



DeepLearning.AI

# Gradients and Gradient Descent

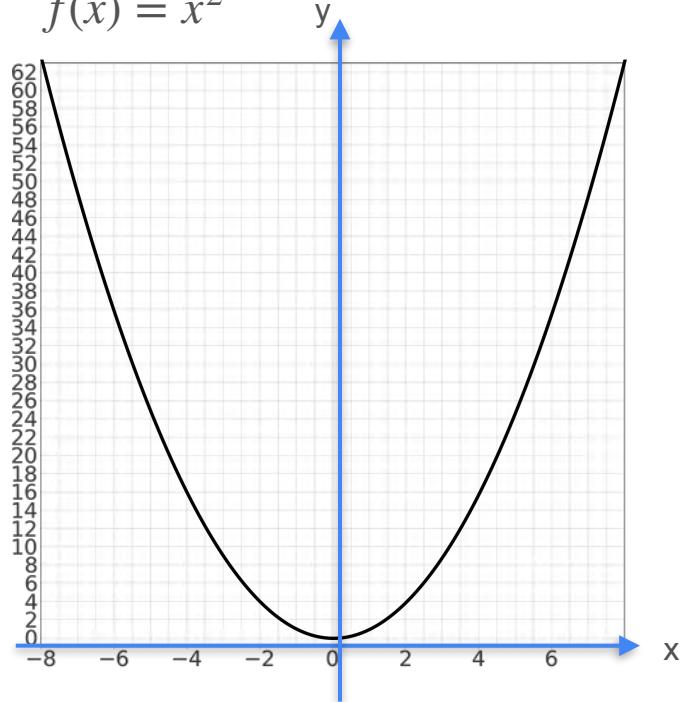
---

## Tangent planes

# Functions of Two Variables

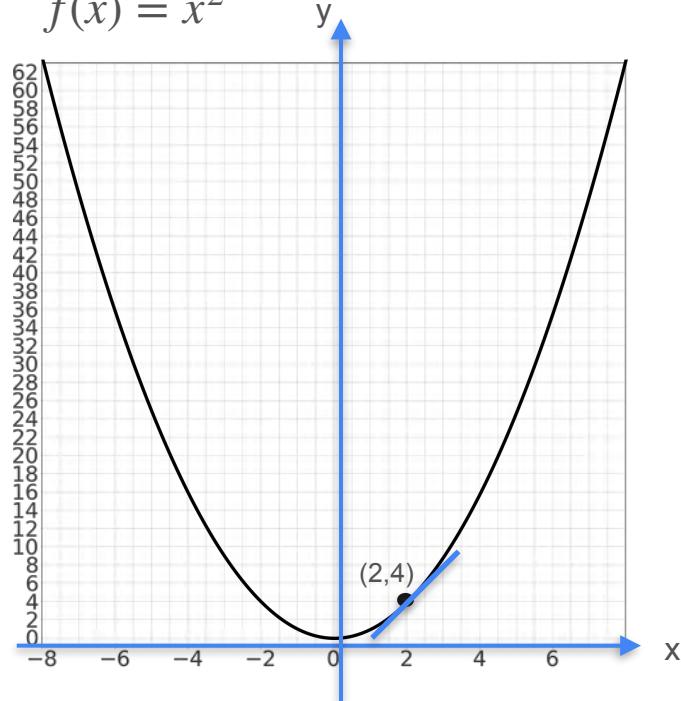
# Functions of Two Variables

$$f(x) = x^2$$



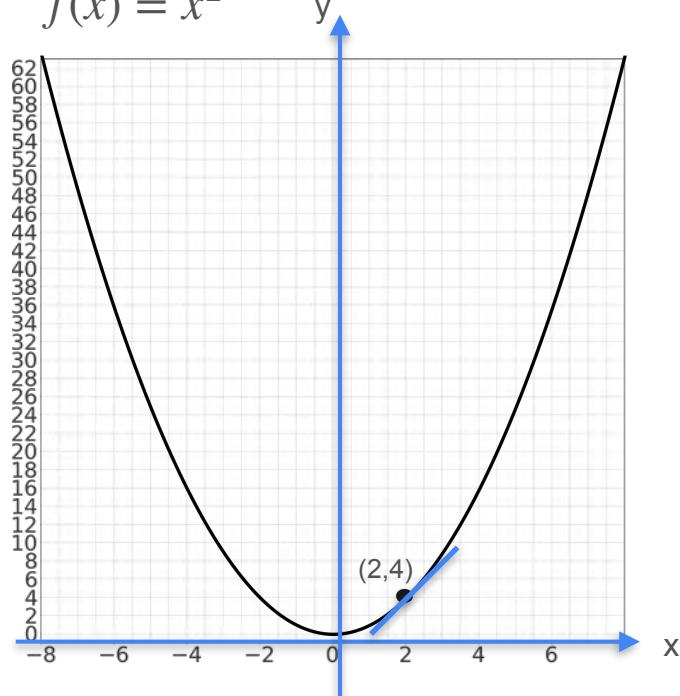
# Functions of Two Variables

$$f(x) = x^2$$

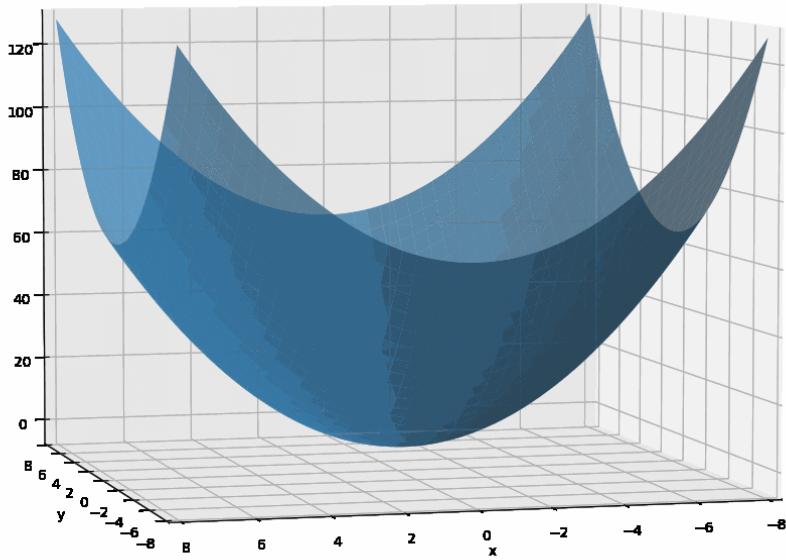


# Functions of Two Variables

$$f(x) = x^2$$

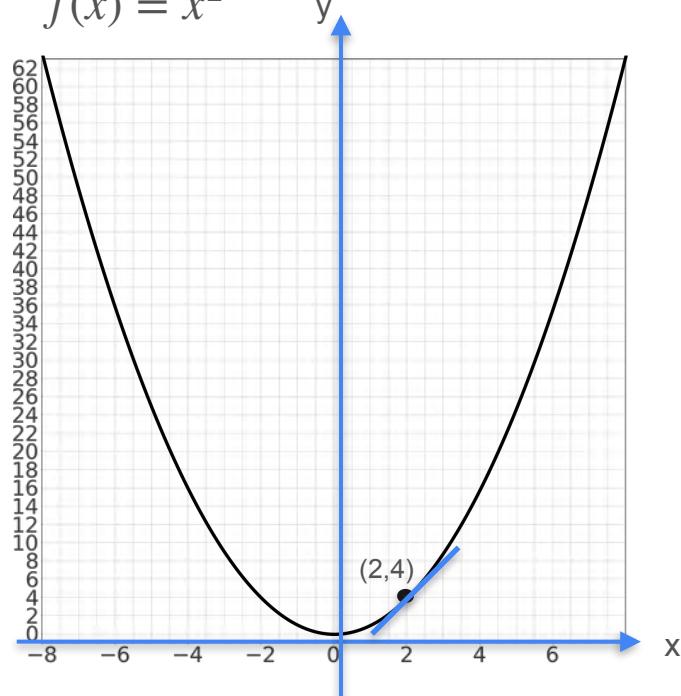


$$f(x, y) = x^2 + y^2$$

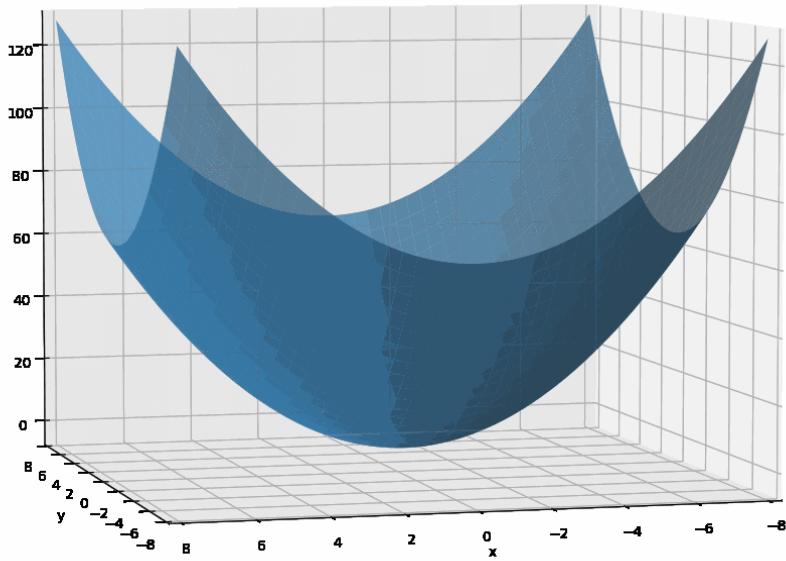


# Functions of Two Variables

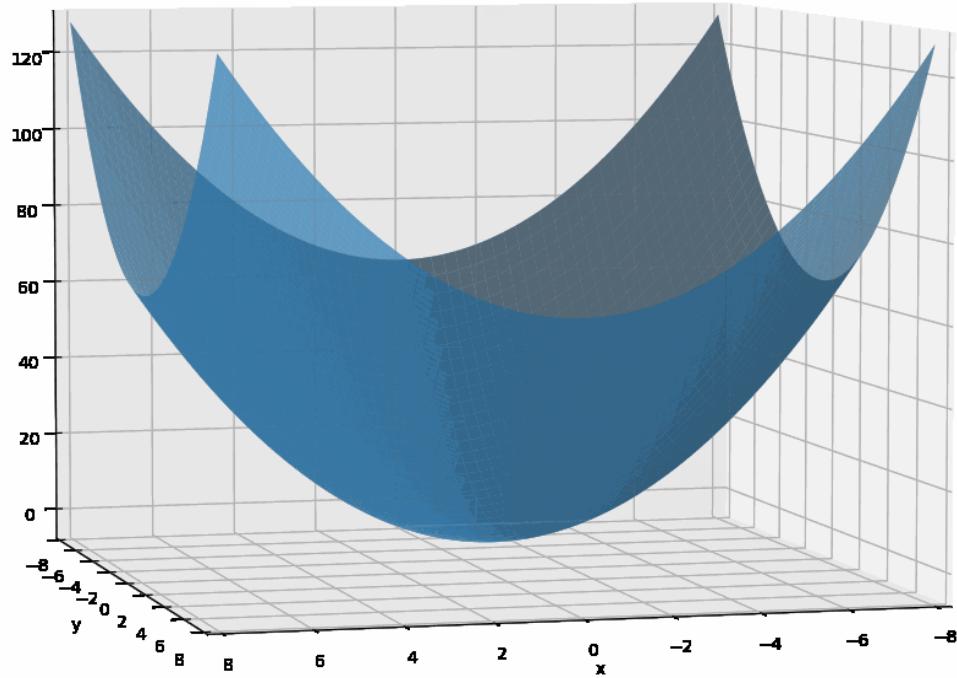
$$f(x) = x^2$$



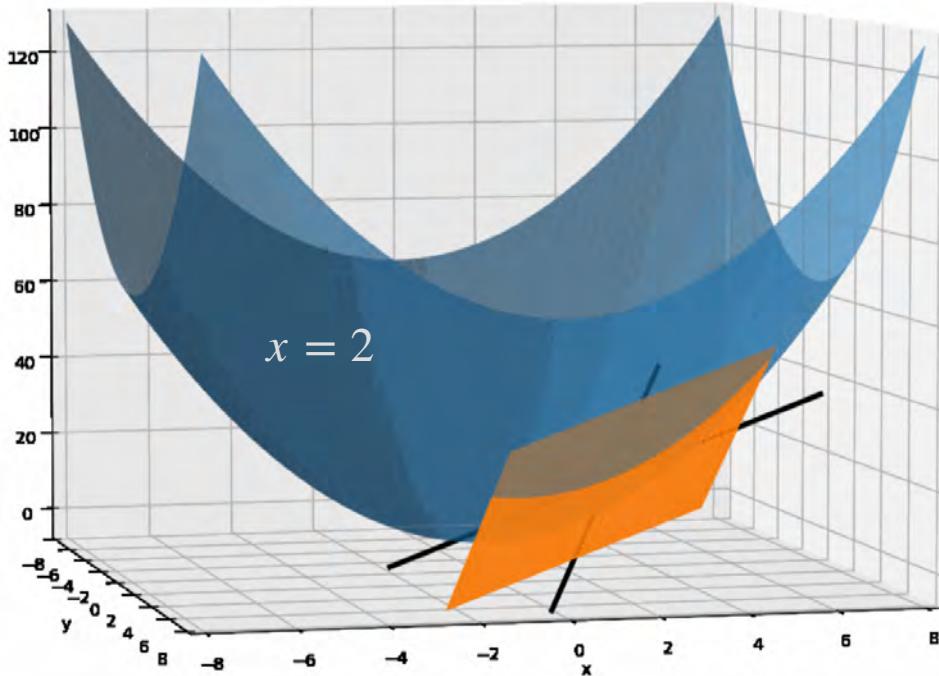
$$f(x, y) = x^2 + y^2$$



# Finding the Tangent Plane



# Finding the Tangent Plane



Fix  $y=4$   $f(x,4) = x^2 + 4^2$

$$\frac{d}{dx} (f(x,4)) = 2x$$

Fix  $x=2$   $f(2,y) = 2^2 + y^2$

$$\frac{d}{dy} (f(2,y)) = 2y$$

The tangent plane contains both tangent lines.

# Video 2: Introduction to Partial Derivatives

Example with the parabola, show tangent plane and slices



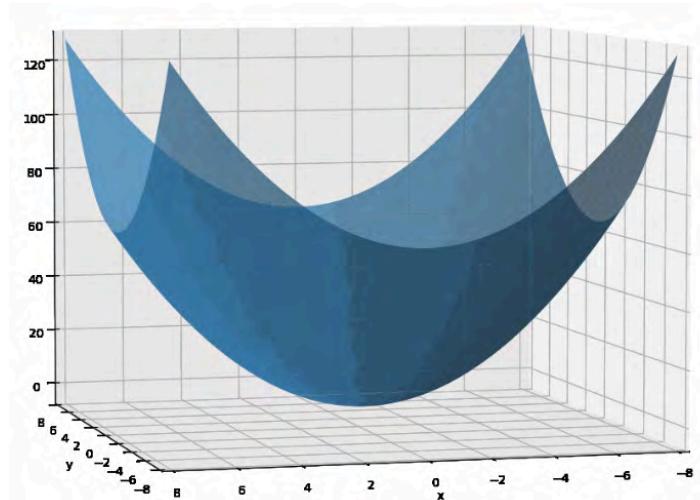
DeepLearning.AI

# Gradients and Gradient Descent

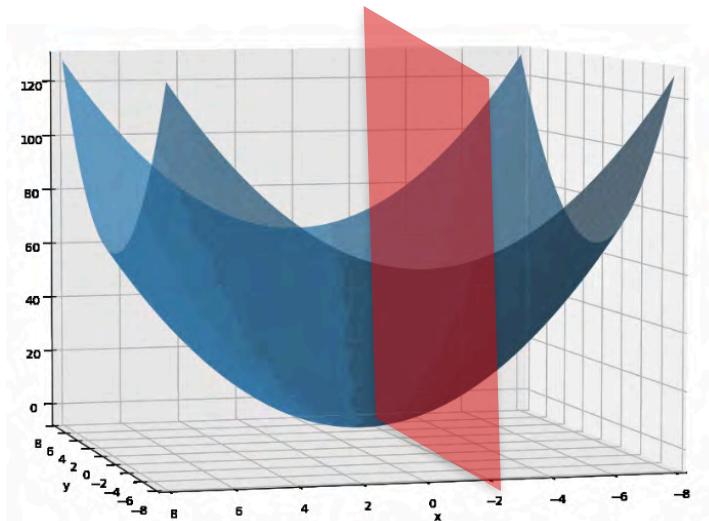
---

## Partial derivatives - Part 1

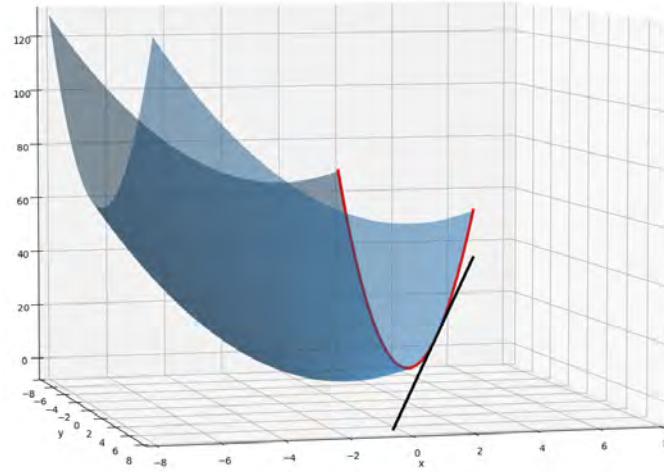
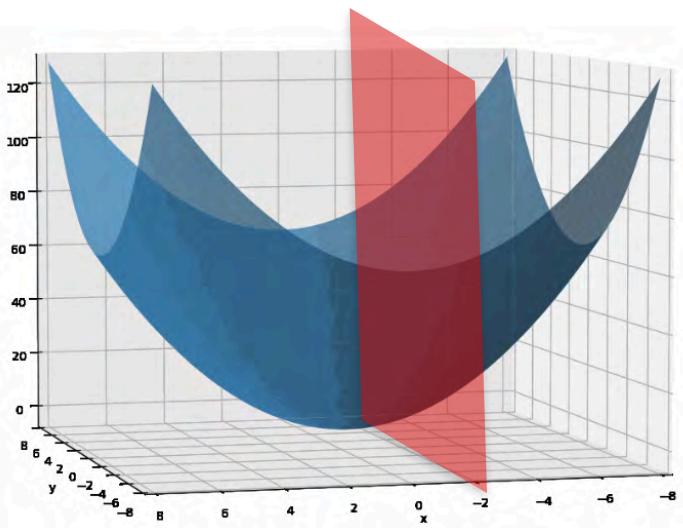
# Slicing the Space



