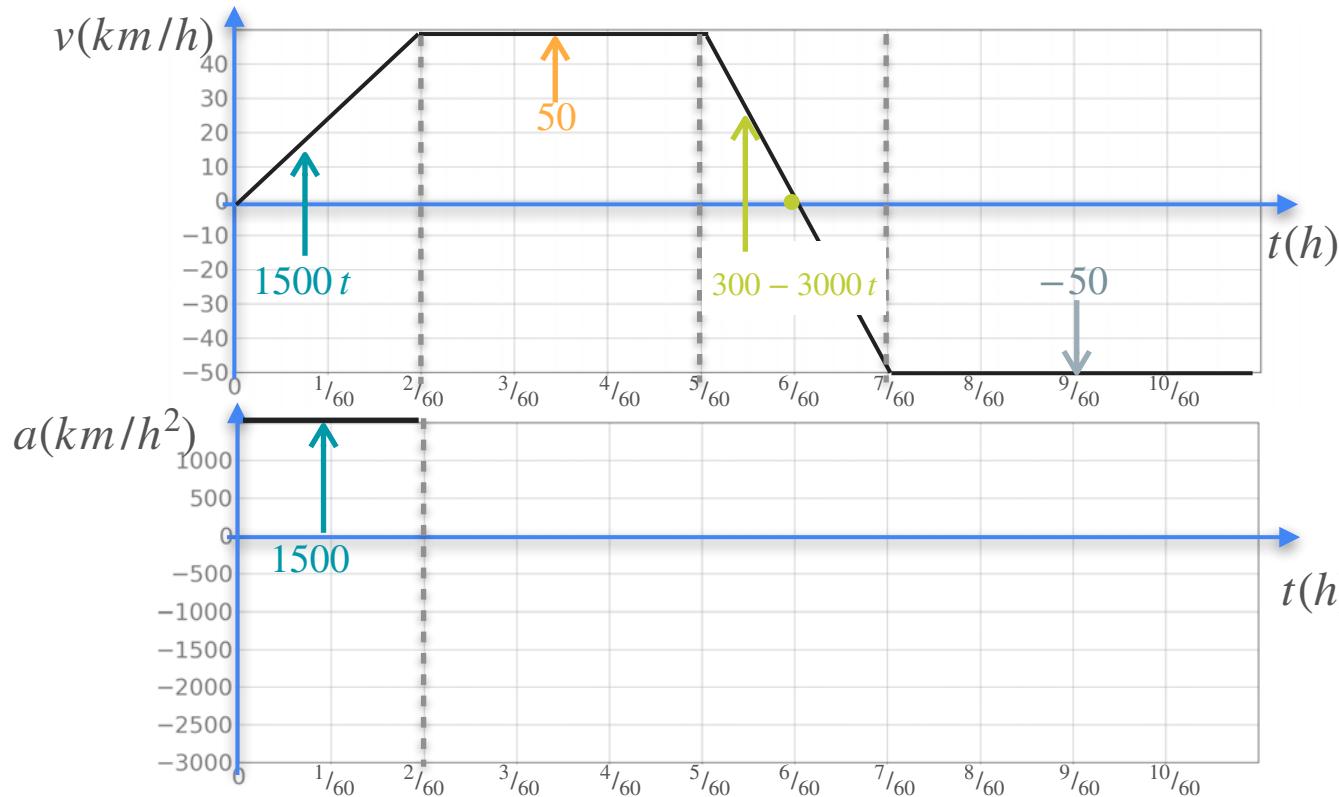
$v$  Velocity

$$\frac{dx}{dt}$$

$a$  Acceleration

$$\frac{dv}{dt}$$

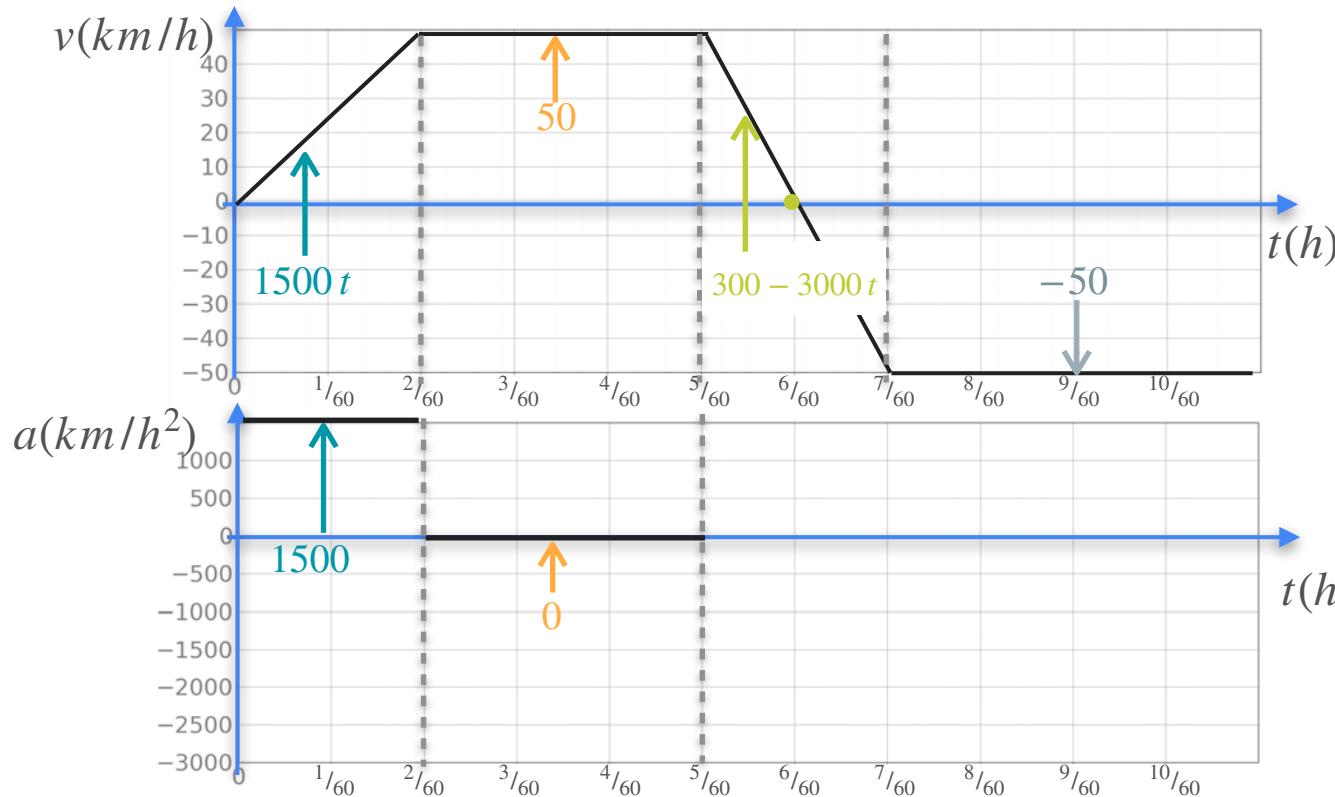
# Understanding Second Derivative



$v$  Velocity  $\frac{dx}{dt}$

$a$  Acceleration  $\frac{dv}{dt}$

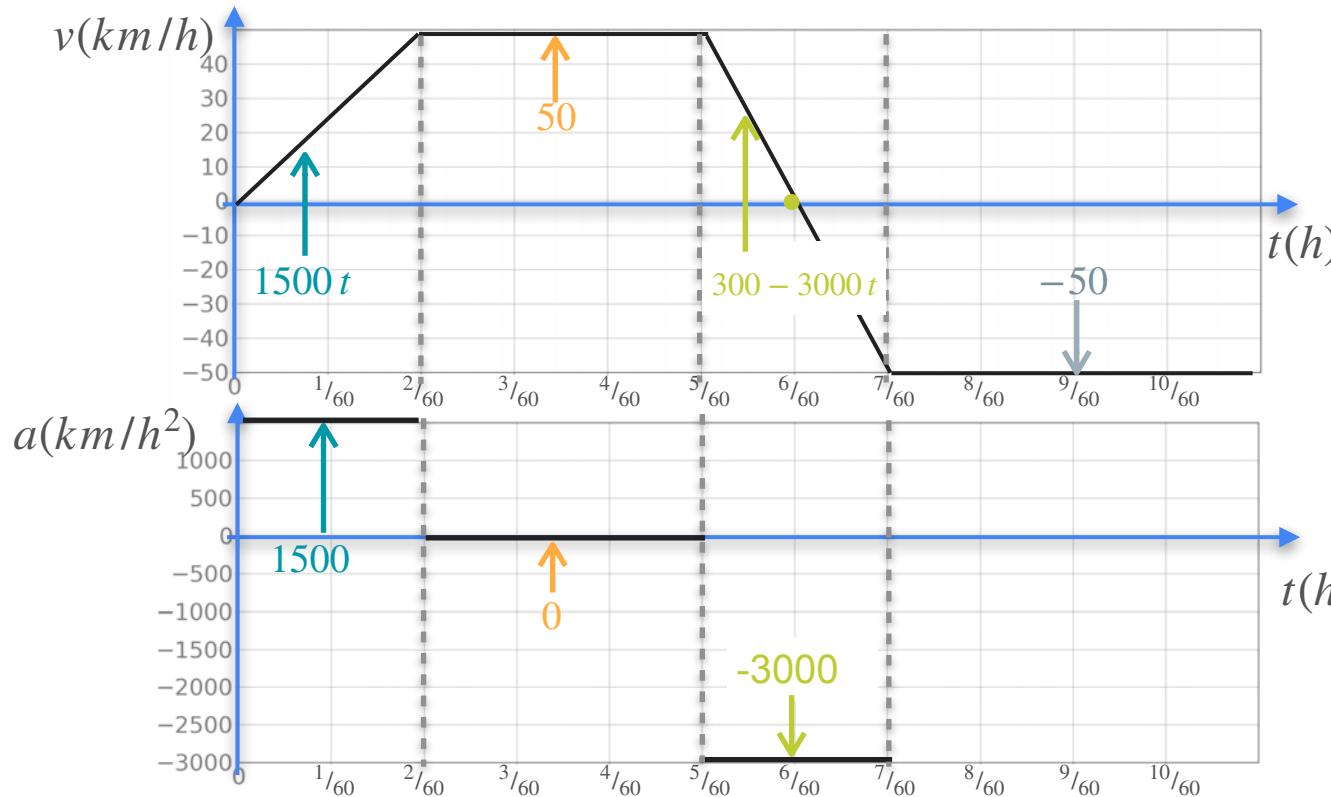
# Understanding Second Derivative



$v$  Velocity  $\frac{dx}{dt}$

$a$  Acceleration  $\frac{dv}{dt}$

# Understanding Second Derivative



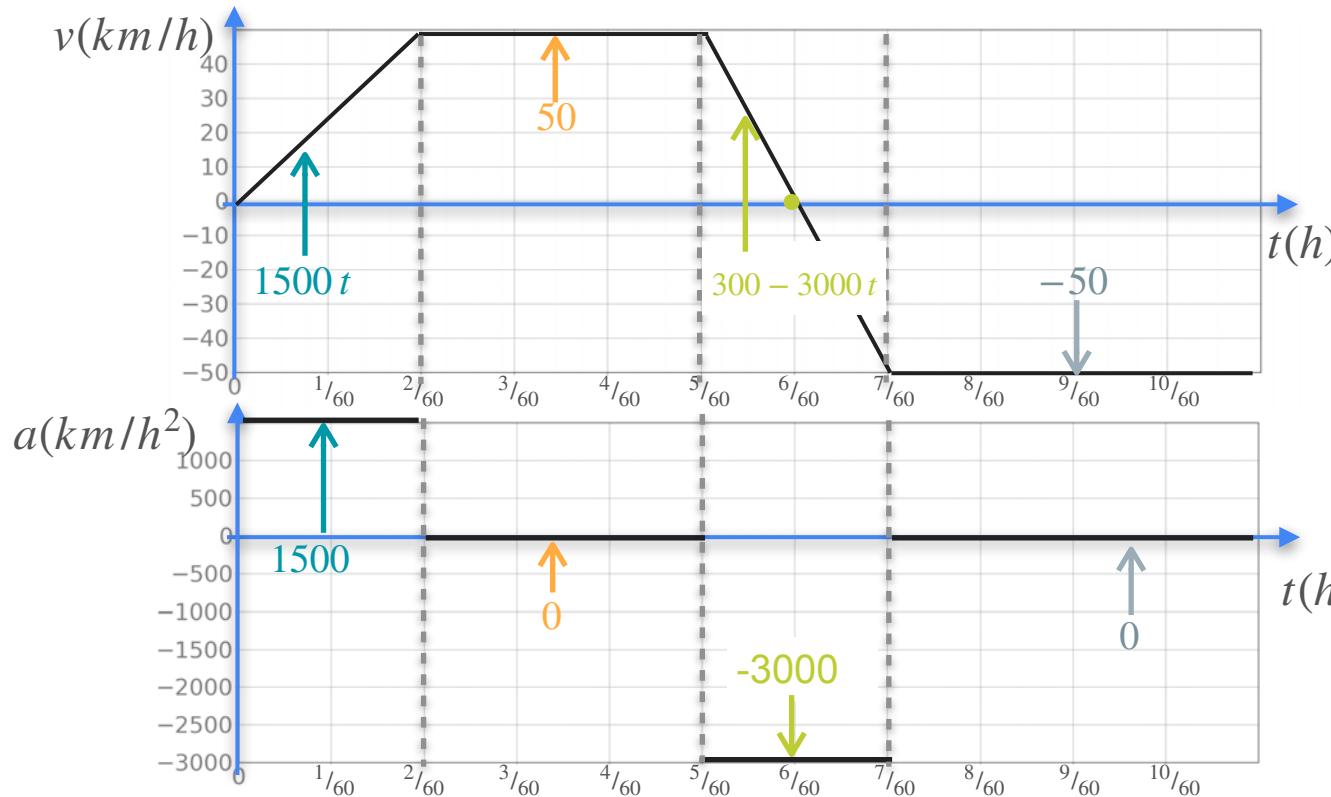
$v$  Velocity

$$\frac{dx}{dt}$$

$a$  Acceleration

$$\frac{dv}{dt}$$

# Understanding Second Derivative



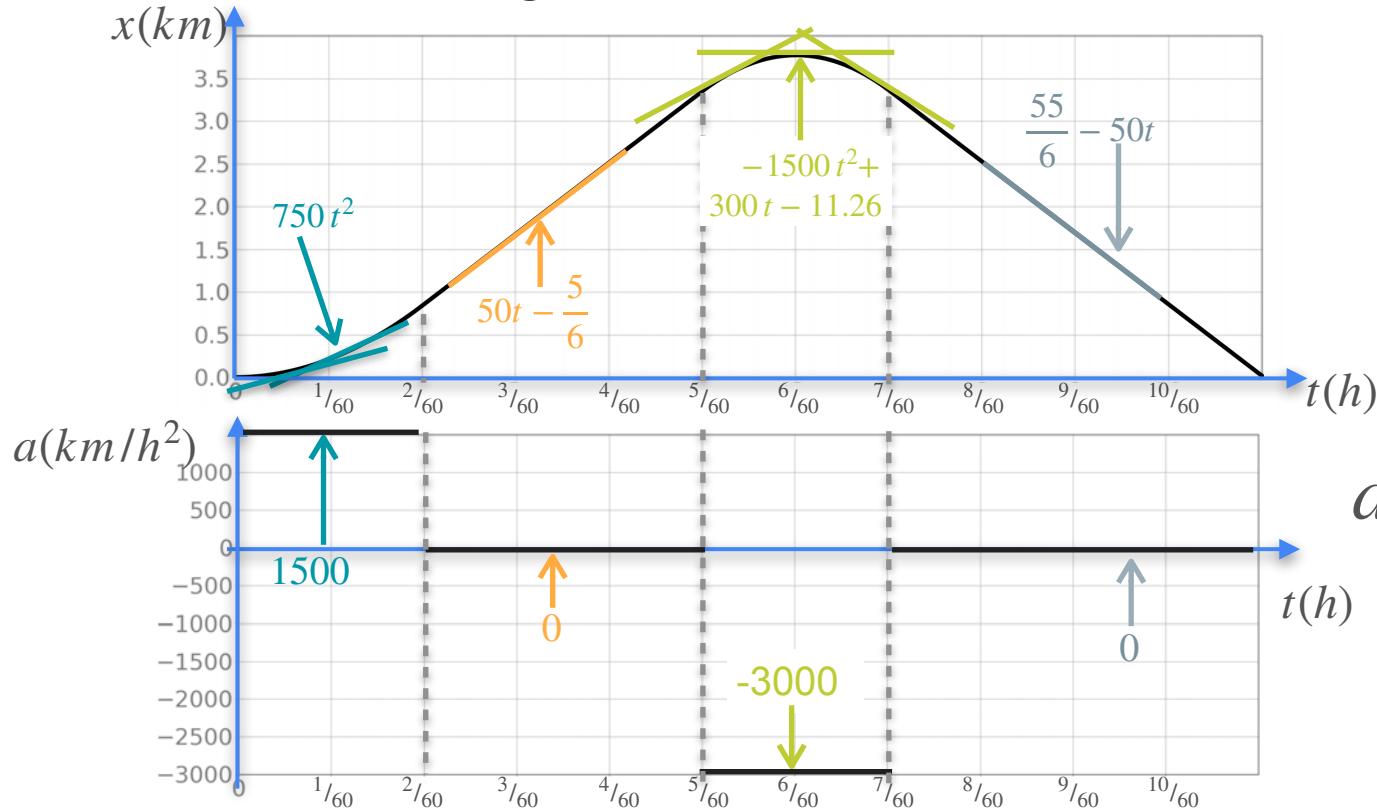
$v$  Velocity

$$\frac{dx}{dt}$$

$a$  Acceleration

$$\frac{dv}{dt}$$

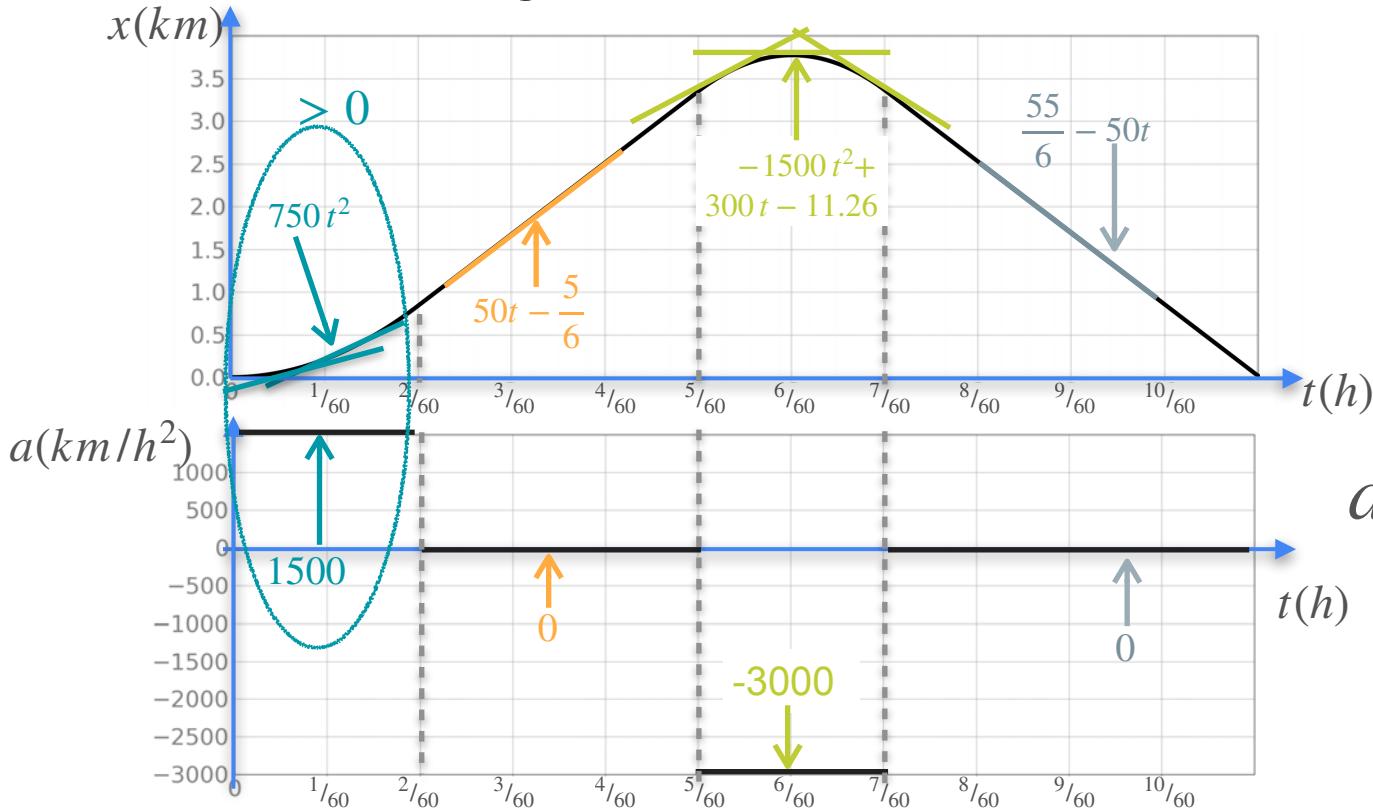
# Understanding Second Derivative



$x$  Distance

$a$  Acceleration  $\frac{d^2x}{dt^2}$

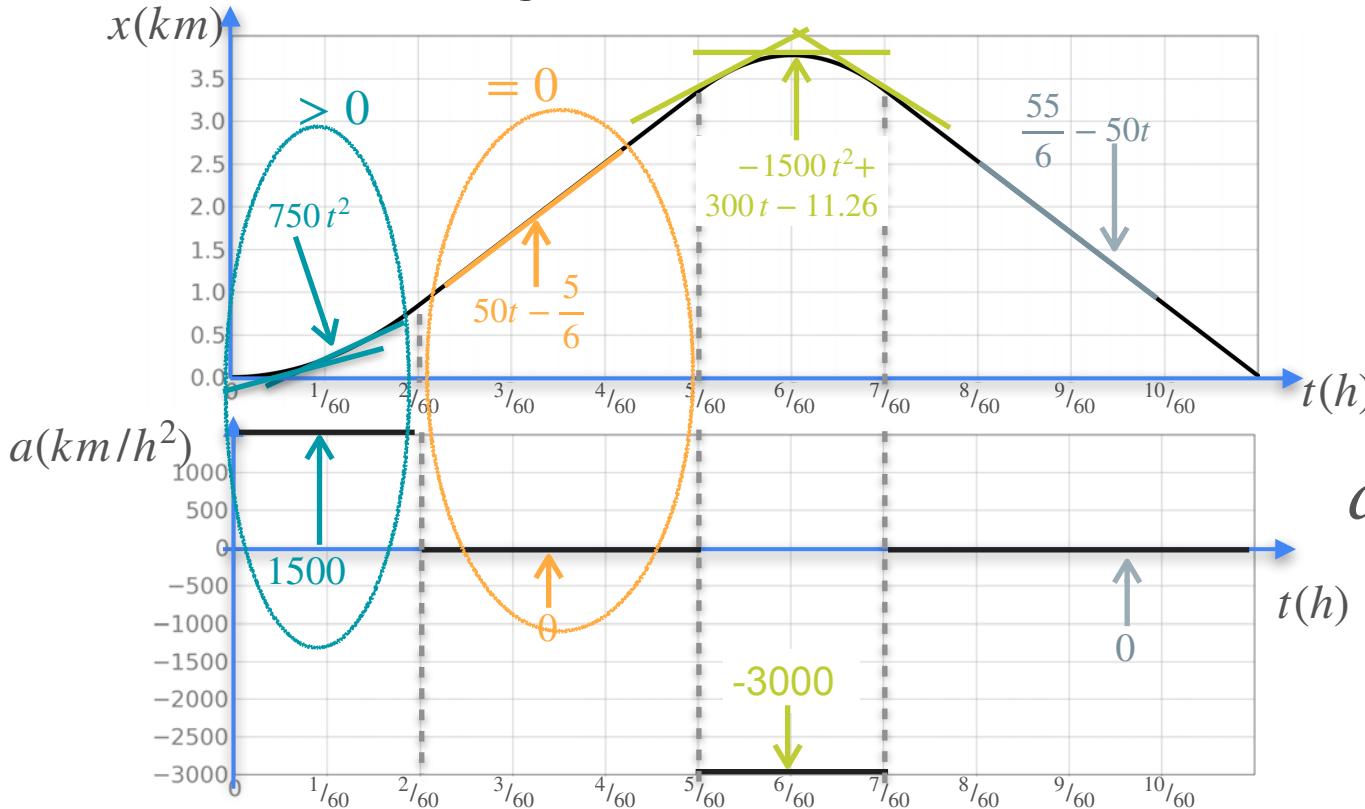
# Understanding Second Derivative



$x$  Distance

$a$  Acceleration  $\frac{d^2x}{dt^2}$

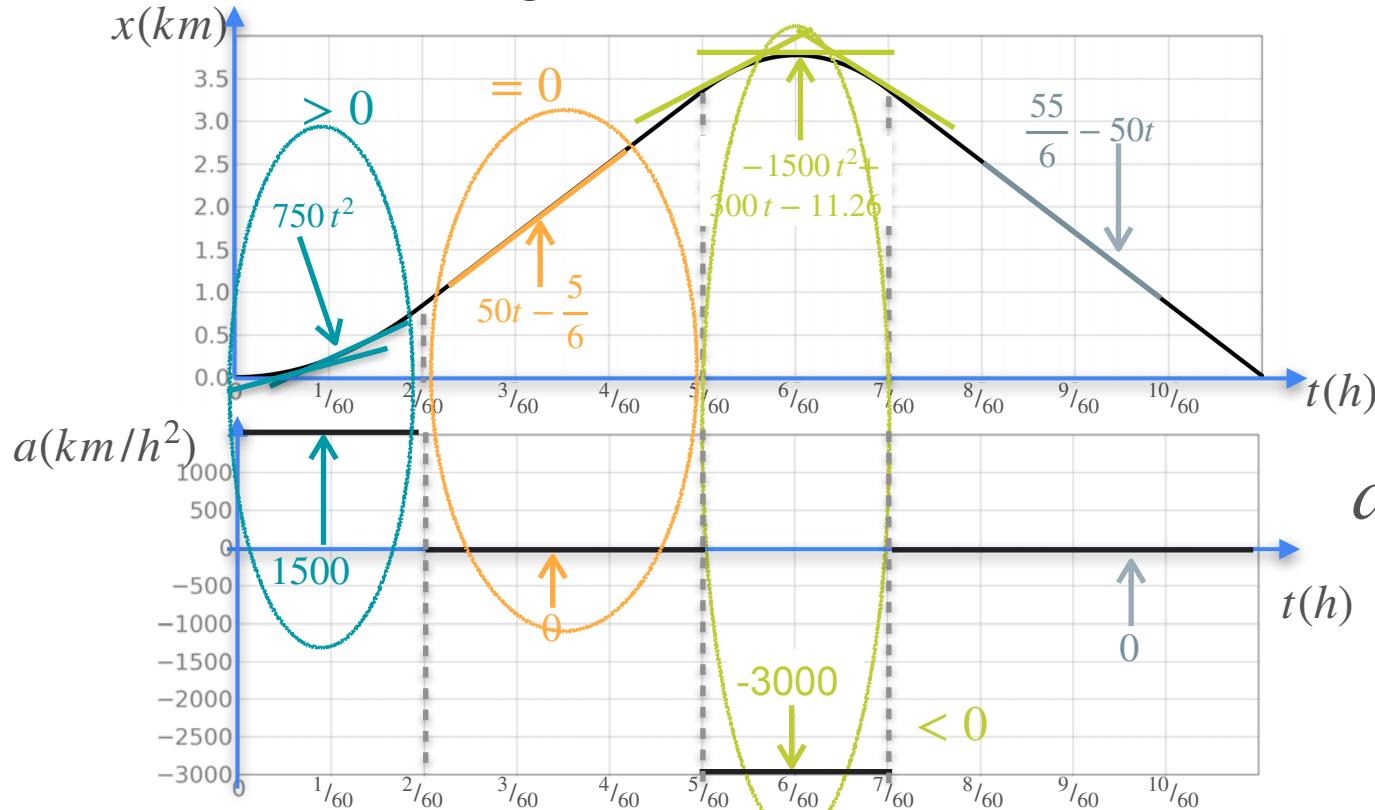
# Understanding Second Derivative



$x$  Distance

$a$  Acceleration  $\frac{d^2x}{dt^2}$

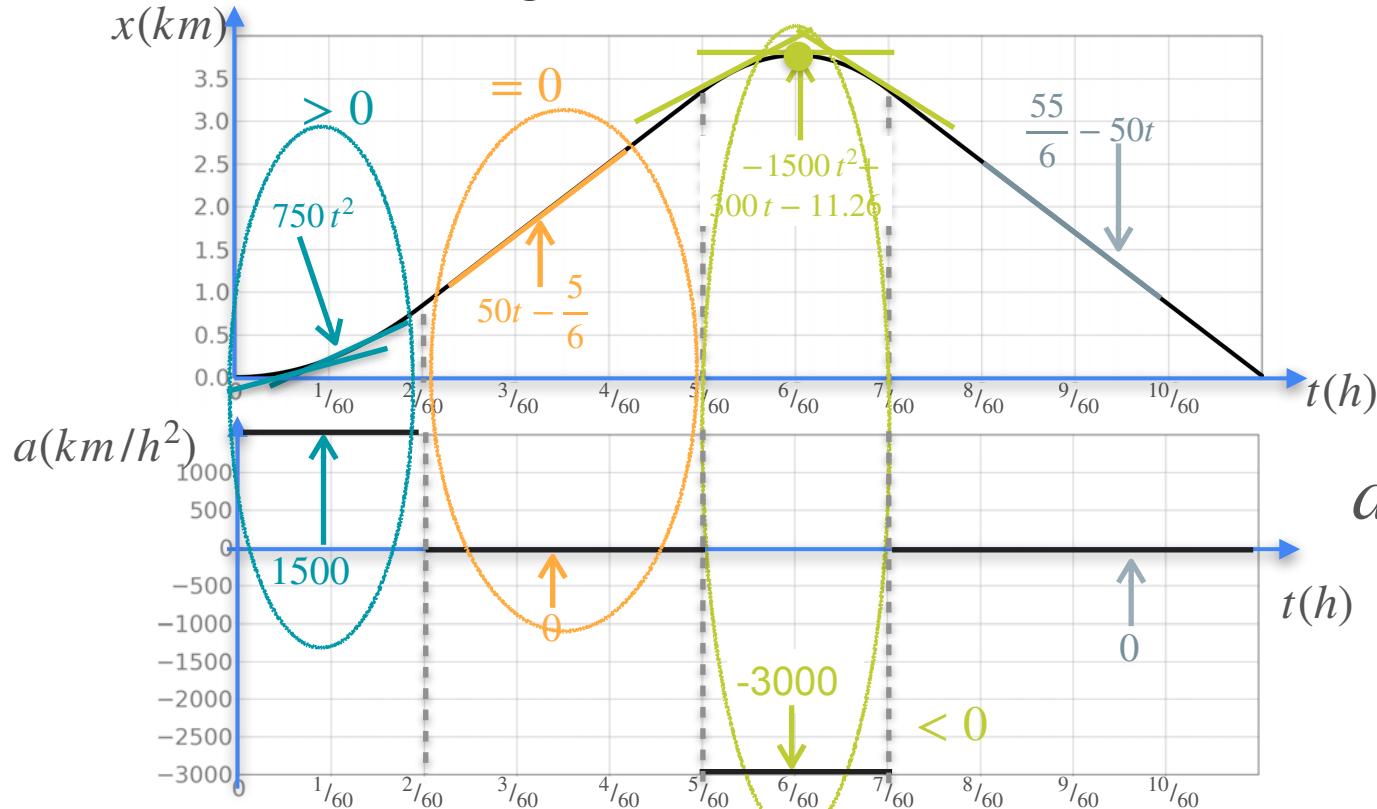
# Understanding Second Derivative



$x$  Distance

$a$  Acceleration  $\frac{d^2x}{dt^2}$

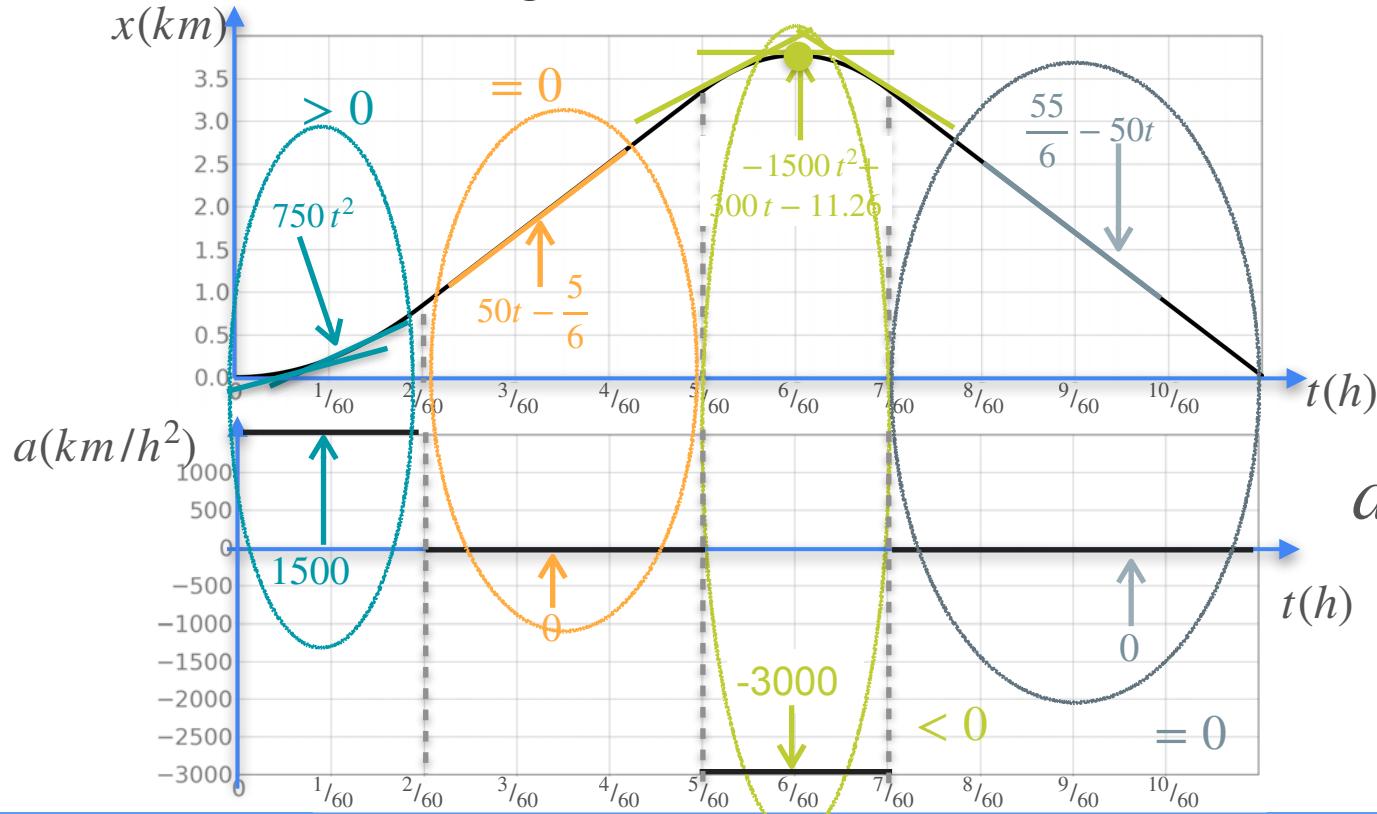
# Understanding Second Derivative



$x$  Distance

$a$  Acceleration  $\frac{d^2x}{dt^2}$

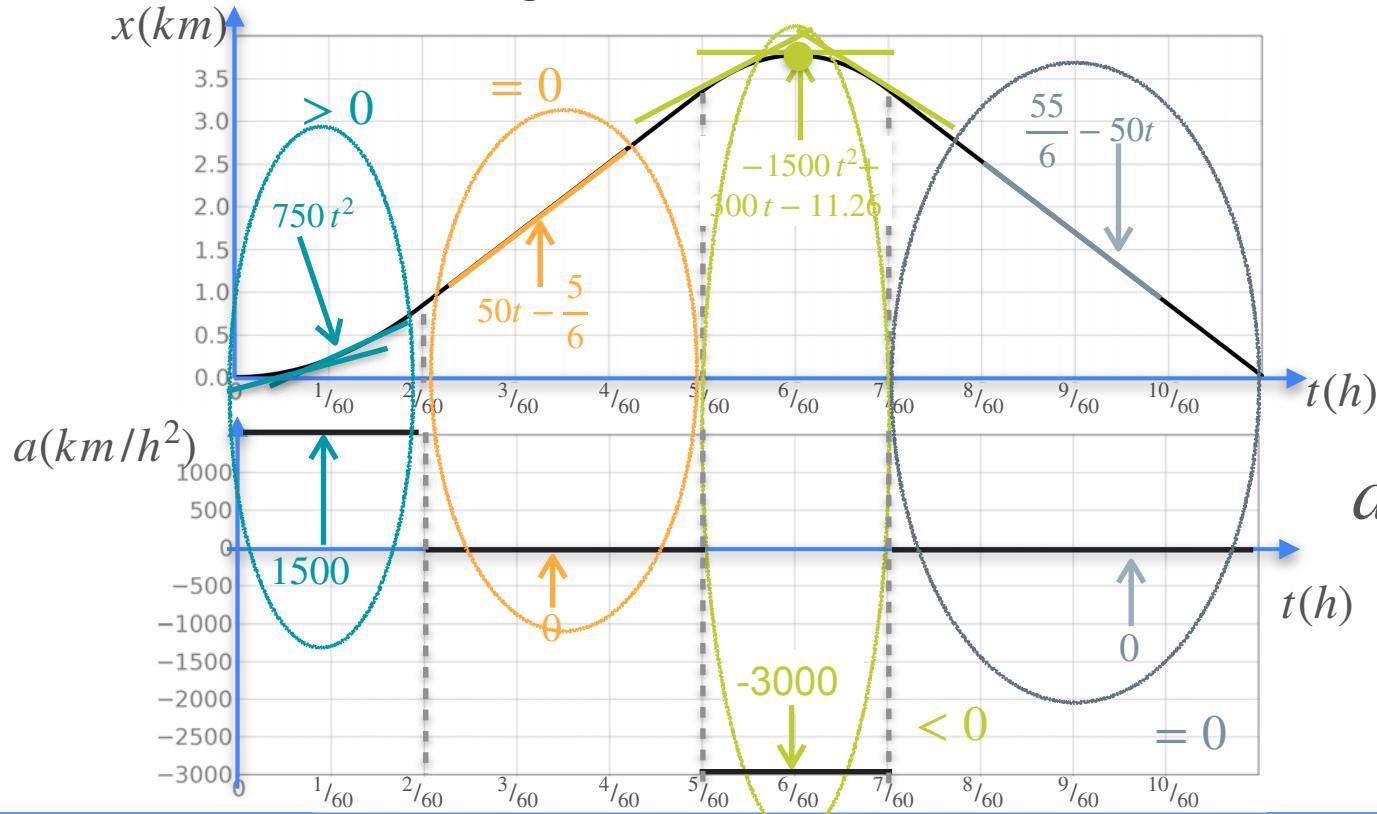
# Understanding Second Derivative



$x$  Distance

$a$  Acceleration  $\frac{d^2x}{dt^2}$

# Understanding Second Derivative



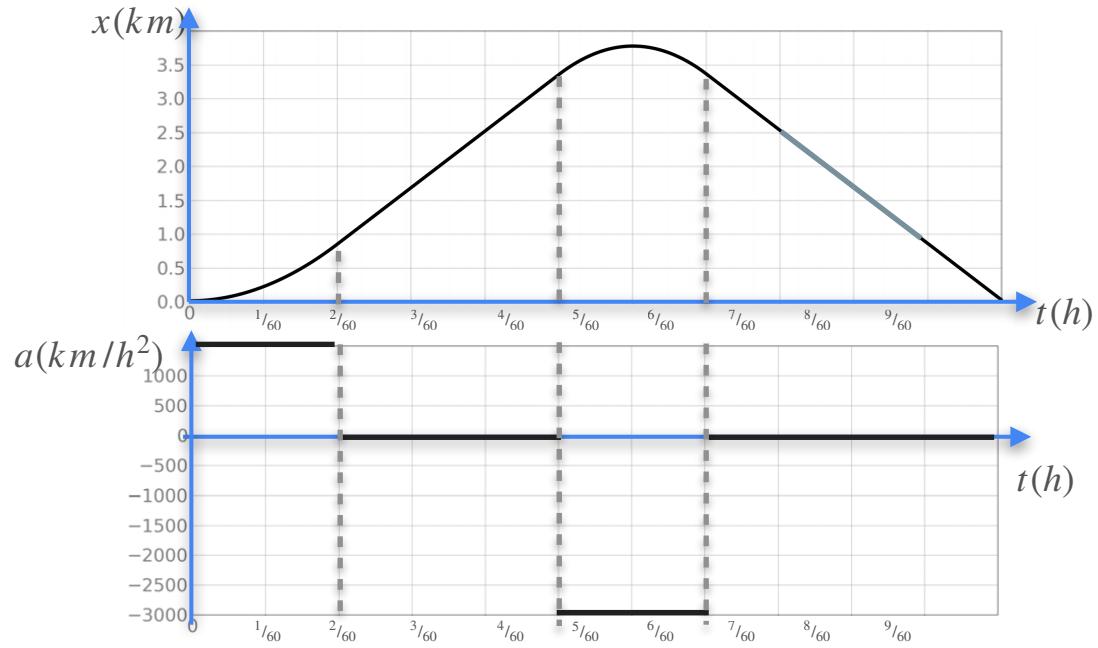
$x$  Distance

Second derivative tells us about the curvature

$a$  Acceleration  $\frac{d^2x}{dt^2}$

# Curvature

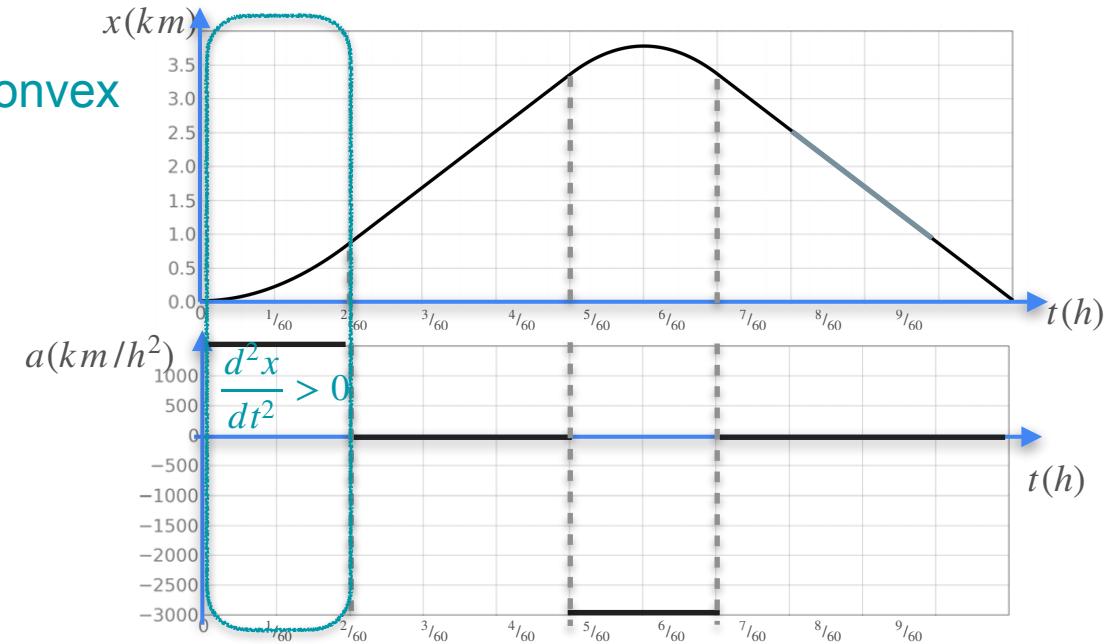
# Curvature



# Curvature

$$\frac{d^2x}{dt^2} > 0$$

Concave up or convex



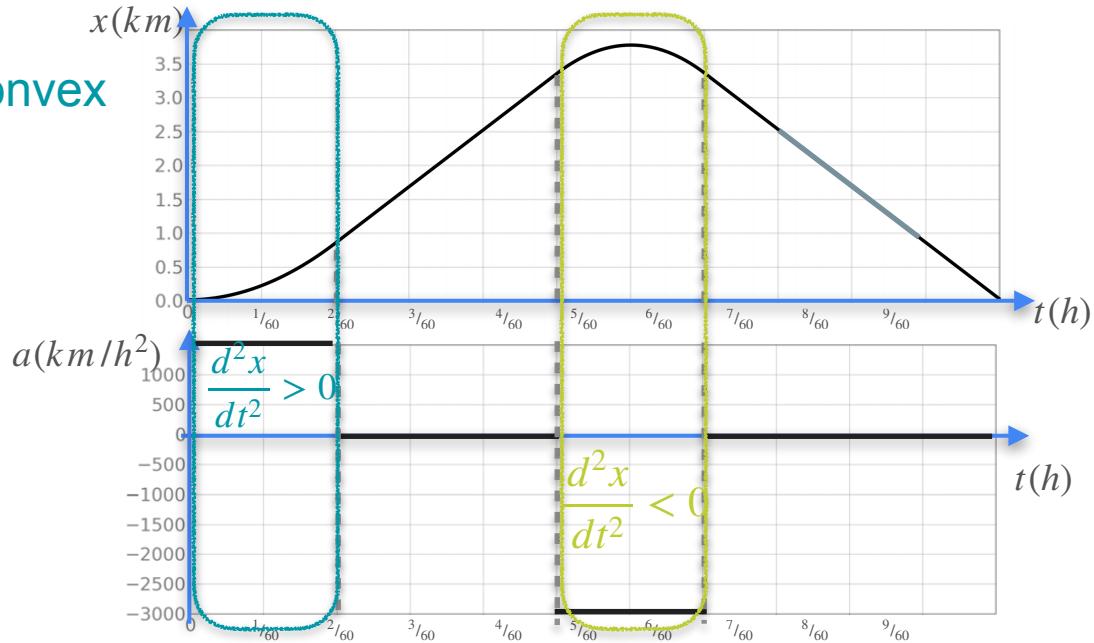
# Curvature

$$\frac{d^2x}{dt^2} > 0$$

Concave up or convex

$$\frac{d^2x}{dt^2} < 0$$

Concave down



# Curvature

$$\frac{d^2x}{dt^2} > 0$$

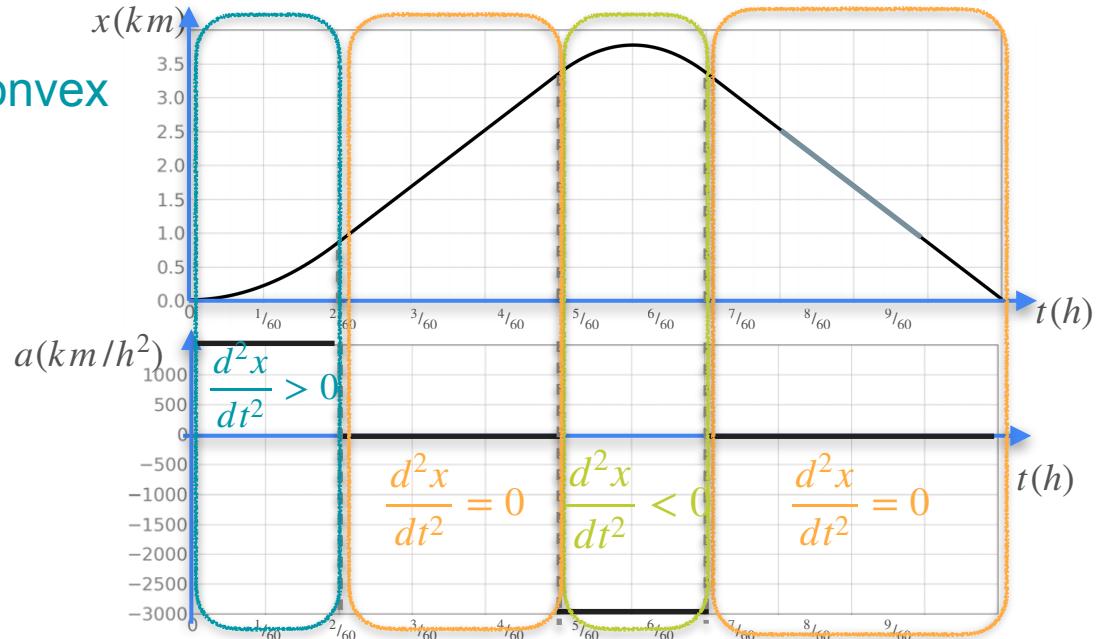
Concave up or convex

$$\frac{d^2x}{dt^2} < 0$$

Concave down

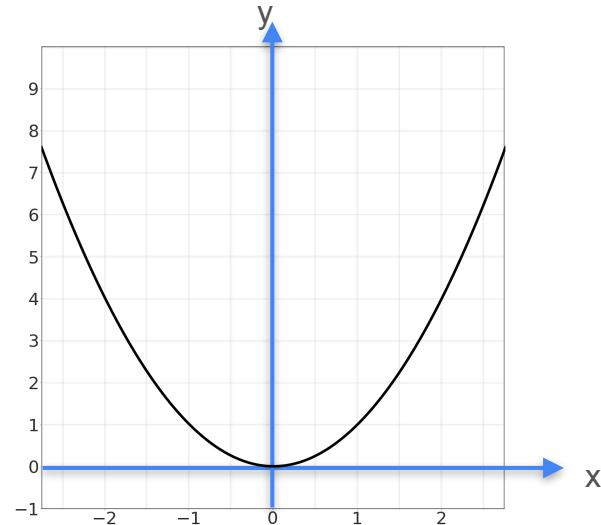
$$\frac{d^2x}{dt^2} = 0$$

Need more information

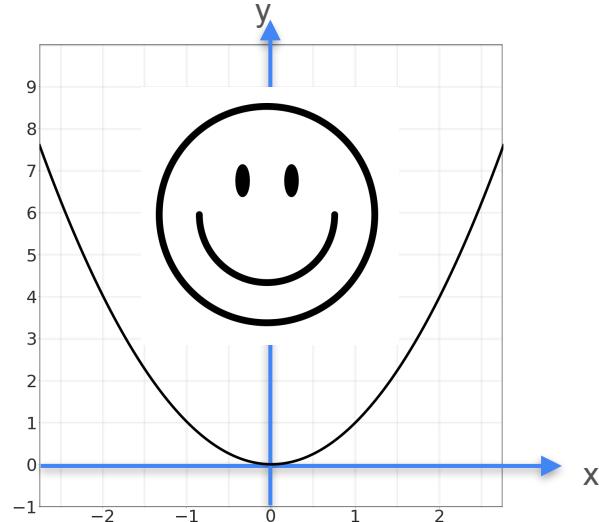


# Curvature

# Curvature



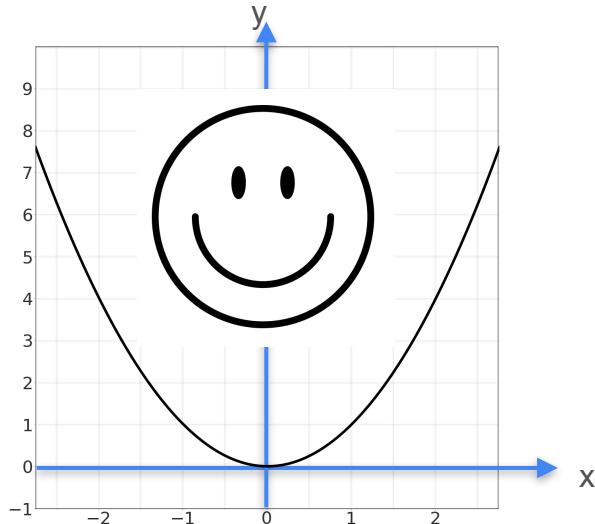
# Curvature



Concave up or convex

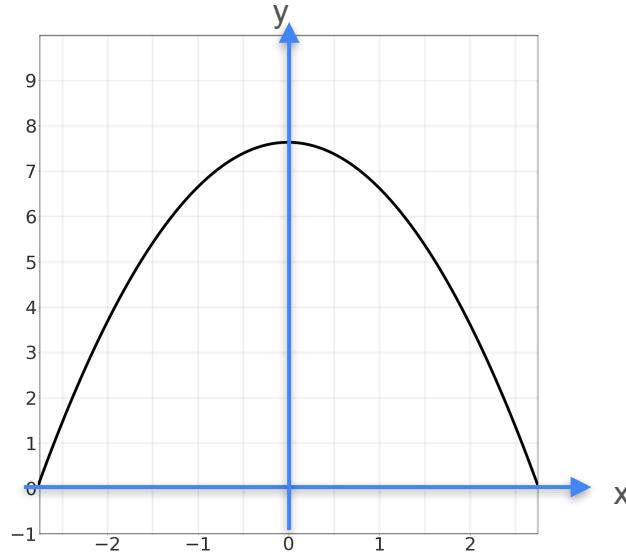
$$f''(0) > 0$$

# Curvature

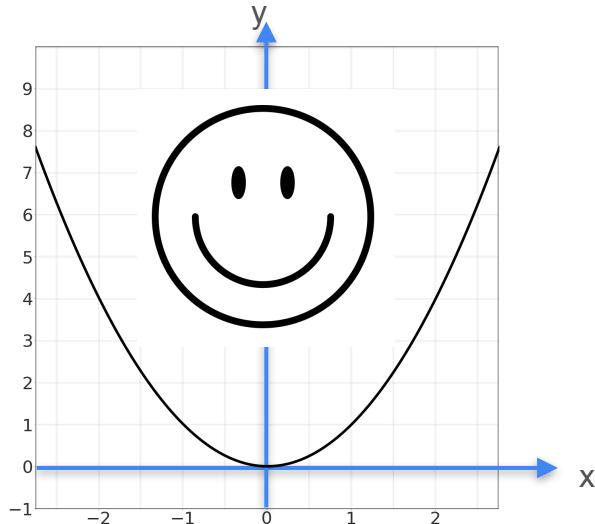


Concave up or convex

$$f''(0) > 0$$

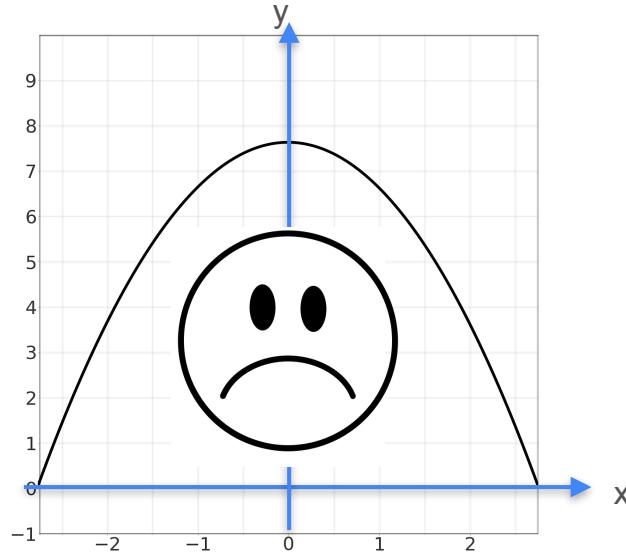


# Curvature



Concave up or convex

$$f''(0) > 0$$



Concave down

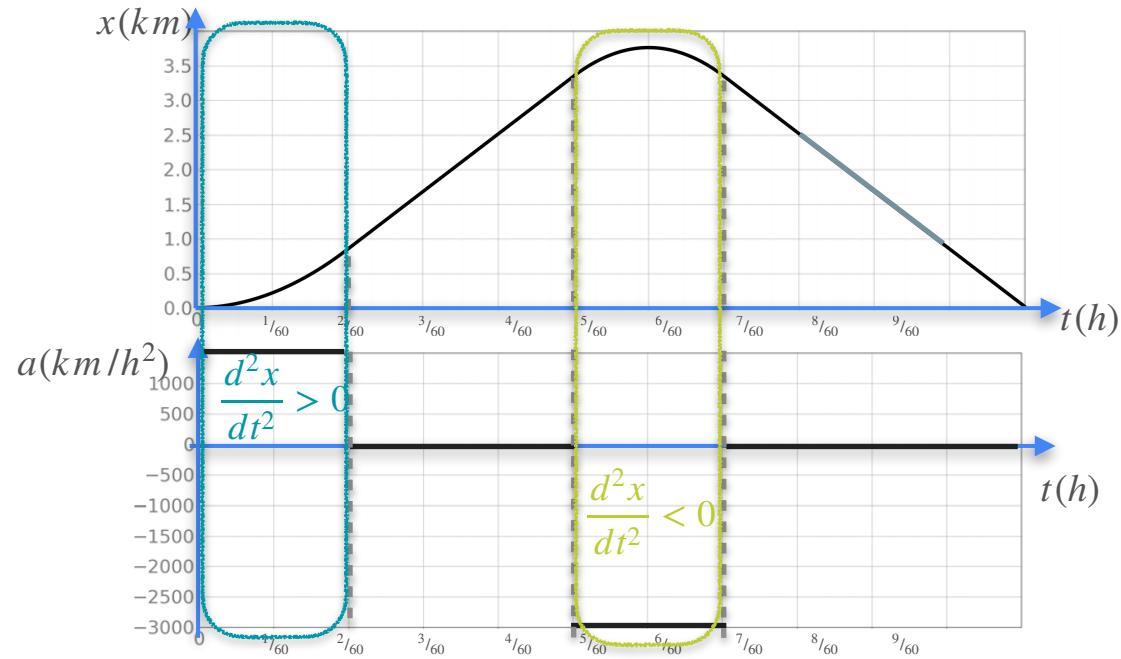
$$f''(0) < 0$$

# Second Derivative and Optimization

$$\frac{d^2x}{dt^2} > 0$$

$$\frac{d^2x}{dt^2} < 0$$

$$\frac{d^2x}{dt^2} = 0$$

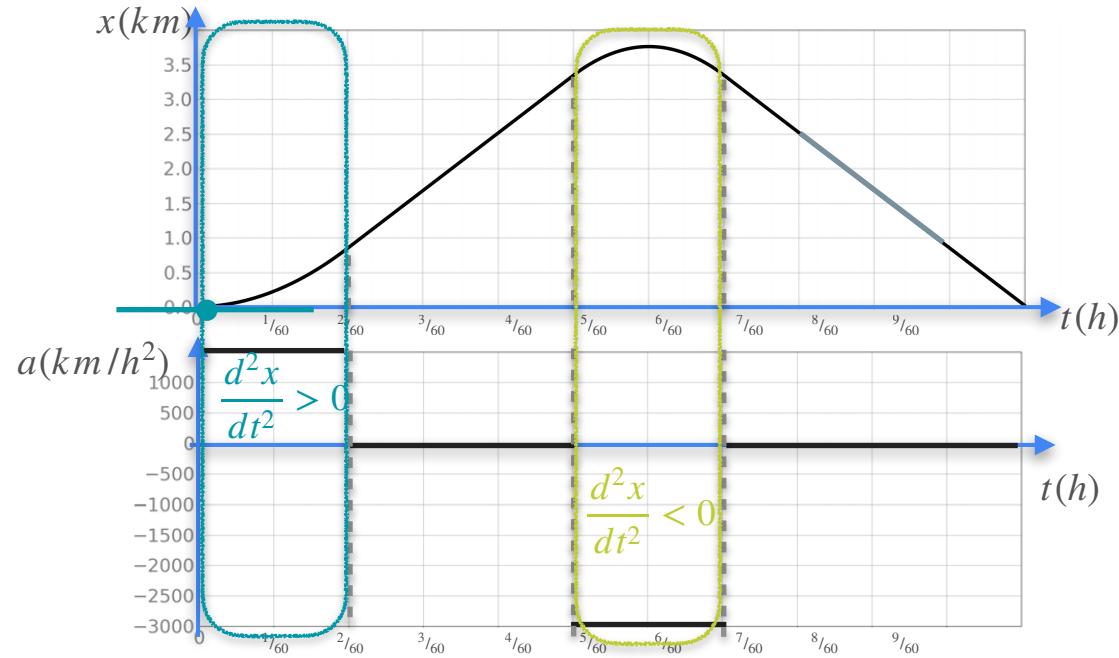


# Second Derivative and Optimization

$\frac{d^2x}{dt^2} > 0$     (Local) Minimum

$\frac{d^2x}{dt^2} < 0$

$\frac{d^2x}{dt^2} = 0$

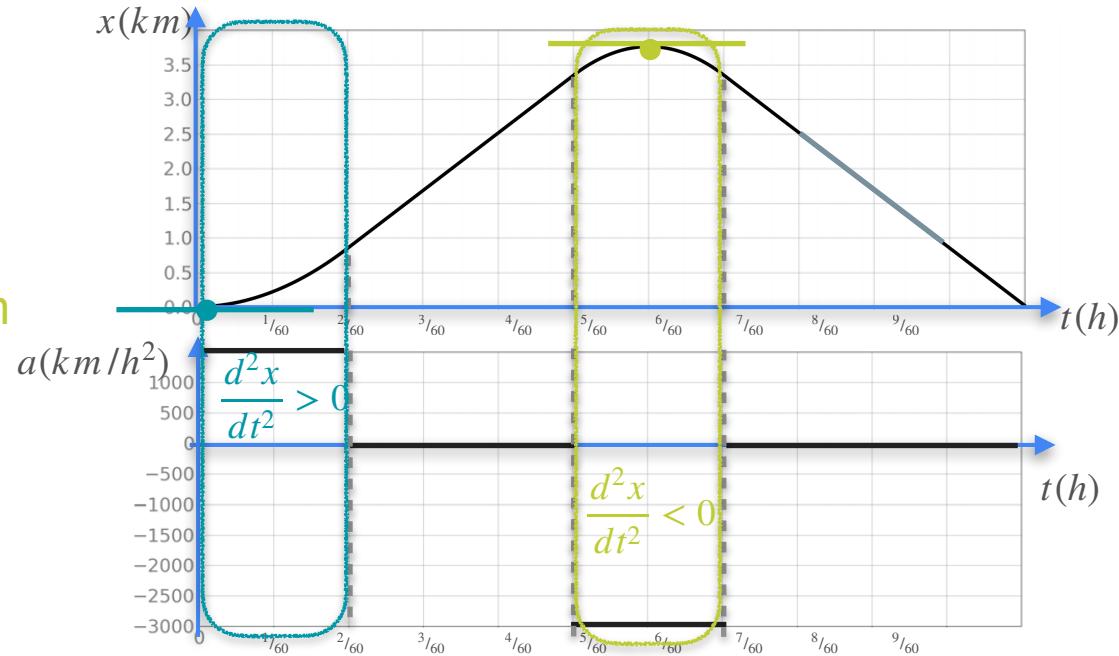