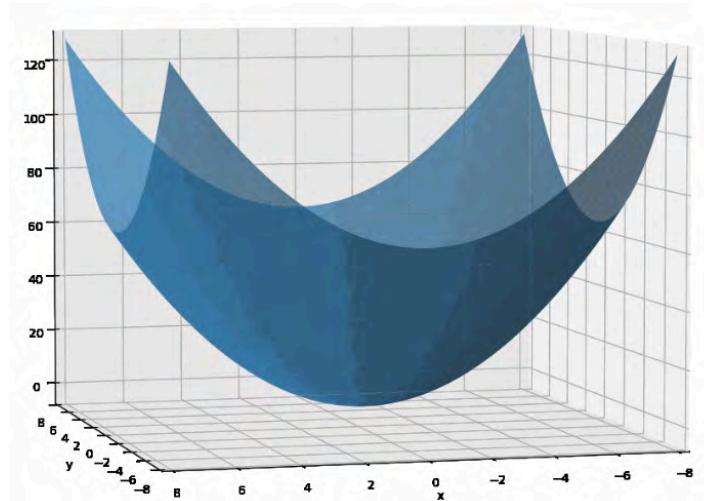
# Slicing the Space



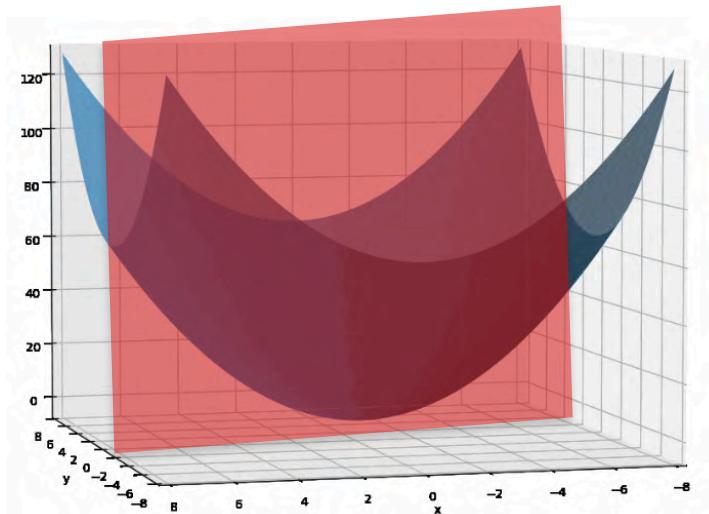
# Slicing the Space



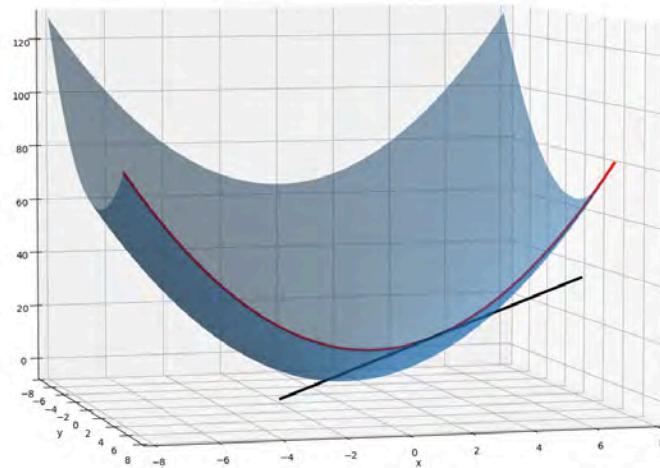
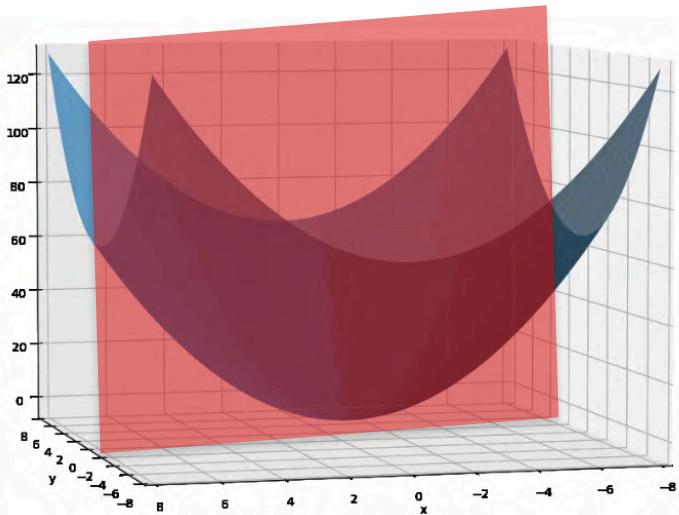
# Slicing the Space



# Slicing the Space

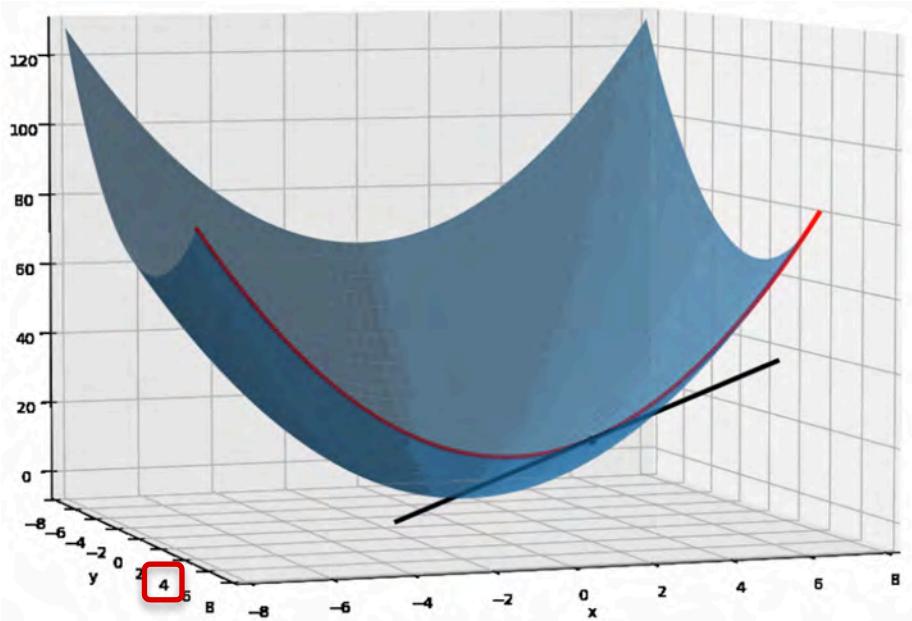


# Slicing the Space



# Partial Derivatives

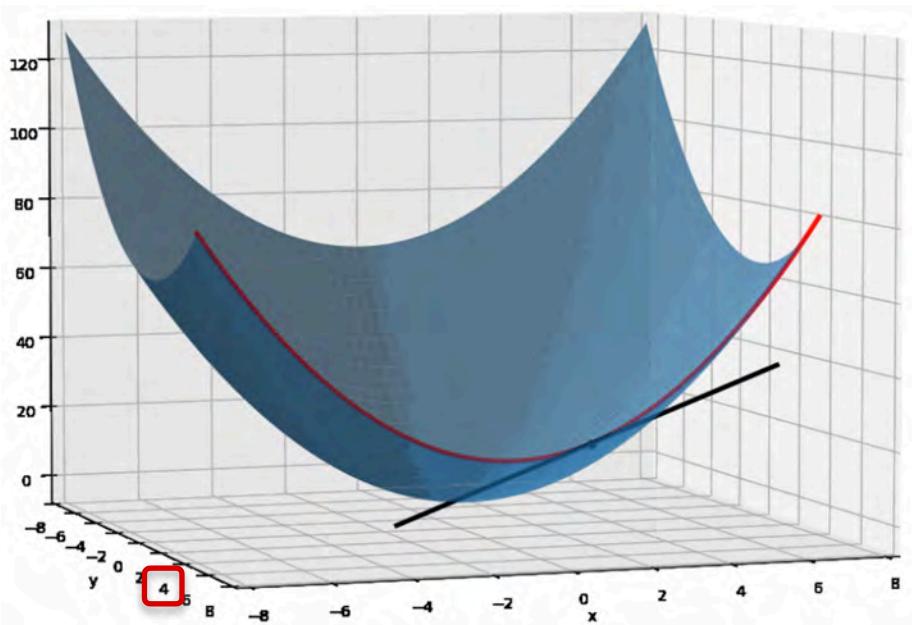
$$f(x, y) = x^2 + y^2$$



# Partial Derivatives

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

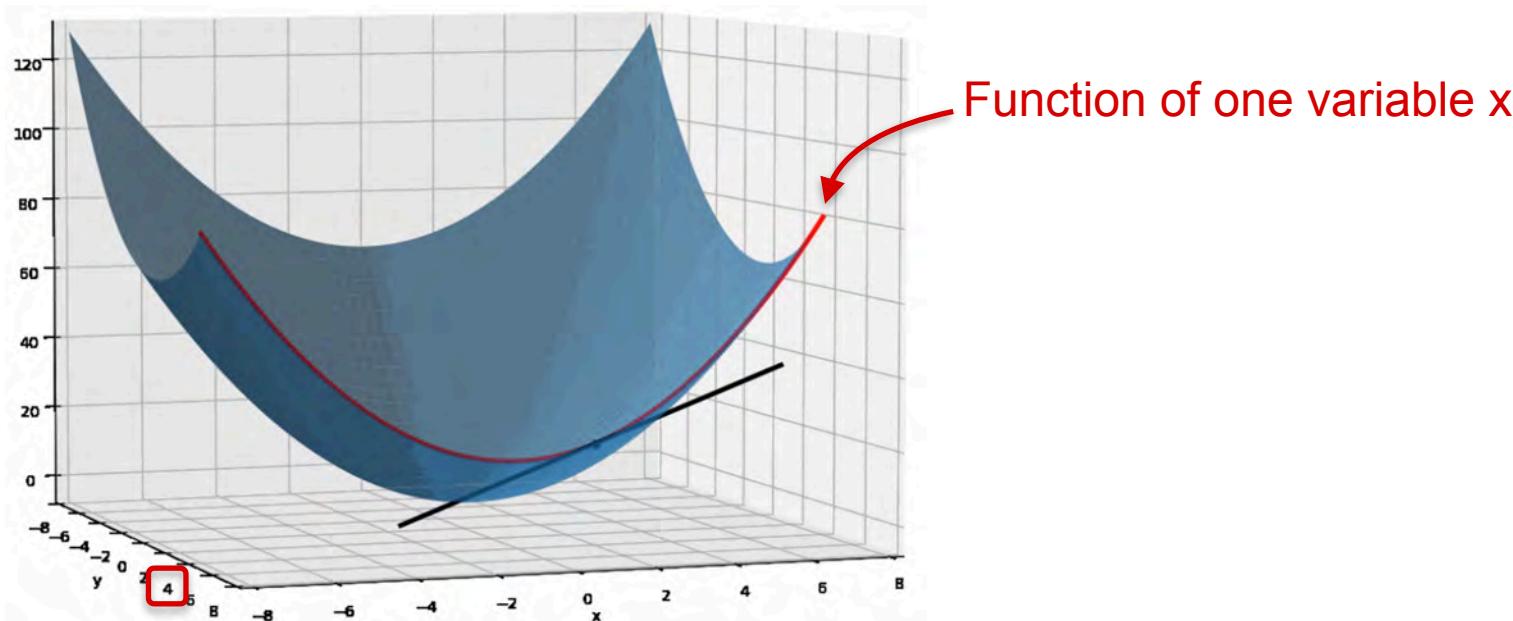
Treat y as a constant



# Partial Derivatives

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

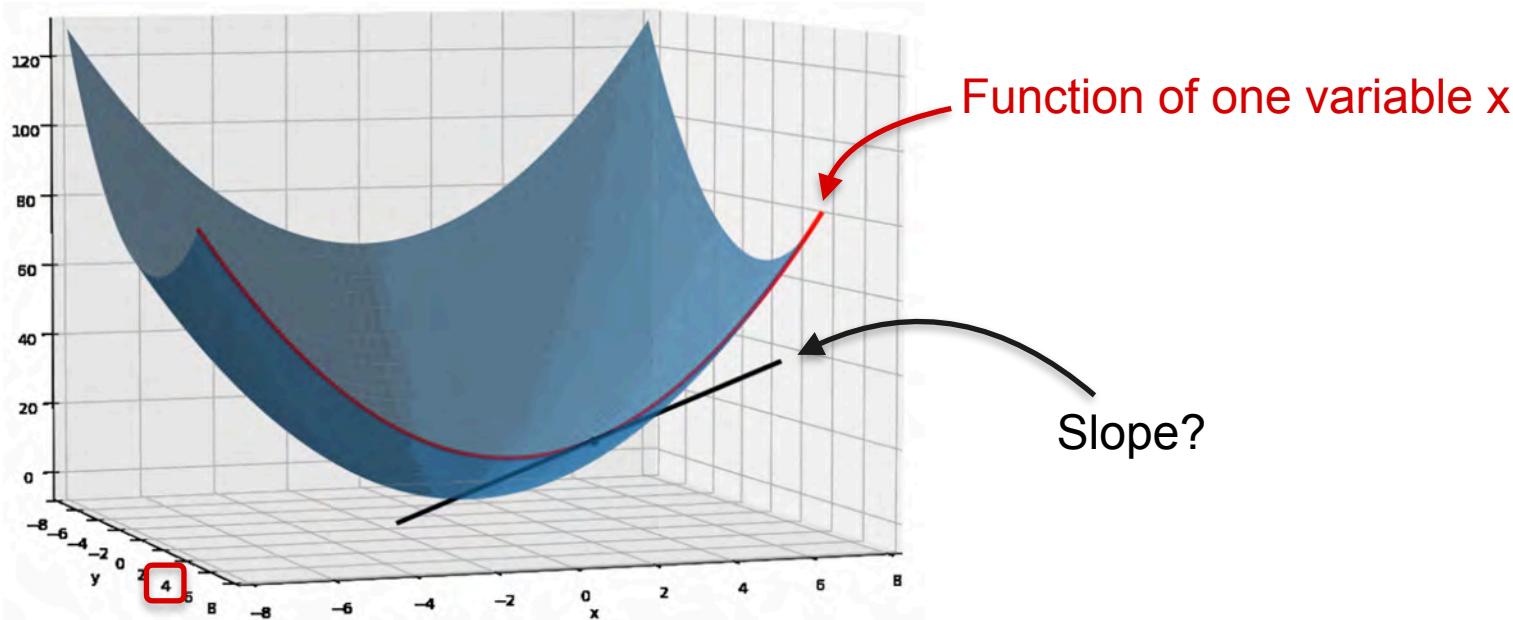
Treat y as a constant



# Partial Derivatives

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

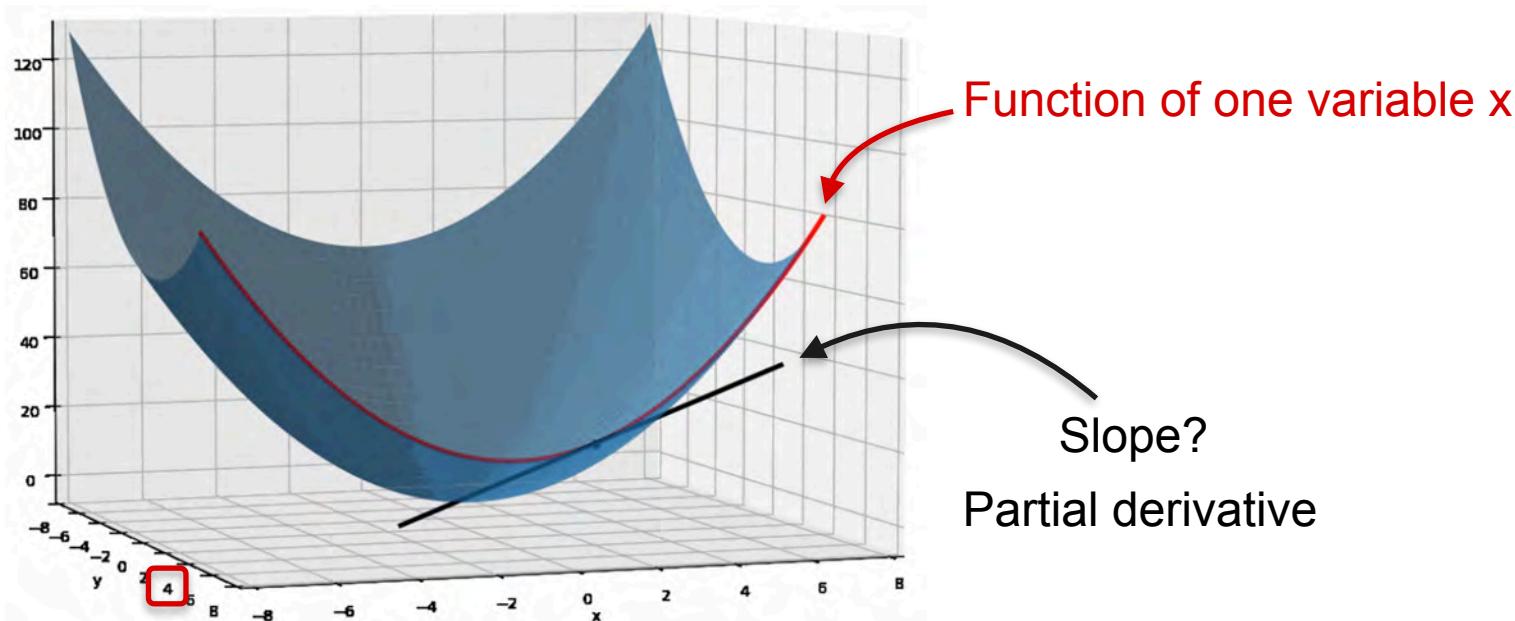
Treat y as a constant



# Partial Derivatives

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

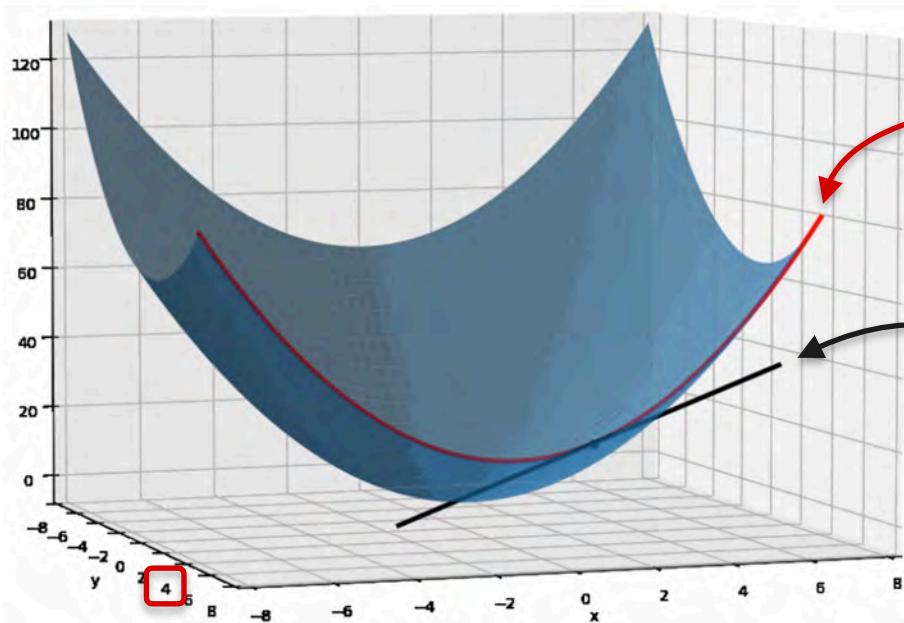
Treat y as a constant



# Partial Derivatives

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

Treat y as a constant



Function of one variable x

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

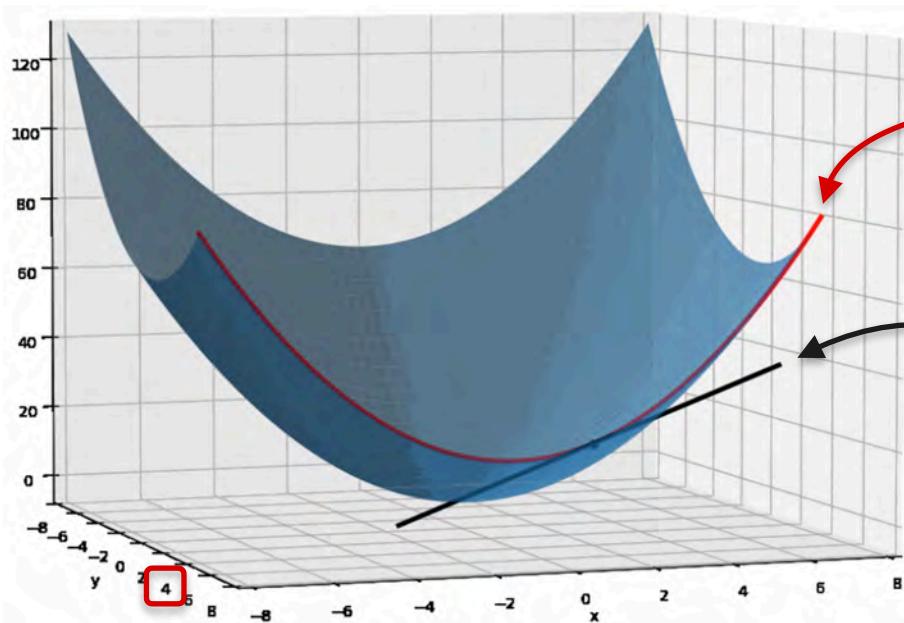
Slope?

Partial derivative

# Partial Derivatives

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

Treat y as a constant



Function of one variable x

Slope?

Partial derivative

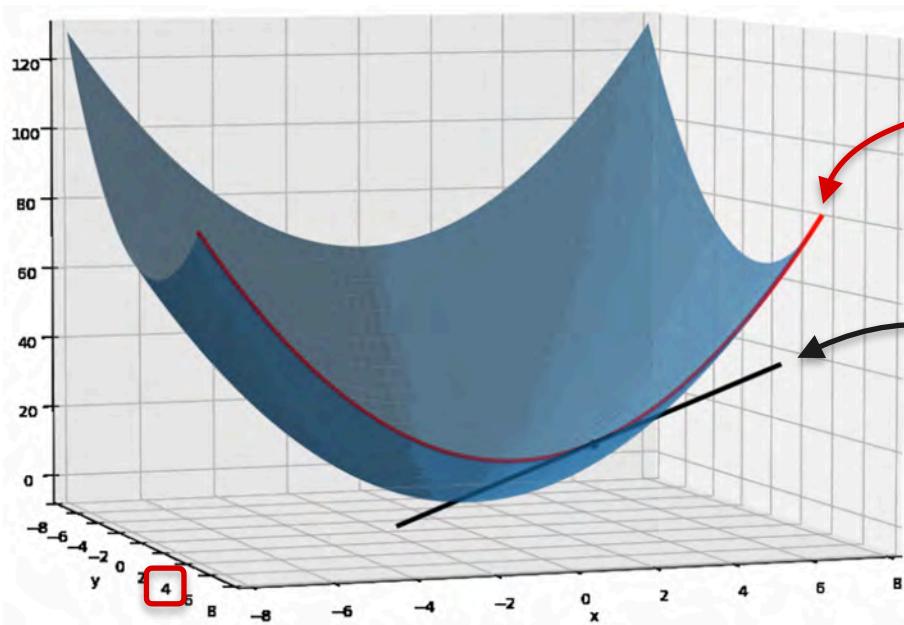
Constant

$$f(x, y) = x^2 + \boxed{y^2}$$

# Partial Derivatives

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

Treat y as a constant



Function of one variable x

Slope?

Partial derivative

Constant

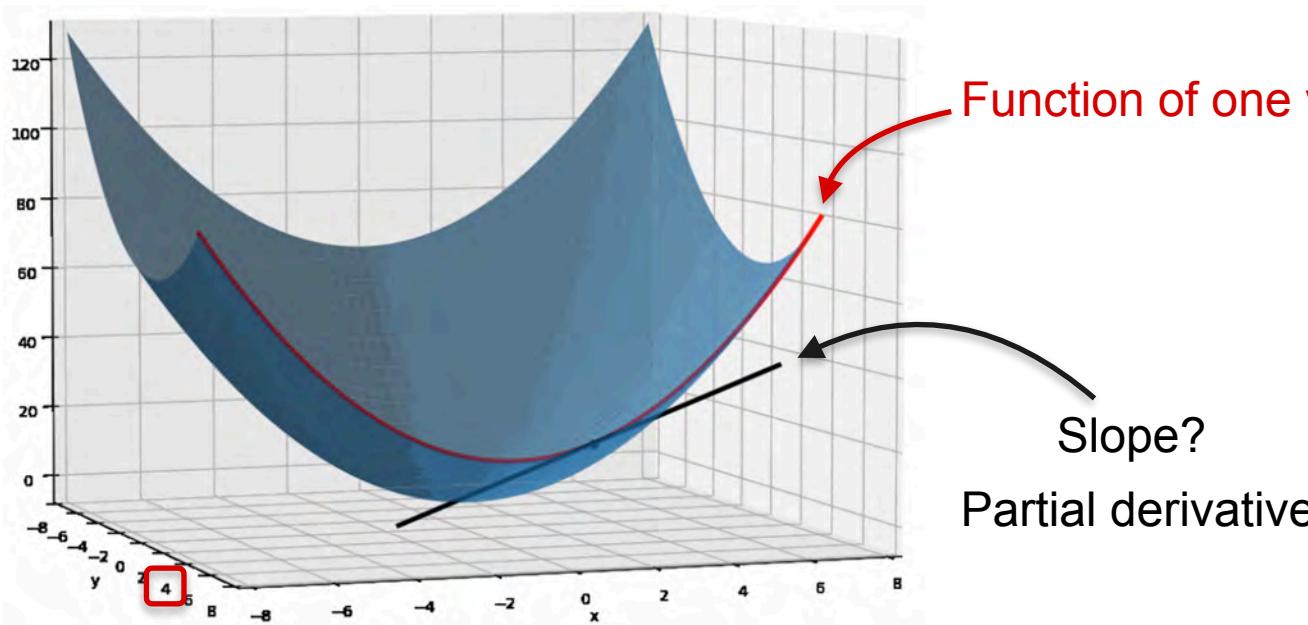
$$f(x, y) = x^2 + \boxed{y^2}$$

$$\frac{\partial f}{\partial x} = 2x + 0$$

# Partial Derivatives

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

Treat y as a constant



Function of one variable x

Constant

$$f(x, y) = x^2 + \boxed{y^2}$$

$$\frac{\partial f}{\partial x} = 2x + \boxed{0}$$

Derivative = 0

# Partial Derivatives

# Partial Derivatives

$$x^2 + y^2$$

# Partial Derivatives

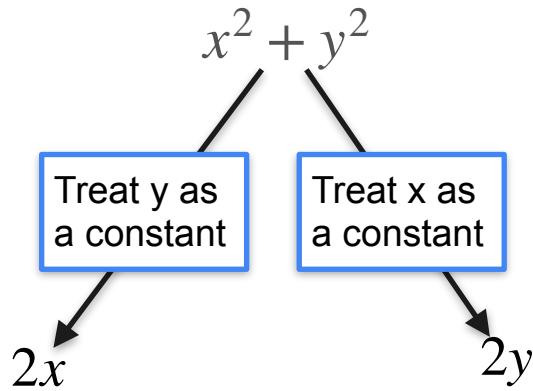
$$x^2 + y^2$$

Treat y as  
a constant

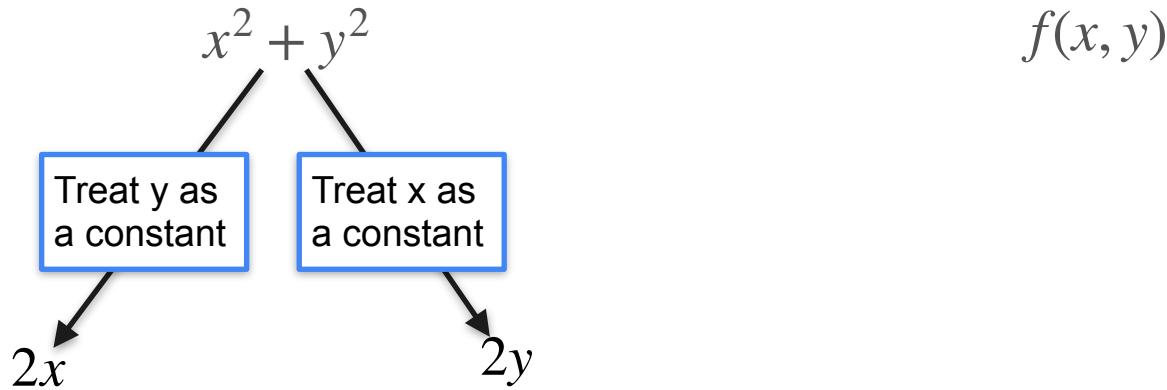
$$2x$$

The diagram illustrates the partial derivative of the function  $x^2 + y^2$  with respect to  $x$ . A blue box contains the instruction "Treat y as a constant". Two arrows point from this box to the terms in the expression: one arrow points from "y as a constant" to the  $y^2$  term, and another arrow points from "y as a constant" to the  $2x$  term.

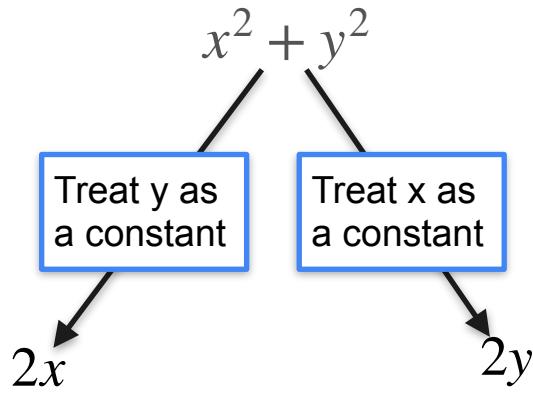
# Partial Derivatives



# Partial Derivatives

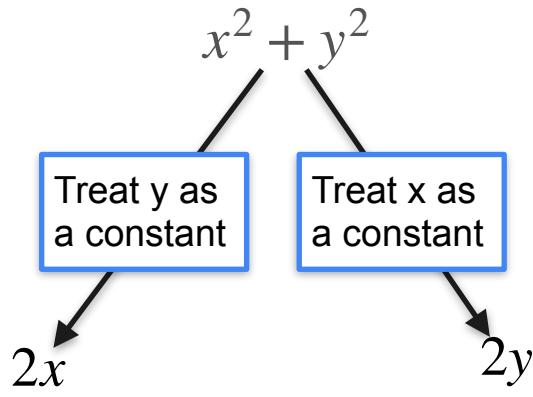


# Partial Derivatives



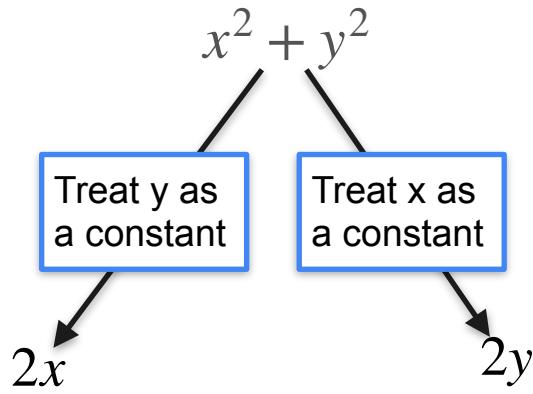
A diagram illustrating the formula for the partial derivative  $f_x$ . At the top, the function  $f(x, y)$  is shown. An arrow points downwards to the formula  $f_x = \frac{\partial f}{\partial x}$ .

# Partial Derivatives



The diagram shows the function  $f(x, y)$  at the top. Two arrows point downwards from the function to the definitions of the partial derivatives. The left arrow points to the equation  $f_x = \frac{\partial f}{\partial x}$ , and the right arrow points to the equation  $f_y = \frac{\partial f}{\partial y}$ .

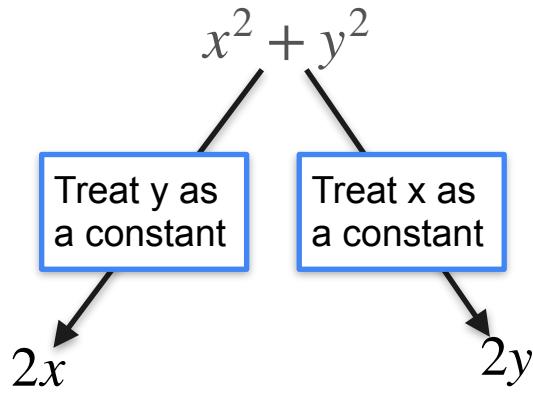
# Partial Derivatives



A diagram illustrating the general concept of partial derivatives. At the top, the function  $f(x, y)$  is shown. Two arrows point downwards from it to the partial derivative formulas  $f_x = \frac{\partial f}{\partial x}$  and  $f_y = \frac{\partial f}{\partial y}$ .

Partial derivative of  
 $f$  with respect to  $x$

# Partial Derivatives



A diagram illustrating partial derivatives. At the top, the function  $f(x, y)$  is shown. Two arrows point downwards from it to the partial derivative formulas  $f_x = \frac{\partial f}{\partial x}$  and  $f_y = \frac{\partial f}{\partial y}$ .

Partial derivative of  
 $f$  with respect to  $x$

Partial derivative of  
 $f$  with respect to  $x$

# Intro To Partial Derivatives

# Intro To Partial Derivatives

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

# Intro To Partial Derivatives

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

**TASK**

# Intro To Partial Derivatives

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

## TASK

Find partial derivatives of  $f$  with respect to  $x$  and  $y$

# Intro To Partial Derivatives

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

## TASK

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

Find partial derivatives of  $f$  with respect to  $x$  and  $y$

# Intro To Partial Derivatives

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

## TASK

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

Find partial derivatives of  $f$  with respect to  $x$  and  $y$

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

# Intro To Partial Derivatives

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

$$\frac{\partial f}{\partial x} = ?$$
  
$$\frac{\partial f}{\partial y} = ?$$

## TASK

Find partial derivatives of  $f$  with respect to  $x$  and  $y$

# Intro To Partial Derivatives

Partial derivative notation

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

$$\frac{\partial f}{\partial x} = ?$$
  
$$\frac{\partial f}{\partial y} = ?$$

## TASK

Find partial derivatives of  $f$  with respect to  $x$  and  $y$

# Intro To Partial Derivatives

Partial derivative notation

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

$$\frac{\partial f}{\partial x} = ?$$
  
$$\frac{\partial f}{\partial y} = ?$$

## TASK

Find partial derivatives of  $f$  with respect to  $x$  and  $y$

# Intro To Partial Derivatives

$$f(x, y) = x^2 + \boxed{y^2}$$

**TASK**

$$\frac{\partial f}{\partial x} = ?$$

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

# Intro To Partial Derivatives

$$f(x, y) = x^2 + \boxed{y^2}$$

## TASK

Find partial derivative of  $f$  with respect to  $x$

$$\frac{\partial f}{\partial x} = ?$$

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

# Intro To Partial Derivatives

$$f(x, y) = x^2 + \boxed{y^2}$$

## TASK

Find partial derivative of  $f$  with respect to  $x$

Step 1:

$$\frac{\partial f}{\partial x} = ?$$

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

# Intro To Partial Derivatives

$$f(x, y) = x^2 + \boxed{y^2}$$

$$\frac{\partial f}{\partial x} = ?$$
  
$$\frac{\partial f}{\partial y} = ?$$

## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

# Intro To Partial Derivatives

$$f(x, y) = x^2 + \boxed{y^2}$$

$$\frac{\partial f}{\partial x} = ?$$
  
$$\frac{\partial f}{\partial y} = ?$$

## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:**

# Intro To Partial Derivatives

$$f(x, y) = x^2 + \boxed{y^2}$$

$$\frac{\partial f}{\partial x} = ?$$
  
$$\frac{\partial f}{\partial y} = ?$$

## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

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

## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

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


## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

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


## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

$$f(x, y) = x^2 + 1$$

$$f(x, y) = x^2 + \boxed{y^2}$$


## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

$$f(x, y) = x^2 + 1$$

$$f(x, y) = x^2 + \boxed{y^2}$$


## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

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

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

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

## TASK

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

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

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

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

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

## TASK

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

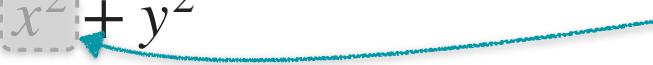
Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

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

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

$$f(x, y) = \boxed{x^2} + y^2$$


## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

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

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

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

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

## TASK

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

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

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

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Intro To Partial Derivatives

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

## TASK

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

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

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

**Step 2:** Differentiate the function using the normal rules of differentiation.



DeepLearning.AI

# Gradients and Gradient Descent

---

## Partial derivatives -Part 2

# Partial Derivatives (More Examples)

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

## TASK

What is the partial derivate of  $f$  with respect to  $x$ ?

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

$$\frac{\partial f}{\partial x} = ?$$

## TASK

What is the partial derivate of  $f$  with respect to  $x$ ?

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

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

**TASK**

Find partial derivate of  $f$  with respect to  $x$

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

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

**TASK**

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$

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

**TASK**

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$

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

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$


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

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$

Constant coefficient

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

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$

Constant coefficient

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

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$

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

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$

Differentiate with respect to  $x$

## TASK

Find partial derivate of  $f$  with respect to  $x$

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

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$

Differentiate with respect to  $x$

## TASK

Find partial derivate of  $f$  with respect to  $x$

$$\frac{\partial f}{\partial x} = 3(2x)$$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$

$$\frac{\partial f}{\partial x} = 3(2x)$$

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$



treat as constant coefficient

## TASK

Find partial derivative of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2$$



treat as constant coefficient

$$\frac{\partial f}{\partial x} = 3(2x)$$

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

$$\frac{\partial f}{\partial x} = 3(2x)y^3$$

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

$$\frac{\partial f}{\partial x} = 3(2x)y^3$$

$$= 6xy^3$$

## TASK

Find partial derivate of  $f$  with respect to  $x$

**Step 1:** Treat all other variables as a constant. In our case  $y$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

## TASK

What is the partial derivate of  $f$  with respect to  $y$ ?

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

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

## TASK

What is the partial derivate of  $f$  with respect to  $y$ ?

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

$$\frac{\partial f}{\partial y} = ?$$

## TASK

What is the partial derivate of  $f$  with respect to  $y$ ?

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

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

**TASK**

What is the partial derivate of  $f$  with respect to  $y$ ?

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3 \text{ } \square y^3$$

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

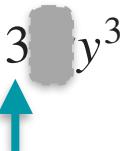
**TASK**

What is the partial derivate of  $f$  with respect to  $y$ ?

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$


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

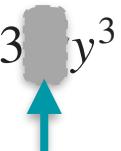
**TASK**

What is the partial derivate of  $f$  with respect to  $y$ ?

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$


$$\frac{\partial f}{\partial y} = 3 \quad \text{[gray square placeholder]}$$

**TASK**

What is the partial derivative of  $f$  with respect to  $y$ ?

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3 \quad y^3$$


$$\frac{\partial f}{\partial y} = 3 \quad (3y^2)$$

## TASK

What is the partial derivate of  $f$  with respect to  $y$ ?

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$



$$\begin{aligned}\frac{\partial f}{\partial y} &= 3(x^2)(3y^2) \\ &= 9x^2y^2\end{aligned}$$

## TASK

What is the partial derivate of  $f$  with respect to  $y$ ?

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.

# Partial Derivatives (More Examples)

$$f(x, y) = 3x^2y^3$$

$$\begin{aligned}\frac{\partial f}{\partial y} &= 3(x^2)(3y^2) \\ &= 9x^2y^2\end{aligned}$$

## TASK

What is the partial derivate of  $f$  with respect to  $y$ ?