# Second Derivative and Optimization

$$\frac{d^2x}{dt^2} > 0 \quad (\text{Local}) \text{ Minimum}$$

$$\frac{d^2x}{dt^2} < 0 \quad (\text{Local}) \text{ maximum}$$

$$\frac{d^2x}{dt^2} = 0$$

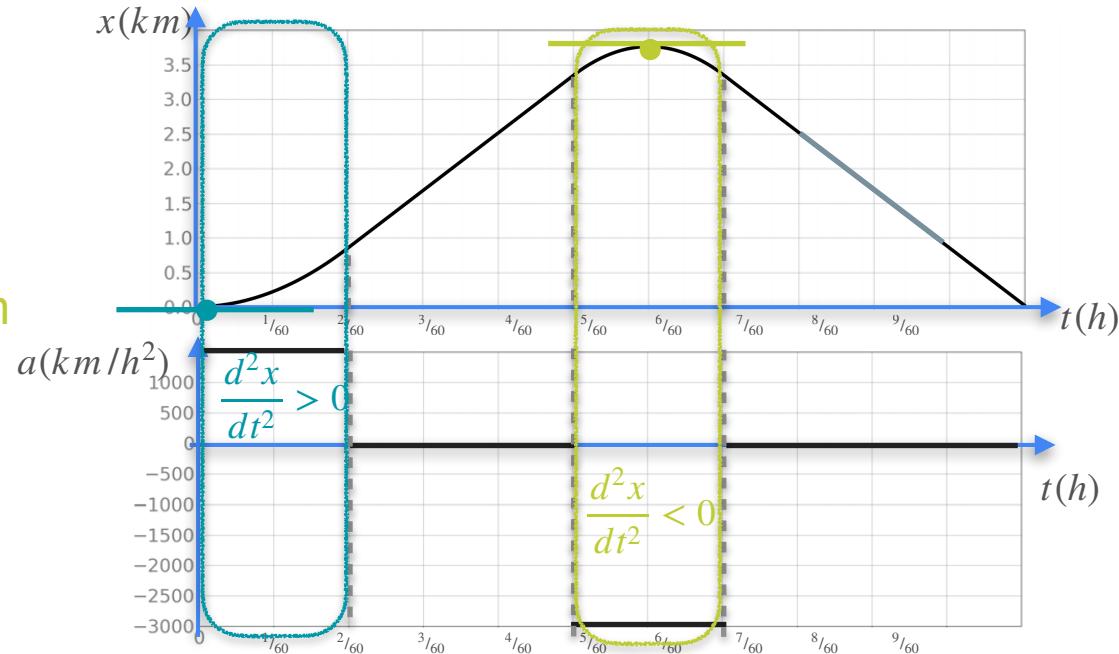


# Second Derivative and Optimization

$\frac{d^2x}{dt^2} > 0$     (Local) Minimum

$\frac{d^2x}{dt^2} < 0$     (Local) maximum

$\frac{d^2x}{dt^2} = 0$     Inconclusive



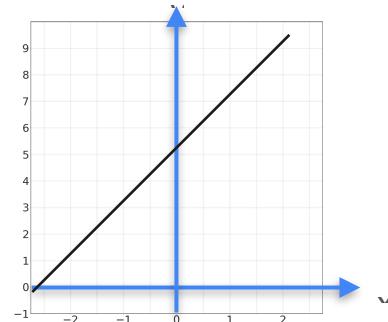
# Curvature

First derivative

Second derivative

# Curvature

First derivative



Increasing

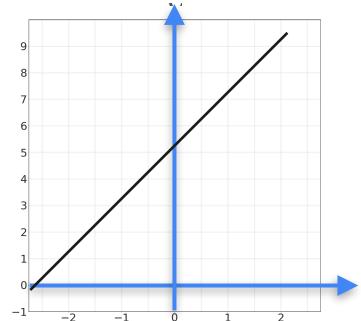
$$f'(0) > 0$$

Second derivative



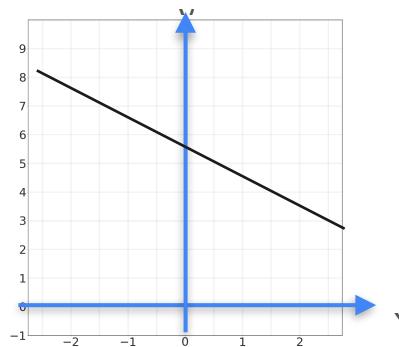
# Curvature

First derivative



Increasing

$$f'(0) > 0$$



Decreasing

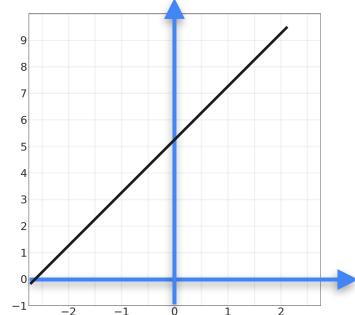
$$f'(0) < 0$$

Second derivative



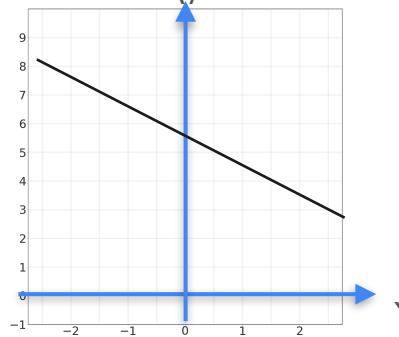
# Curvature

First derivative



Increasing

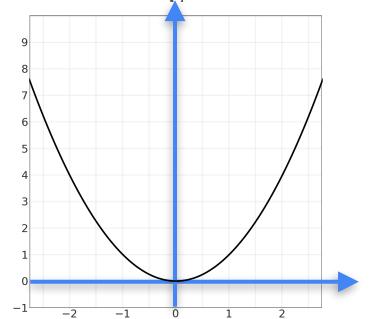
$$f'(0) > 0$$



Decreasing

$$f'(0) < 0$$

Second derivative

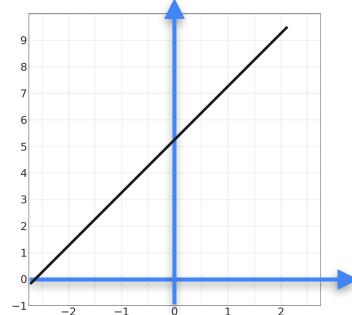


Concave up

$$f''(0) > 0$$

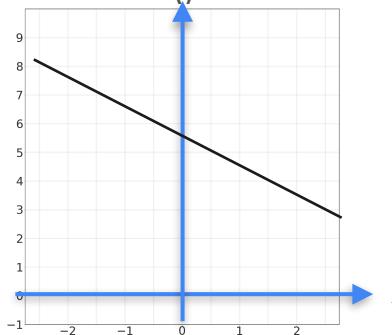
# Curvature

First derivative



Increasing

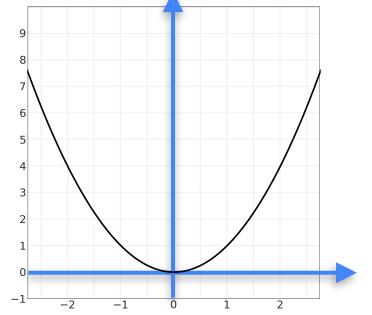
$$f'(0) > 0$$



Decreasing

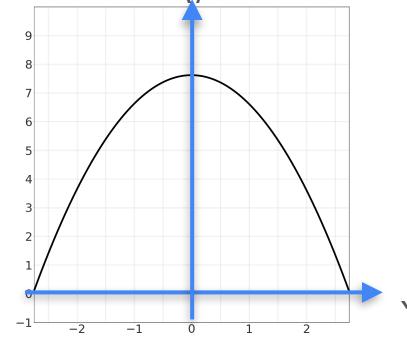
$$f'(0) < 0$$

Second derivative



Concave up

$$f''(0) > 0$$



Concave down

$$f''(0) < 0$$



DeepLearning.AI

# Optimization in Neural Networks and Newton's Method

---

## The Hessian

# Second Derivative

# Second Derivative

1 variable

2 variables

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t $x$

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t $x$ $f_y(x, y)$ Rate of change w.r.t $y$

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ $f_y(x, y)$ $\nabla f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$

# Second Derivative

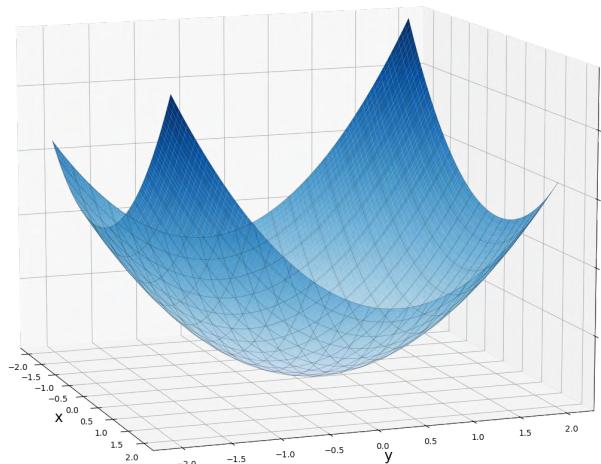
	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ $f_y(x, y)$ $\nabla f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$
Second derivative	$f''(x)$ Rate of change of the rate of change of $f(x)$	

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t $x$ $f_y(x, y)$ Rate of change w.r.t $y$ $\nabla f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$
Second derivative	$f''(x)$ Rate of change of the rate of change of $f(x)$	???

# Second Derivative

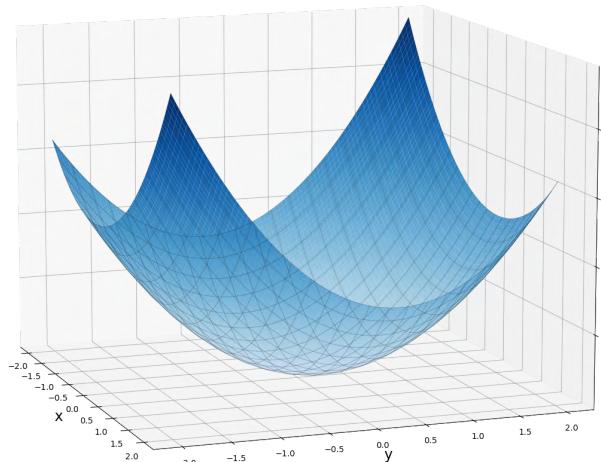
# Second Derivative



$$f(x, y) =$$

$$2x^2 + 3y^2 - xy$$

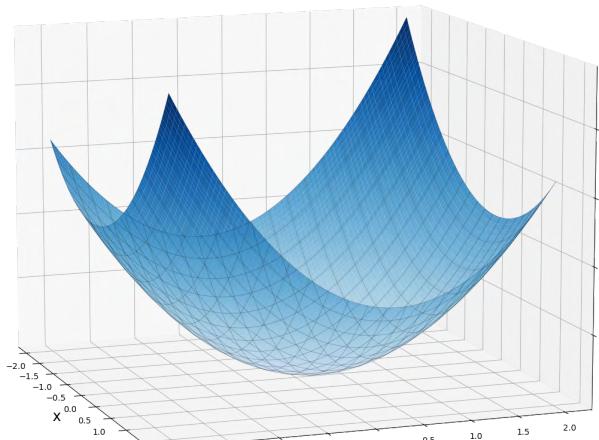
# Second Derivative



$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$\begin{matrix} 4x - y \\ x \end{matrix}$$

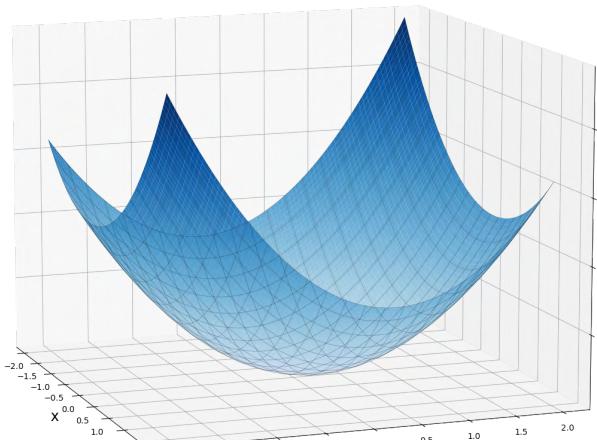
# Second Derivative



$$f(x, y) = 2x^2 + 3y^2 - xy$$

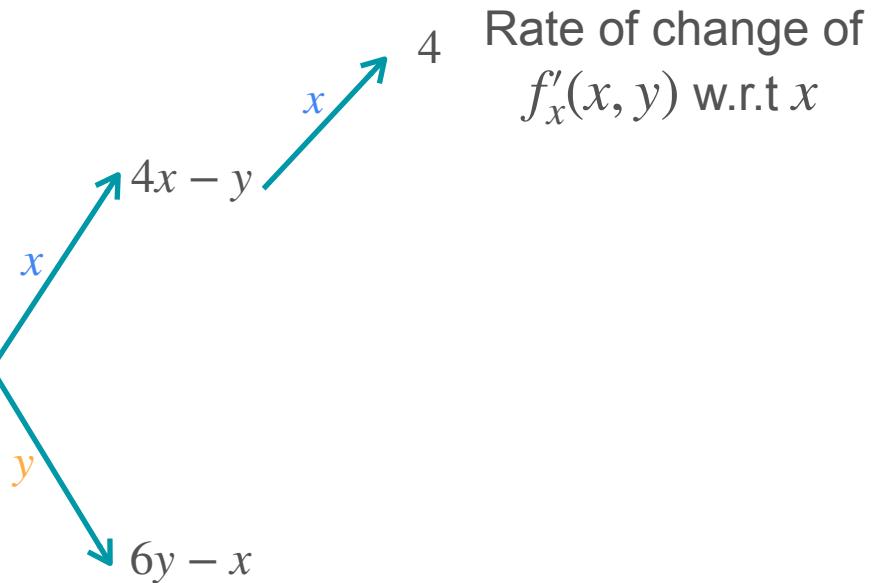
A diagram illustrating the second derivatives of the function  $f(x, y) = 2x^2 + 3y^2 - xy$ . A central point is labeled  $x$  above a horizontal axis and  $y$  to its right. Two arrows originate from this point: one pointing up and to the left labeled  $4x - y$ , and another pointing down and to the right labeled  $6y - x$ .