**Step 1:** Treat all other variables as a constant. In our case  $x$ .

**Step 2:** Differentiate the function using the normal rules of differentiation.



DeepLearning.AI

# Gradients and Gradient Descent

---

## Gradients

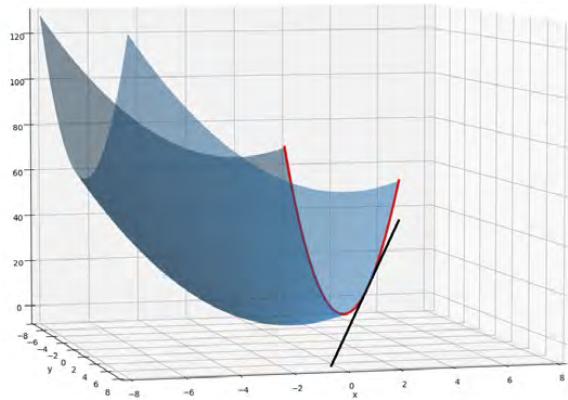
# Gradient

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

# Gradient

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

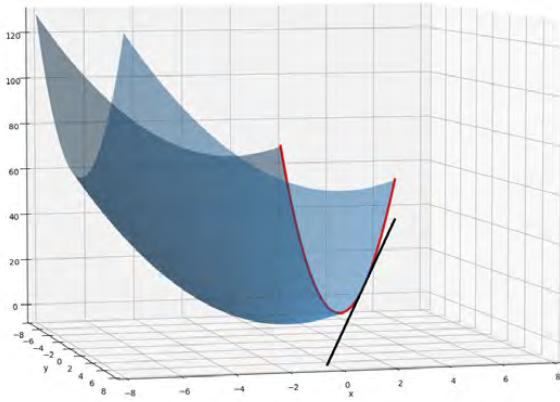
Treat y as a constant



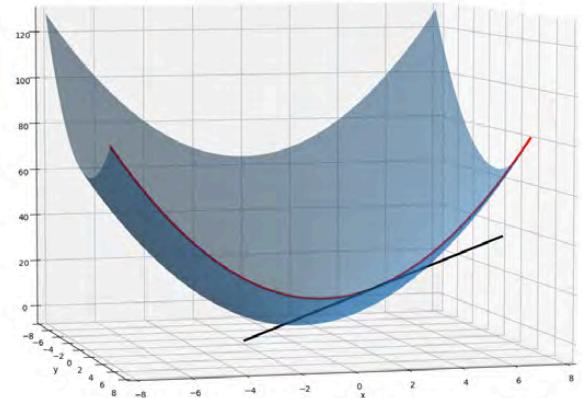
# Gradient

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

Treat y as a constant



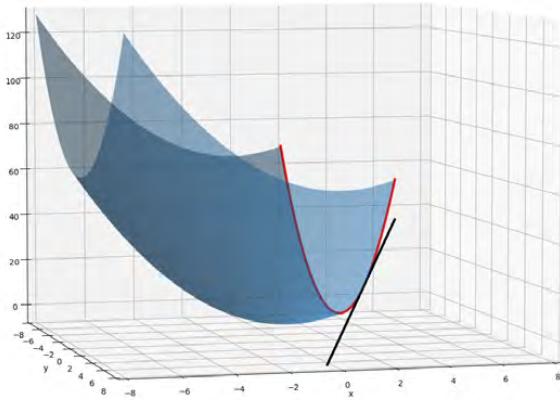
Treat x as a constant



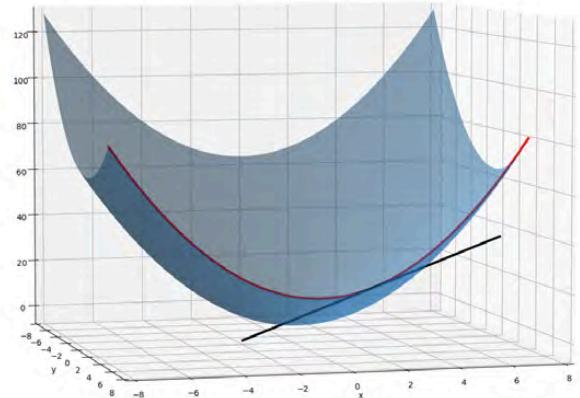
# Gradient

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

Treat y as a constant



Treat x as a constant

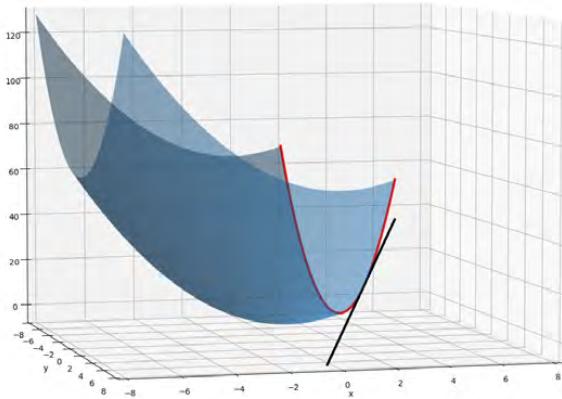


$$\frac{\partial f}{\partial x} = 2x$$

# Gradient

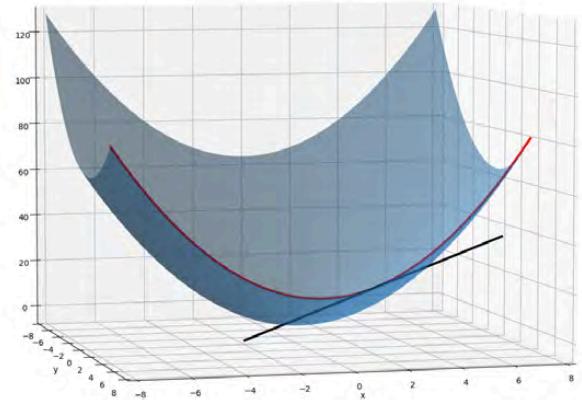
$$f(x, y) = x^2 + y^2$$

Treat y as a constant



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

Treat x as a constant



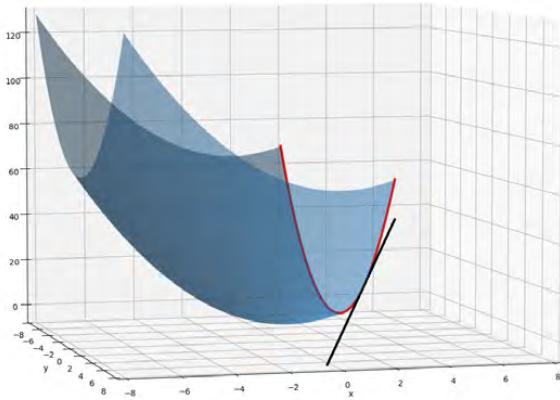
$$\frac{\partial f}{\partial y} = 2y$$

# Gradient

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

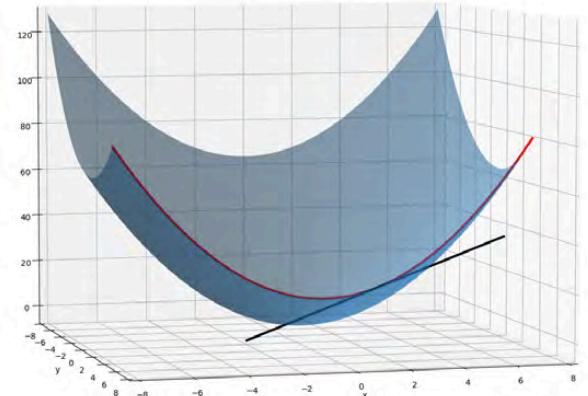
Gradient

Treat y as a constant



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

Treat x as a constant

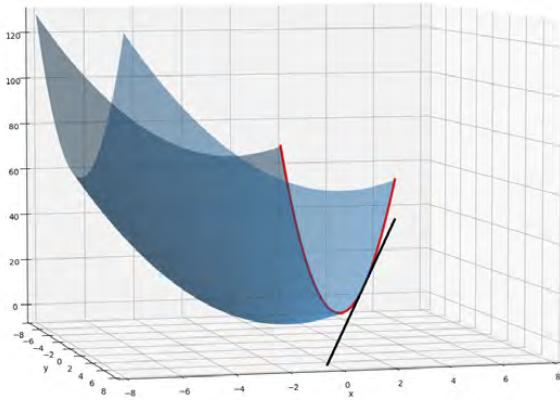


$$\frac{\partial f}{\partial y} = 2y$$

# Gradient

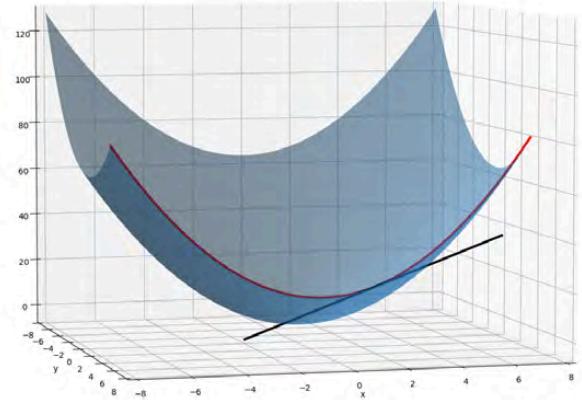
$$f(x, y) = x^2 + y^2$$

Treat y as a constant



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

Treat x as a constant



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

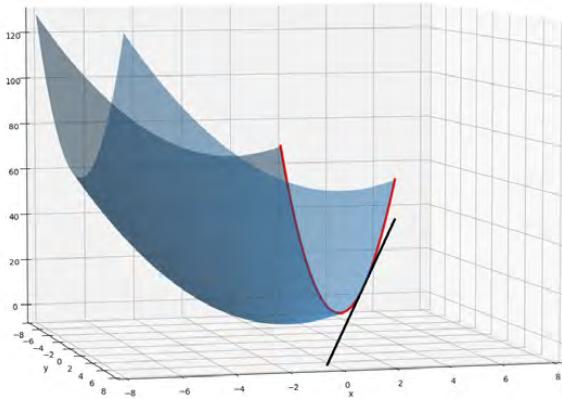
Gradient

$$\begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

# Gradient

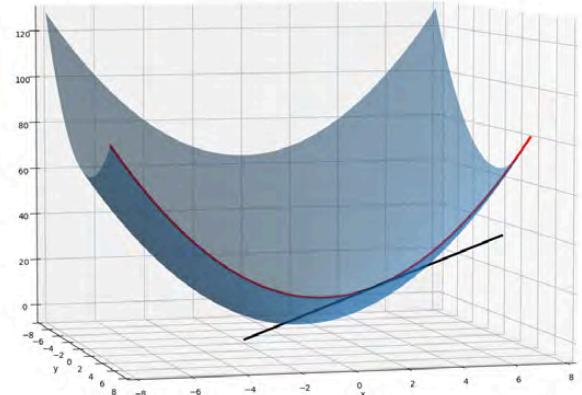
$$f(x, y) = x^2 + y^2$$

Treat y as a constant



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

Treat x as a constant



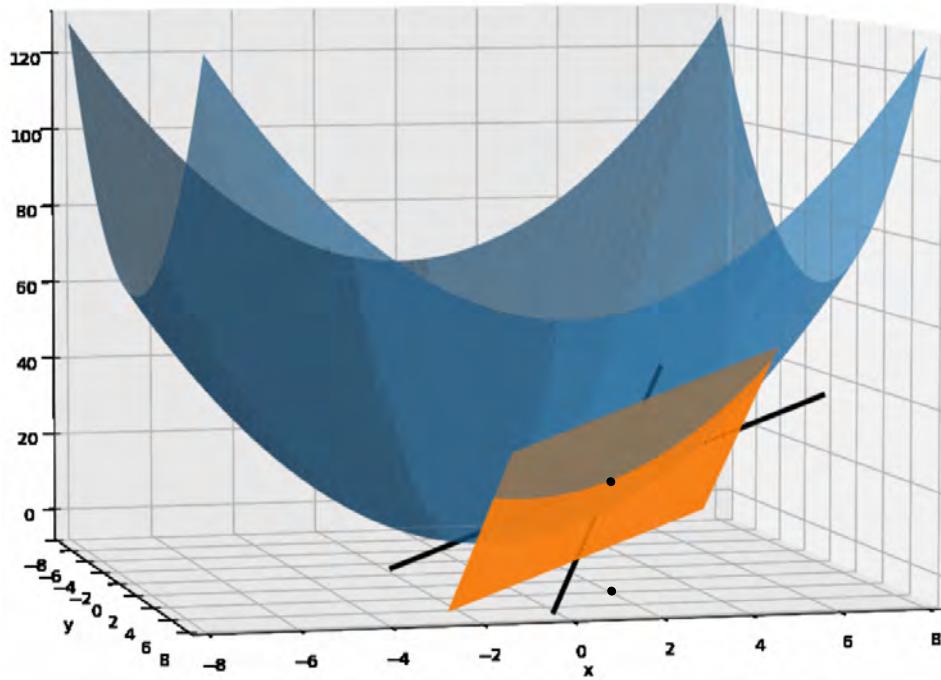
$$\frac{\partial f}{\partial y} = 2y$$

Gradient

$$\begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

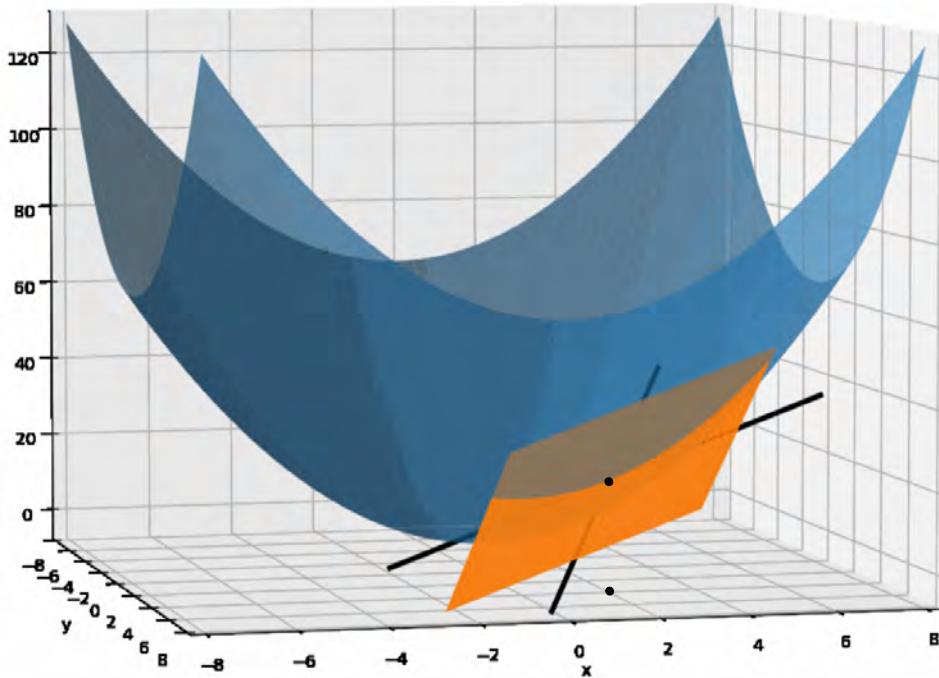
# Gradient



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

The gradient of  $f(x, y)$  is:  $\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$

# Gradient



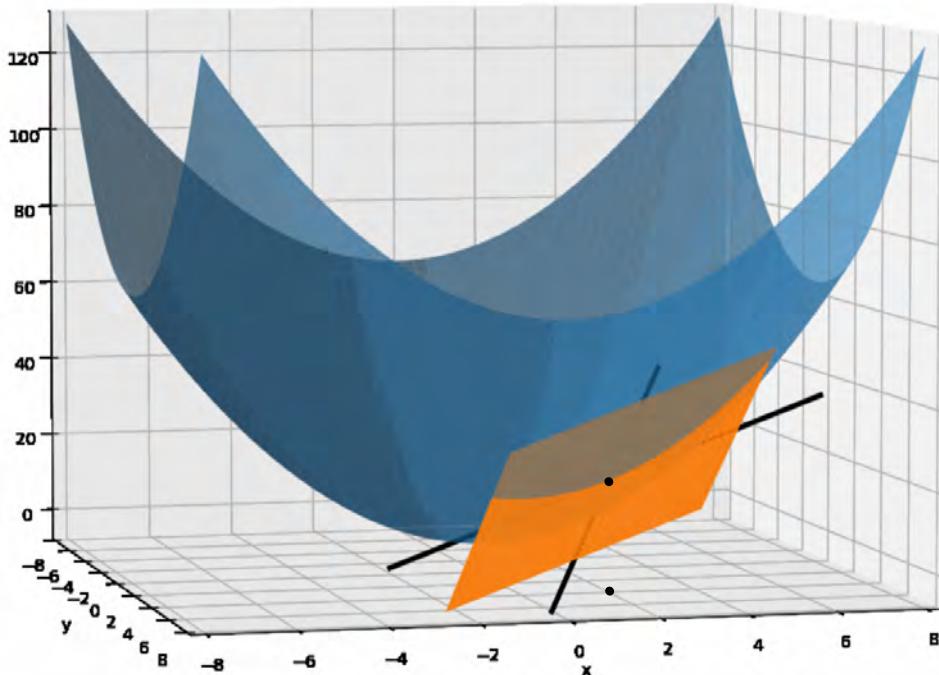
$$f(x, y) = x^2 + y^2$$

The gradient of  $f(x, y)$  is:  $\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$

## TASK

Find the gradient of  $f(x, y)$  at  $(2, 3)$

# Gradient



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

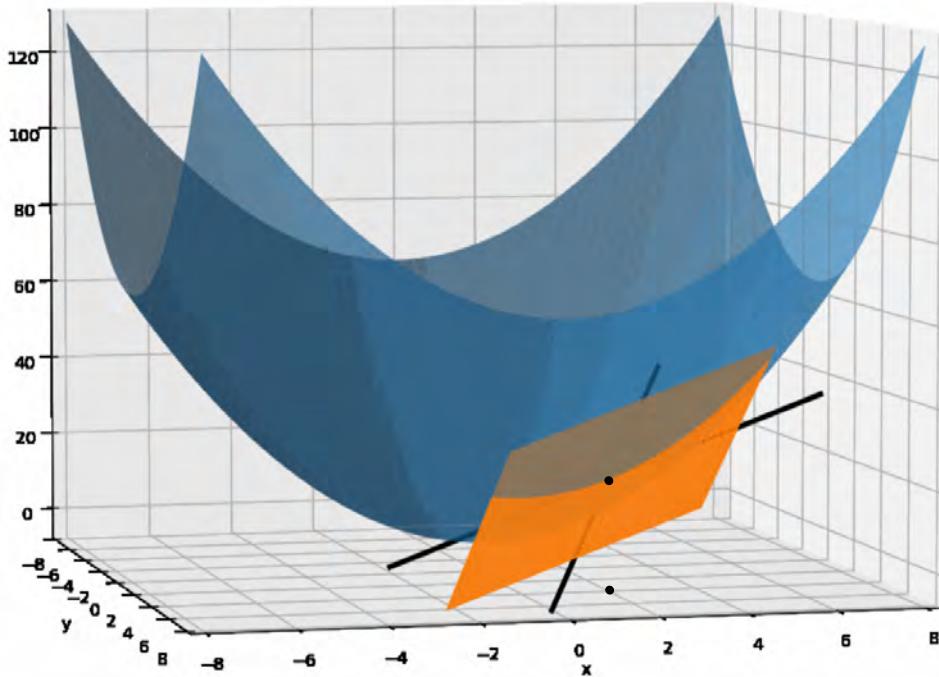
The gradient of  $f(x, y)$  is:  $\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$

## TASK

Find the gradient of  $f(x, y)$  at  $(2, 3)$

The gradient of  $f(x, y)$  is given as:

# Gradient



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

The gradient of  $f(x, y)$  is:  $\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$

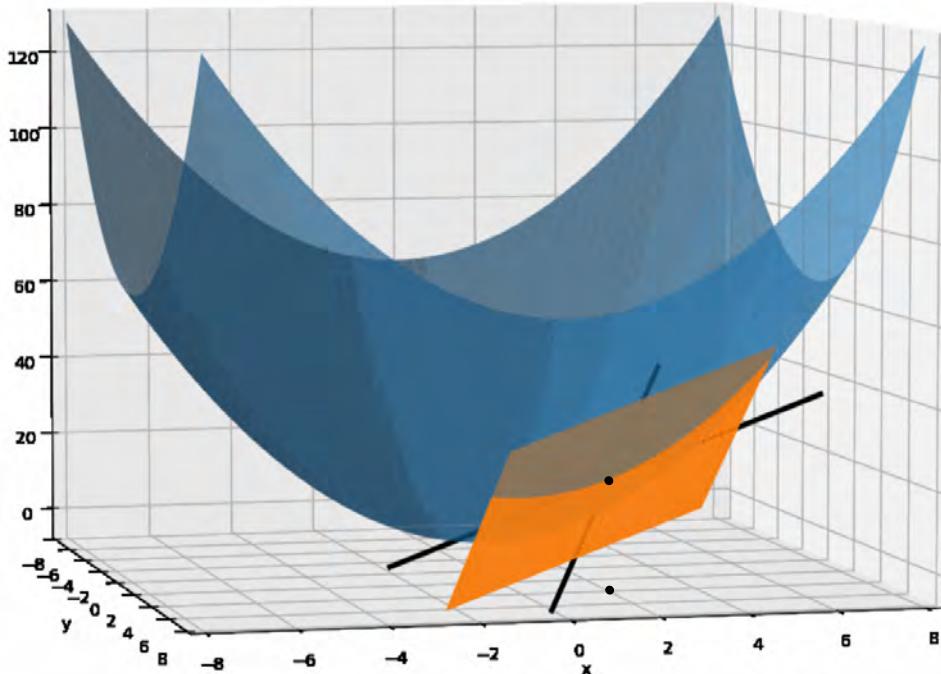
## TASK

Find the gradient of  $f(x, y)$  at  $(2, 3)$

The gradient of  $f(x, y)$  is given as:

$$\nabla f = \begin{bmatrix} 2 \cdot 2 \\ 2 \cdot 3 \end{bmatrix}$$

# Gradient



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

The gradient of  $f(x, y)$  is:  $\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$