# Second Derivative



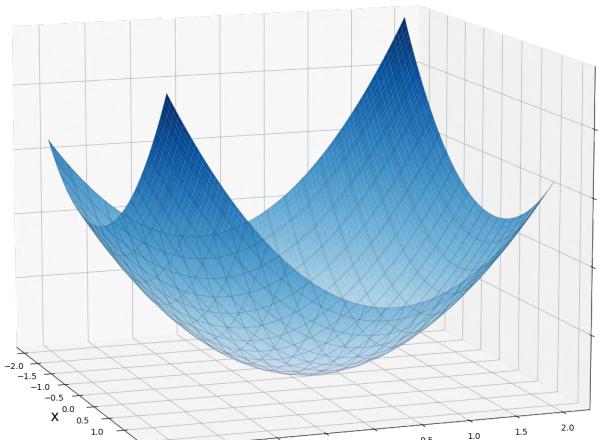
$$f(x, y) =$$

$$2x^2 + 3y^2 - xy$$



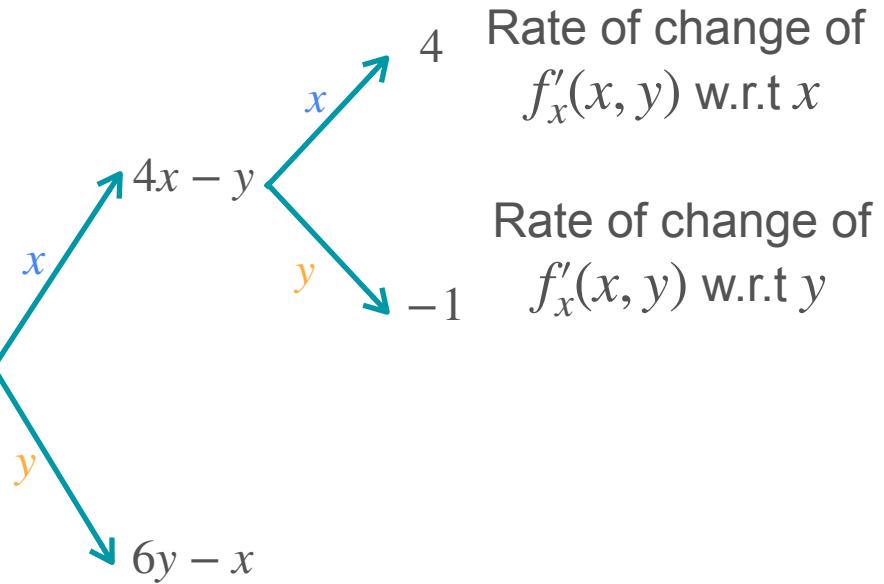
Rate of change of  
 $f'_x(x, y)$  w.r.t  $x$

# Second Derivative

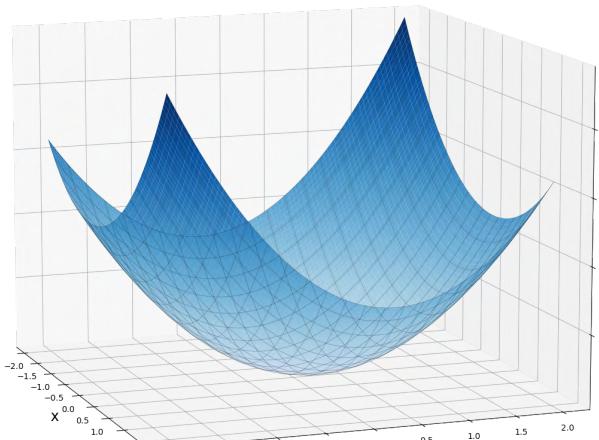


$$f(x, y) =$$

$$2x^2 + 3y^2 - xy$$

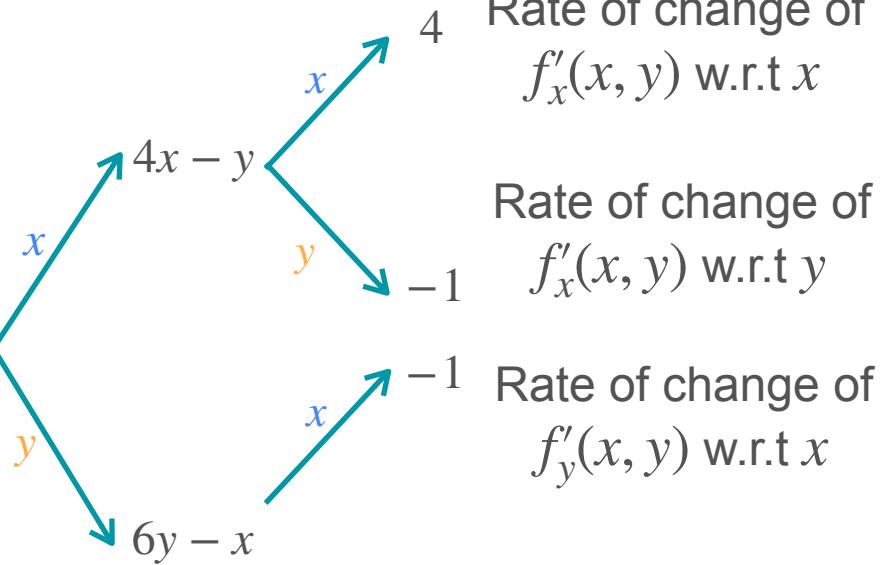


# Second Derivative

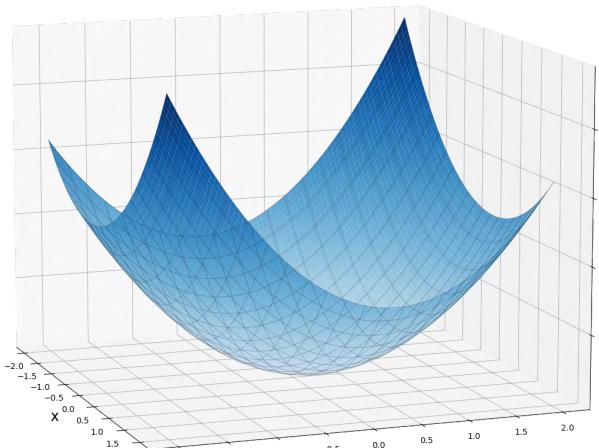


$$f(x, y) =$$

$$2x^2 + 3y^2 - xy$$

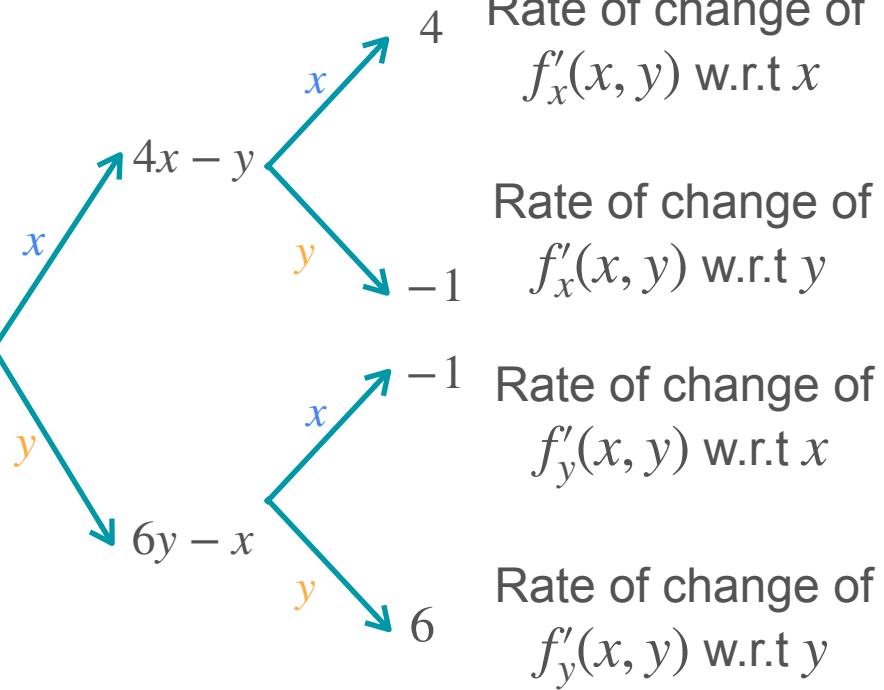


# Second Derivative



$$f(x, y) =$$

$$2x^2 + 3y^2 - xy$$



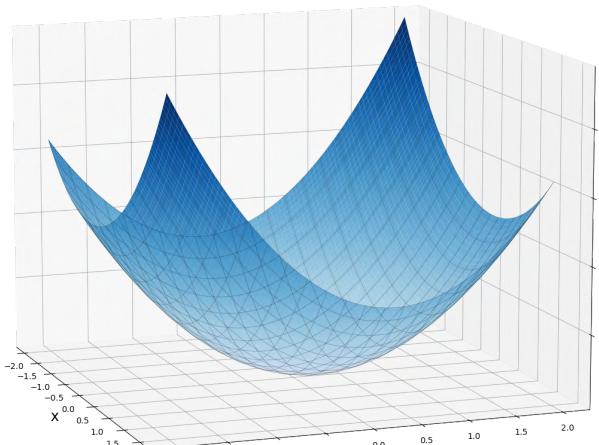
Rate of change of  
 $f'_x(x, y)$  w.r.t  $x$

Rate of change of  
 $f'_x(x, y)$  w.r.t  $y$

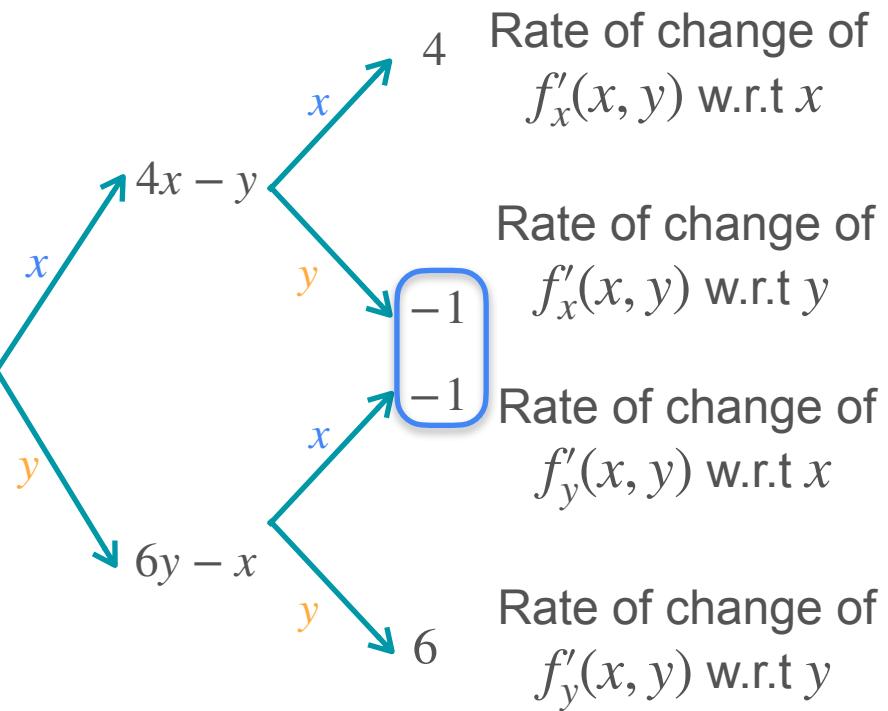
Rate of change of  
 $f'_y(x, y)$  w.r.t  $x$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $y$

# Second Derivative



$$f(x, y) =$$
$$2x^2 + 3y^2 - xy$$



# What Do These Mean?

Rate of change of  
 $f_x(x, y)$  w.r.t  $x$

Rate of change of  
 $f_y(x, y)$  w.r.t  $y$

Rate of change of  
 $f_x(x, y)$  w.r.t  $y$

Rate of change of  
 $f_y(x, y)$  w.r.t  $x$

# What Do These Mean?

Rate of change of  
 $f_x(x, y)$  w.r.t  $x$

Rate of change of  
 $f_y(x, y)$  w.r.t  $y$

Change in the change in the function  
w.r.t tiny changes in  $x$  and  $y$

Rate of change of  
 $f_x(x, y)$  w.r.t  $y$

Rate of change of  
 $f_y(x, y)$  w.r.t  $x$

# What Do These Mean?

Rate of change of

$f_x(x, y)$  w.r.t  $x$

Rate of change of

$f_y(x, y)$  w.r.t  $y$

Rate of change of

$f_x(x, y)$  w.r.t  $y$

Rate of change of

$f_y(x, y)$  w.r.t  $x$

Change in the change in the function  
w.r.t tiny changes in  $x$  and  $y$

Same idea as  
with one  
variable!

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis

# What Do These Mean?

Rate of change of

$f_x(x, y)$  w.r.t  $x$

Rate of change of

$f_y(x, y)$  w.r.t  $y$

Rate of change of

$f_x(x, y)$  w.r.t  $y$

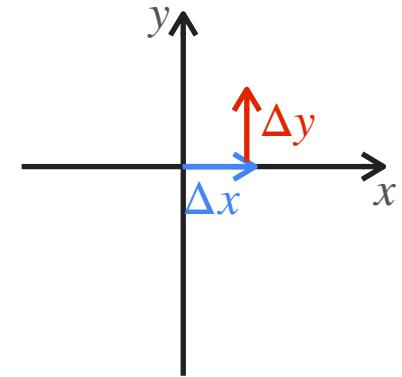
Rate of change of

$f_y(x, y)$  w.r.t  $x$

Change in the change in the function  
w.r.t tiny changes in  $x$  and  $y$

Same idea as  
with one  
variable!

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis



# What Do These Mean?

Rate of change of

$f_x(x, y)$  w.r.t  $x$

Rate of change of

$f_y(x, y)$  w.r.t  $y$

Rate of change of

$f_x(x, y)$  w.r.t  $y$

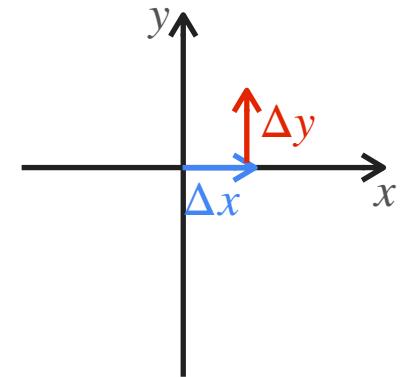
Rate of change of

$f_y(x, y)$  w.r.t  $x$

Change in the change in the function  
w.r.t tiny changes in  $x$  and  $y$

Same idea as  
with one  
variable!

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis
2. They are the same!



# What Do These Mean?

Rate of change of

$f_x(x, y)$  w.r.t  $x$

Rate of change of

$f_y(x, y)$  w.r.t  $y$

Rate of change of

$f_x(x, y)$  w.r.t  $y$

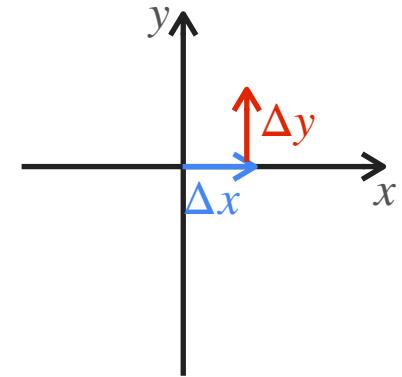
Rate of change of

$f_y(x, y)$  w.r.t  $x$

Change in the change in the function  
w.r.t tiny changes in  $x$  and  $y$

Same idea as  
with one  
variable!

1. Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis
2. They are the same!  
*(In most cases)*



# Notation

Rate of change of

$$f'_x(x, y) \text{ w.r.t } x$$

Rate of change of

$$f'_y(x, y) \text{ w.r.t } y$$

Rate of change of

$$f'_x(x, y) \text{ w.r.t } y$$

Rate of change of

$$f'_y(x, y) \text{ w.r.t } x$$

# Notation

## Leibniz's notation

Rate of change of  
 $f'_x(x, y)$  w.r.t  $x$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $y$

Rate of change of  
 $f'_x(x, y)$  w.r.t  $y$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $x$

# Notation

## Leibniz's notation

Rate of change of  
 $f'_x(x, y)$  w.r.t  $x$

$$\frac{\partial^2 f}{\partial x^2}$$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $y$

$$\frac{\partial^2 f}{\partial y^2}$$

Rate of change of  
 $f'_x(x, y)$  w.r.t  $y$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $x$

# Notation

Rate of change of  
 $f'_x(x, y)$  w.r.t  $x$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $y$

Rate of change of  
 $f'_x(x, y)$  w.r.t  $y$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $x$

## Leibniz's notation

$$\frac{\partial^2 f}{\partial x^2}$$

$$\frac{\partial^2 f}{\partial y^2}$$

$$\frac{\partial^2 f}{\partial x \partial y}$$

$$\frac{\partial^2 f}{\partial y \partial x}$$

# Notation

Rate of change of  
 $f'_x(x, y)$  w.r.t  $x$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $y$

Rate of change of  
 $f'_x(x, y)$  w.r.t  $y$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $x$

## Leibniz's notation

$$\frac{\partial^2 f}{\partial x^2}$$

$$\frac{\partial^2 f}{\partial y^2}$$

$$\frac{\partial^2 f}{\partial x \partial y}$$

$$\frac{\partial^2 f}{\partial y \partial x}$$

## Lagrange's notation

# Notation

Rate of change of  
 $f'_x(x, y)$  w.r.t  $x$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $y$

Rate of change of  
 $f''_x(x, y)$  w.r.t  $y$

Rate of change of  
 $f''_y(x, y)$  w.r.t  $x$

## Leibniz's notation

$$\frac{\partial^2 f}{\partial x^2}$$

$$\frac{\partial^2 f}{\partial y^2}$$

$$\frac{\partial^2 f}{\partial x \partial y}$$

$$\frac{\partial^2 f}{\partial y \partial x}$$

## Lagrange's notation

$$f_{xx}(x, y)$$

$$f_{yy}(x, y)$$

# Notation

Rate of change of  
 $f'_x(x, y)$  w.r.t  $x$

Rate of change of  
 $f'_y(x, y)$  w.r.t  $y$

Rate of change of  
 $f''_x(x, y)$  w.r.t  $y$

Rate of change of  
 $f''_y(x, y)$  w.r.t  $x$

## Leibniz's notation

$$\frac{\partial^2 f}{\partial x^2}$$

$$\frac{\partial^2 f}{\partial y^2}$$

$$\frac{\partial^2 f}{\partial x \partial y}$$

$$\frac{\partial^2 f}{\partial y \partial x}$$

## Lagrange's notation

$$f_{xx}(x, y)$$

$$f_{yy}(x, y)$$

$$f_{xy}(x, y)$$

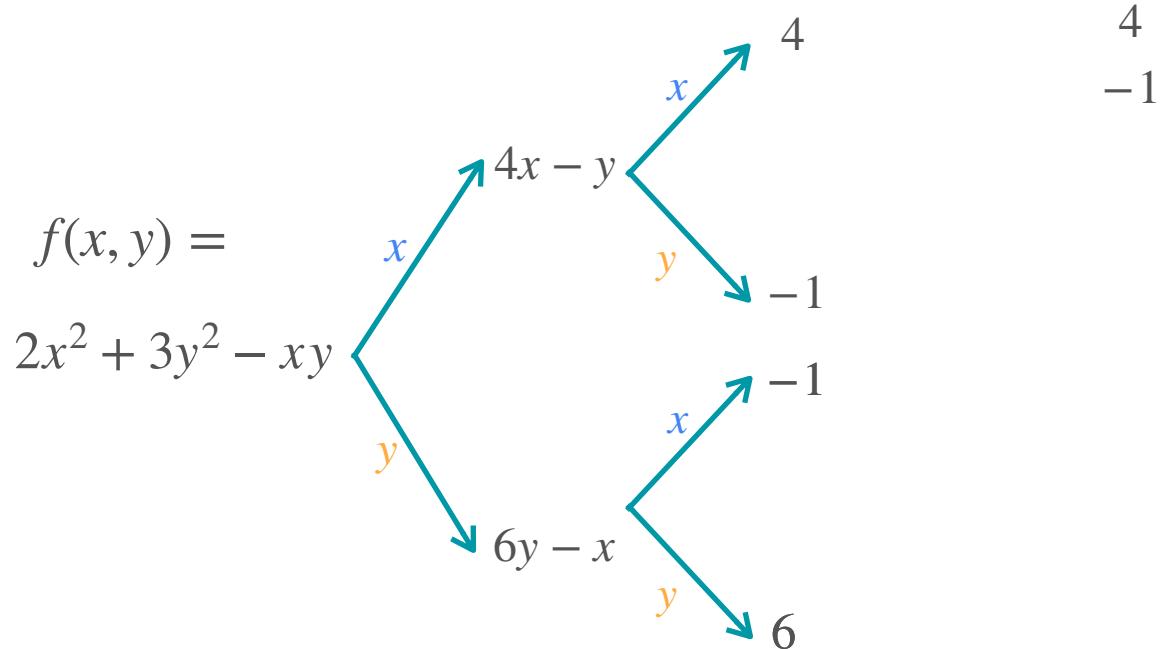
$$f_{yx}(x, y)$$

# Hessian Matrix

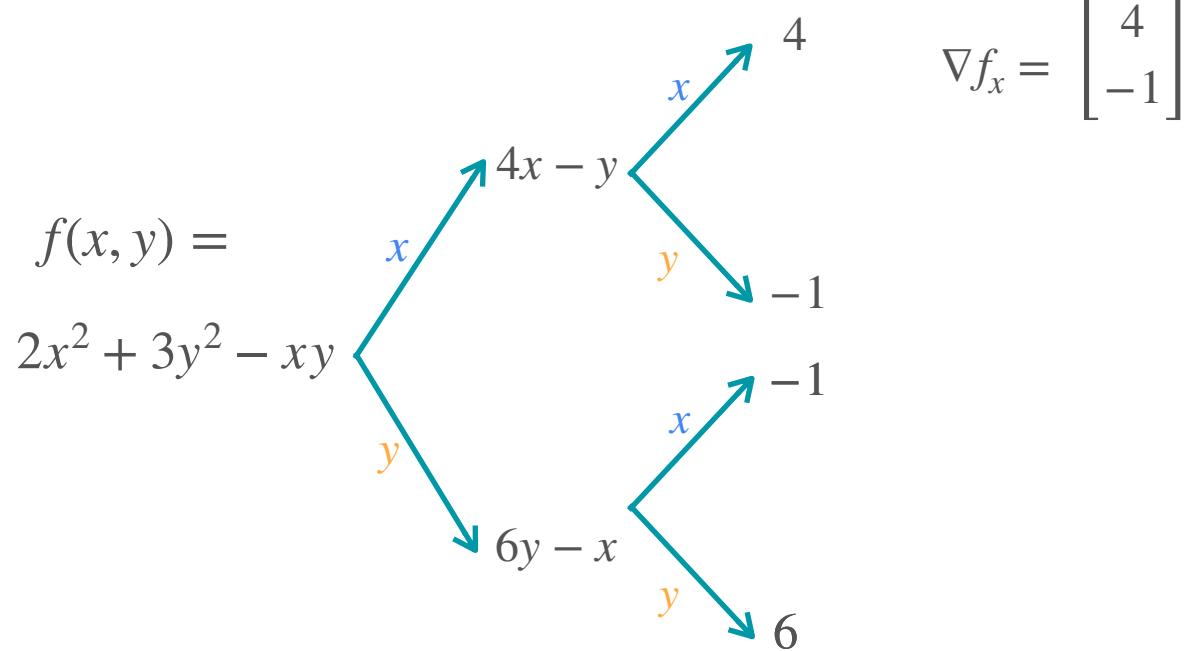
# Hessian Matrix

$$f(x, y) = 2x^2 + 3y^2 - xy$$
$$\begin{matrix} & \begin{matrix} 4 & \\ & -1 \end{matrix} \\ \begin{matrix} x \\ y \end{matrix} & \begin{pmatrix} 4x - y & \\ & 6y - x \end{pmatrix} \\ & \begin{matrix} -1 & \\ & 6 \end{matrix} \end{matrix}$$

# Hessian Matrix



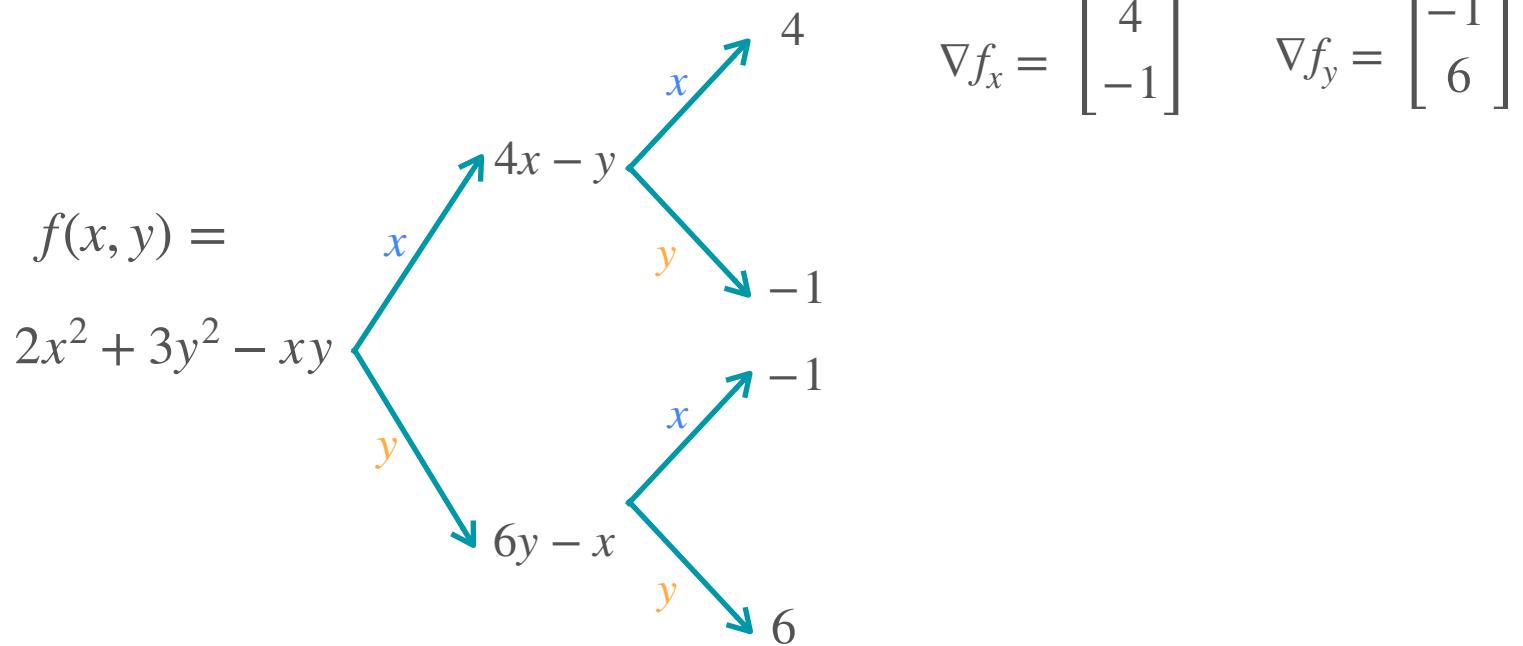
# Hessian Matrix



# Hessian Matrix

$$f(x, y) = 2x^2 + 3y^2 - xy$$
$$\nabla f_x = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$
$$\begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

# Hessian Matrix



# Hessian Matrix

$$f(x, y) = 2x^2 + 3y^2 - xy$$
$$\begin{matrix} & \begin{matrix} x \\ y \end{matrix} & \\ \begin{matrix} x \\ y \end{matrix} & \begin{matrix} 4x - y \\ 6y - x \end{matrix} & \begin{matrix} 4 \\ -1 \\ -1 \\ 6 \end{matrix} \end{matrix}$$

$$\nabla f_x = \begin{bmatrix} 4 \\ -1 \end{bmatrix} \quad \nabla f_y = \begin{bmatrix} -1 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

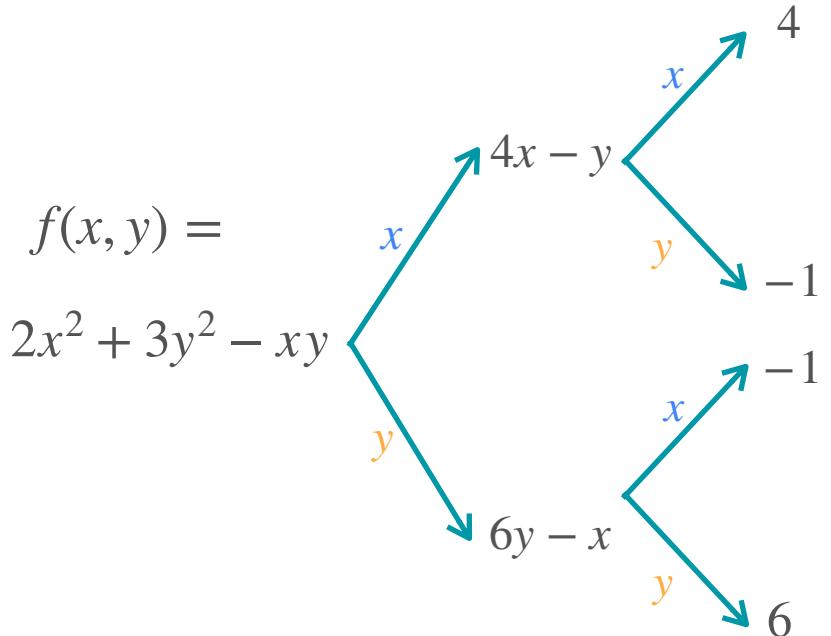
# Hessian Matrix

$$f(x, y) = 2x^2 + 3y^2 - xy$$
$$\begin{matrix} & \begin{matrix} 4x - y & 4 \\ -1 & 6y - x \end{matrix} \\ \begin{matrix} x \\ y \end{matrix} & \end{matrix}$$

$$\nabla f_x = \begin{bmatrix} 4 \\ -1 \end{bmatrix} \quad \nabla f_y = \begin{bmatrix} -1 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix} = \begin{bmatrix} \nabla f_x^T \\ \nabla f_y^T \end{bmatrix}$$

# Hessian Matrix



$$\nabla f_x = \begin{bmatrix} 4 \\ -1 \end{bmatrix} \quad \nabla f_y = \begin{bmatrix} -1 \\ 6 \end{bmatrix}$$

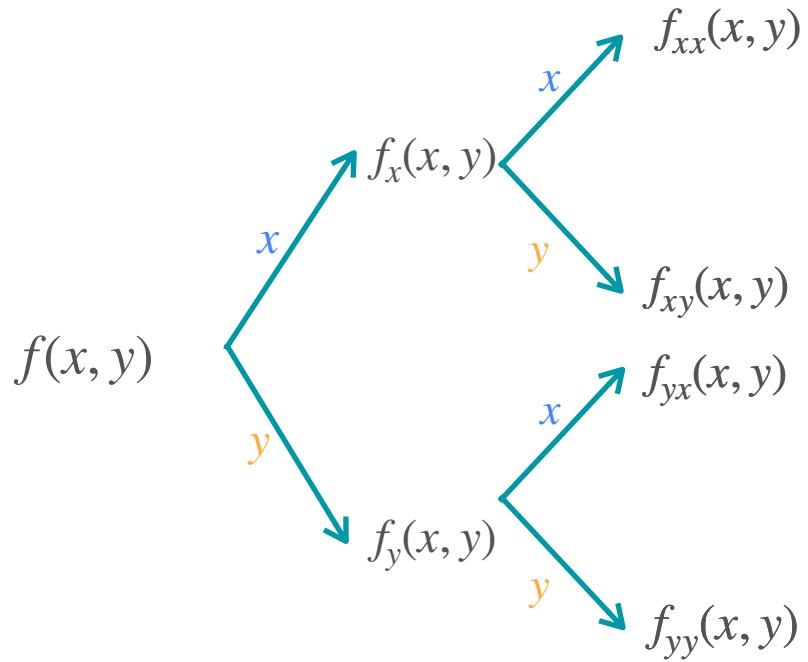
$$H = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix} = \begin{bmatrix} \nabla f_x^T \\ \nabla f_y^T \end{bmatrix}$$

**Hessian  
matrix**

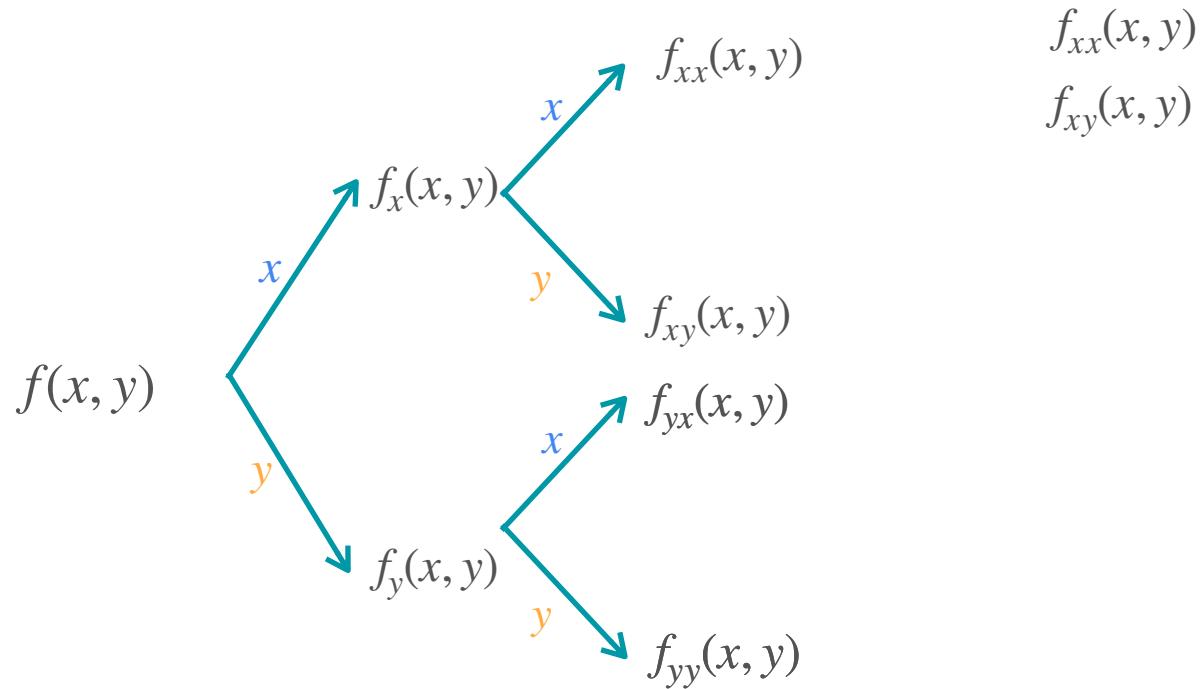
All information  
about second  
derivatives

# Hessian Matrix - General Case

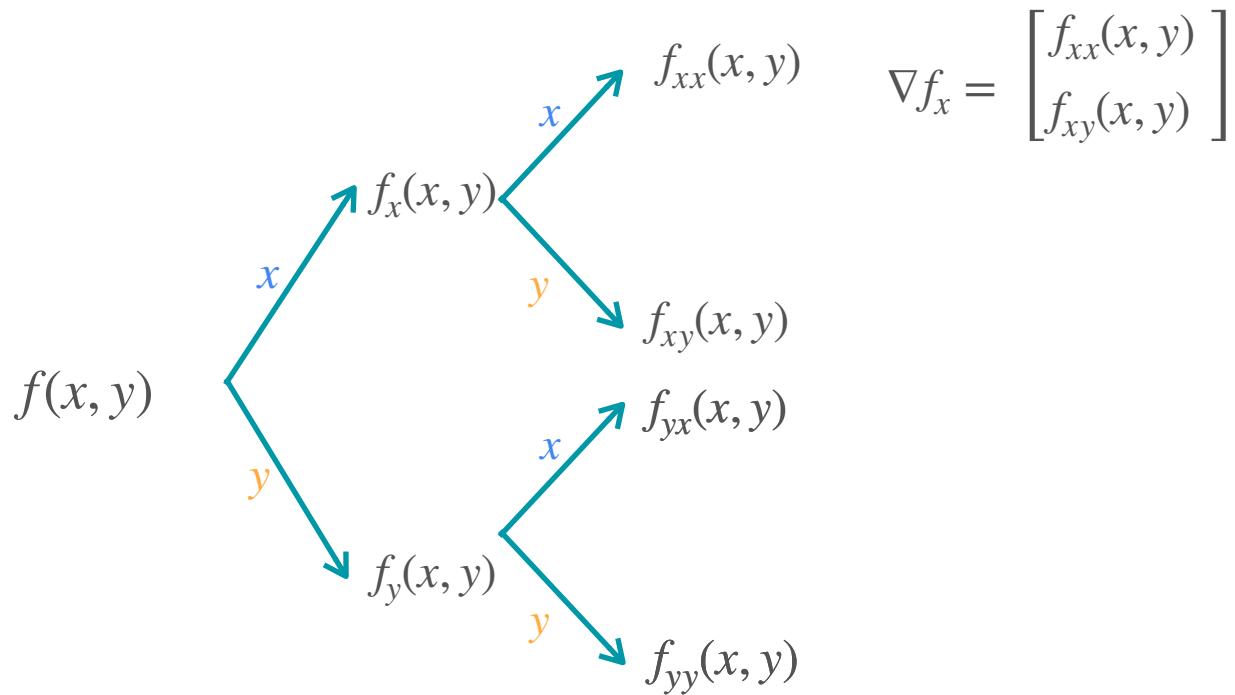
# Hessian Matrix - General Case



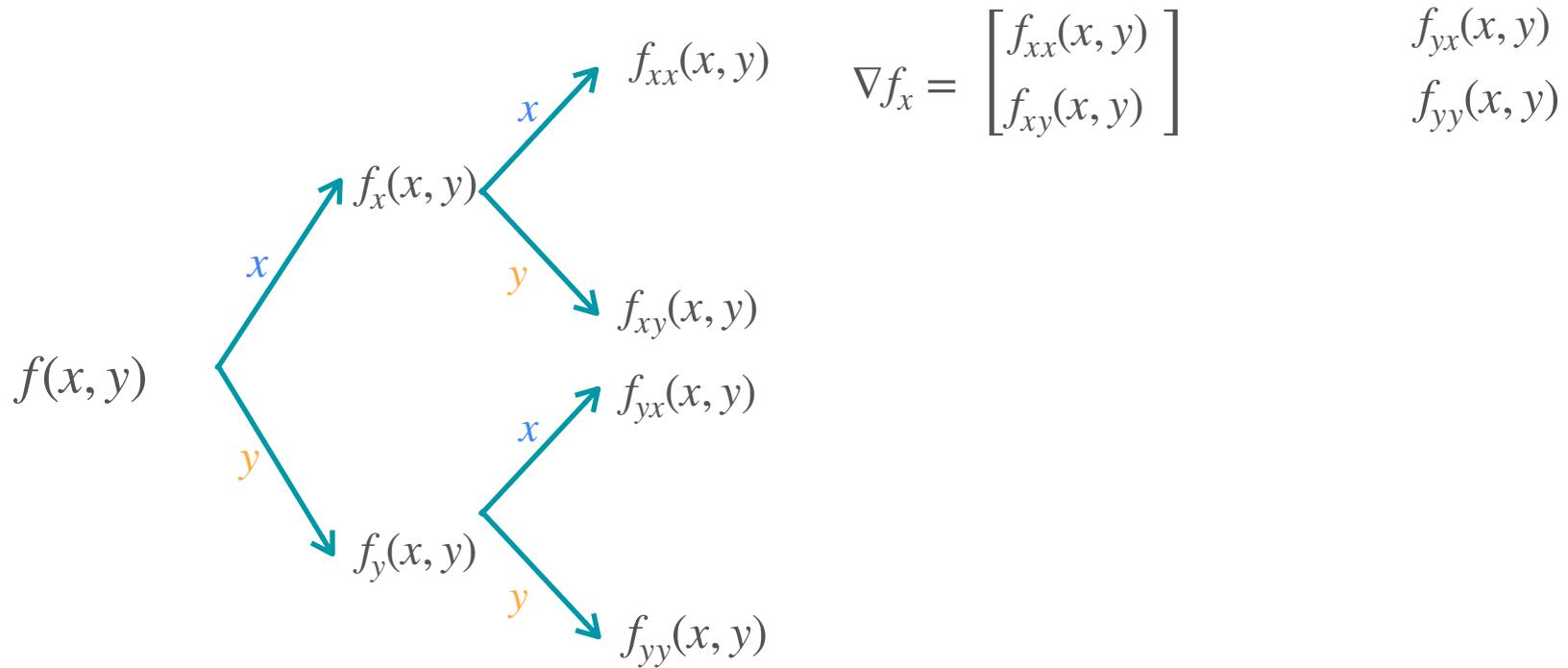
# Hessian Matrix - General Case



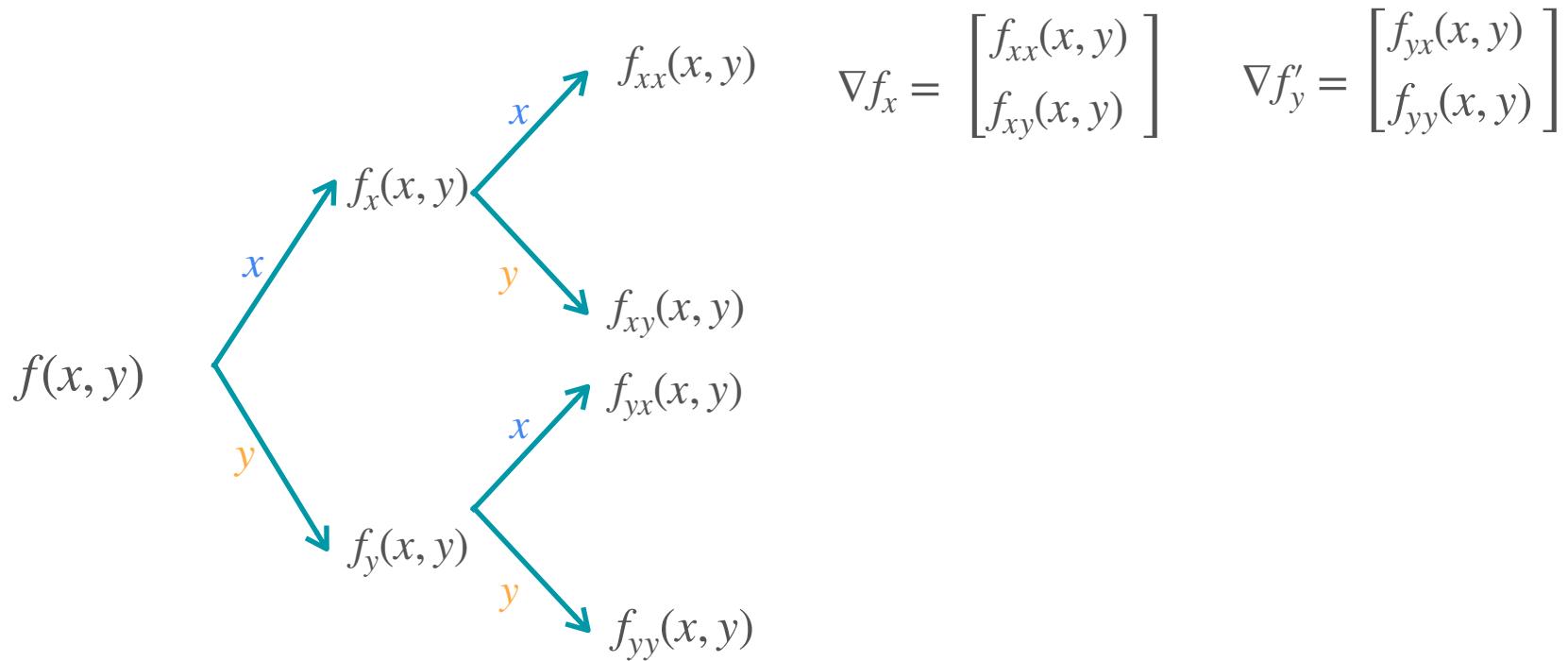
# Hessian Matrix - General Case



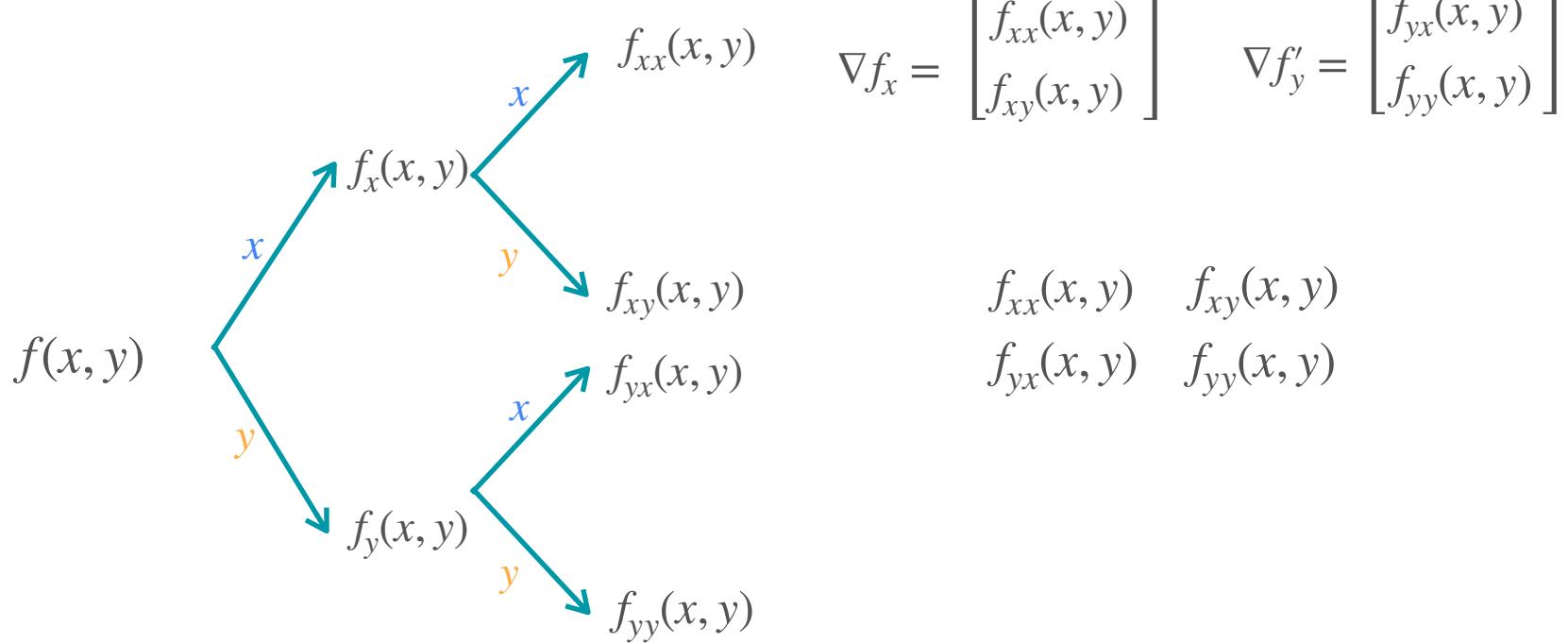
# Hessian Matrix - General Case



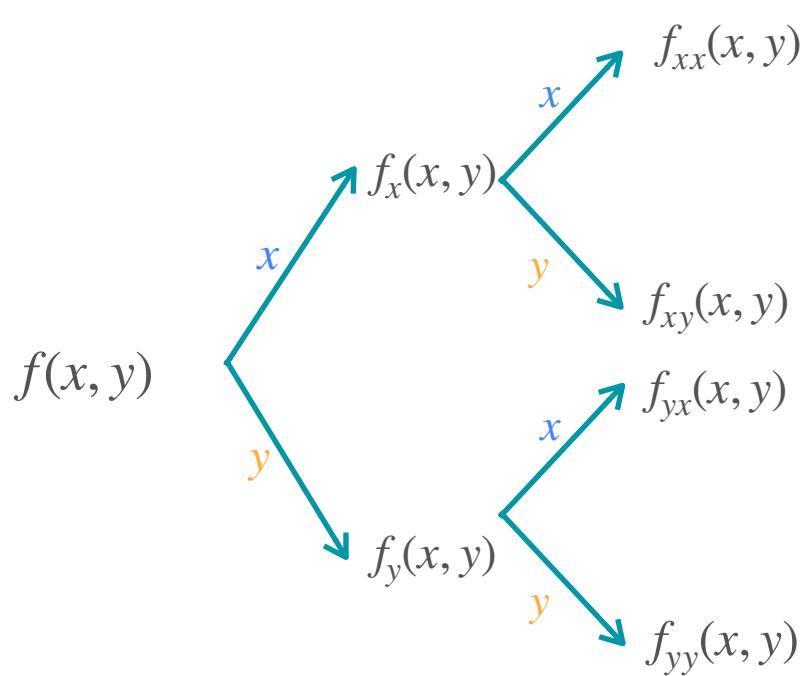
# Hessian Matrix - General Case



# Hessian Matrix - General Case



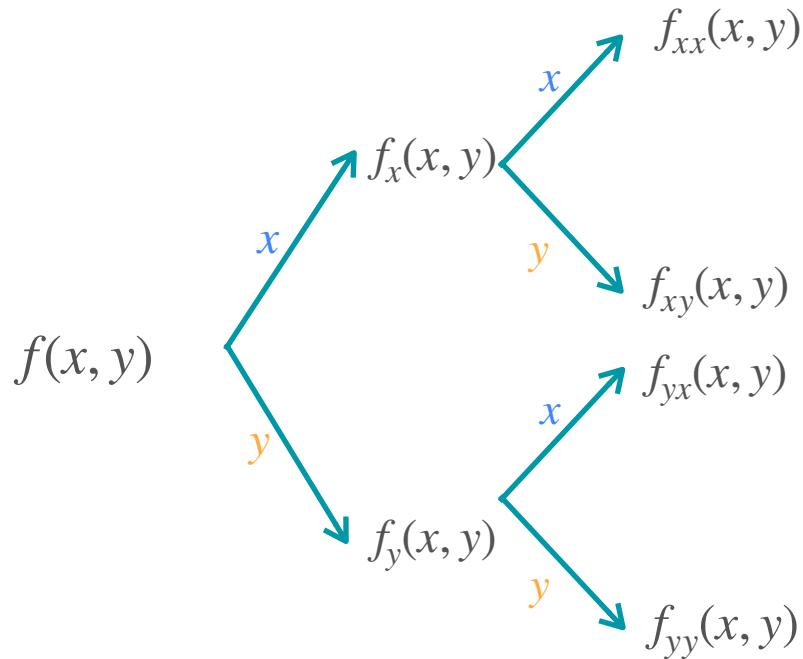
# Hessian Matrix - General Case



$$\nabla f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix} \quad \nabla f'_y = \begin{bmatrix} f_{yx}(x, y) \\ f_{yy}(x, y) \end{bmatrix}$$

$$\begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix} = \begin{bmatrix} \nabla f_x^T \\ \nabla f_y^T \end{bmatrix}$$

# Hessian Matrix - General Case



$$\nabla f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix} \quad \nabla f'_y = \begin{bmatrix} f_{yx}(x, y) \\ f_{yy}(x, y) \end{bmatrix}$$

$$H = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix} = \begin{bmatrix} \nabla f_x^T \\ \nabla f_y^T \end{bmatrix}$$

**Hessian  
matrix**

All information  
about second  
derivatives

# Second Derivative

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t $x$ $f_y(x, y)$ Rate of change w.r.t $y$ $\nabla f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$
Second derivative	$f''(x)$ Rate of change of the rate of change of $f(x)$	

# Second Derivative

	1 variable	2 variables
Function	$f(x)$	$f(x, y)$
First derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t $x$ $f_y(x, y)$ Rate of change w.r.t $y$ $\nabla f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$
Second derivative	$f''(x)$ Rate of change of the rate of change of $f(x)$	$H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$



DeepLearning.AI

# Optimization in Neural Networks and Newton's Method

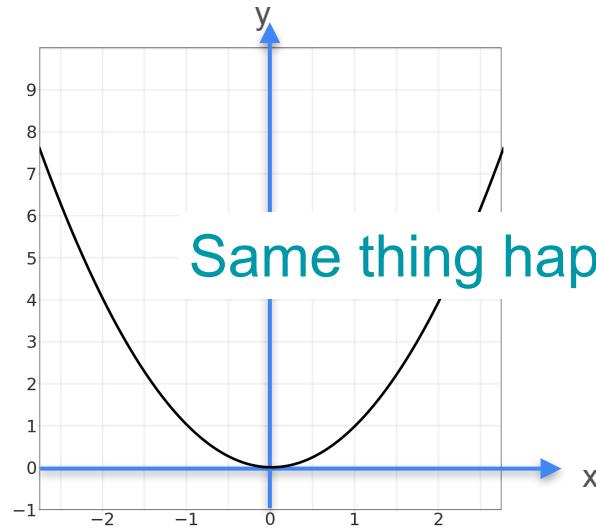
---

## Hessians and concavity

# Remember...

Same thing happens for many variables!

# Remember...

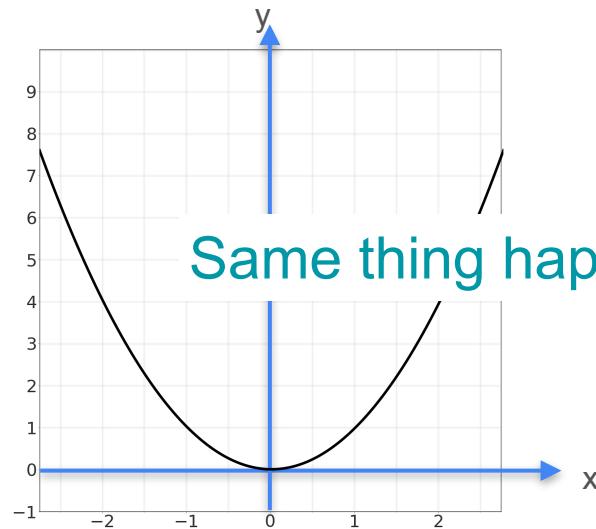


Same thing happens for many variables!

Concave up or convex

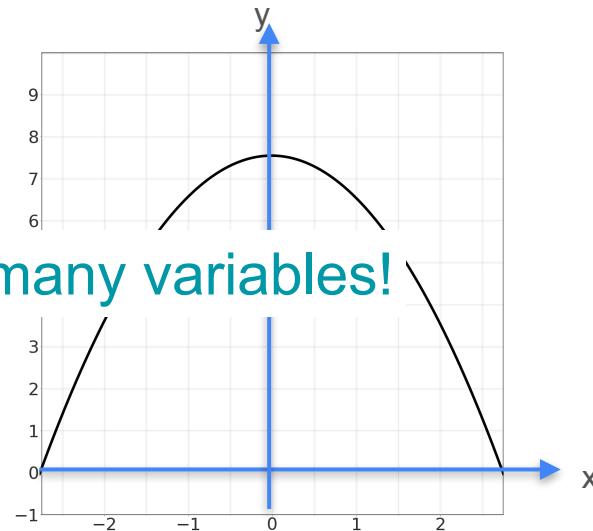
$$f''(0) > 0$$

# Remember...



Concave up or convex

$$f''(0) > 0$$

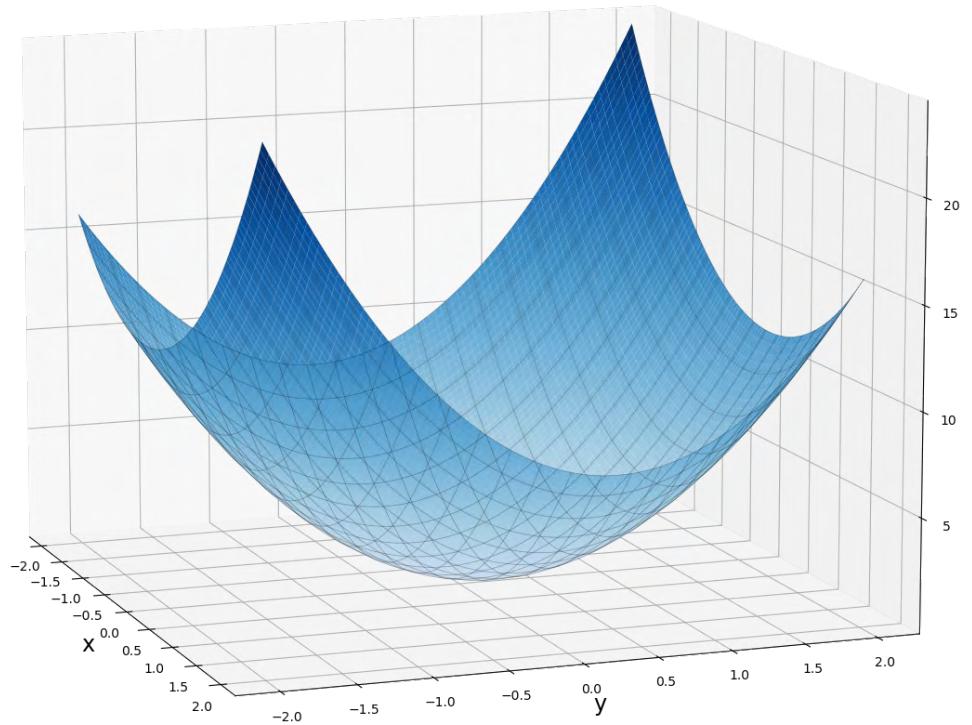


Concave down

$$f''(0) < 0$$

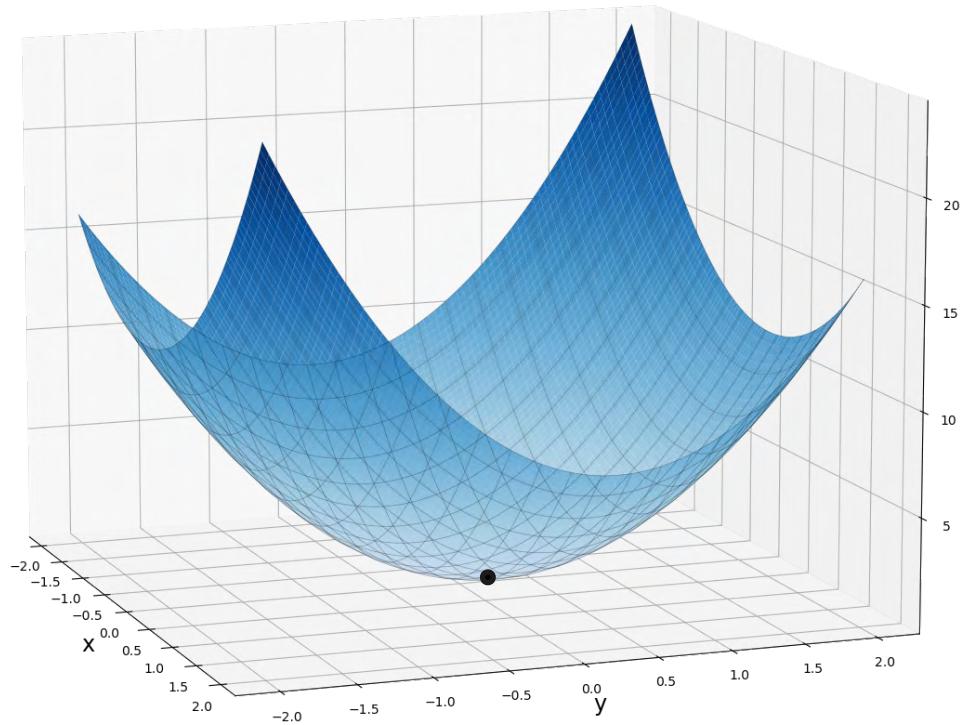
# Concave Up

# Concave Up



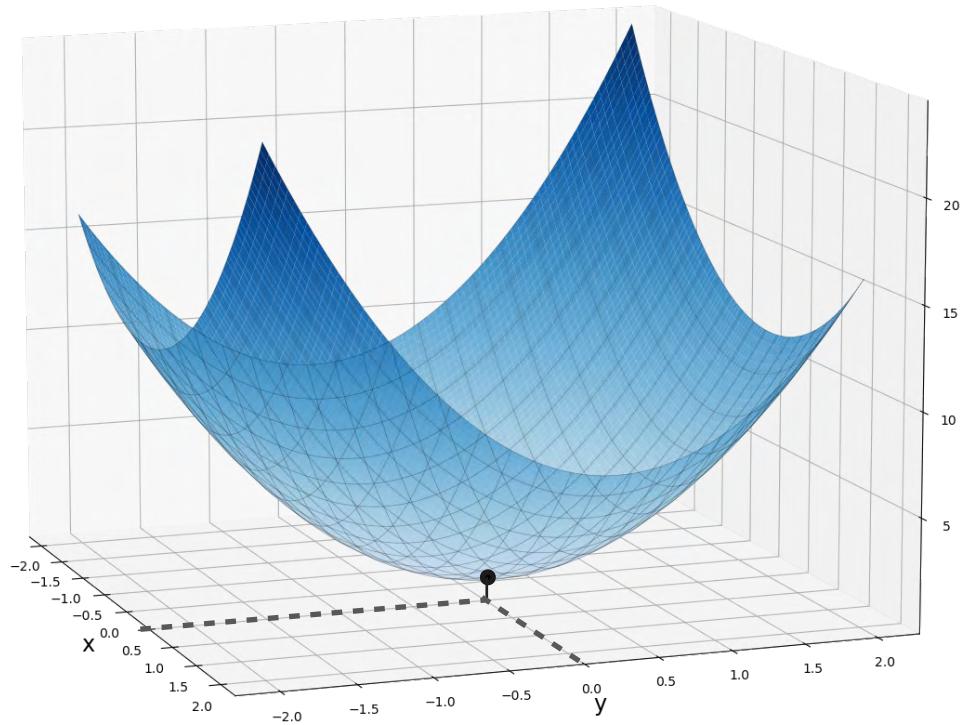
$$f(x, y) = 2x^2 + 3y^2 - xy$$

# Concave Up



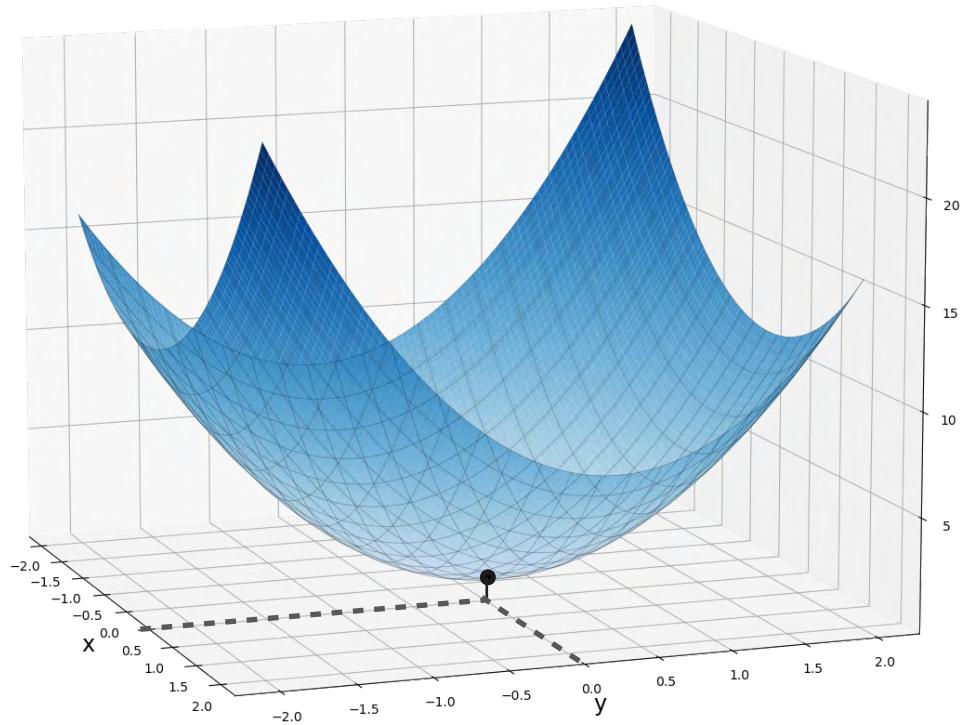
$$f(x, y) = 2x^2 + 3y^2 - xy$$

# Concave Up



$$f(x, y) = 2x^2 + 3y^2 - xy$$

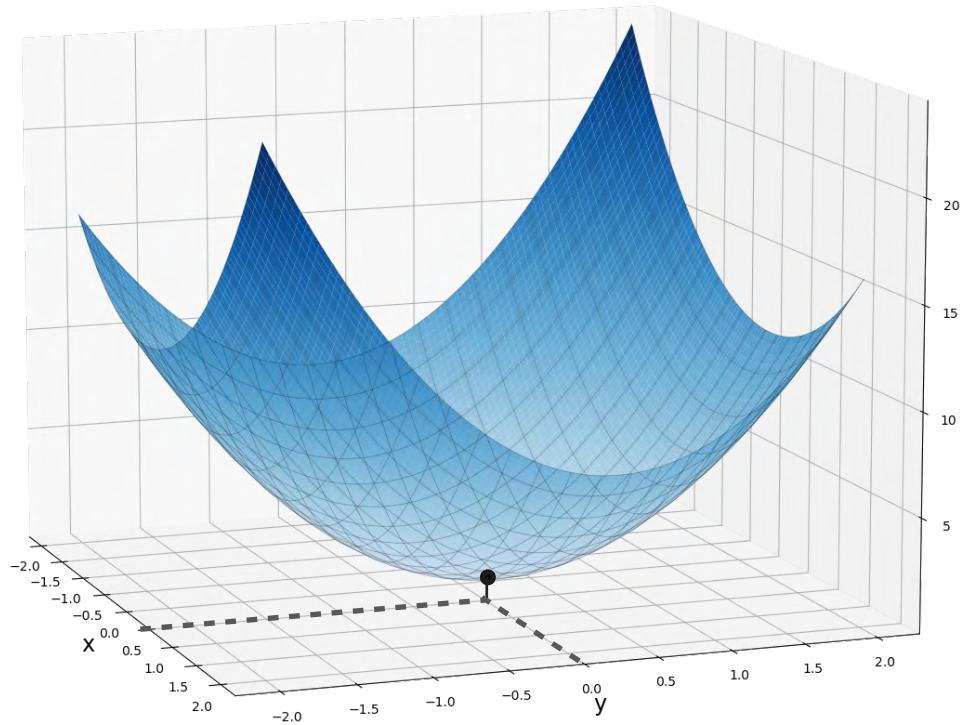
# Concave Up



$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

# Concave Up

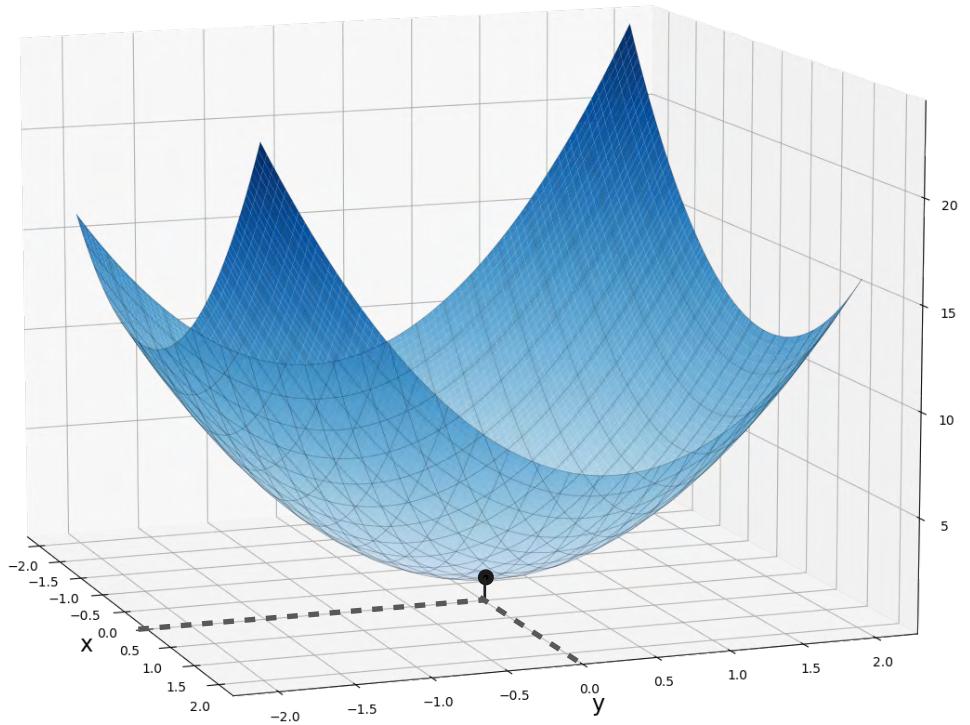


$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

# Concave Up

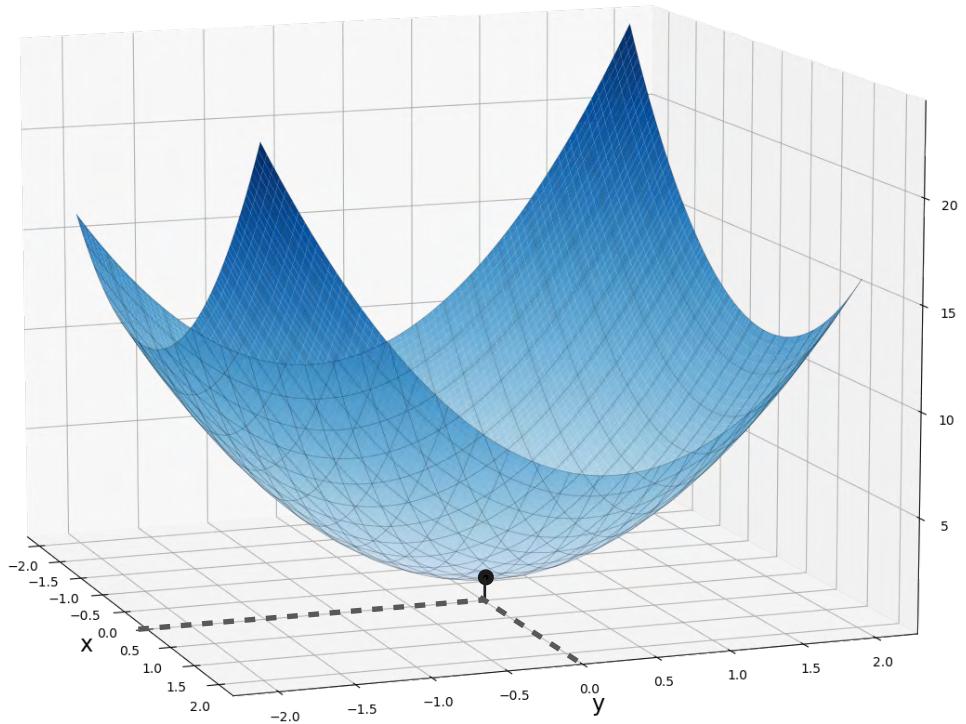


$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det \left( \begin{bmatrix} 4 - \lambda & -1 \\ -1 & 6 - \lambda \end{bmatrix} \right)$$

# Concave Up

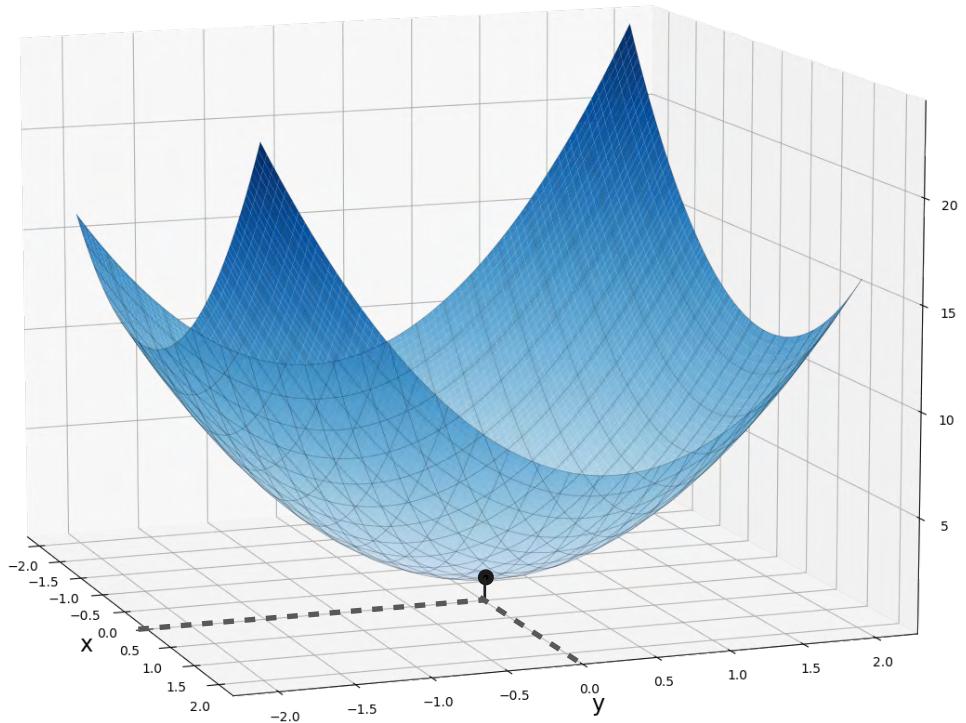


$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det \left( \begin{bmatrix} 4 - \lambda & -1 \\ -1 & 6 - \lambda \end{bmatrix} \right)$$
$$= (4 - \lambda)(6 - \lambda) - (-1)(-1)$$

# Concave Up



$$f(x, y) = 2x^2 + 3y^2 - xy$$

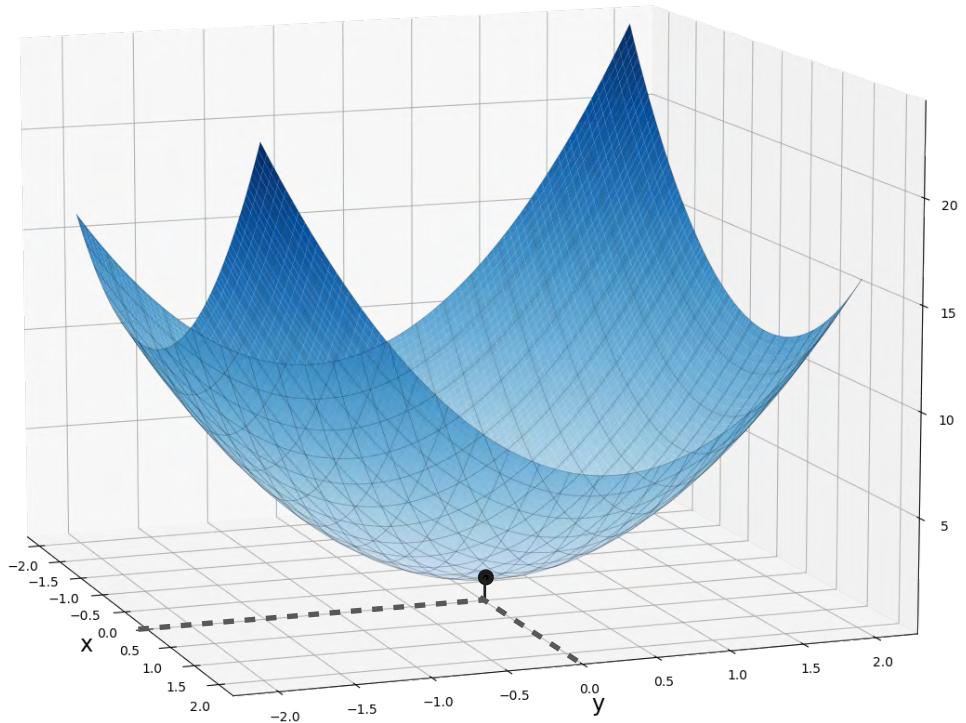
$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det \left( \begin{bmatrix} 4 - \lambda & -1 \\ -1 & 6 - \lambda \end{bmatrix} \right)$$

$$= (4 - \lambda)(6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23$$

# Concave Up



$$f(x, y) = 2x^2 + 3y^2 - xy$$

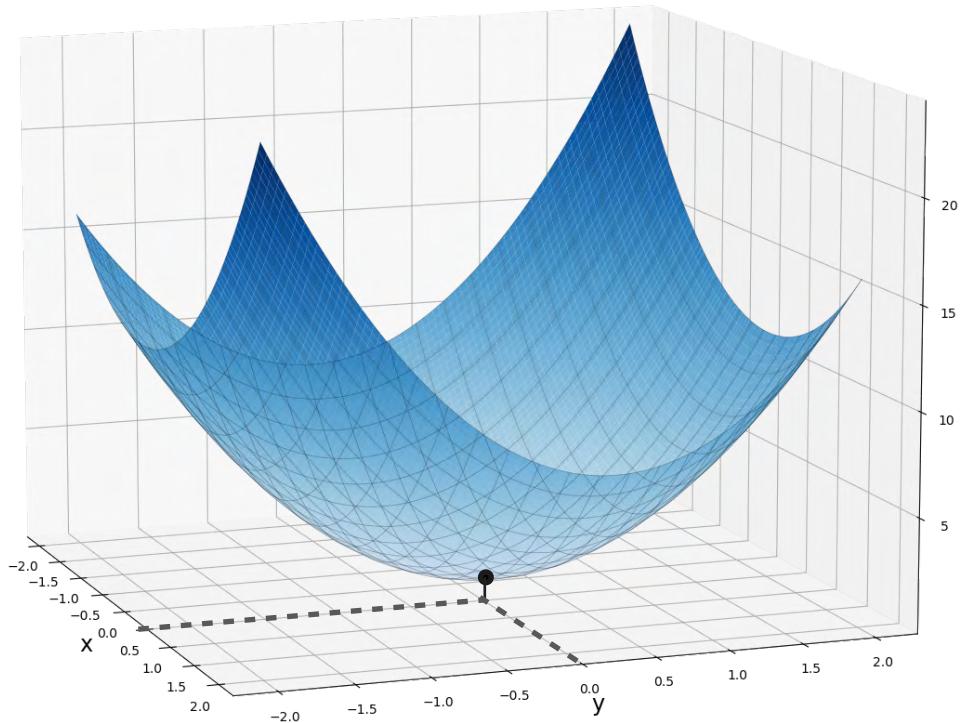
$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det \left( \begin{bmatrix} 4 - \lambda & -1 \\ -1 & 6 - \lambda \end{bmatrix} \right)$$

$$= (4 - \lambda)(6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23 \rightarrow \lambda_1 = 6.41$$

# Concave Up



$$f(x, y) = 2x^2 + 3y^2 - xy$$

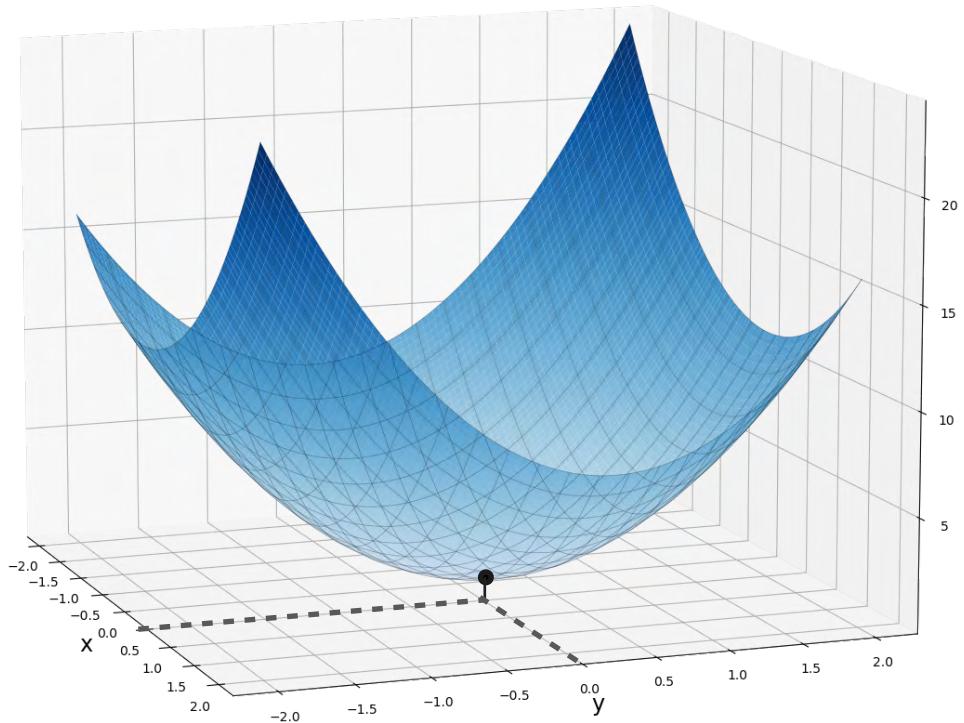
$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det \left( \begin{bmatrix} 4 - \lambda & -1 \\ -1 & 6 - \lambda \end{bmatrix} \right)$$

$$= (4 - \lambda)(6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23 \quad \begin{matrix} \xrightarrow{\text{blue}} \\ \xrightarrow{\text{blue}} \end{matrix} \begin{matrix} \lambda_1 = 6.41 \\ \lambda_2 = 3.59 \end{matrix}$$

# Concave Up



$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det \left( \begin{bmatrix} 4 - \lambda & -1 \\ -1 & 6 - \lambda \end{bmatrix} \right)$$

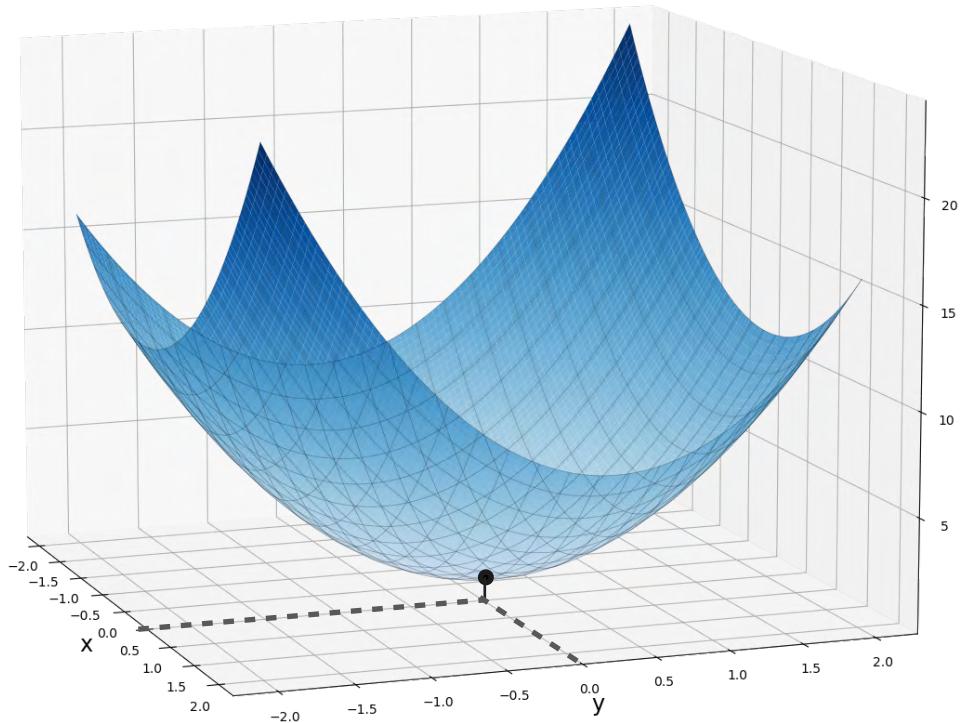
$$= (4 - \lambda)(6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23$$

$$\lambda_1 = 6.41$$
$$\lambda_2 = 3.59$$

$> 0$

# Concave Up



$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$H(0,0) = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) = \det \left( \begin{bmatrix} 4 - \lambda & -1 \\ -1 & 6 - \lambda \end{bmatrix} \right)$$

$$= (4 - \lambda)(6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 - 10\lambda + 23$$

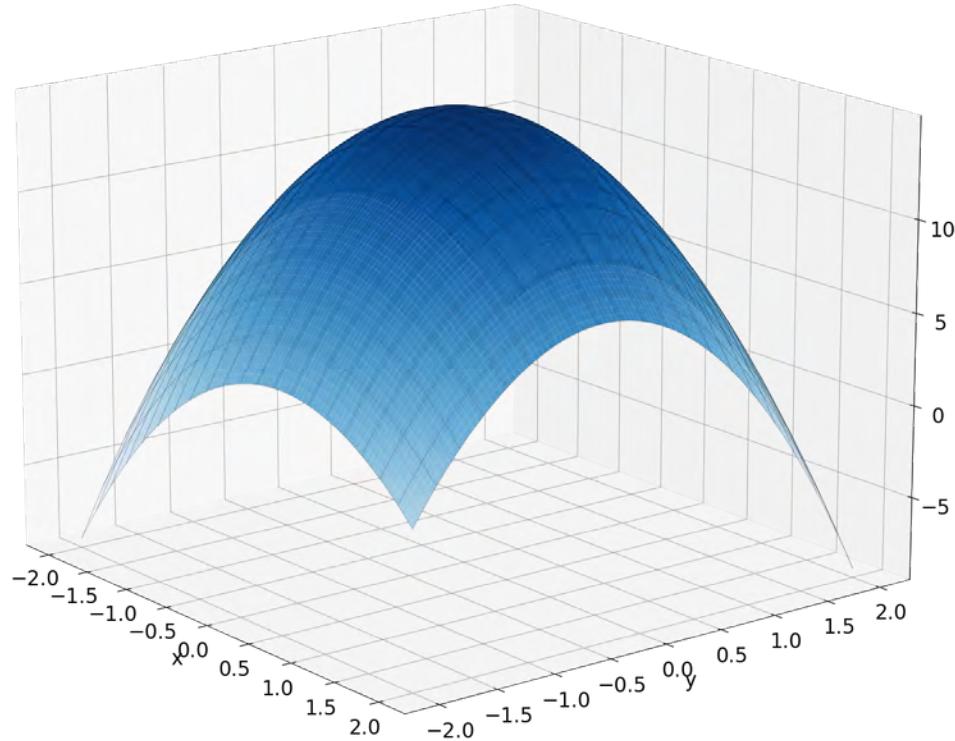
$$\lambda_1 = 6.41$$
$$\lambda_2 = 3.59$$

(0,0) is a minimum!