## TASK

Find the gradient of  $f(x, y)$  at  $(2, 3)$

The gradient of  $f(x, y)$  is given as:

$$\nabla f = \begin{bmatrix} 2 \cdot 2 \\ 2 \cdot 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$



DeepLearning.AI

# Gradients and Gradient Descent

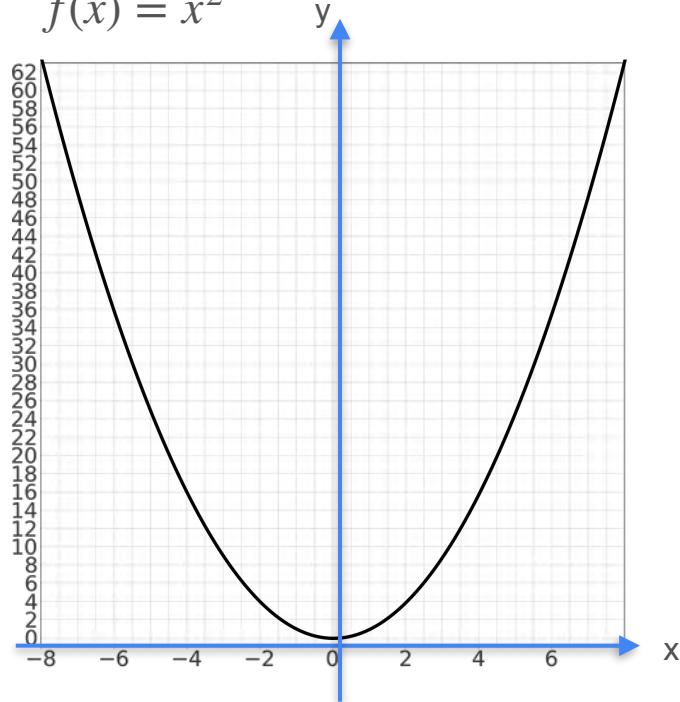
---

**Gradients and maxima/  
minima**

# Functions of Two Variables

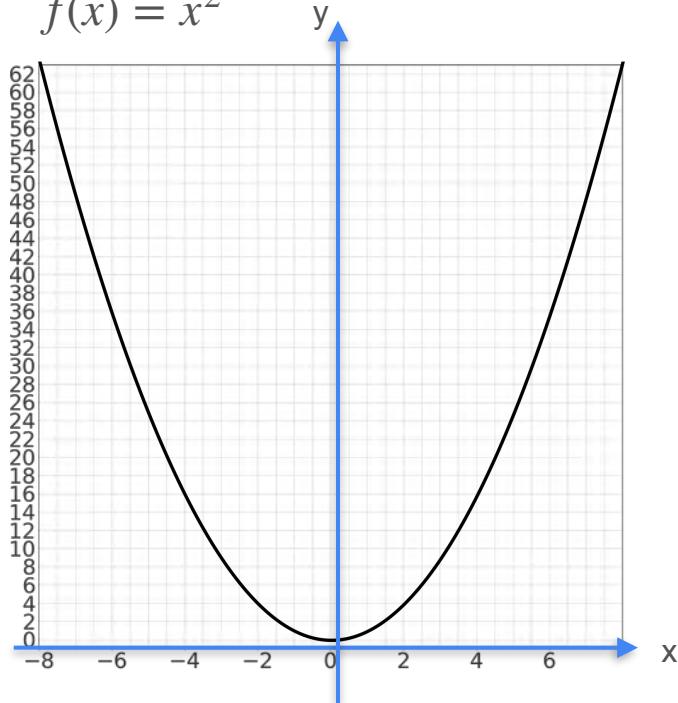
# Functions of Two Variables

$$f(x) = x^2$$

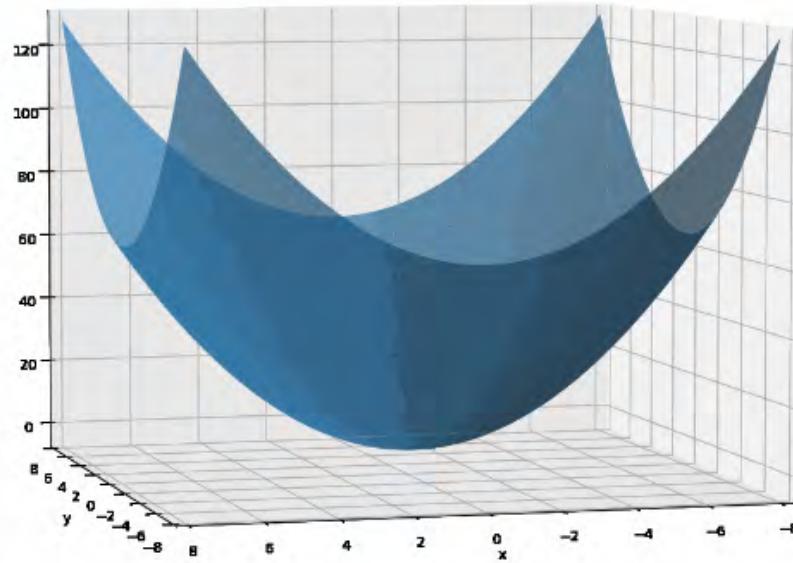


# Functions of Two Variables

$$f(x) = x^2$$

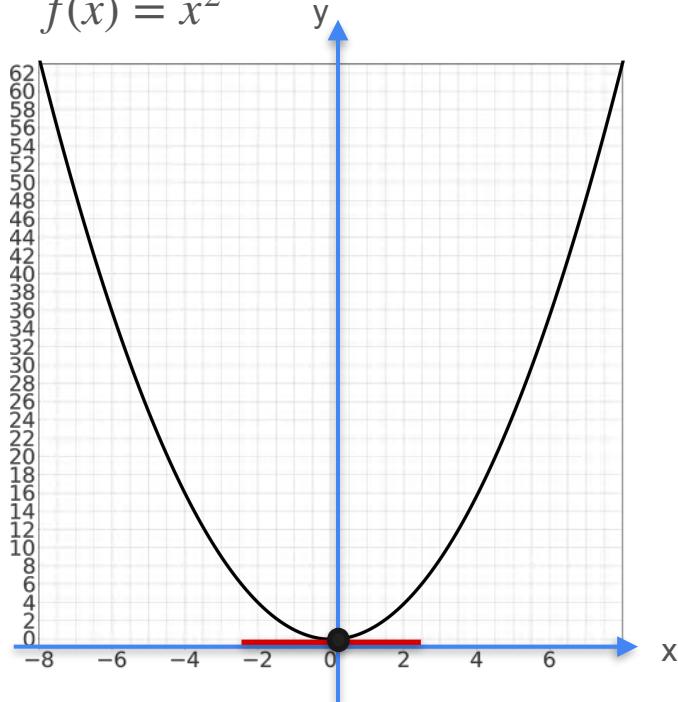


$$f(x, y) = x^2 + y^2$$

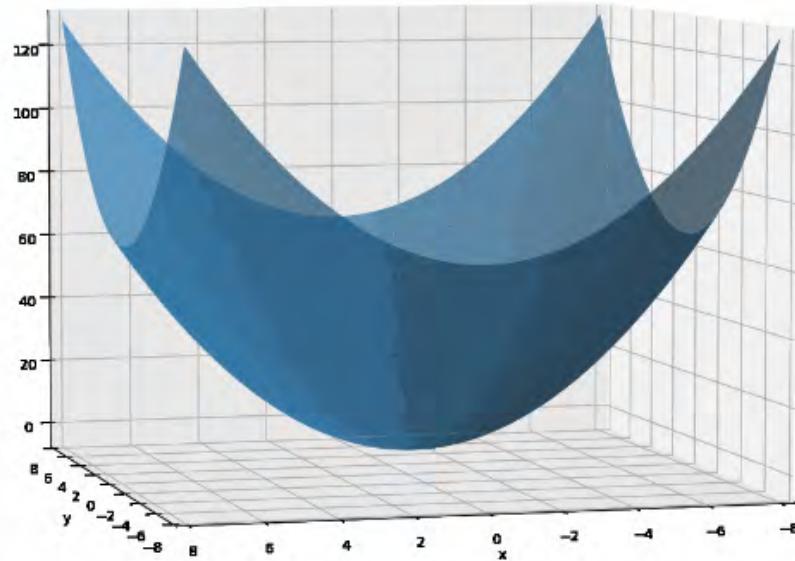


# Functions of Two Variables

$$f(x) = x^2$$

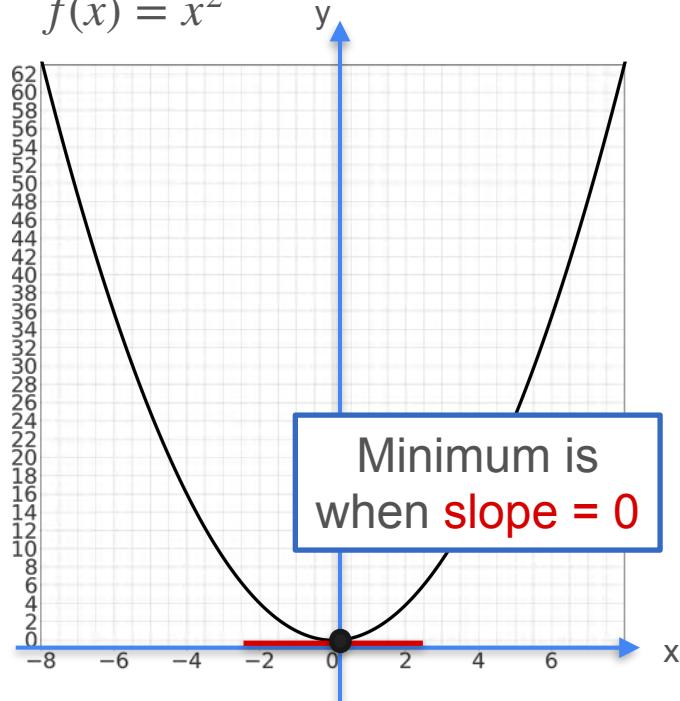


$$f(x, y) = x^2 + y^2$$

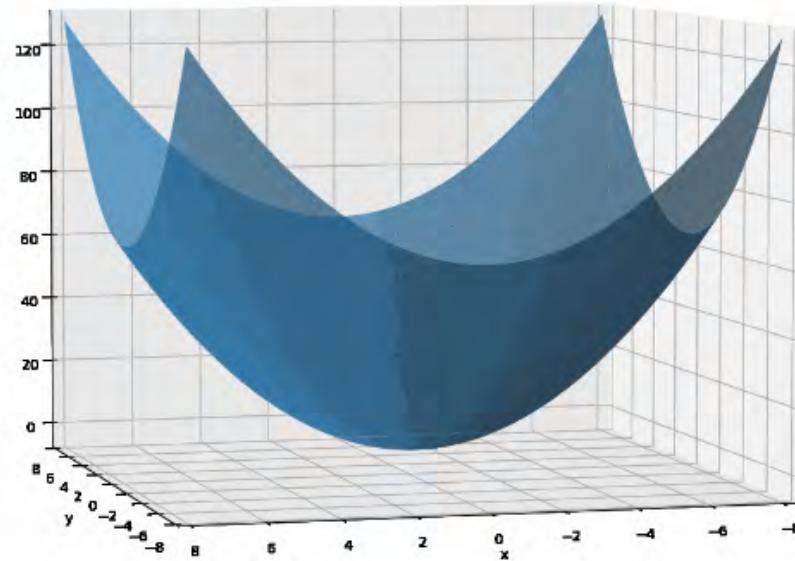


# Functions of Two Variables

$$f(x) = x^2$$

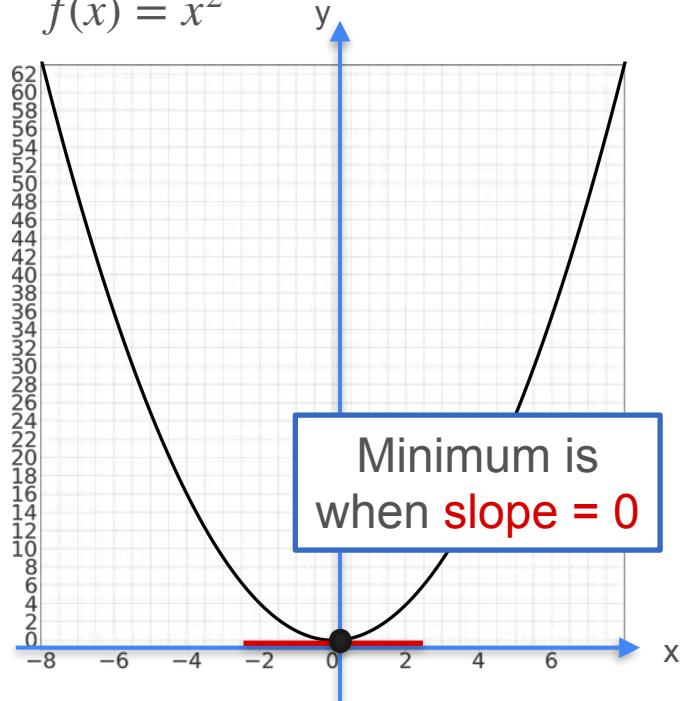


$$f(x, y) = x^2 + y^2$$

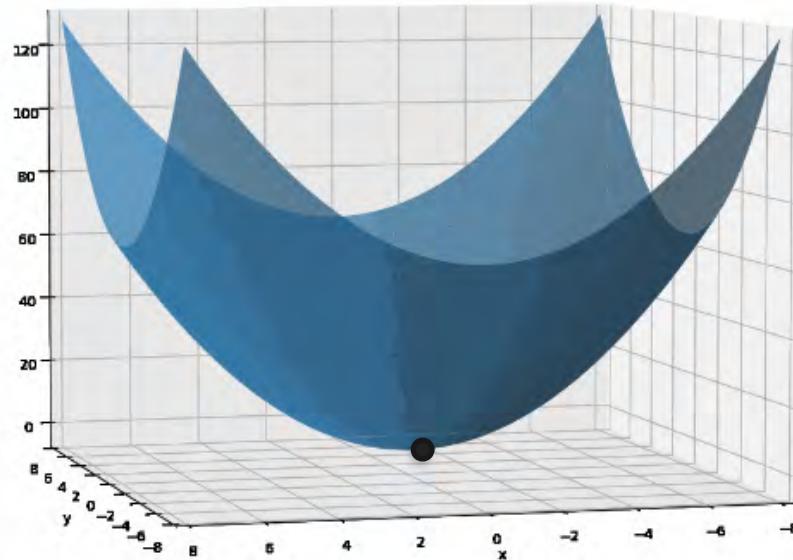


# Functions of Two Variables

$$f(x) = x^2$$

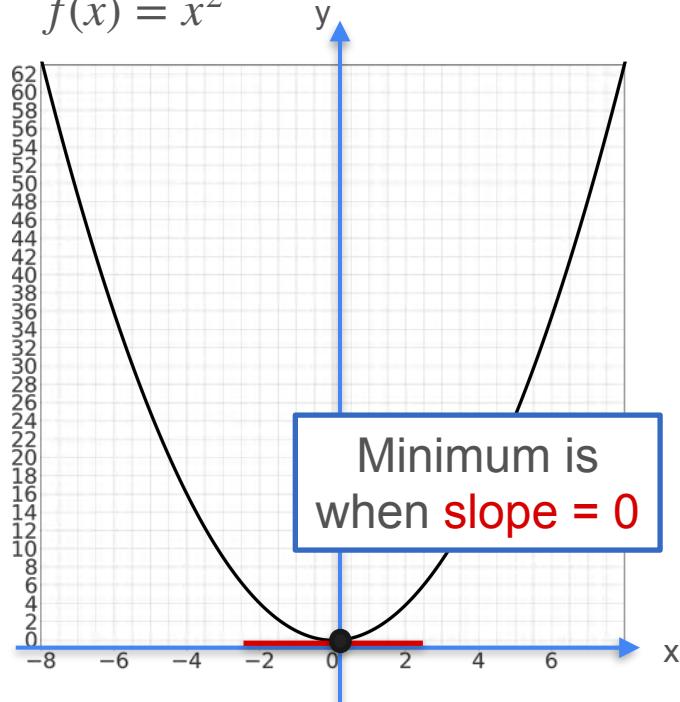


$$f(x, y) = x^2 + y^2$$

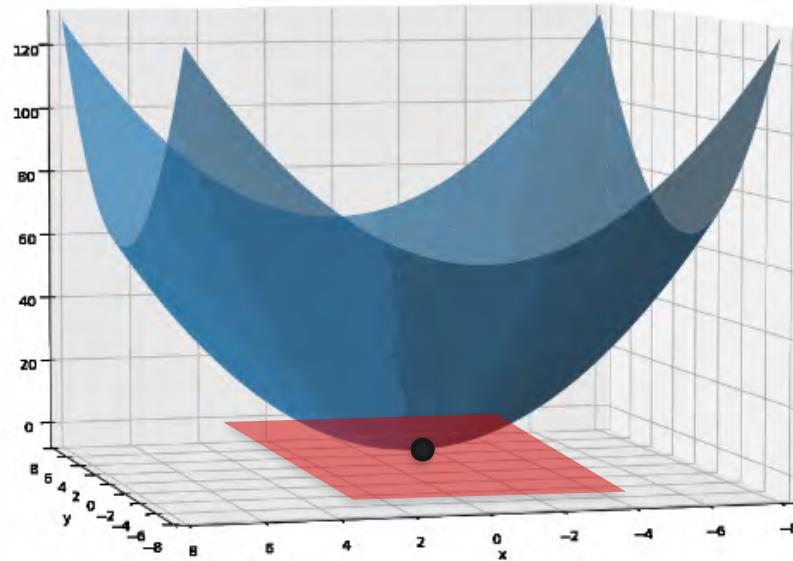


# Functions of Two Variables

$$f(x) = x^2$$

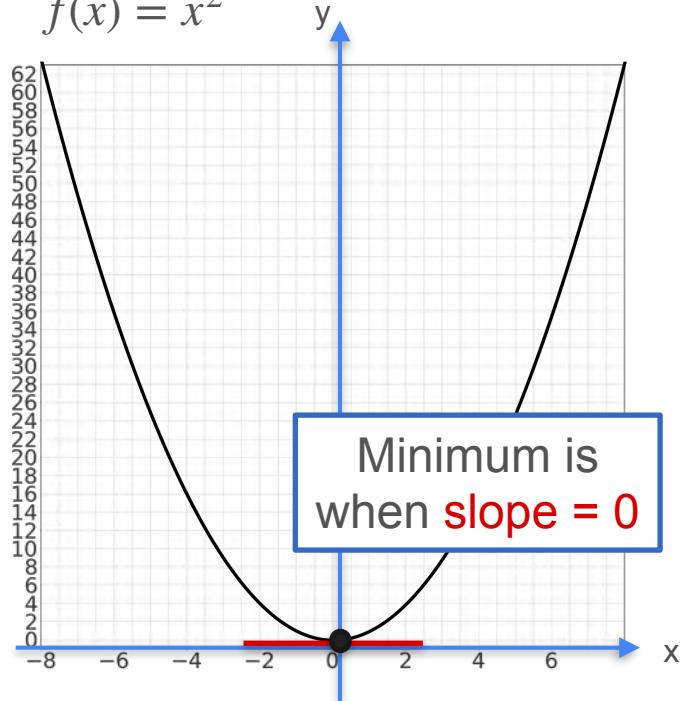


$$f(x, y) = x^2 + y^2$$

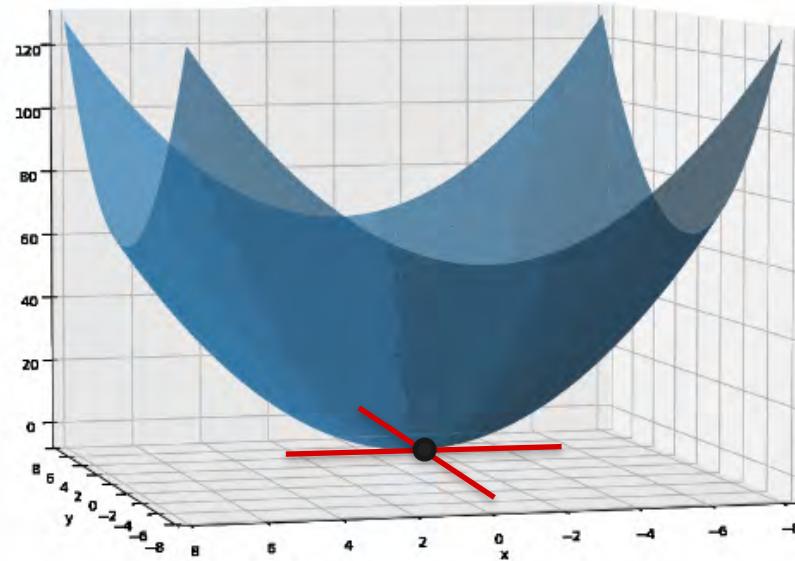


# Functions of Two Variables

$$f(x) = x^2$$

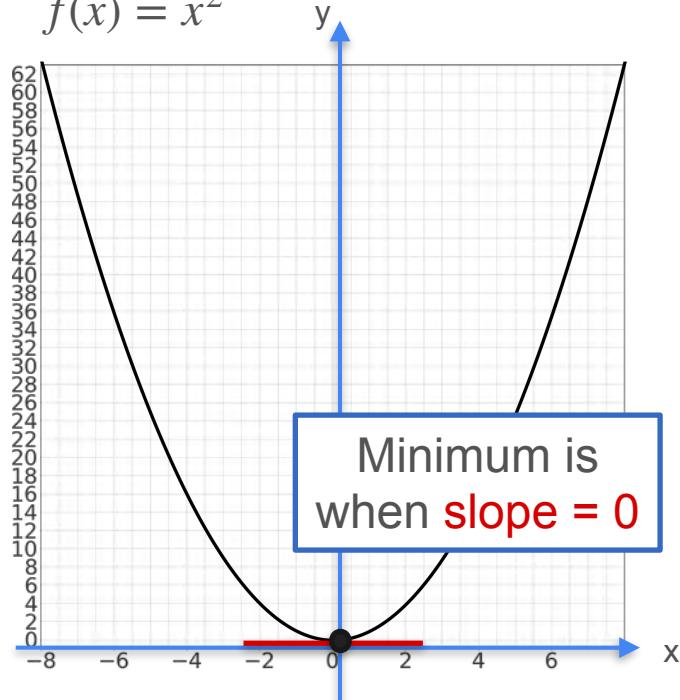


$$f(x, y) = x^2 + y^2$$

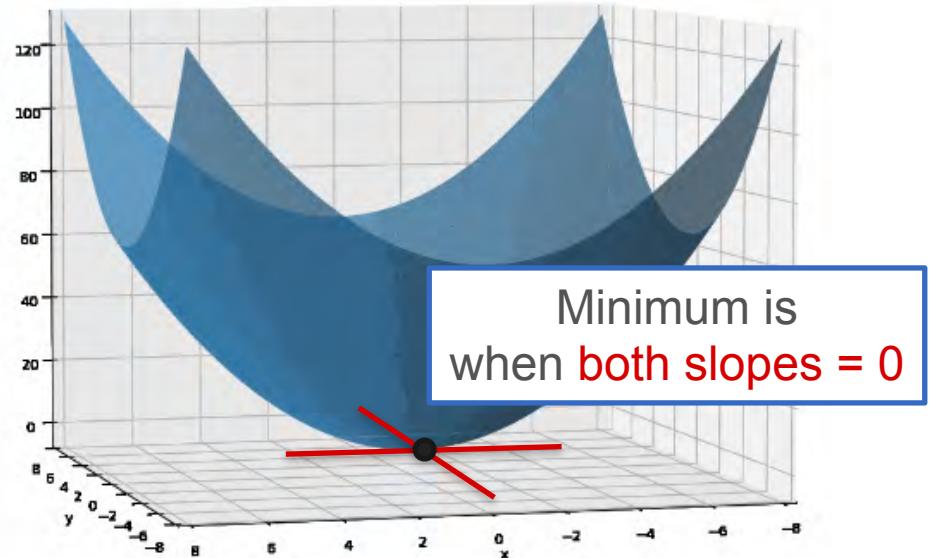


# Functions of Two Variables

$$f(x) = x^2$$

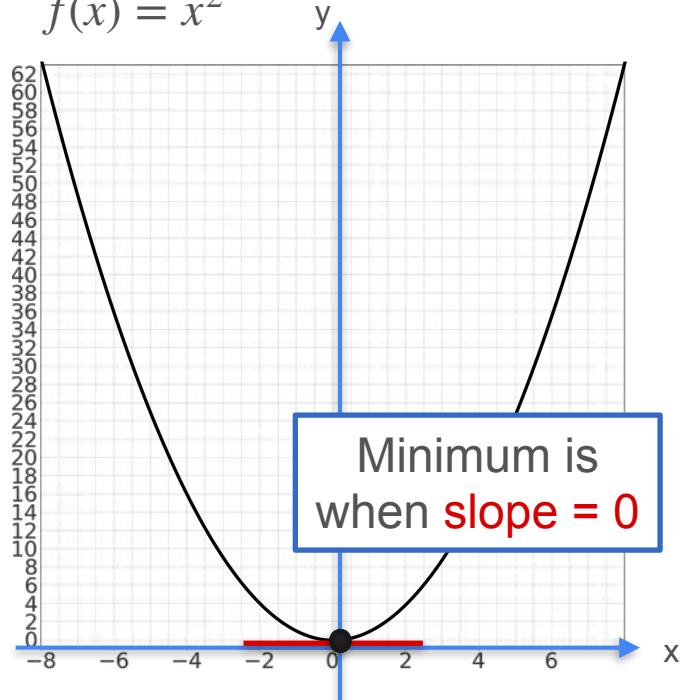


$$f(x, y) = x^2 + y^2$$

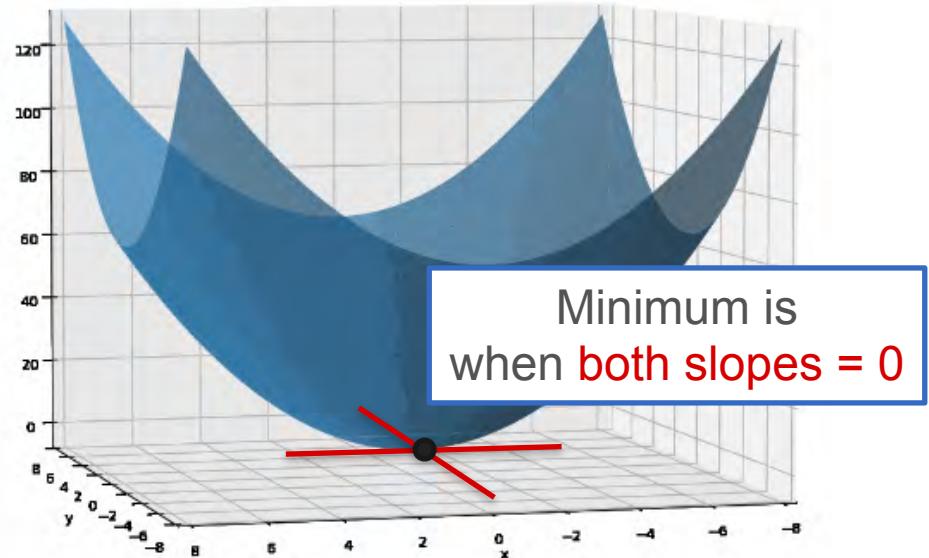


# Functions of Two Variables

$$f(x) = x^2$$

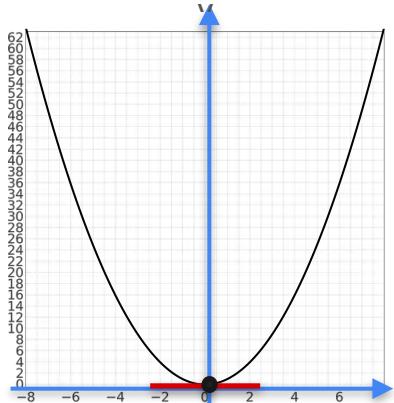


$$f(x, y) = x^2 + y^2$$



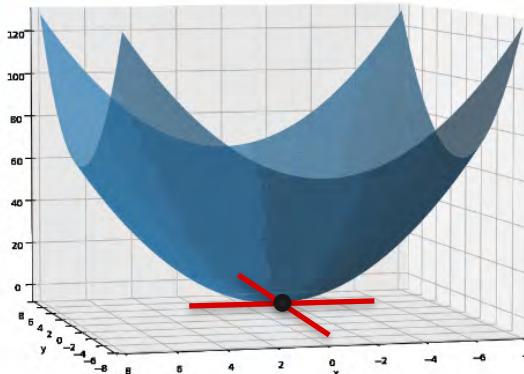
# Functions of Two Variables

$$f(x) = x^2$$



Minimum is  
when **slope = 0**

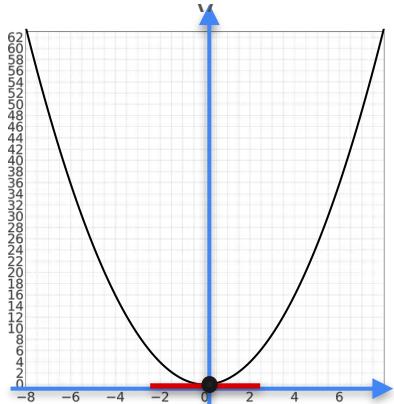
$$f(x, y) = x^2 + y^2$$



Minimum is  
when **both slopes = 0**

# Functions of Two Variables

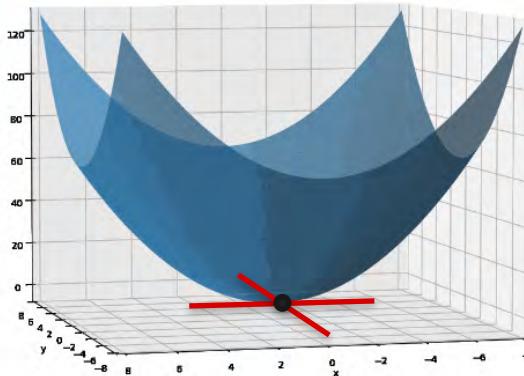
$$f(x) = x^2$$



Minimum is  