> 0

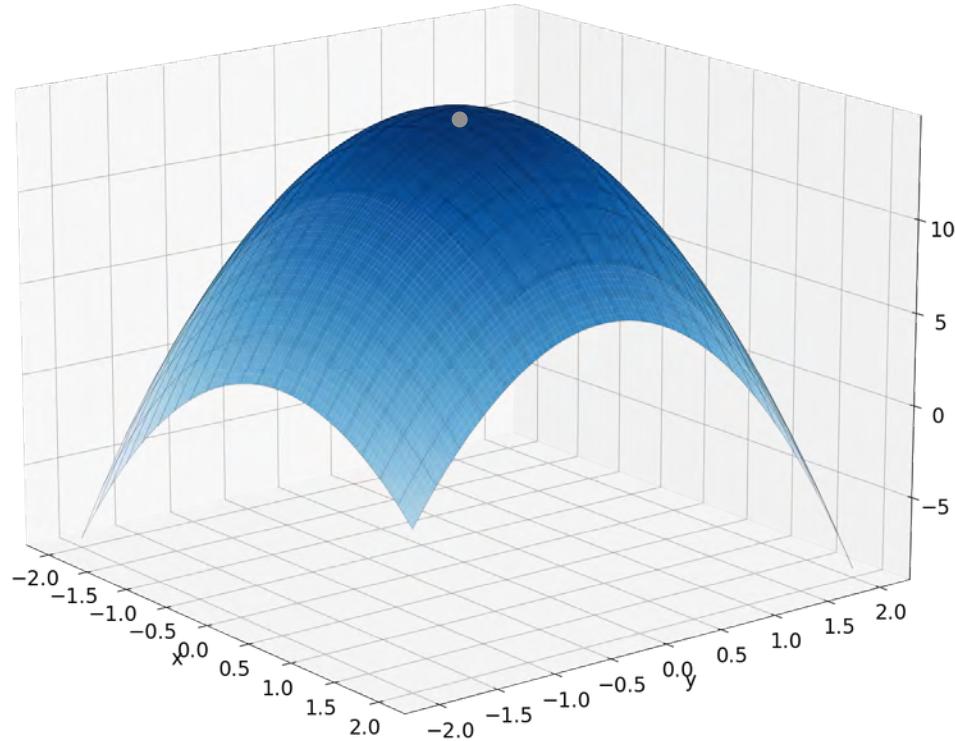
# Concave Down

# Concave Down



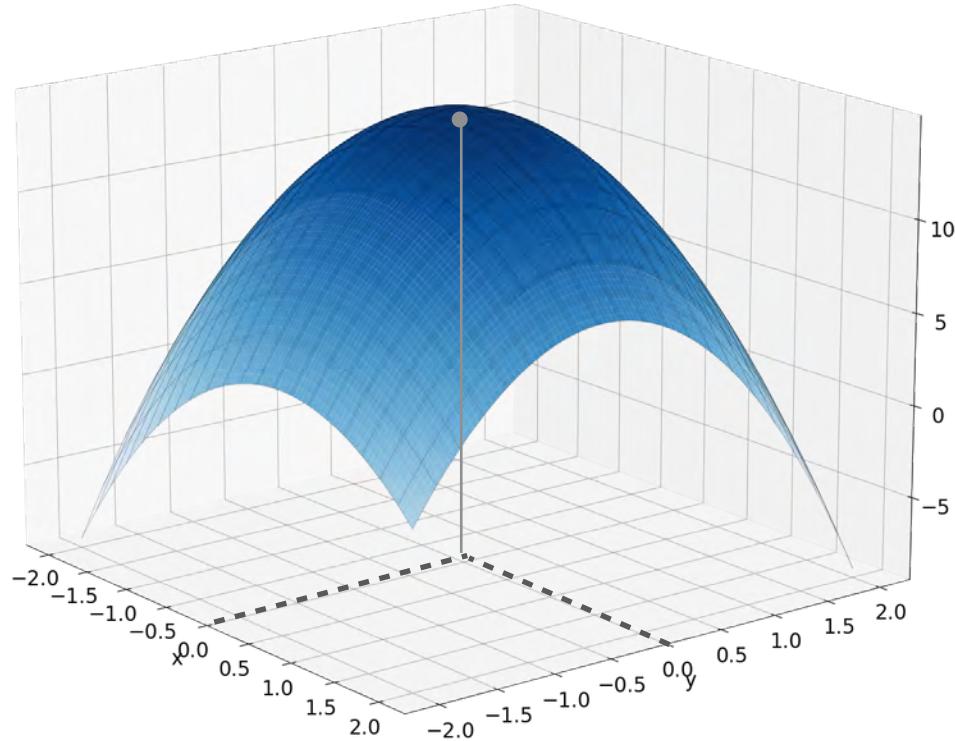
$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

# Concave Down



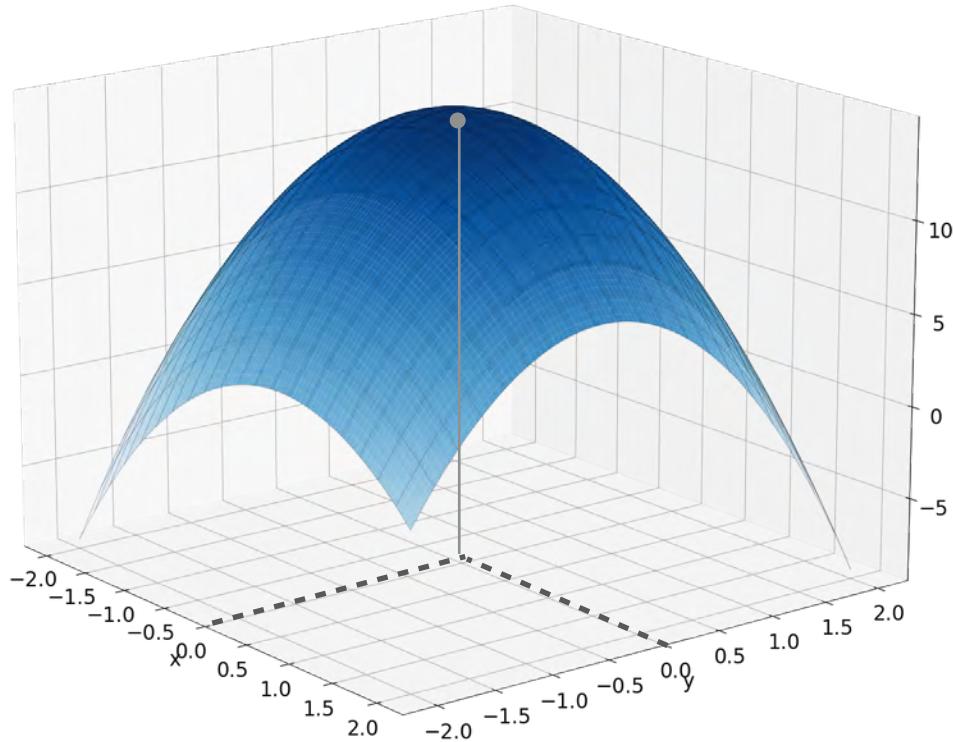
$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

# Concave Down



$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

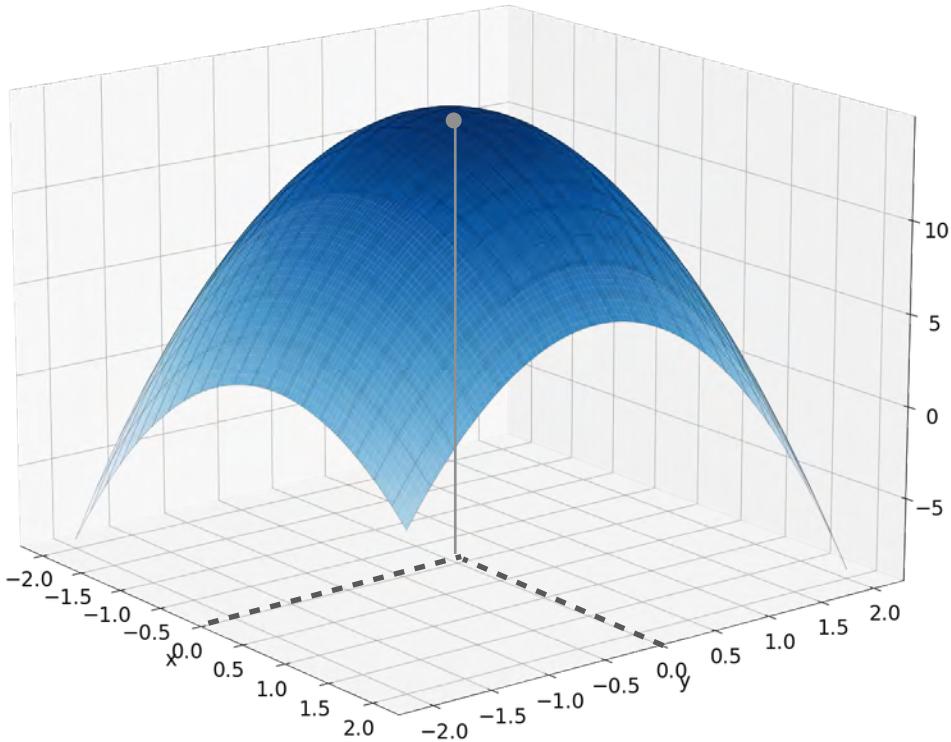
# Concave Down



$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x, y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

# Concave Down

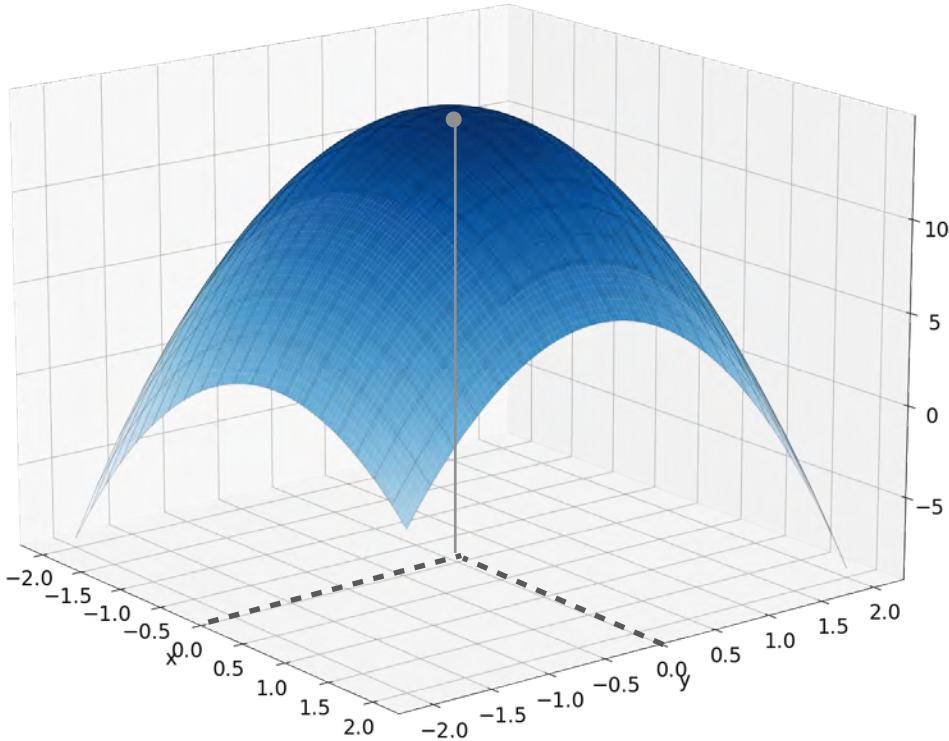


$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x, y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

# Concave Down



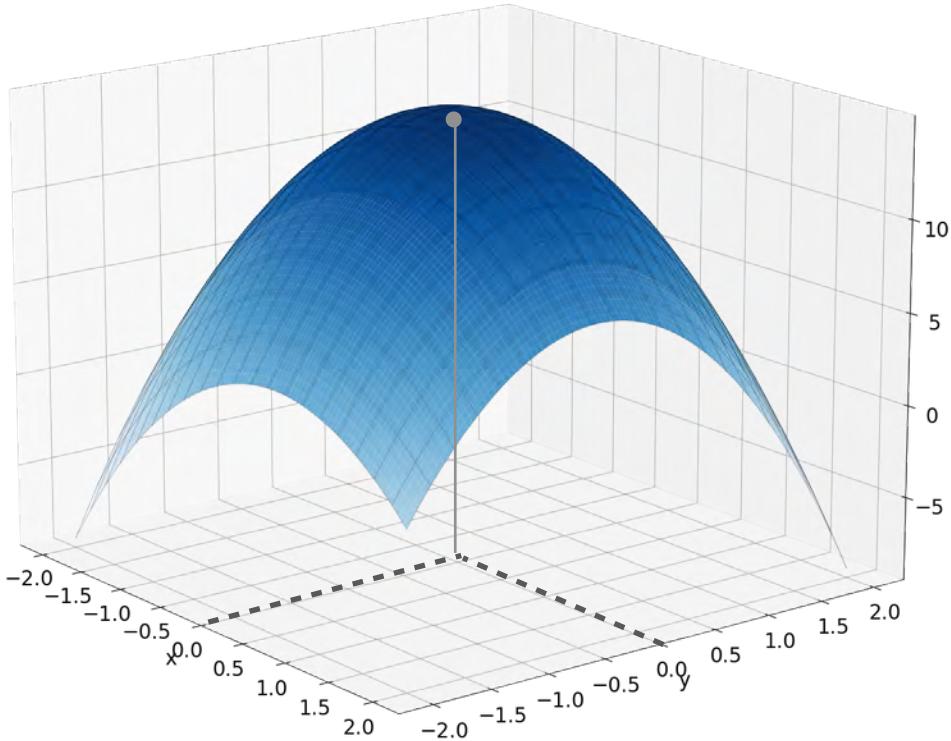
$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x, y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

# Concave Down



$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

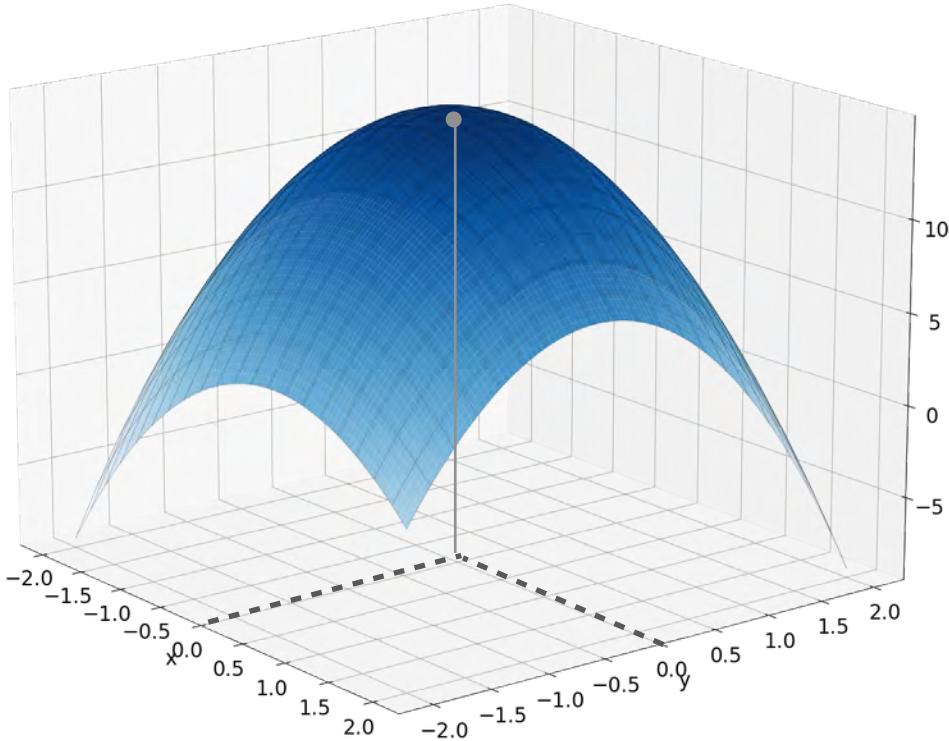
$$\nabla f(x, y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

# Concave Down



$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x, y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

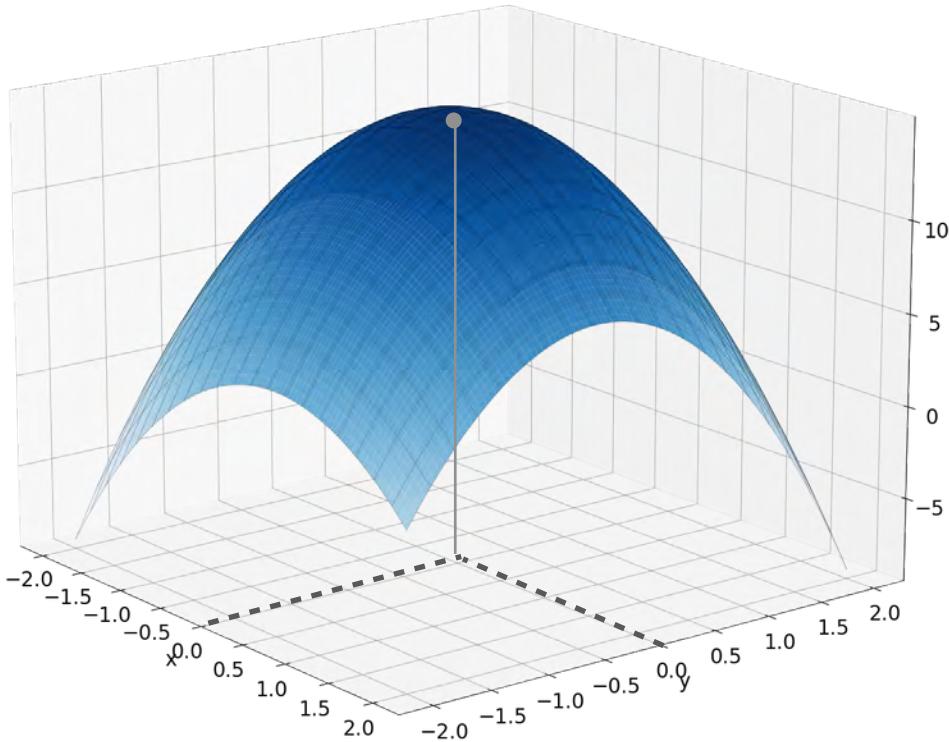
$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 + 10\lambda + 23$$

# Concave Down



$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x, y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

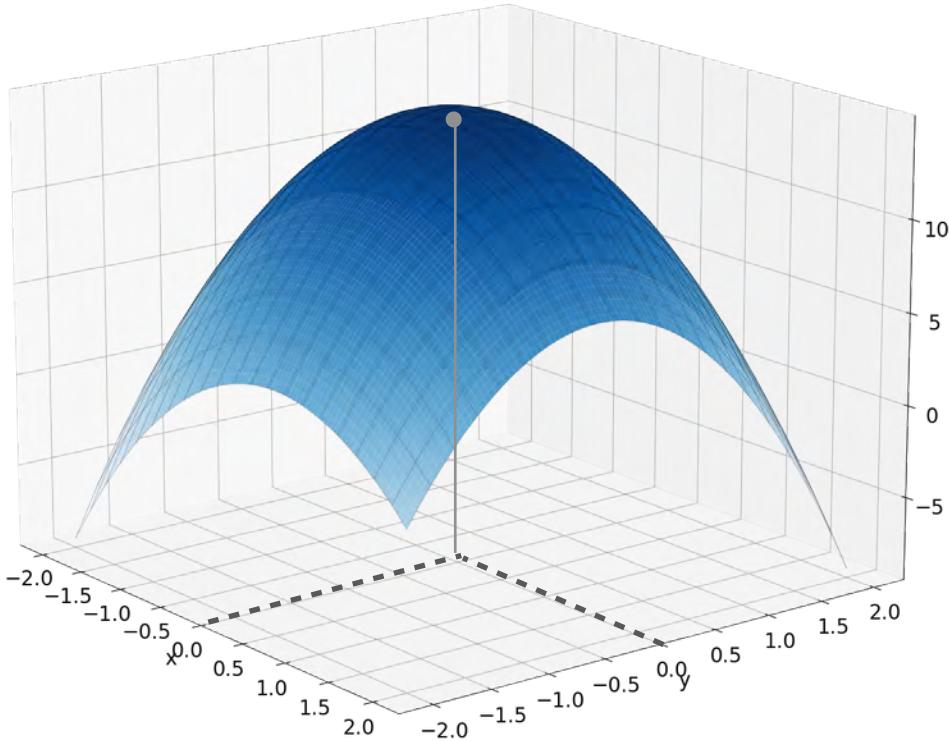
$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 + 10\lambda + 23 \rightarrow \lambda_1 = -3.49$$

# Concave Down



$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x, y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

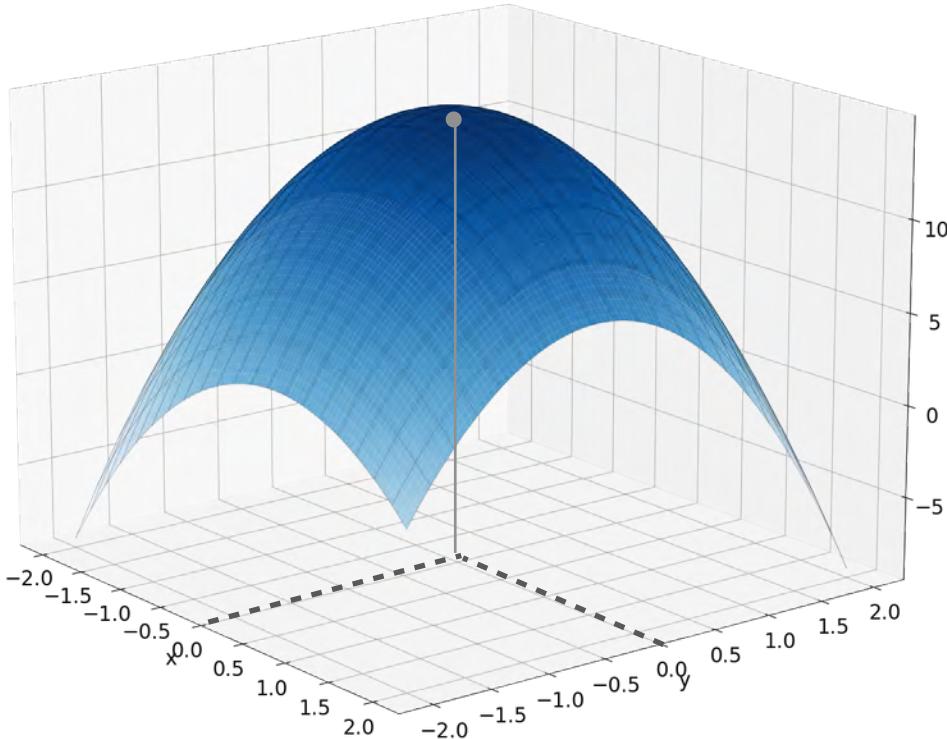
$$\det(H(0,0) - \lambda I) =$$

$$(-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 + 10\lambda + 23$$

$$\begin{array}{l} \lambda_1 = -3.49 \\ \lambda_2 = -6.41 \end{array}$$

# Concave Down



$$f(x, y) = -2x^2 - 3y^2 - xy + 15$$

$$\nabla f(x, y) = \begin{bmatrix} -4x - y \\ -x - 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} -4 & -1 \\ -1 & -6 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(-4 - \lambda)(-6 - \lambda) - (-1)(-1)$$

$$= \lambda^2 + 10\lambda + 23$$

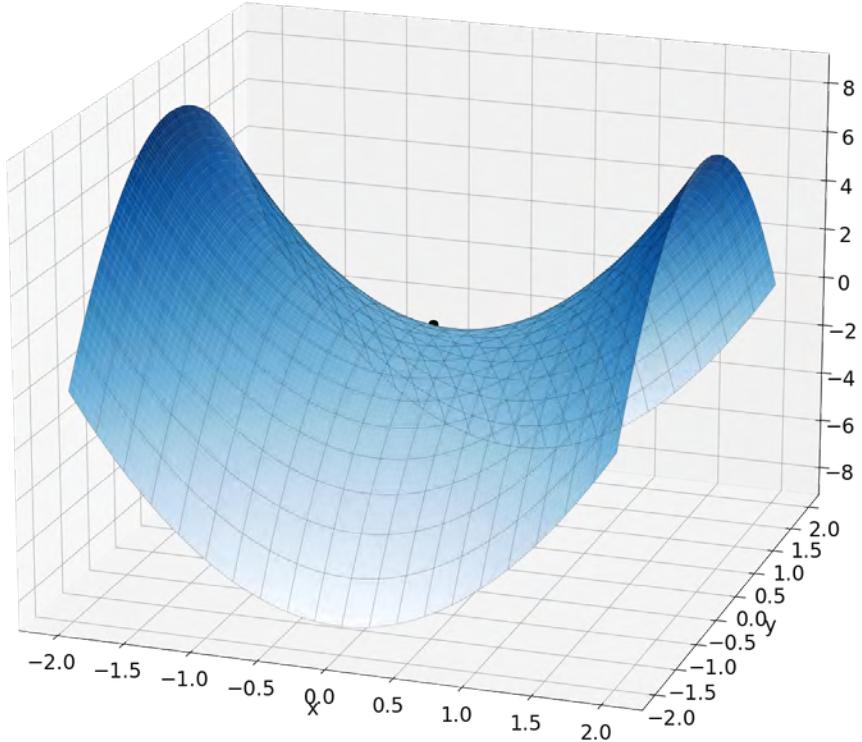
$$\lambda_1 = -3.49$$
$$\lambda_2 = -6.41$$

(0,0) is a maximum!

< 0

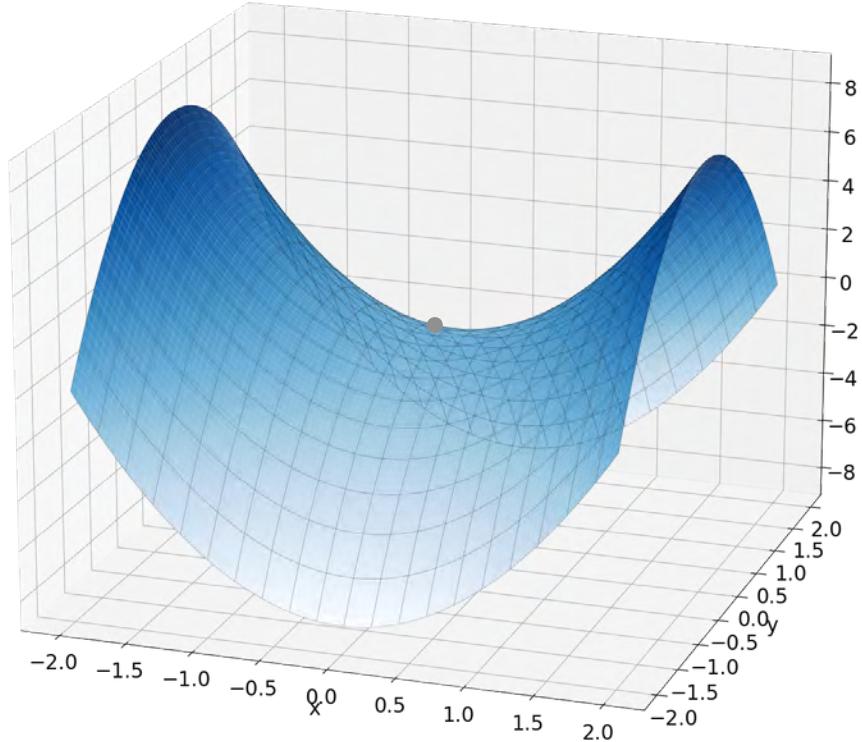
# Saddle Point

# Saddle Point



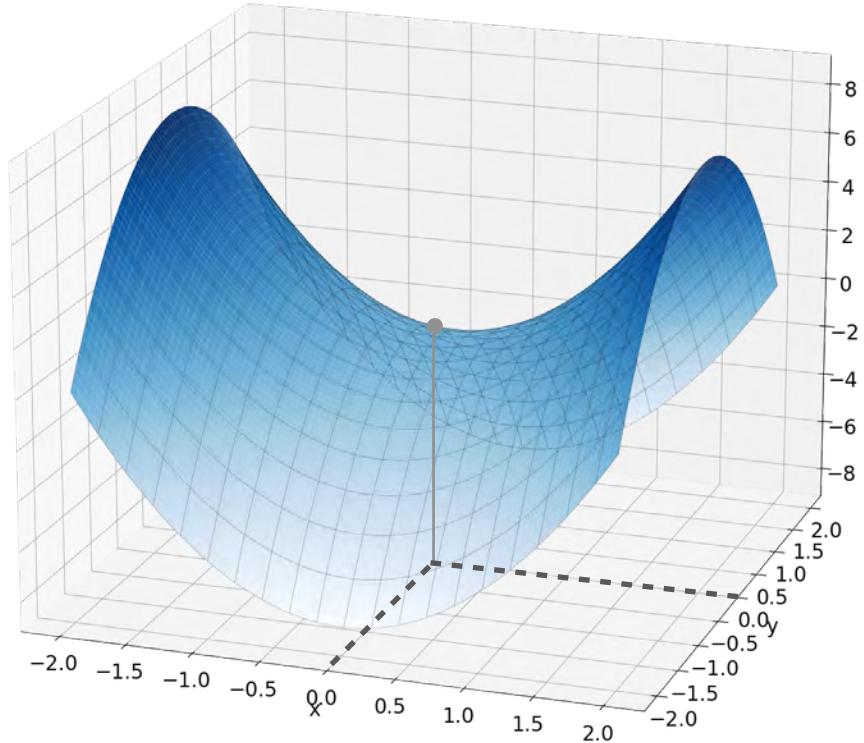
$$f(x, y) = 2x^2 - 2y^2$$

# Saddle Point



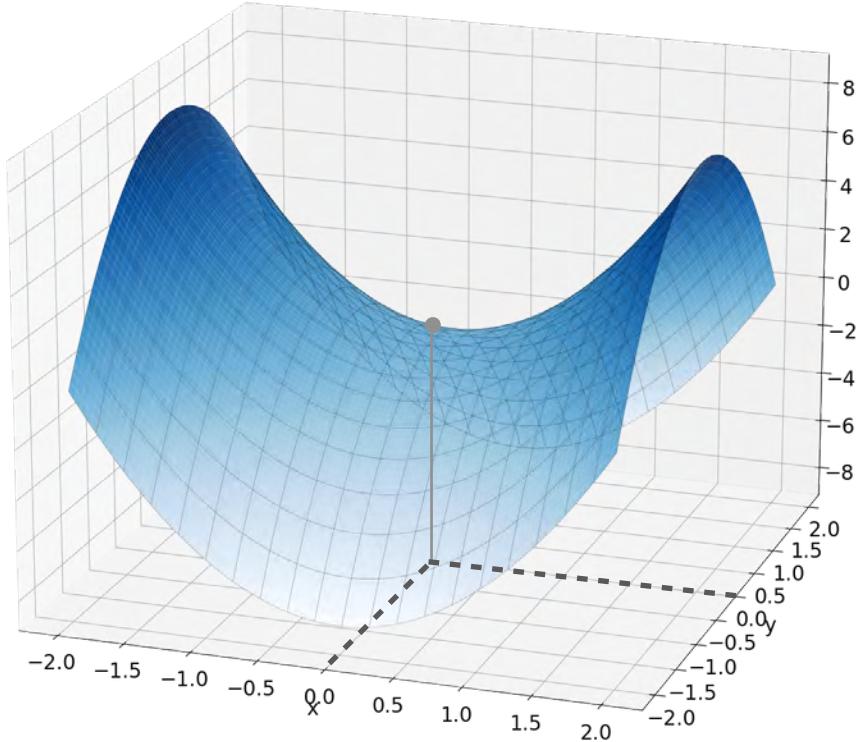
$$f(x, y) = 2x^2 - 2y^2$$

# Saddle Point



$$f(x, y) = 2x^2 - 2y^2$$

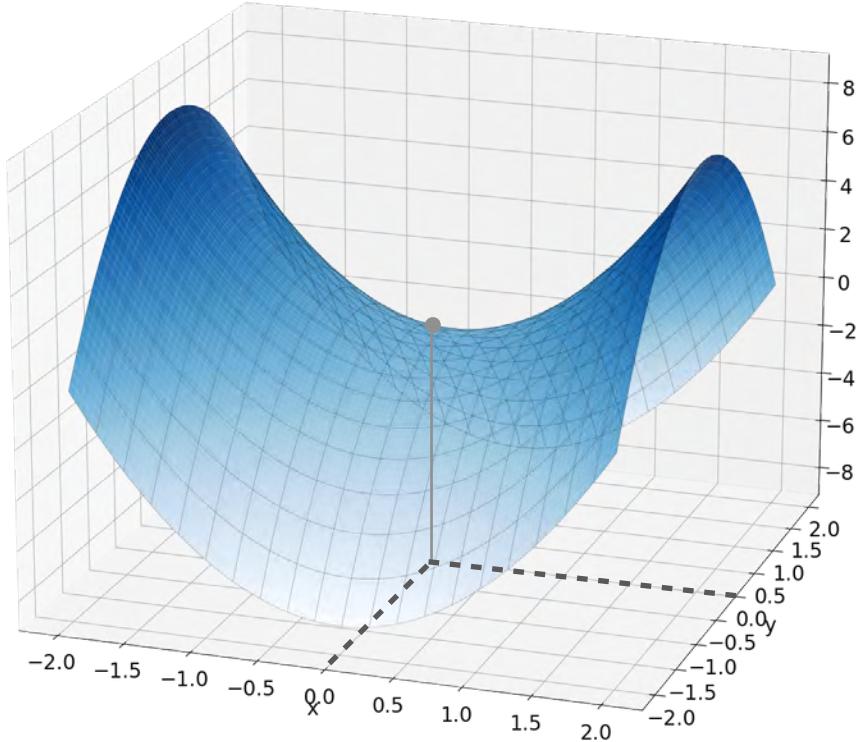
# Saddle Point



$$f(x, y) = 2x^2 - 2y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

# Saddle Point

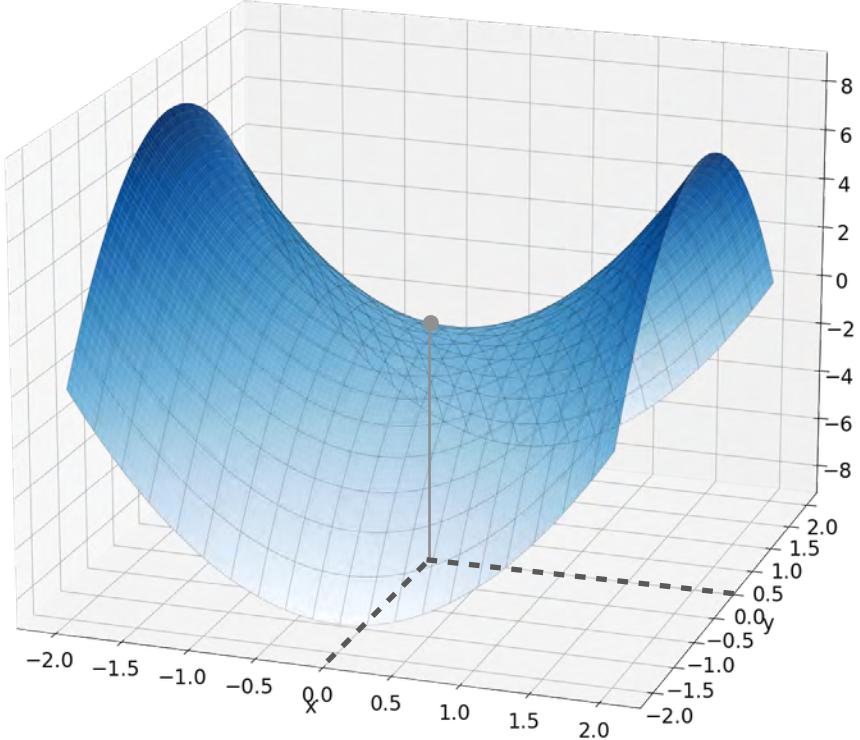


$$f(x, y) = 2x^2 - 2y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

# Saddle Point



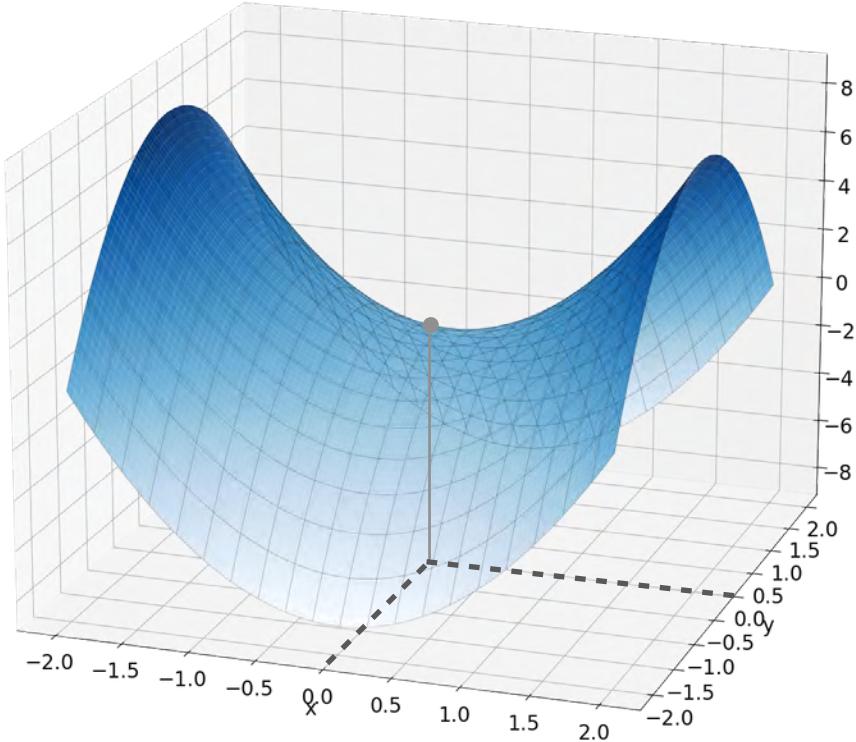
$$f(x, y) = 2x^2 - 2y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

# Saddle Point



$$f(x, y) = 2x^2 - 2y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

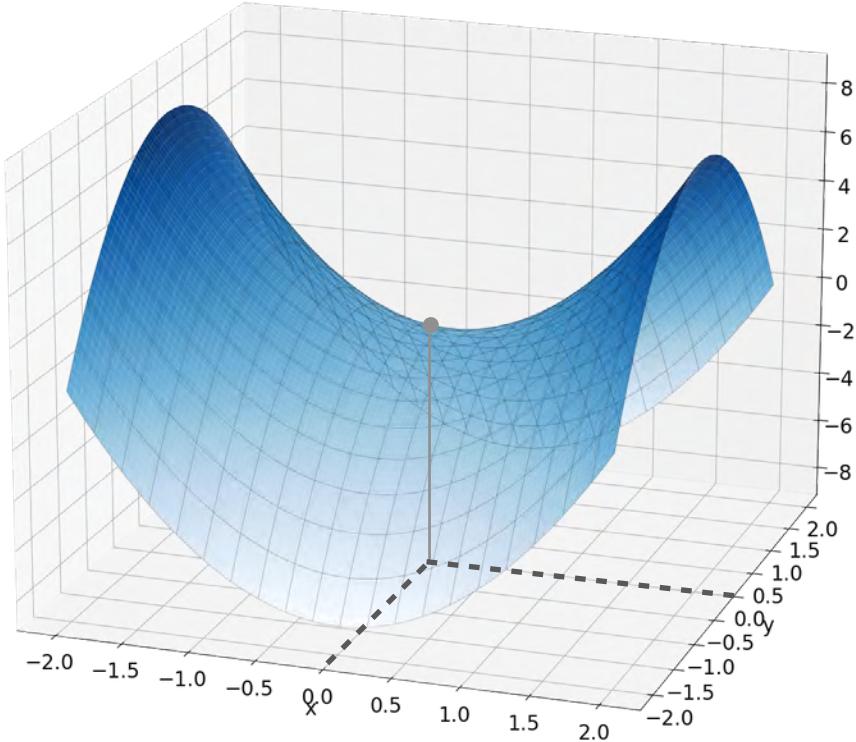
$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(4 - \lambda)(-4 - \lambda) - 0$$

$$\lambda_1 = -4$$

# Saddle Point



$$f(x, y) = 2x^2 - 2y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

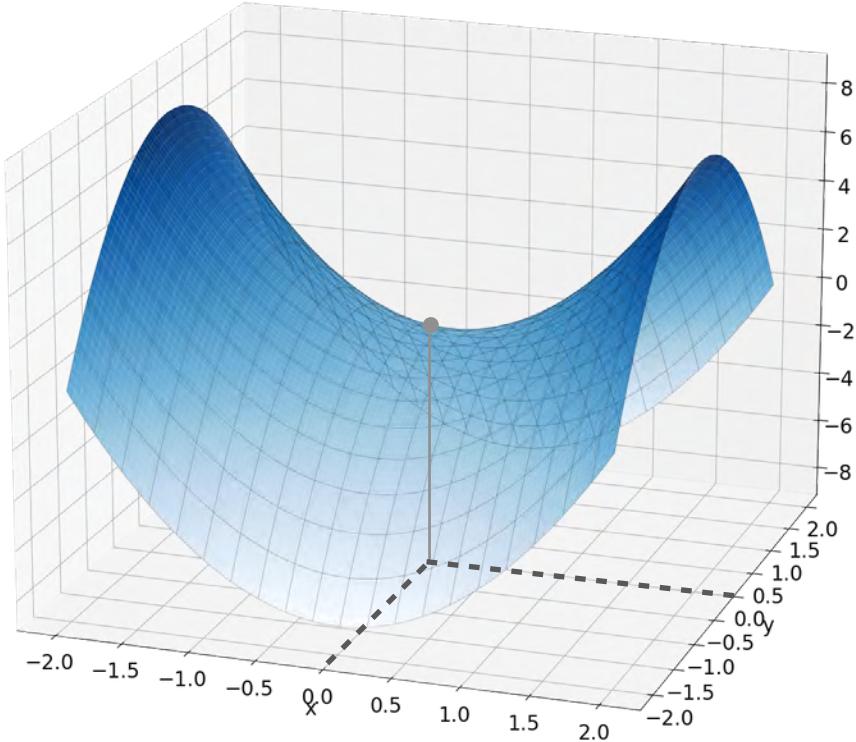
$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(4 - \lambda)(-4 - \lambda) - 0$$

$$\begin{array}{l} \xrightarrow{\hspace{1cm}} \lambda_1 = -4 \\ \xrightarrow{\hspace{1cm}} \lambda_2 = 4 \end{array}$$

# Saddle Point



$$f(x, y) = 2x^2 - 2y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

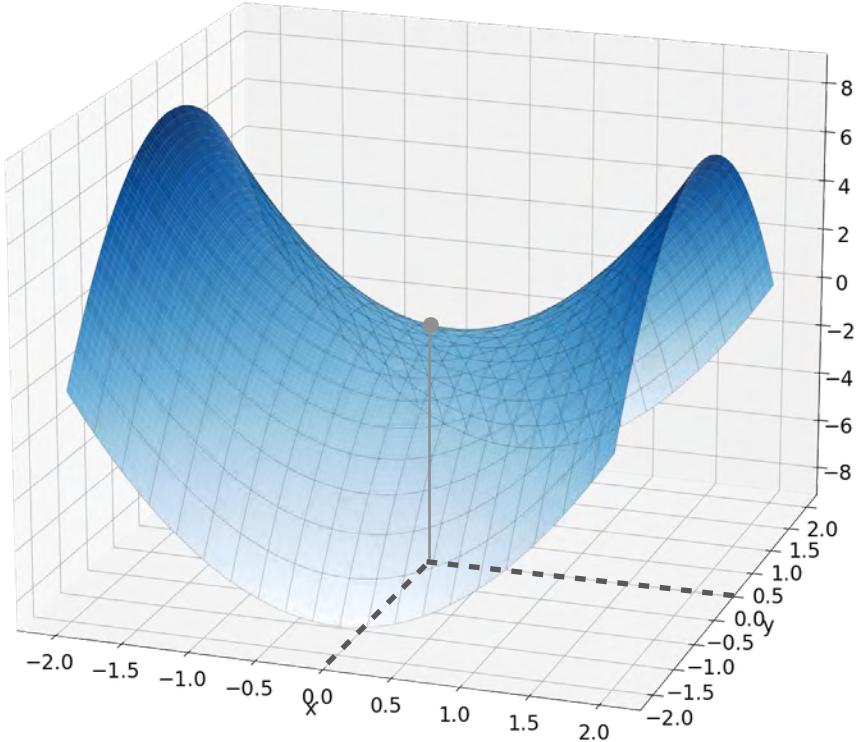
$$\det(H(0,0) - \lambda I) =$$

$$(4 - \lambda)(-4 - \lambda) - 0 < 0$$

$\lambda_1 = -4$   
 $\lambda_2 = 4$

(0,0) is saddle point

# Saddle Point



$$f(x, y) = 2x^2 - 2y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 4x \\ -4y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\det(H(0,0) - \lambda I) =$$

$$(4 - \lambda)(-4 - \lambda) - 0 < 0$$

$\lambda_1 = -4$

$$> 0$$

$\lambda_2 = 4$

(0,0) is saddle point

# Summary

# Summary

1 variable  
 $f(x)$

2 variables  
 $f(x, y)$

More variables  
 $f(x_1, x_2, \dots, x_n)$

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$		

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	All $\lambda_i > 0$

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$		

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Down paraboloid $\lambda_1 < 0 \text{ & } \lambda_2 < 0$	

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Down paraboloid $\lambda_1 < 0 \text{ & } \lambda_2 < 0$	All $\lambda_i < 0$

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Down paraboloid $\lambda_1 < 0 \text{ & } \lambda_2 < 0$	All $\lambda_i < 0$
Need more information	$f''(x) = 0$		

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Down paraboloid $\lambda_1 < 0 \text{ & } \lambda_2 < 0$	All $\lambda_i < 0$
Need more information	$f''(x) = 0$	Saddle point $\lambda_1 > 0 \text{ & } \lambda_2 < 0$ $\lambda_1 < 0 \text{ & } \lambda_2 > 0$	

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Down paraboloid $\lambda_1 < 0 \text{ & } \lambda_2 < 0$	All $\lambda_i < 0$
Need more information	$f''(x) = 0$	Saddle point $\lambda_1 > 0 \text{ & } \lambda_2 < 0$ $\lambda_1 < 0 \text{ & } \lambda_2 > 0$ Or some $\lambda_i = 0$	

# Summary

	1 variable $f(x)$	2 variables $f(x, y)$	More variables $f(x_1, x_2, \dots, x_n)$
(Local) minima	Happy face $f''(x) > 0$	Upper paraboloid $\lambda_1 > 0 \text{ & } \lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Down paraboloid $\lambda_1 < 0 \text{ & } \lambda_2 < 0$	All $\lambda_i < 0$
Need more information	$f''(x) = 0$	Saddle point $\lambda_1 > 0 \text{ & } \lambda_2 < 0$ $\lambda_1 < 0 \text{ & } \lambda_2 > 0$ Or some $\lambda_i = 0$	Some $\lambda_i > 0$ and some $\lambda_j < 0$ OR At least one $\lambda_i = 0$



DeepLearning.AI

# Optimization in Neural Networks and Newton's Method

---

**Newton's method for two  
variables**

# Newton's Method

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

2 variables

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix}$$

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} -$$

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k)$$

# Newton's Method

1 variable

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$$

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k) \nabla f(x_k, y_k)$$

# Newton's Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k) \nabla f(x_k, y_k)$$

# Newton's Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \textcolor{orange}{H^{-1}(x_k, y_k)} \quad \textcolor{teal}{\nabla f(x_k, y_k)}$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \nabla f(x_k, y_k) \quad H^{-1}(x_k, y_k)$$

# Newton's Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k) \nabla f(x_k, y_k)$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \nabla f(x_k, y_k) \cancel{H^{-1}(x_k, y_k)}$$

# Newton's Method

2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \underbrace{\mathbf{H}^{-1}(x_k, y_k)}_{2 \times 2} \underbrace{\nabla f(x_k, y_k)}_{2 \times 1}$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \nabla f(x_k, y_k) \cancel{- \mathbf{H}^{-1}(x_k, y_k)}$$

# Newton's Method

2 variables

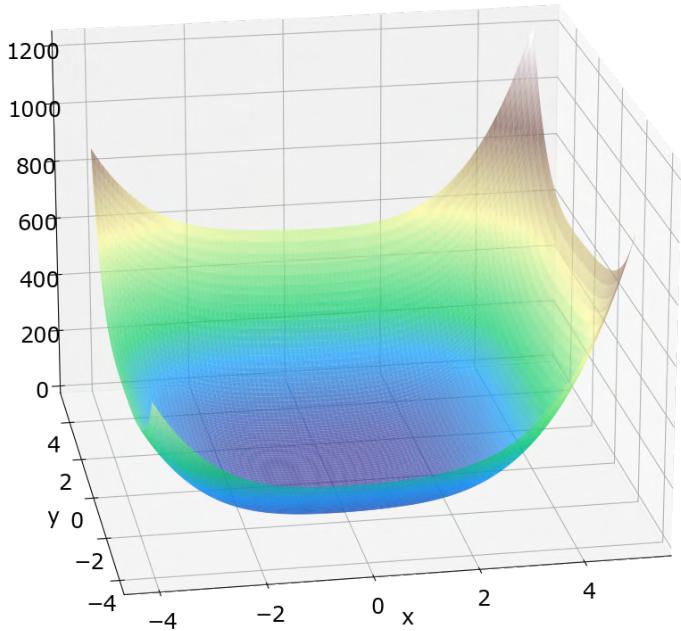
$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \underbrace{\mathbf{H}^{-1}(x_k, y_k)}_{2 \times 2} \underbrace{\nabla f(x_k, y_k)}_{2 \times 1}$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \nabla f(x_k, y_k) \cancel{- H^{-1}(x_k, y_k)}$$

When working with 2 variables the order is crucial!

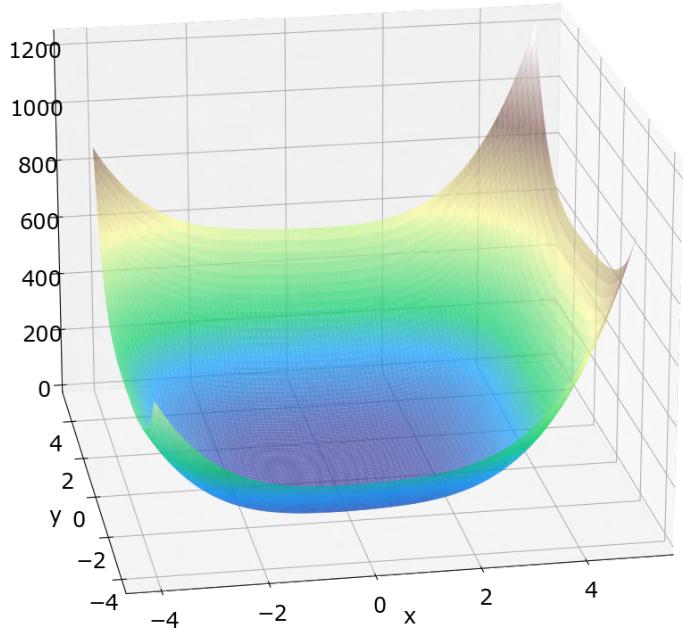
# An Example

# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

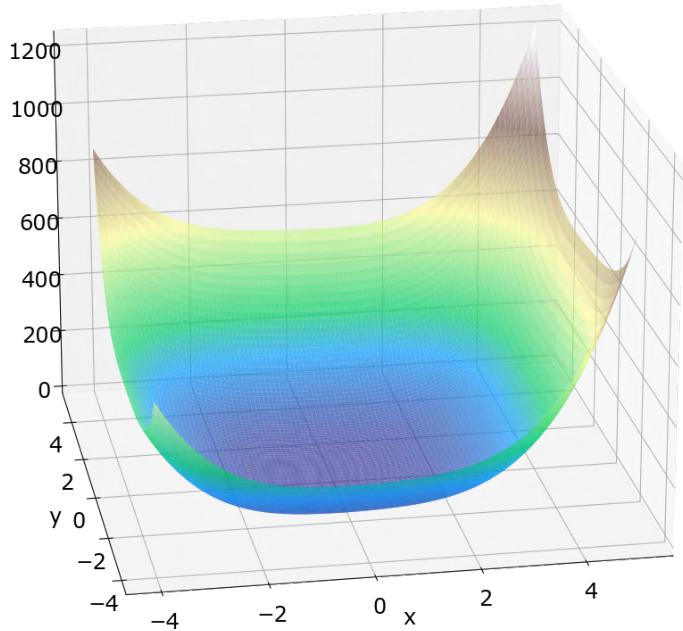
# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$f(x, y) \rightarrow 4x^3 + 8x - y - 0.4xy$$

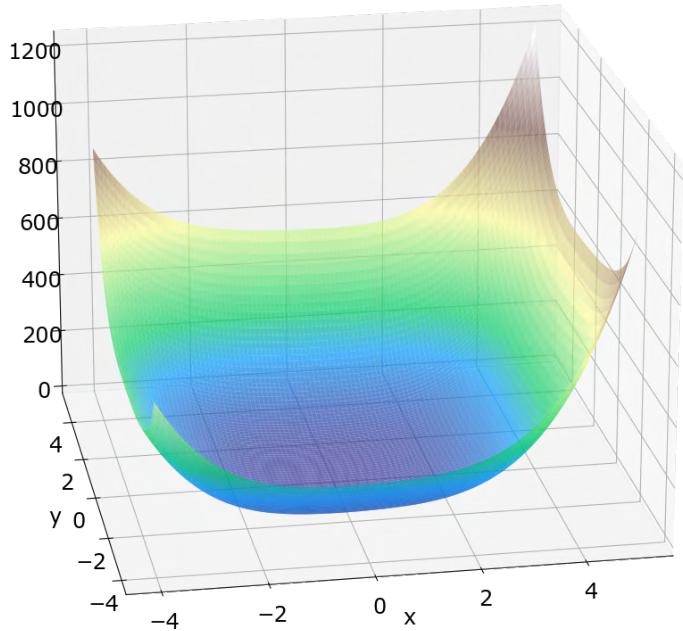
# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\begin{array}{l} \textcolor{blue}{x} \nearrow 4x^3 + 8x - y - 0.4xy \\ f(x, y) \\ \textcolor{orange}{y} \searrow 3.2y^3 + 4y - x - 0.2x^2 \end{array}$$

# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

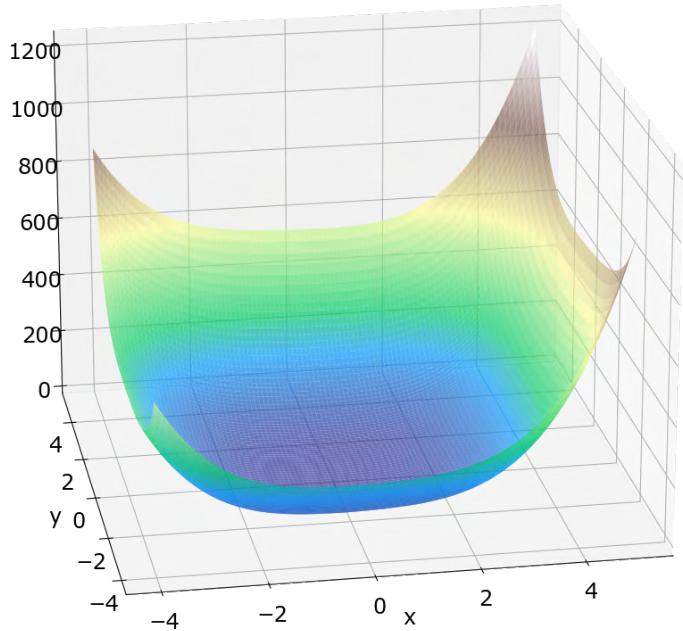
$f(x, y)$

$x$   $12x^2 + 8 - 0.4y$

$4x^3 + 8x - y - 0.4xy$

$y$   $3.2y^3 + 4y - x - 0.2x^2$

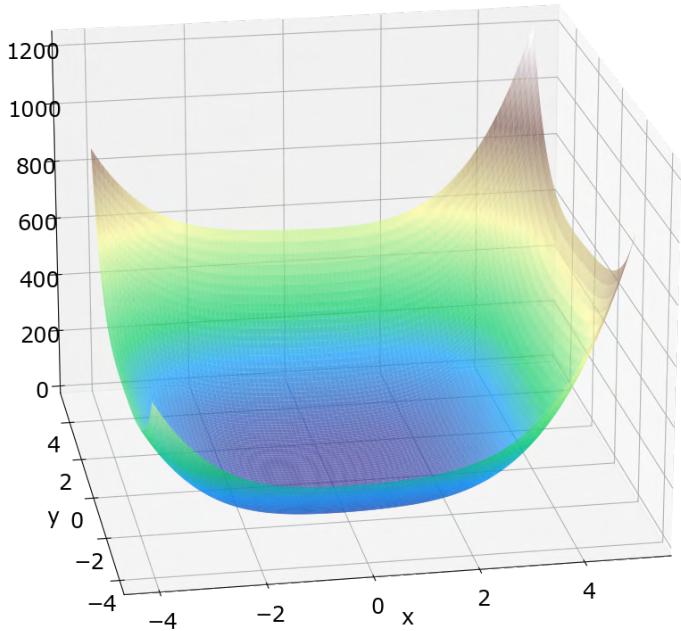
# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\begin{aligned} f(x, y) & \xrightarrow{x} 4x^3 + 8x - y - 0.4xy \\ & \xrightarrow{y} 3.2y^3 + 4y - x - 0.2x^2 \\ & \quad \quad \quad \xrightarrow{x} 12x^2 + 8 - 0.4y \\ & \quad \quad \quad \xrightarrow{y} -1 - 0.4x \end{aligned}$$

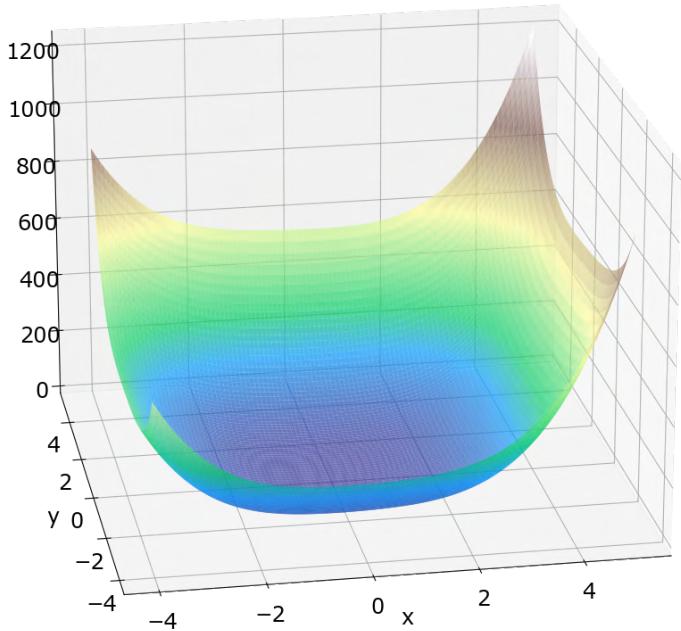
# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$f(x, y) \begin{matrix} \nearrow x \\ \searrow y \end{matrix} \begin{array}{l} 4x^3 + 8x - y - 0.4xy \\ 3.2y^3 + 4y - x - 0.2x^2 \end{array} \begin{matrix} \nearrow x \\ \searrow y \end{matrix} \begin{array}{l} 12x^2 + 8 - 0.4y \\ -1 - 0.4x \\ -1 - 0.4x \end{array}$$

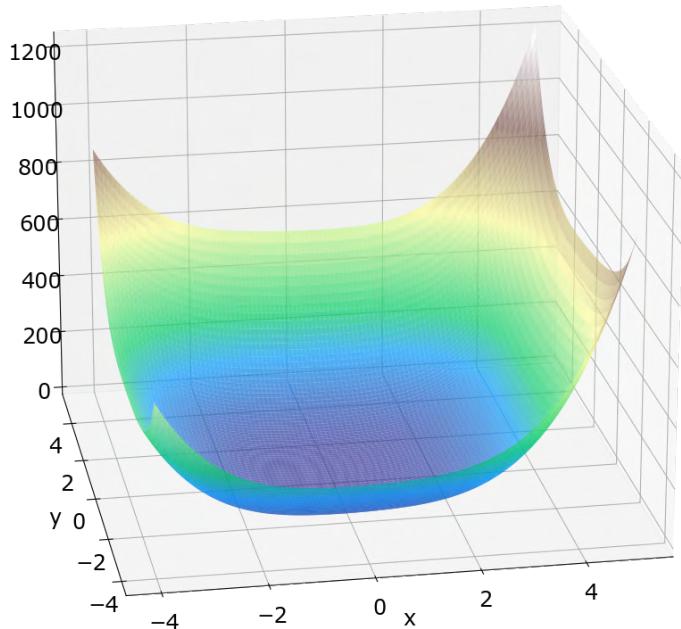
# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