when **slope = 0**

$$f'(x) = 0$$

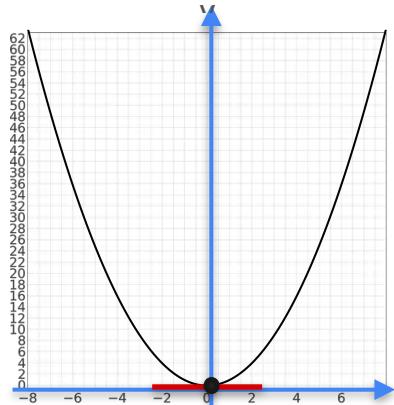
$$f(x, y) = x^2 + y^2$$



Minimum is  
when **both slopes = 0**

# Functions of Two Variables

$$f(x) = x^2$$

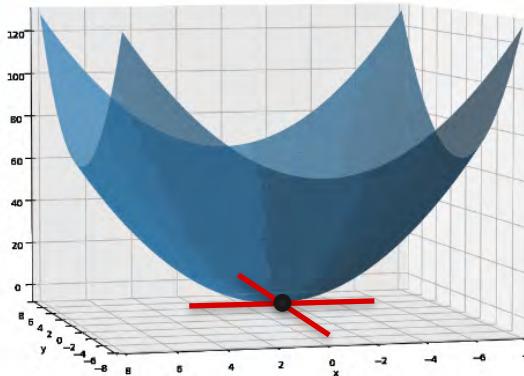


Minimum is  
when **slope = 0**

$$f'(x) = 0$$

$$2x = 0$$

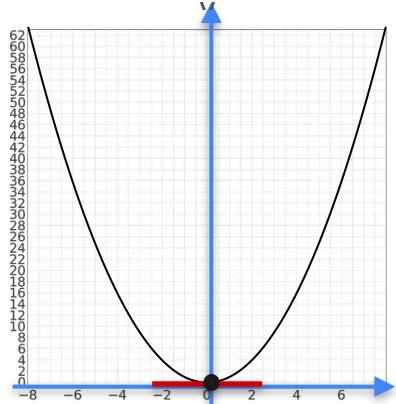
$$f(x, y) = x^2 + y^2$$



Minimum is  
when **both slopes = 0**

# Functions of Two Variables

$$f(x) = x^2$$



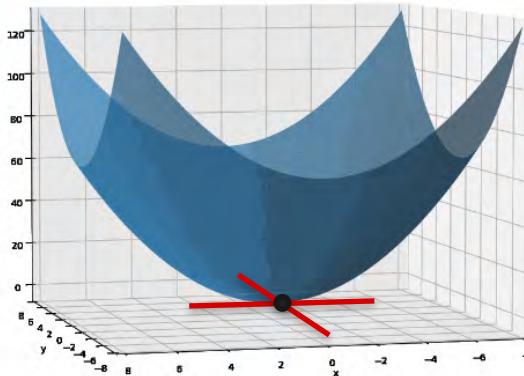
Minimum is  
when **slope = 0**

$$f'(x) = 0$$

$$2x = 0$$

$$x = 0$$

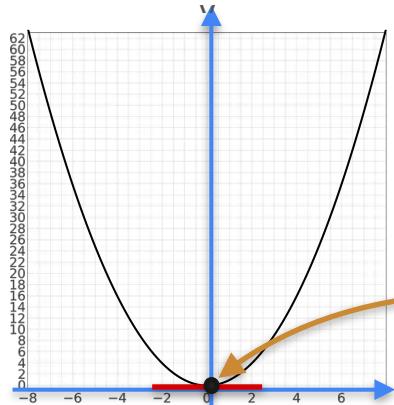
$$f(x, y) = x^2 + y^2$$



Minimum is  
when **both slopes = 0**

# Functions of Two Variables

$$f(x) = x^2$$



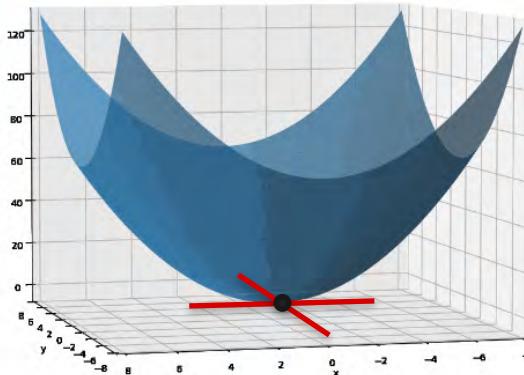
Minimum is  
when **slope = 0**

$$f'(x) = 0$$

$$2x = 0$$

$$x = 0$$

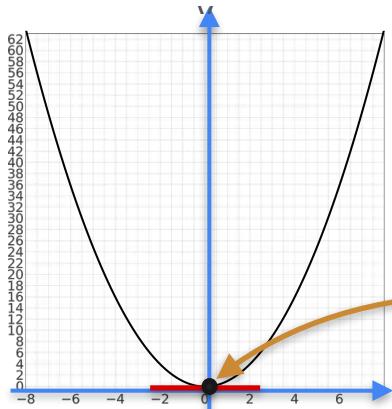
$$f(x, y) = x^2 + y^2$$



Minimum is  
when **both slopes = 0**

# Functions of Two Variables

$$f(x) = x^2$$



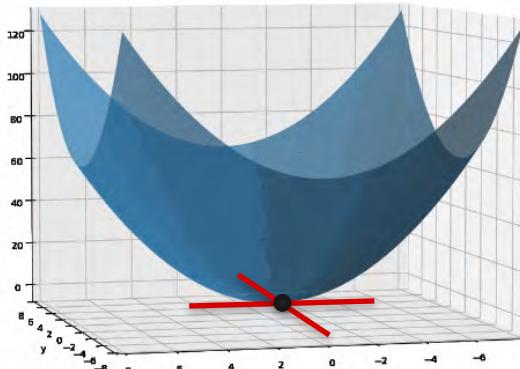
Minimum is  
when **slope = 0**

$$f'(x) = 0$$

$$2x = 0$$

$$x = 0$$

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

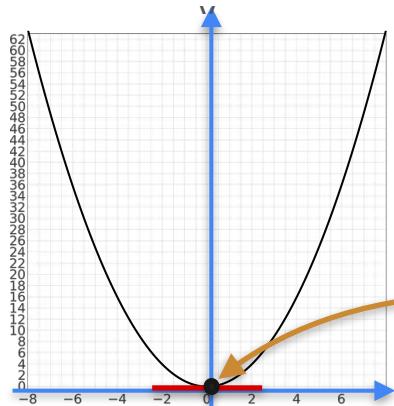


Minimum is  
when **both slopes = 0**

$$\frac{\partial f}{\partial x} = 0 \text{ and } \frac{\partial f}{\partial y} = 0$$

# Functions of Two Variables

$$f(x) = x^2$$



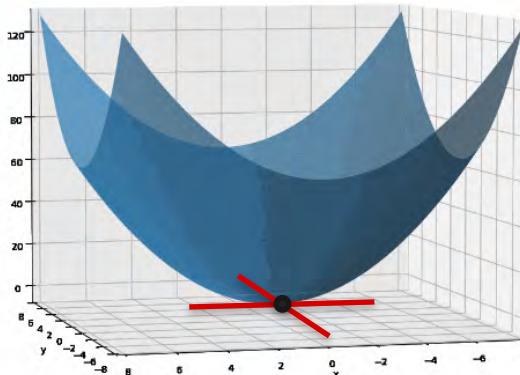
Minimum is  
when **slope = 0**

$$f'(x) = 0$$

$$2x = 0$$

$$x = 0$$

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



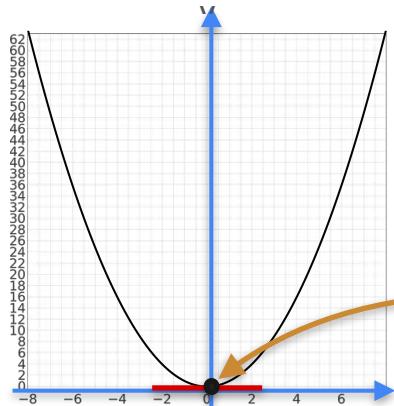
Minimum is  
when **both slopes = 0**

$$\frac{\partial f}{\partial x} = 0 \text{ and } \frac{\partial f}{\partial y} = 0$$

$$2x = 0 \text{ and } 2y = 0$$

# Functions of Two Variables

$$f(x) = x^2$$



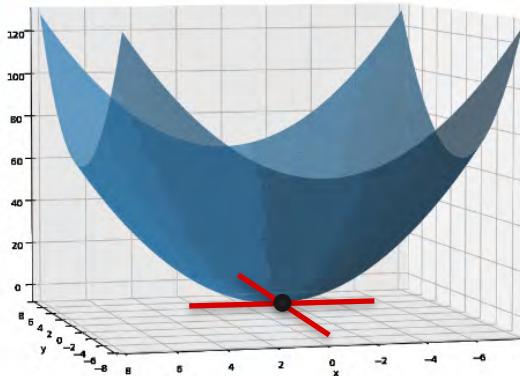
Minimum is  
when **slope = 0**

$$f'(x) = 0$$

$$2x = 0$$

$$x = 0$$

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



$$\frac{\partial f}{\partial x} = 0 \text{ and } \frac{\partial f}{\partial y} = 0$$