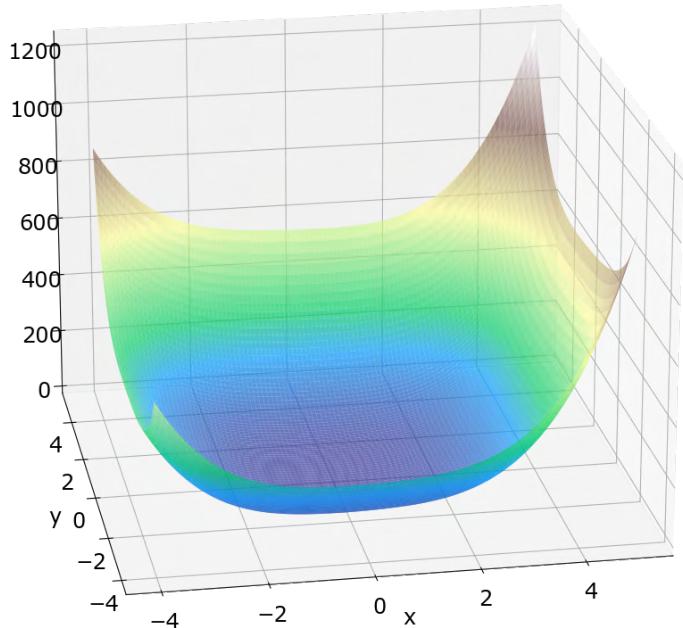
$$\begin{aligned} f(x, y) &\quad \begin{array}{l} \nearrow x \\ \searrow y \end{array} \\ &= 4x^3 + 8x - y - 0.4xy && \begin{array}{l} \nearrow x \\ \searrow y \end{array} & 12x^2 + 8 - 0.4y \\ &\quad \begin{array}{l} \nearrow x \\ \searrow y \end{array} && & -1 - 0.4x \\ &= 3.2y^3 + 4y - x - 0.2x^2 && \begin{array}{l} \nearrow x \\ \searrow y \end{array} & -1 - 0.4x \\ &\quad \begin{array}{l} \nearrow x \\ \searrow y \end{array} && & 9.6y^2 + 4 \end{aligned}$$

# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

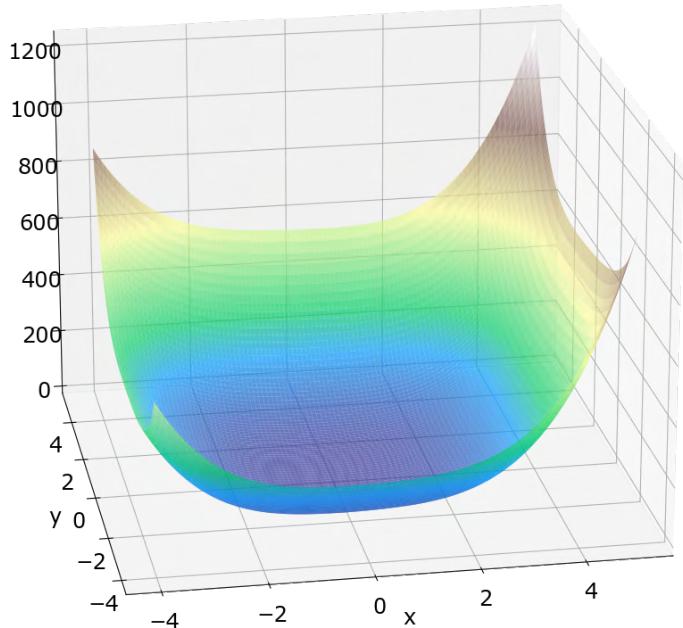
# An Example



$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

$$\nabla f(x, y) = \begin{bmatrix} 4x^3 + 8x - y - 0.4xy \\ 3.2y^3 + 4y - x - 0.2x^2 \end{bmatrix}$$

# An Example

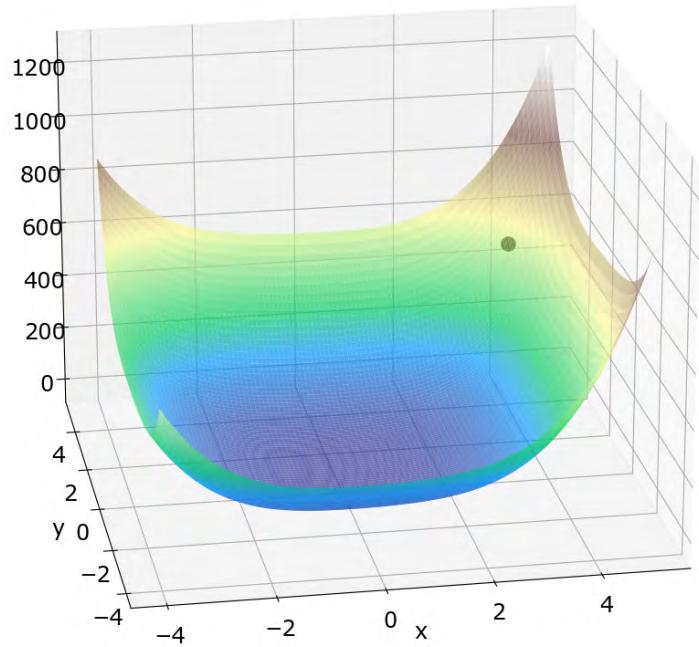


$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

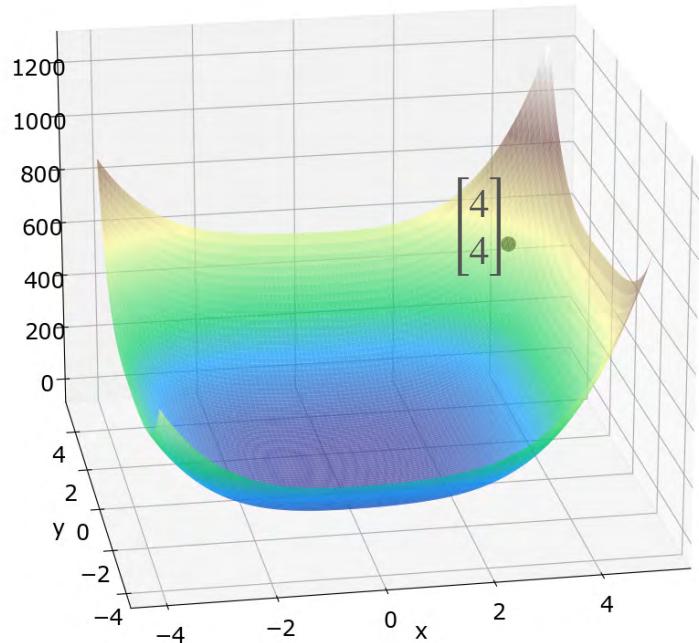
$$\nabla f(x, y) = \begin{bmatrix} 4x^3 + 8x - y - 0.4xy \\ 3.2y^3 + 4y - x - 0.2x^2 \end{bmatrix}$$

$$H(x, y) = \begin{bmatrix} 12x^2 + 8 - 0.4y & -1 - 0.4x \\ -1 - 0.4x & 9.6y^2 + 4 \end{bmatrix}$$

# An Example

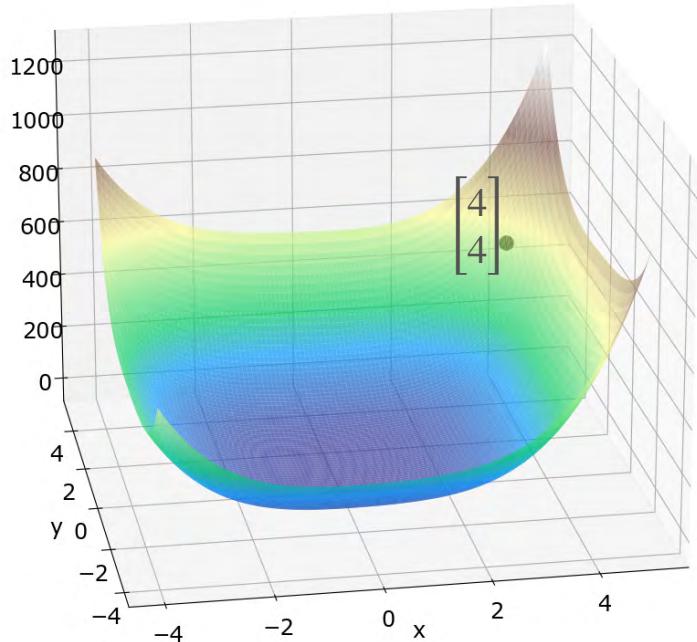


# An Example



Start at some point  $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

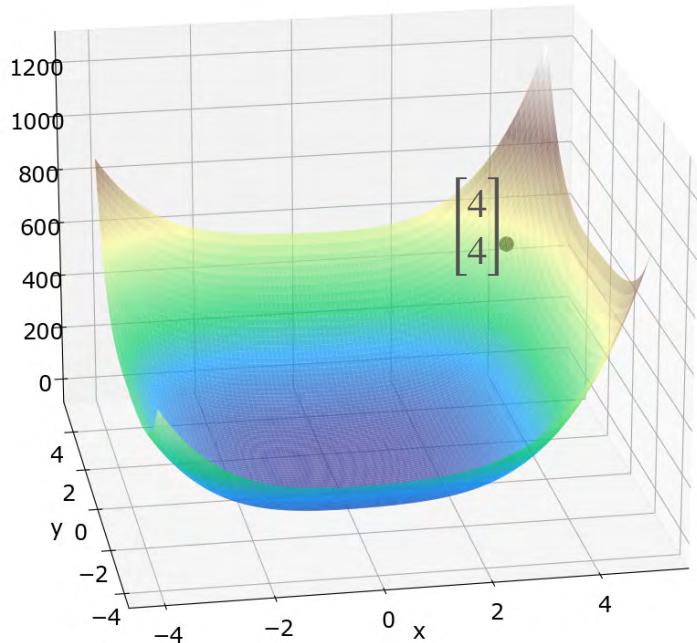
# An Example



Start at some point  $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$$\nabla f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$$

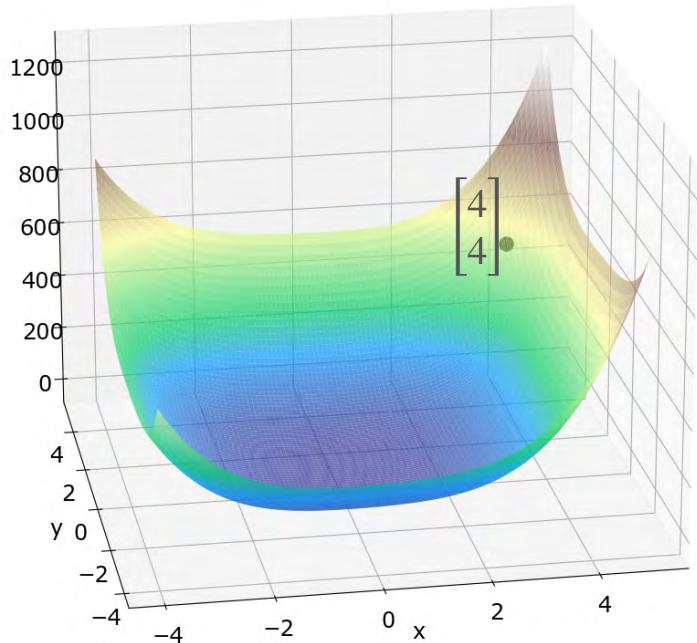
# An Example



Start at some point  $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$$\nabla f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix} \quad H(4,4) = \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}$$

# An Example

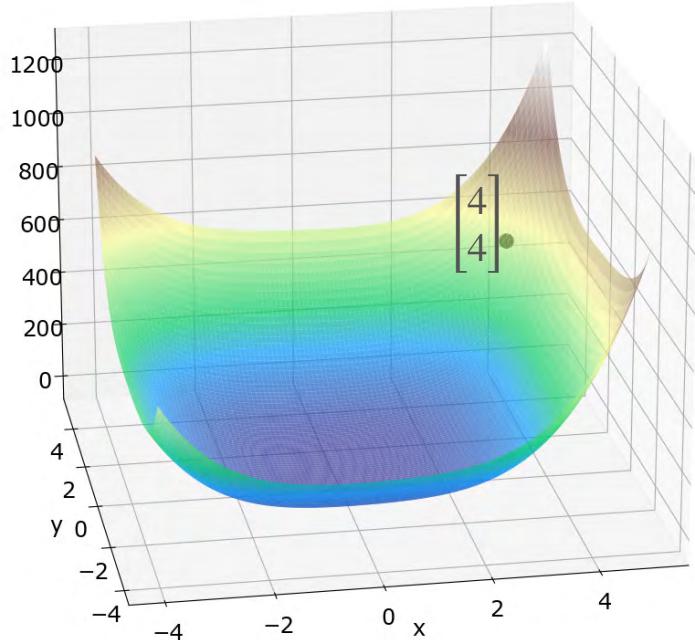


Start at some point  $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$$\nabla f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix} \quad H(4,4) = \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} -$$

# An Example

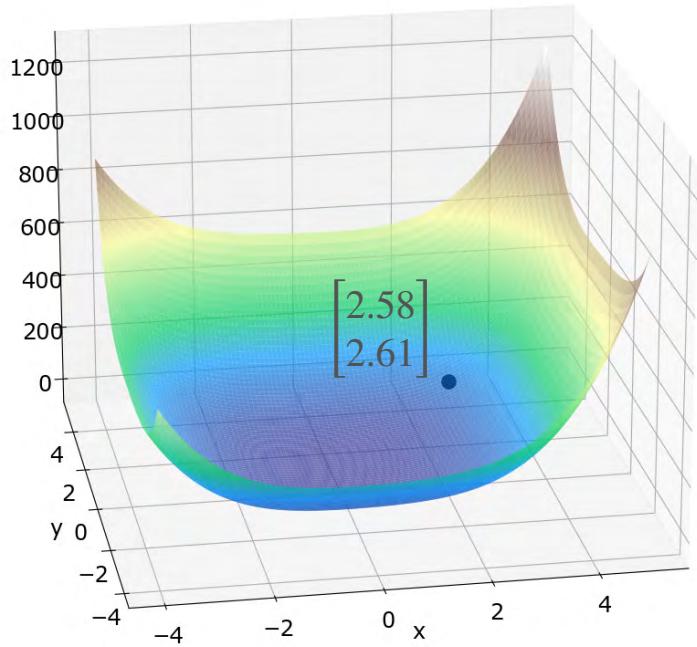


Start at some point  $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$$\nabla f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix} \quad H(4,4) = \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}^{-1} \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$$

# An Example



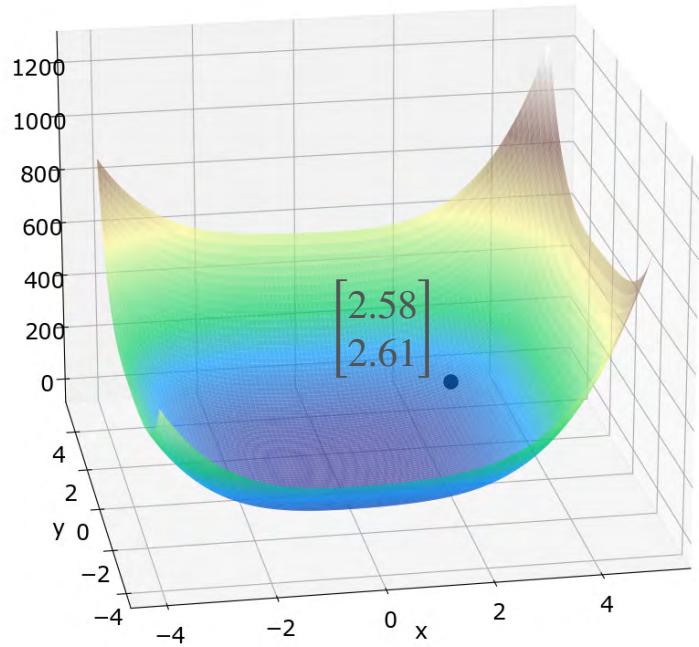
Start at some point  $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$$\nabla f(4,4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix} \quad H(4,4) = \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}^{-1} \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$$

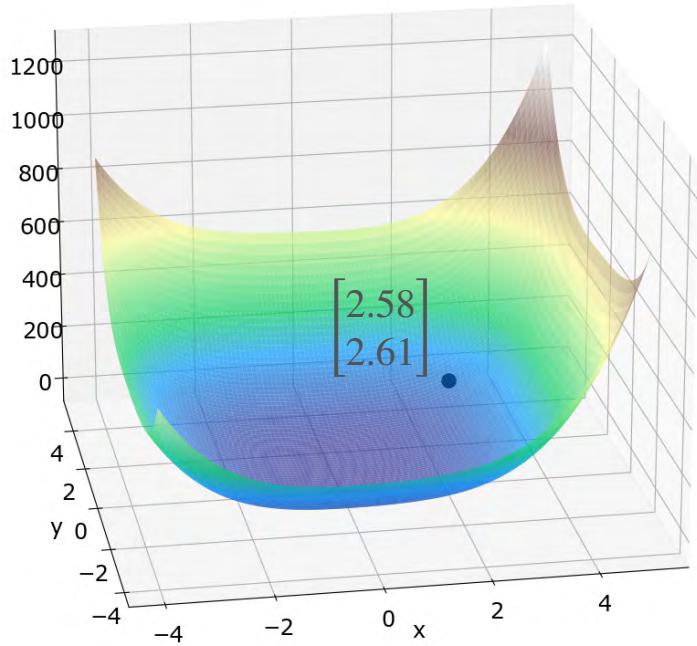
$$= \begin{bmatrix} 2.58 \\ 2.62 \end{bmatrix}$$

# An Example



$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

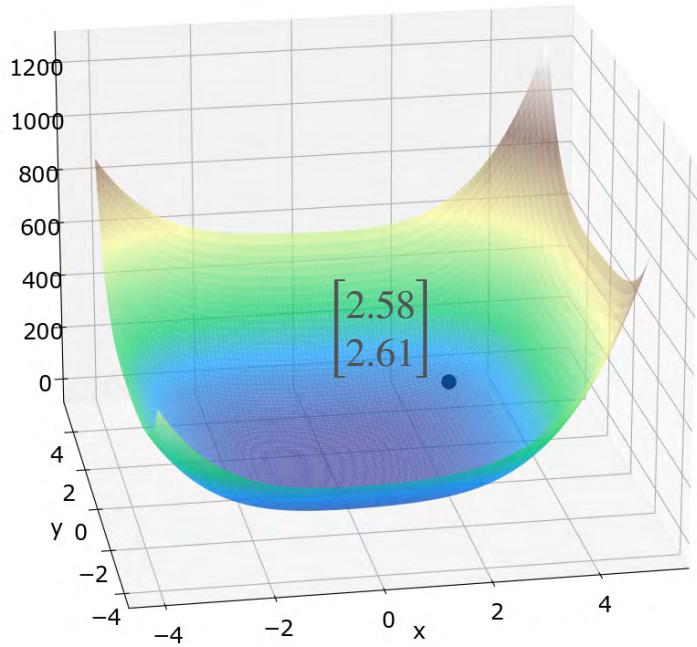
# An Example



$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\nabla f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

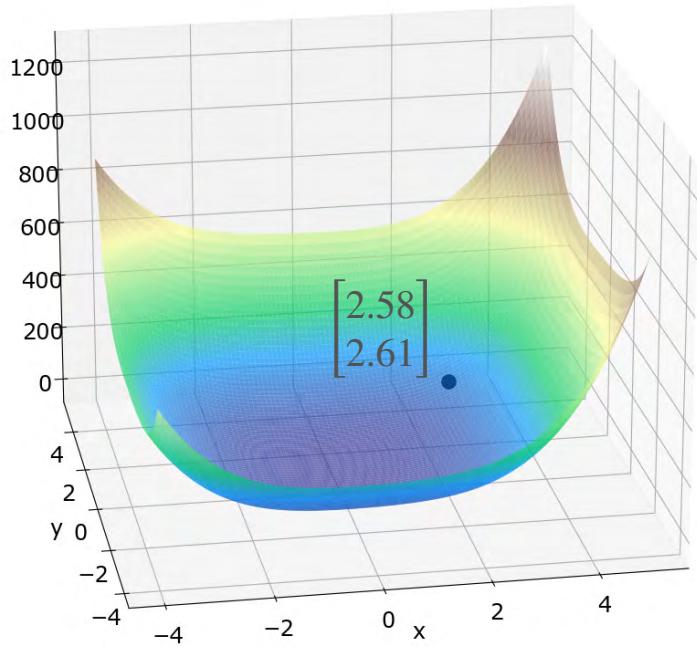
# An Example



$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\nabla f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} \quad H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

# An Example

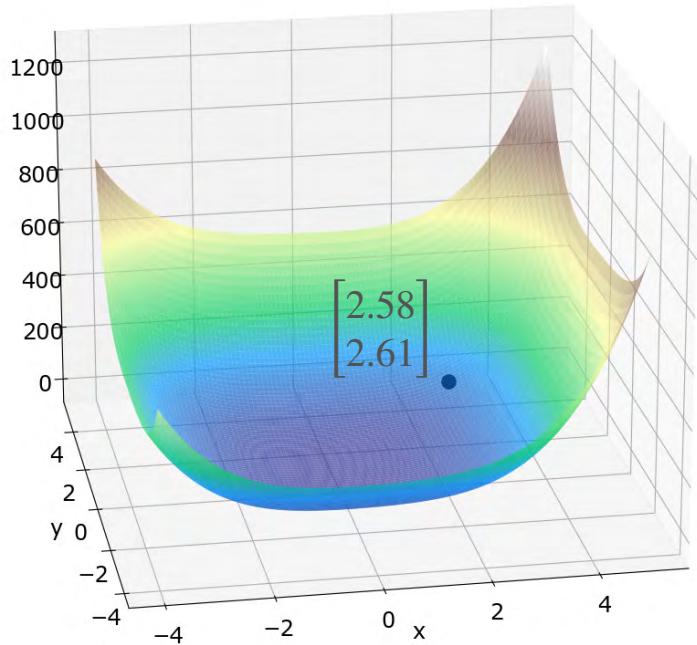


$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\nabla f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} \quad H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix} -$$

# An Example

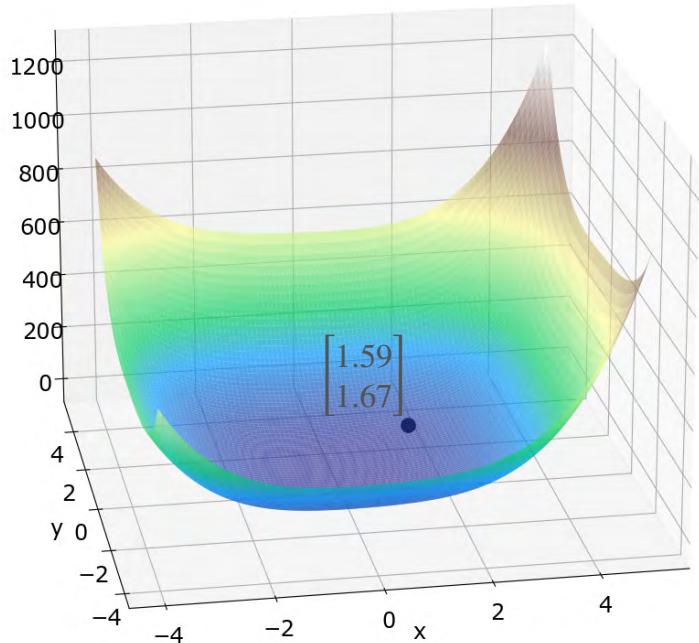


$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\nabla f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix} - \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}^{-1} \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

# An Example



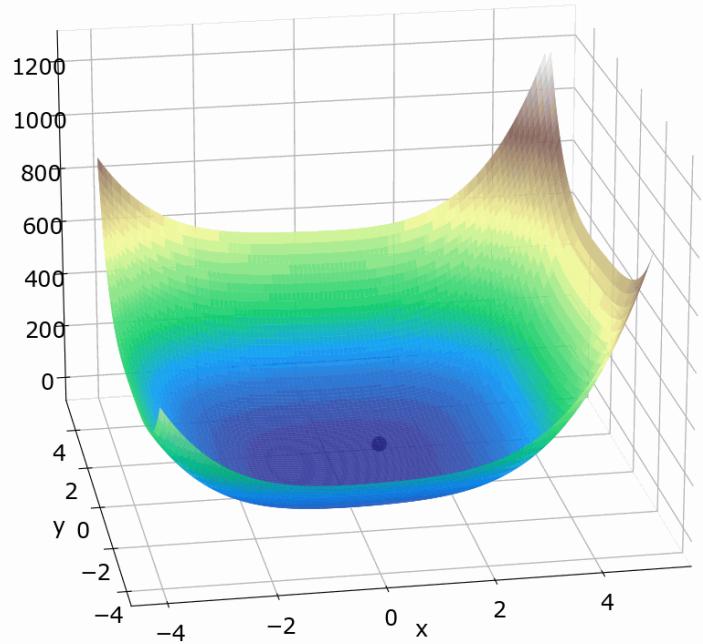
$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\nabla f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix} H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

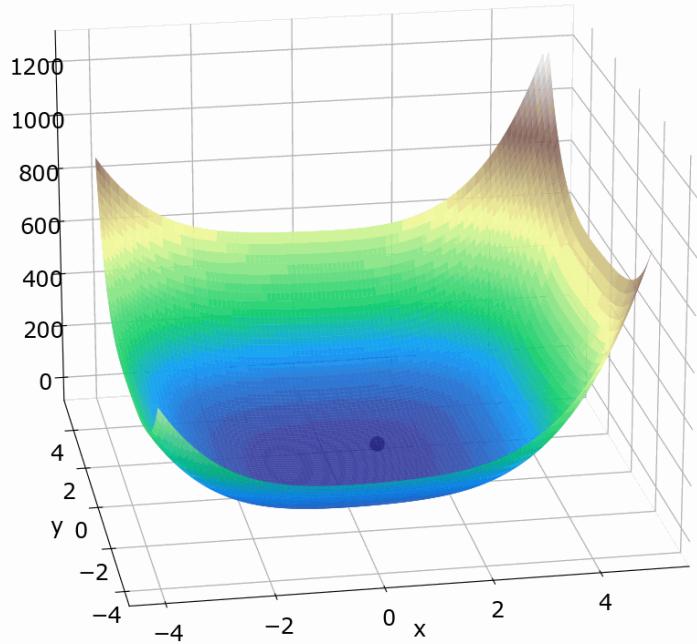
$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix} - \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}^{-1} \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

$$= \begin{bmatrix} 1.59 \\ 1.67 \end{bmatrix}$$

# An Example

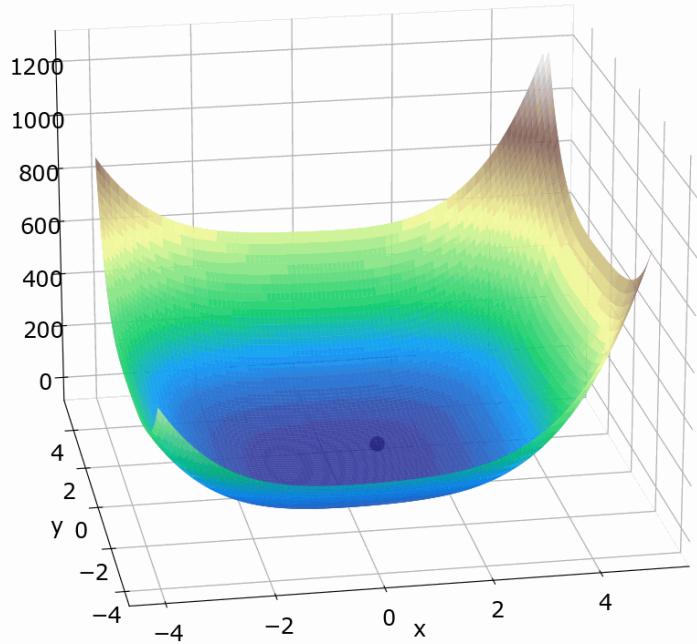


# An Example



Repeat until you are close enough to the actual zero!

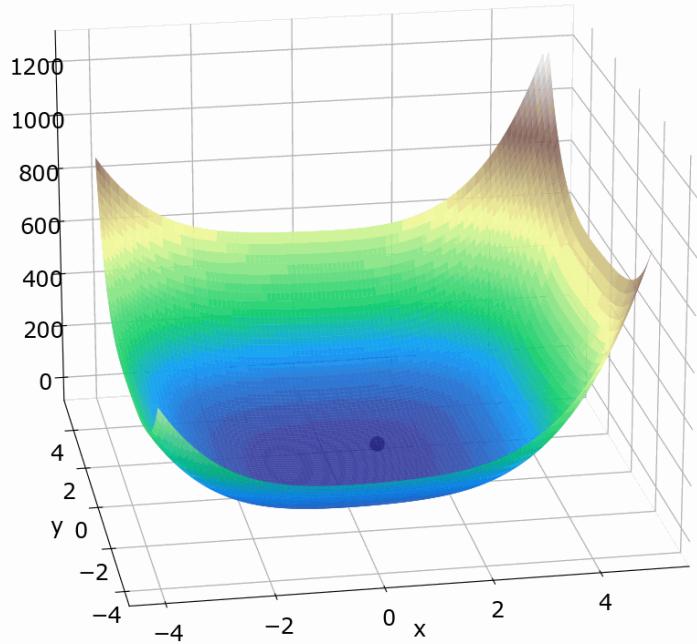
# An Example



Repeat until you are close enough to the actual zero!

Needed  $k = 8$  steps

# An Example

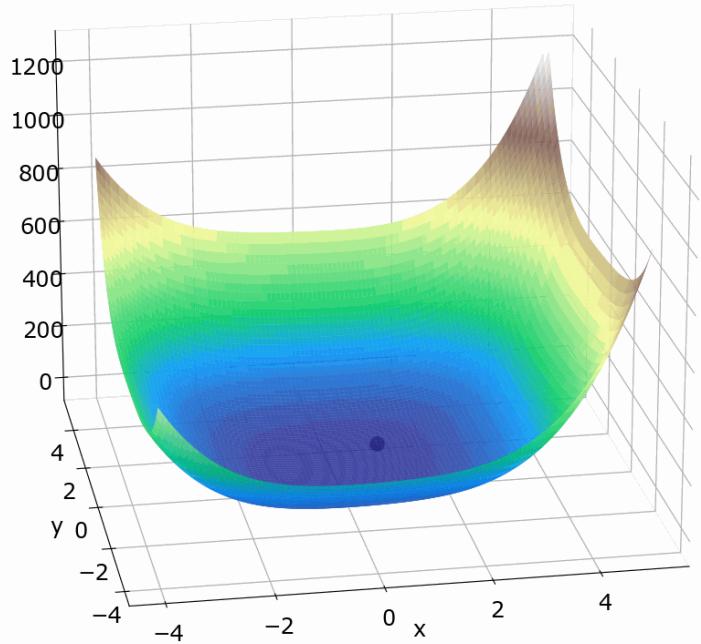


Repeat until you are close enough to the actual zero!

Needed  $k = 8$  steps

$$\begin{bmatrix} x_8 \\ y_8 \end{bmatrix} = \begin{bmatrix} 4.15 \cdot 10^{-17} \\ -2.05 \cdot 10^{-17} \end{bmatrix}$$

# An Example



Repeat until you are close enough to the actual zero!

Needed  $k = 8$  steps

$$\begin{bmatrix} x_8 \\ y_8 \end{bmatrix} = \begin{bmatrix} 4.15 \cdot 10^{-17} \\ -2.05 \cdot 10^{-17} \end{bmatrix}$$

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$



DeepLearning.AI

# Optimization in Neural Networks and Newton's Method

---

## Conclusion