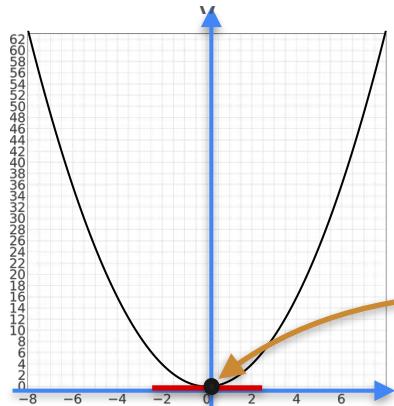
$$2x = 0 \text{ and } 2y = 0$$

$$(x, y) = (0,0)$$

Minimum is  
when **both slopes = 0**

# Functions of Two Variables

$$f(x) = x^2$$



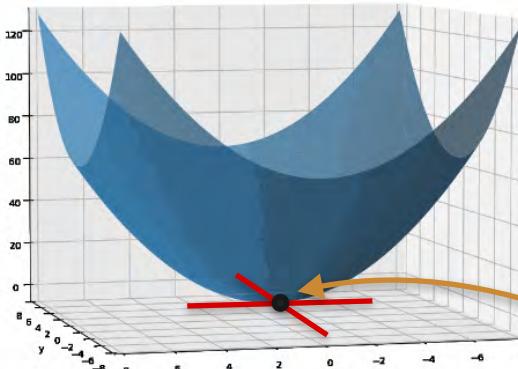
Minimum is  
when **slope = 0**

$$f'(x) = 0$$

$$2x = 0$$

$$x = 0$$

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



Minimum is  
when **both slopes = 0**

$$\frac{\partial f}{\partial x} = 0 \text{ and } \frac{\partial f}{\partial y} = 0$$

$$2x = 0 \text{ and } 2y = 0$$

$$(x, y) = (0,0)$$



DeepLearning.AI

# Gradients and Gradient Descent

---

**Optimization with gradients:  
An example**

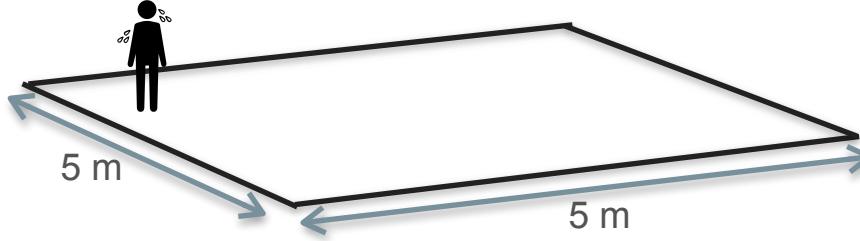
# Motivation for Optimization in Two Variables



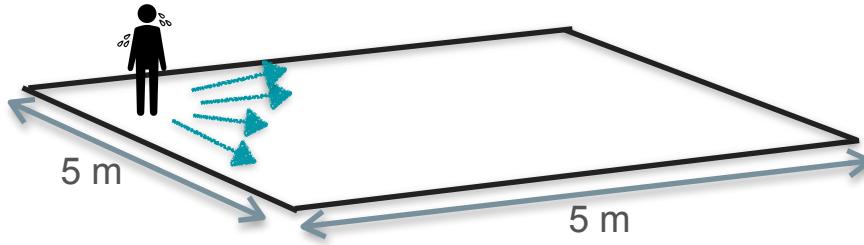
# Motivation for Optimization in Two Variables



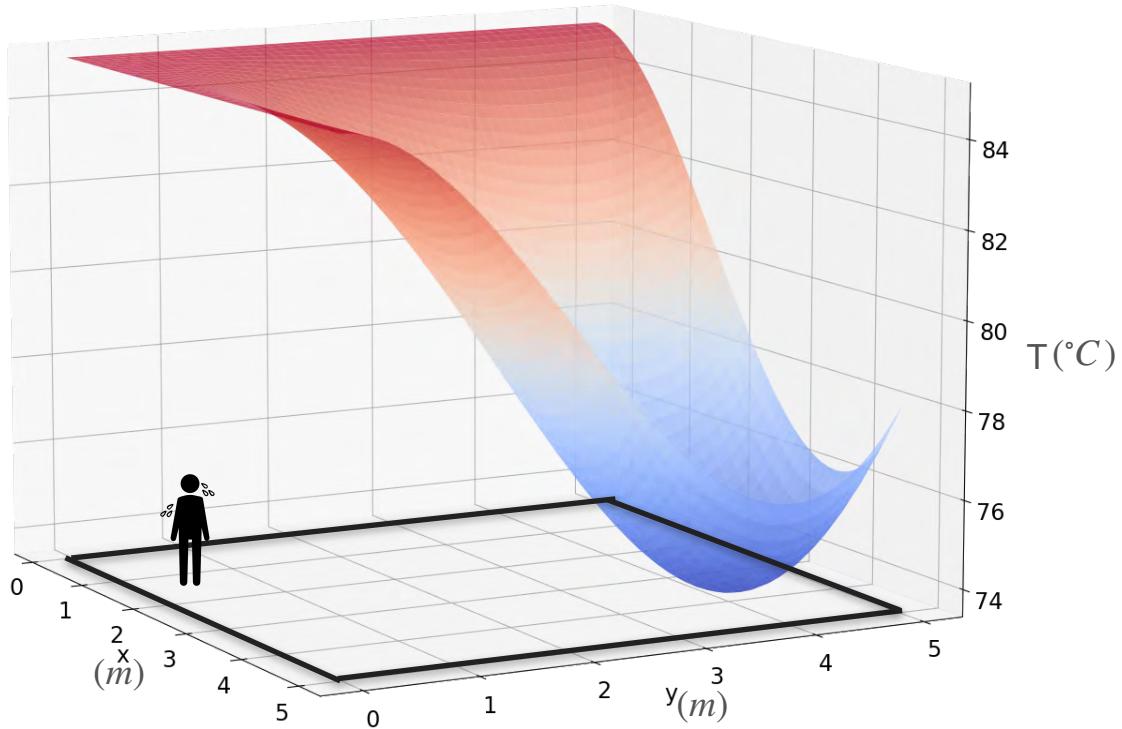
# Motivation for Optimization in Two Variables



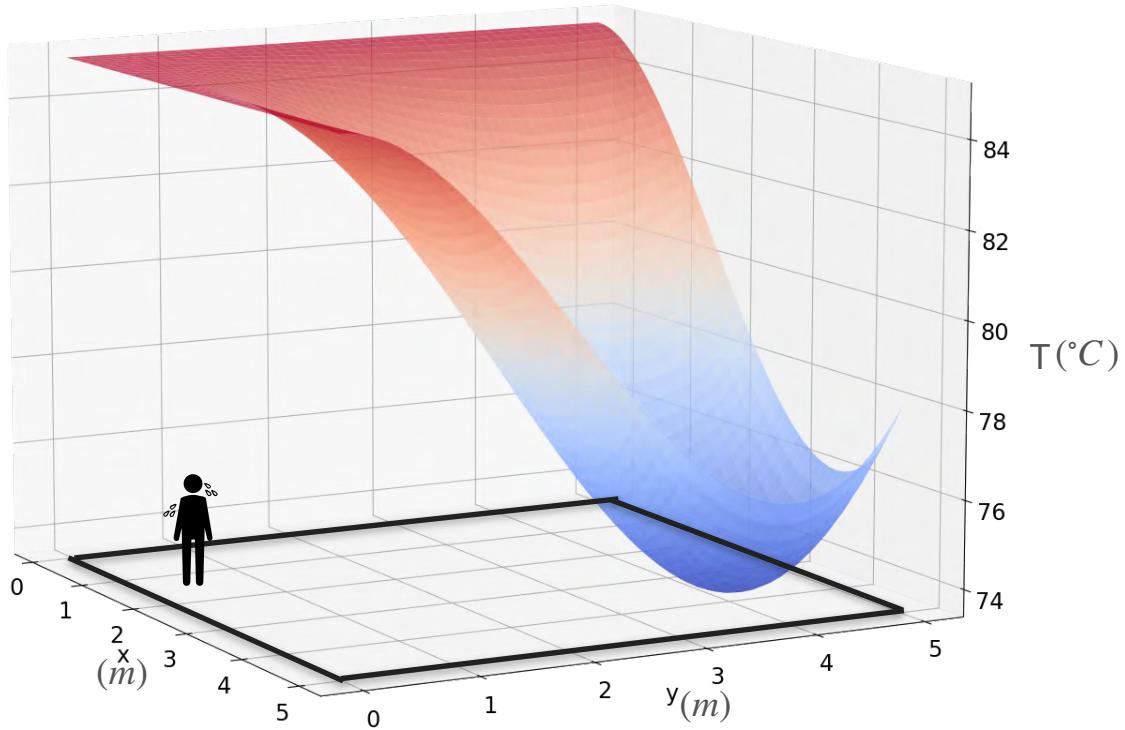
# Motivation for Optimization in Two Variables



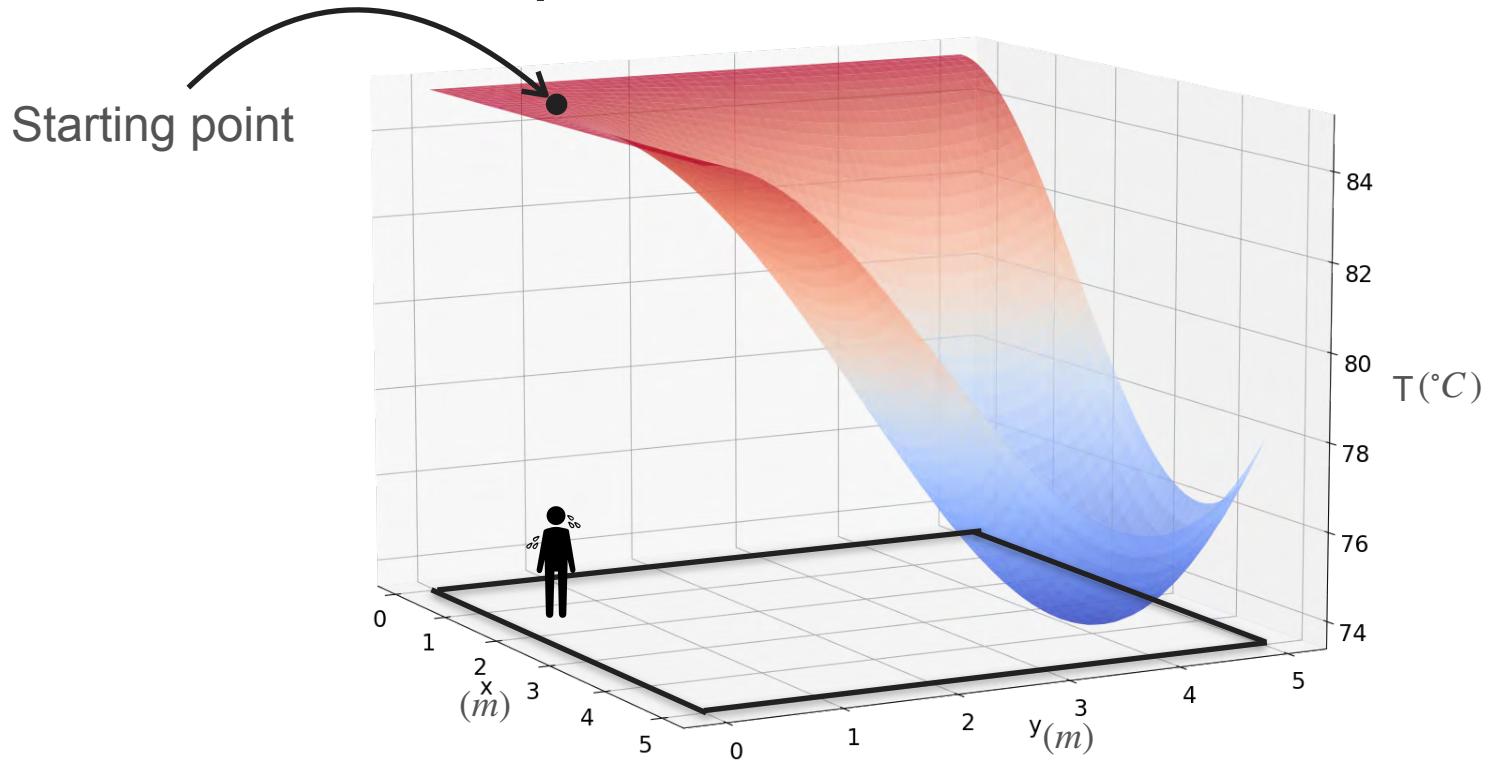
# Motivation for Optimization in Two Variables



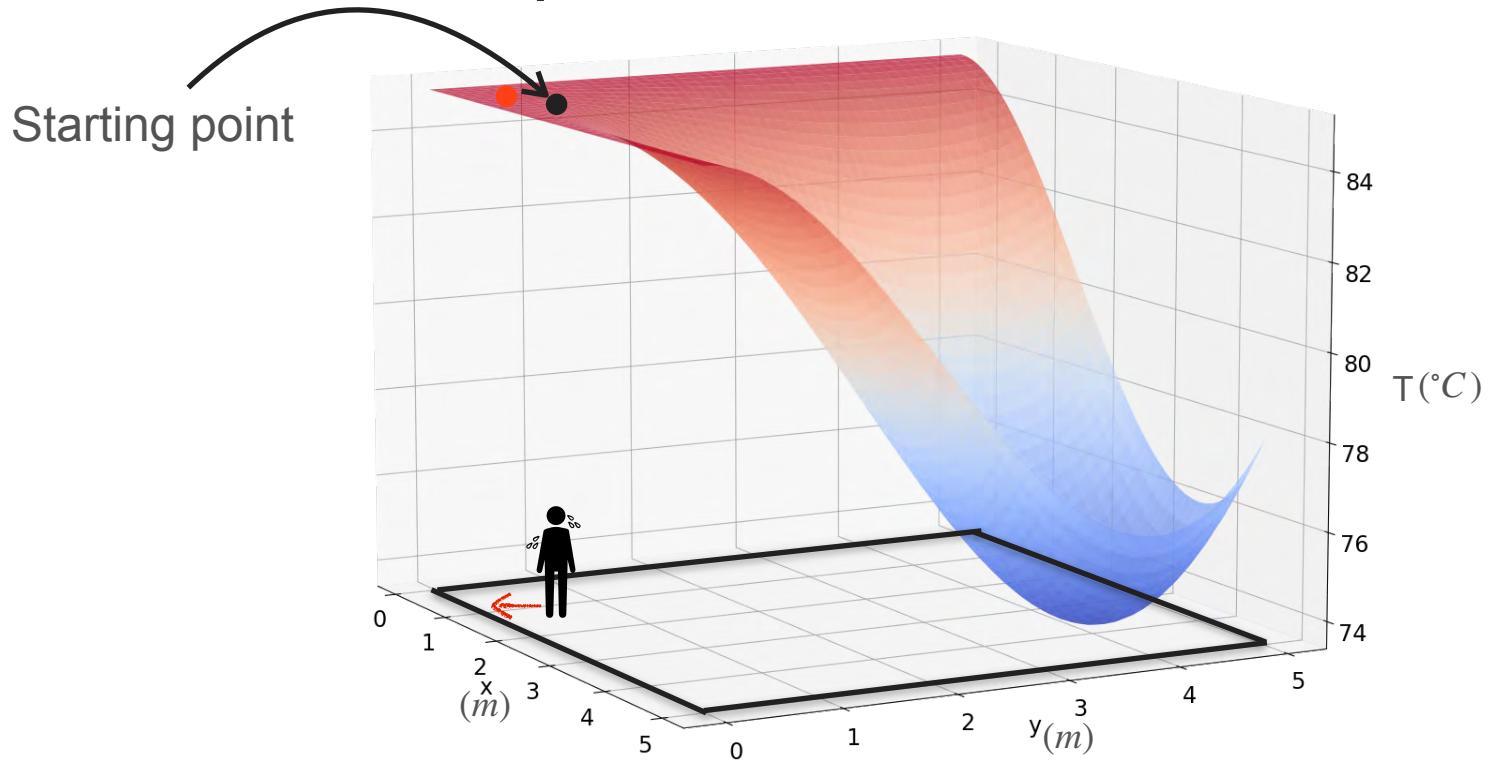
# Motivation for Optimization in Two Variables



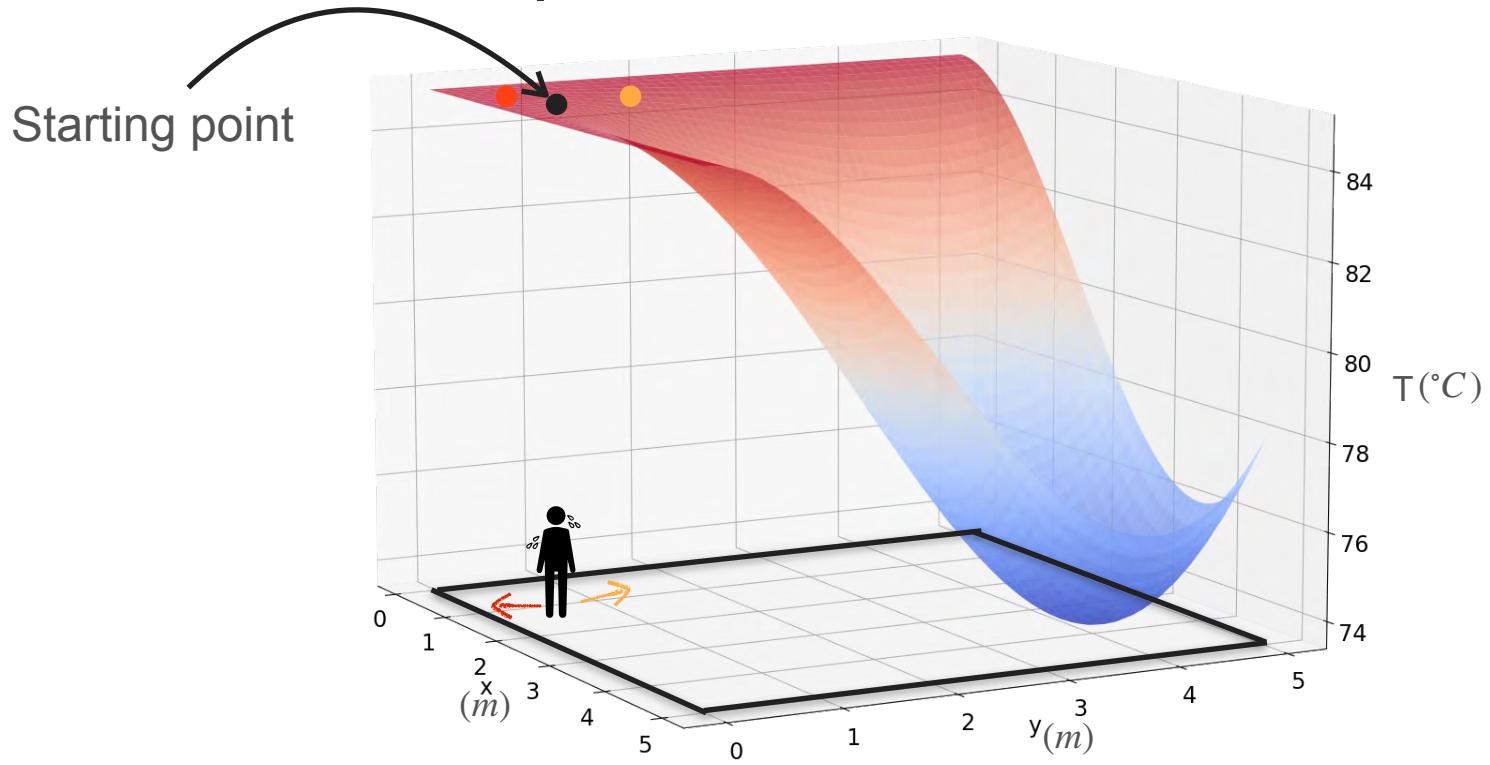
# Motivation for Optimization in Two Variables



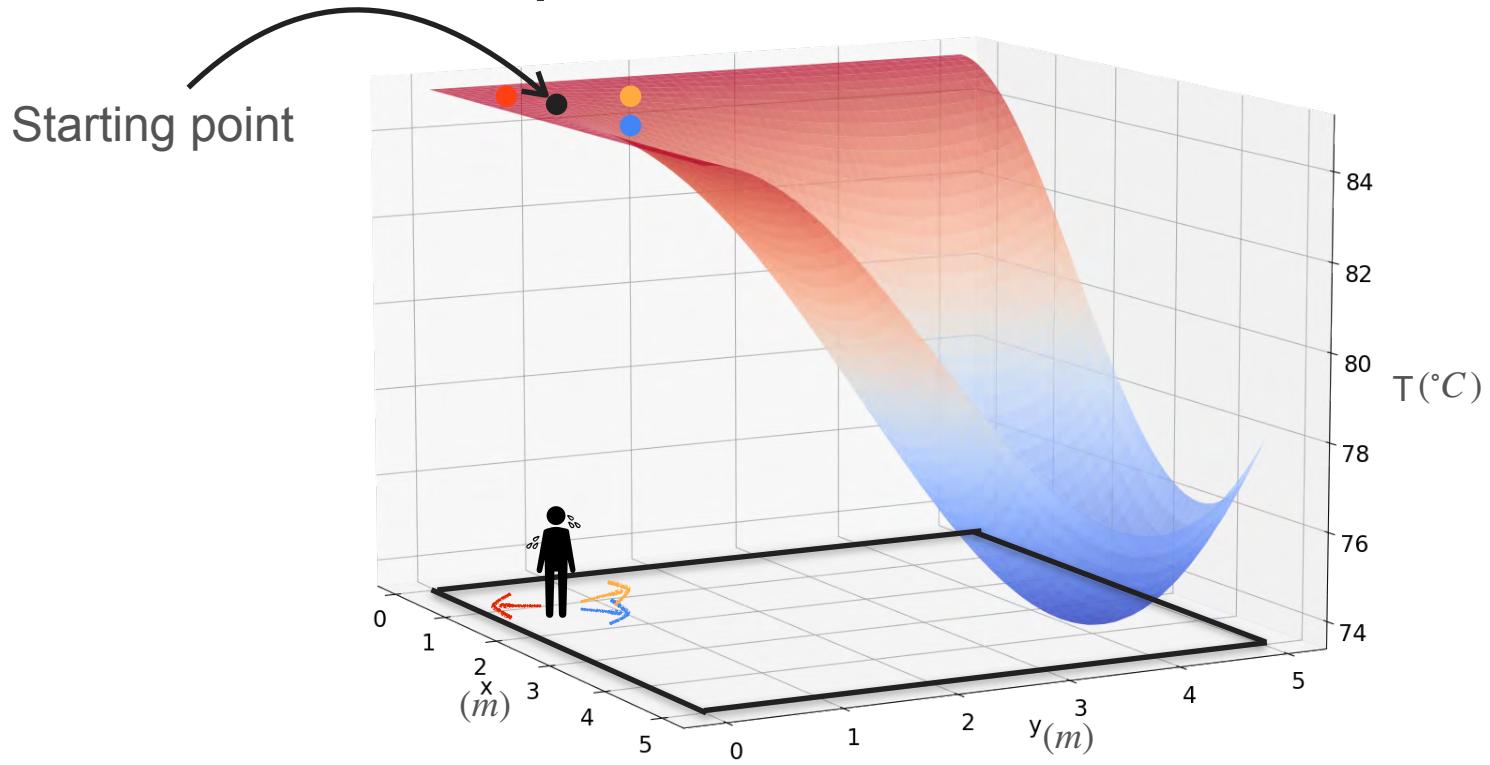
# Motivation for Optimization in Two Variables



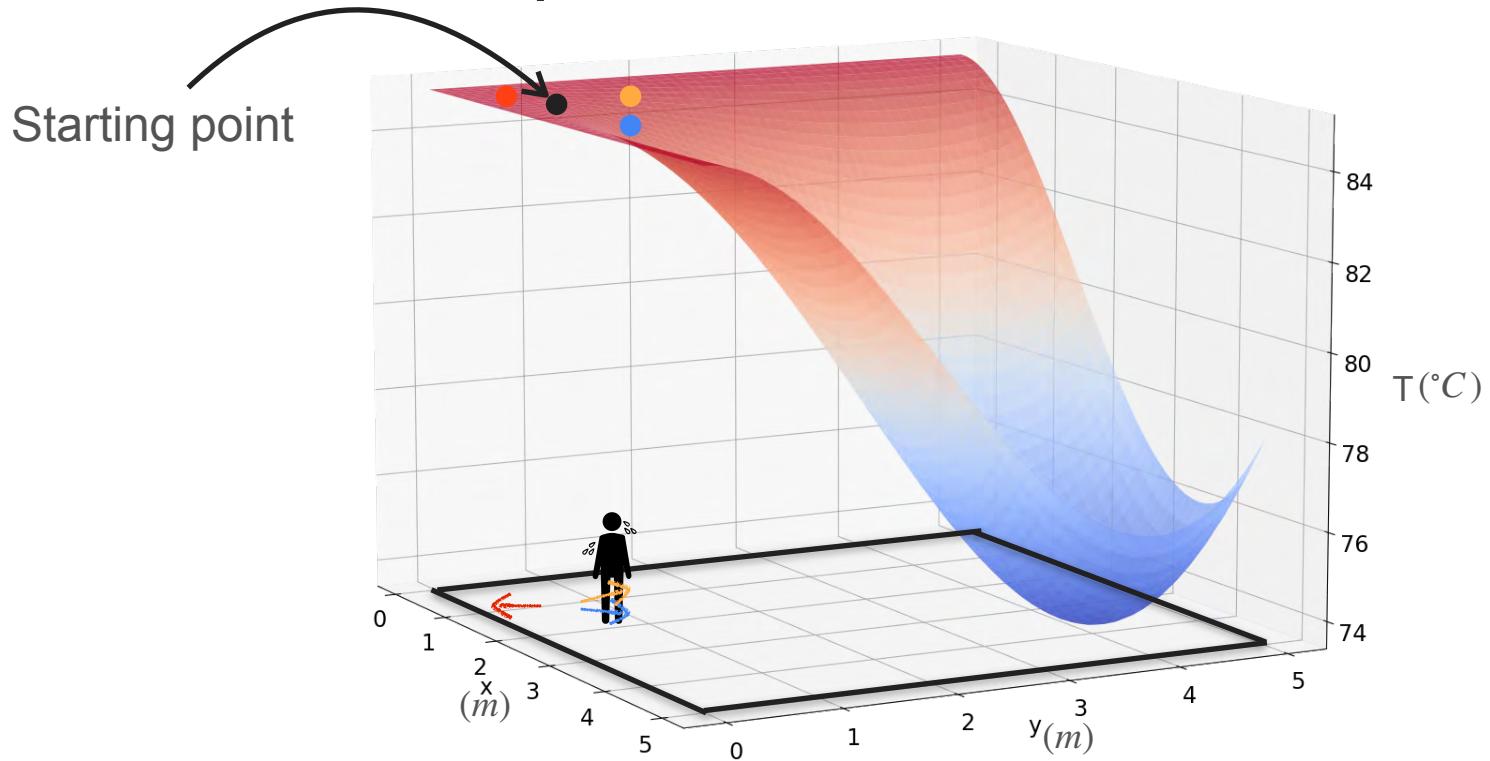
# Motivation for Optimization in Two Variables



# Motivation for Optimization in Two Variables

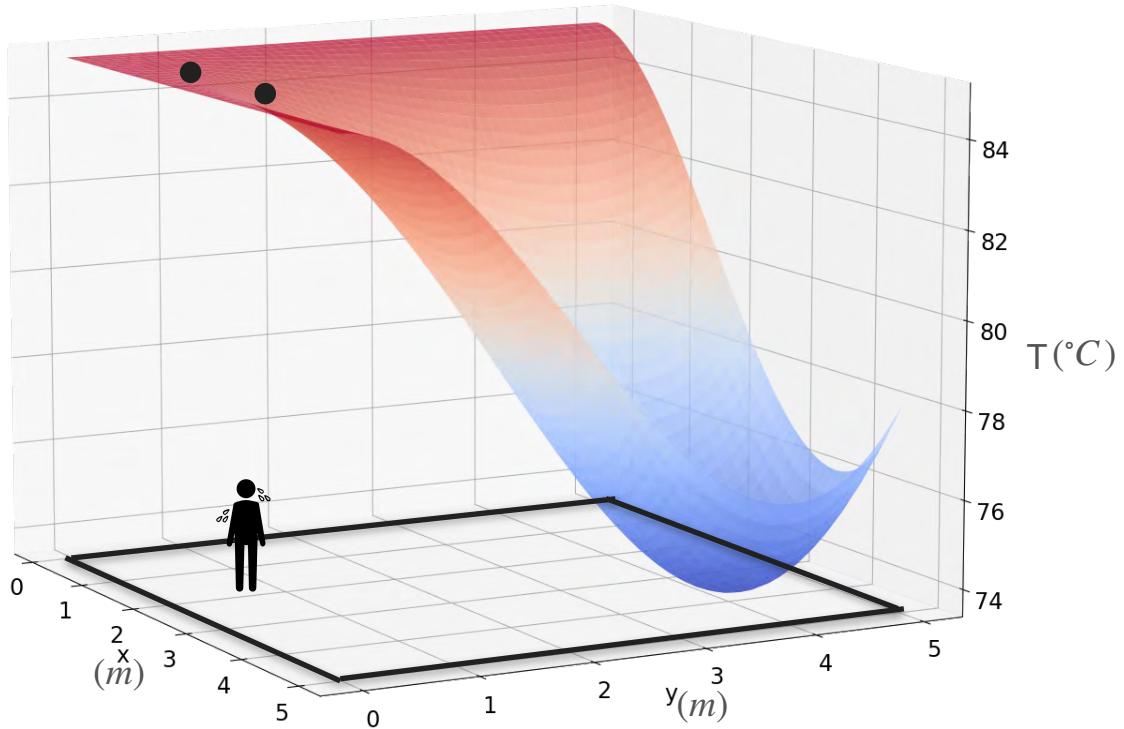


# Motivation for Optimization in Two Variables

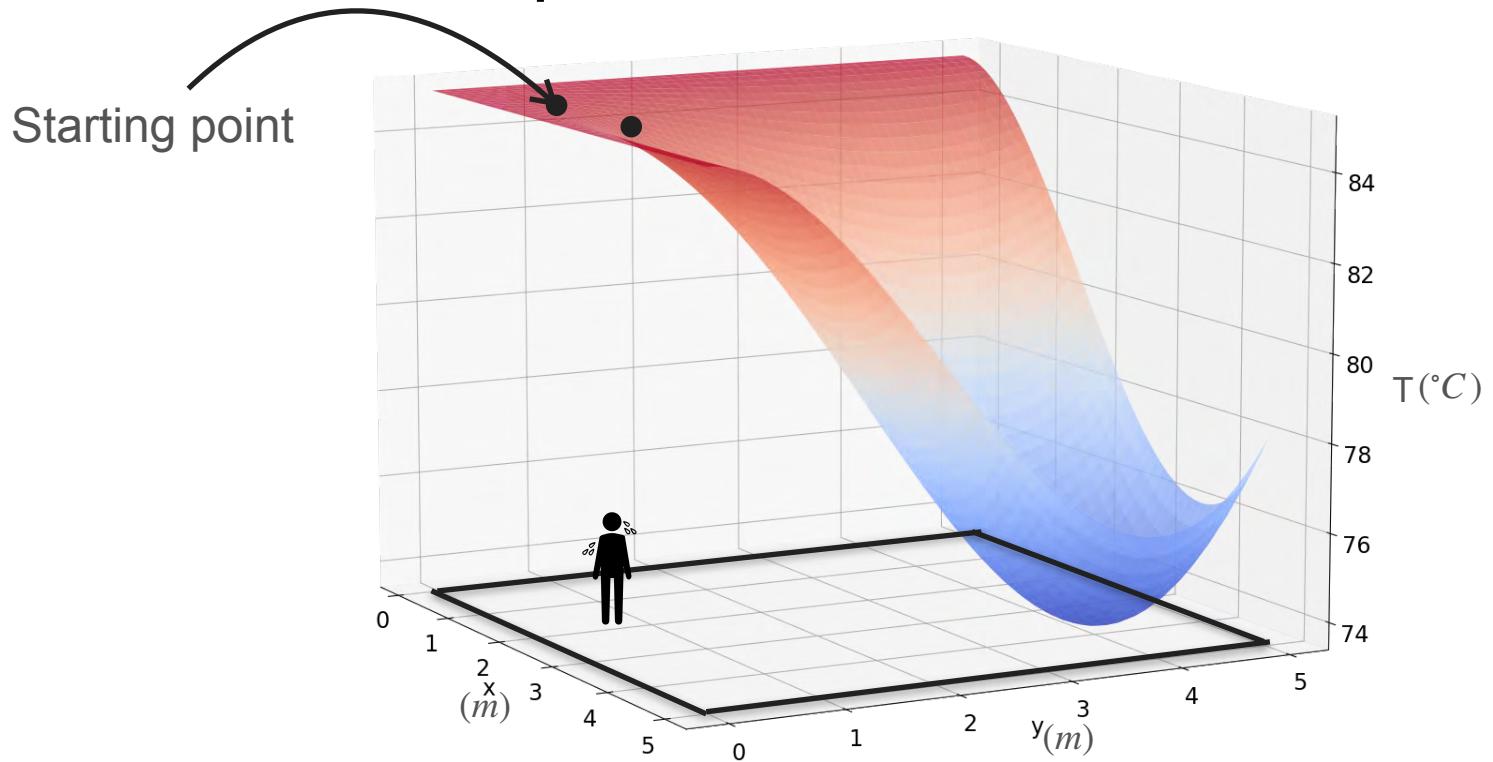


# Motivation for Optimization in Two Variables

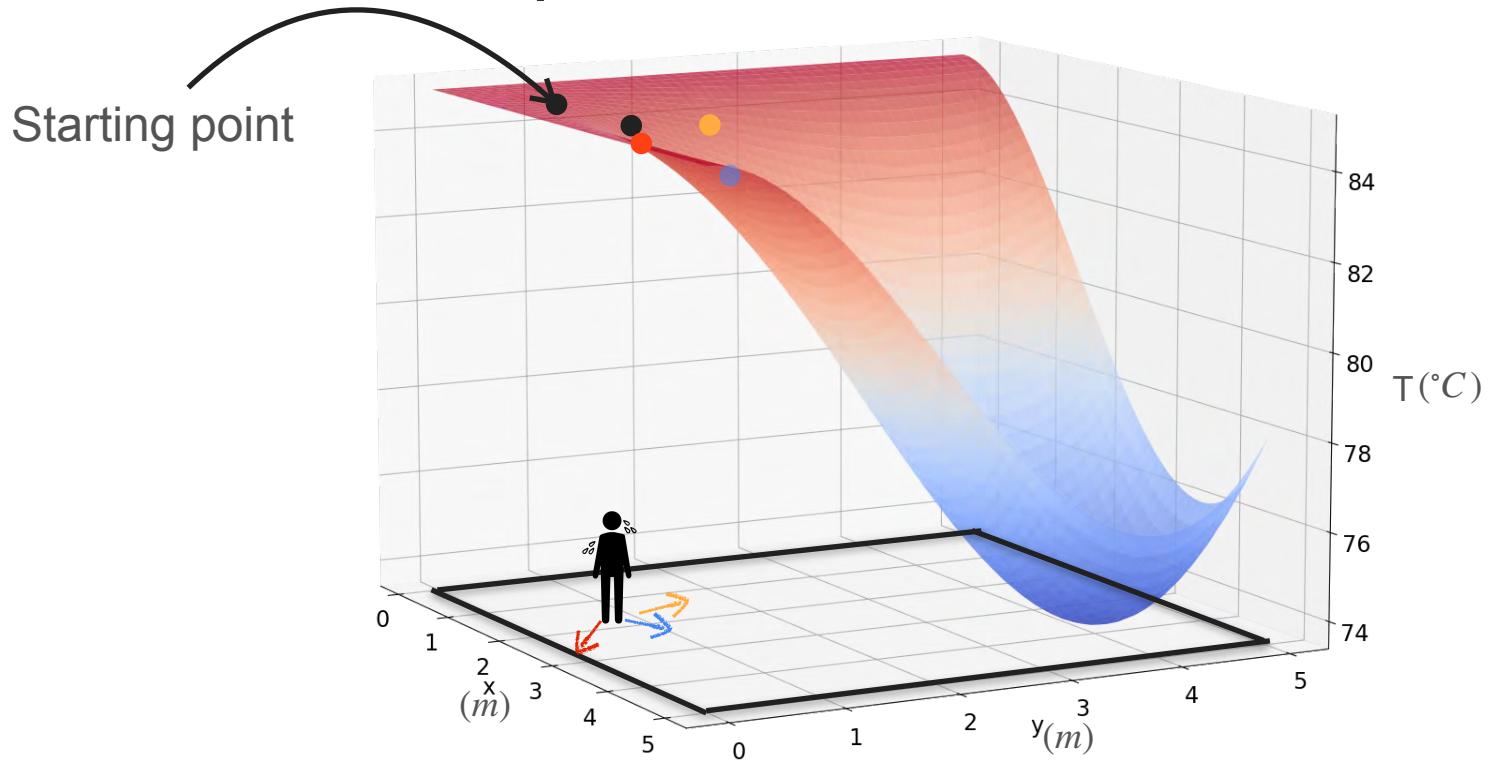
Starting point



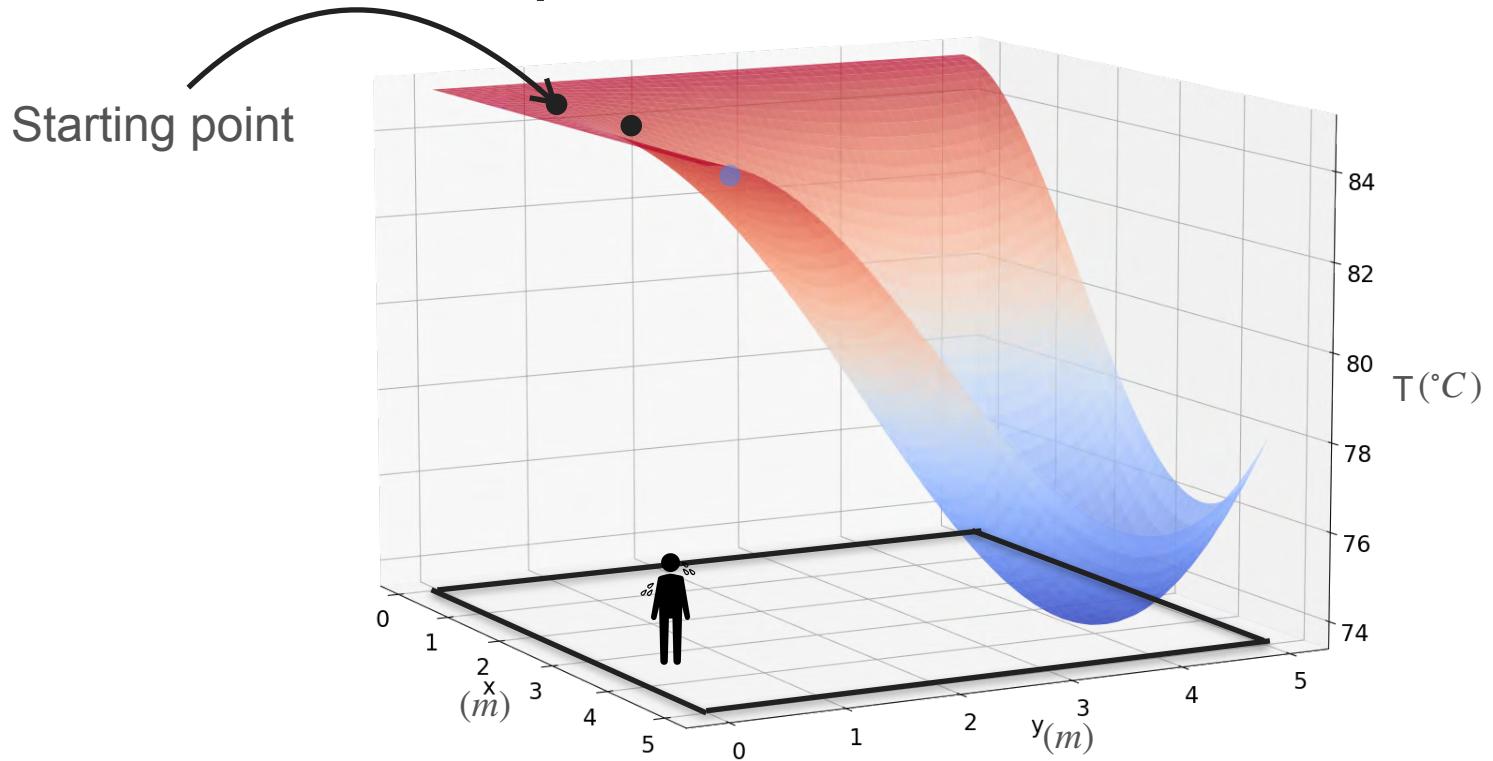
# Motivation for Optimization in Two Variables



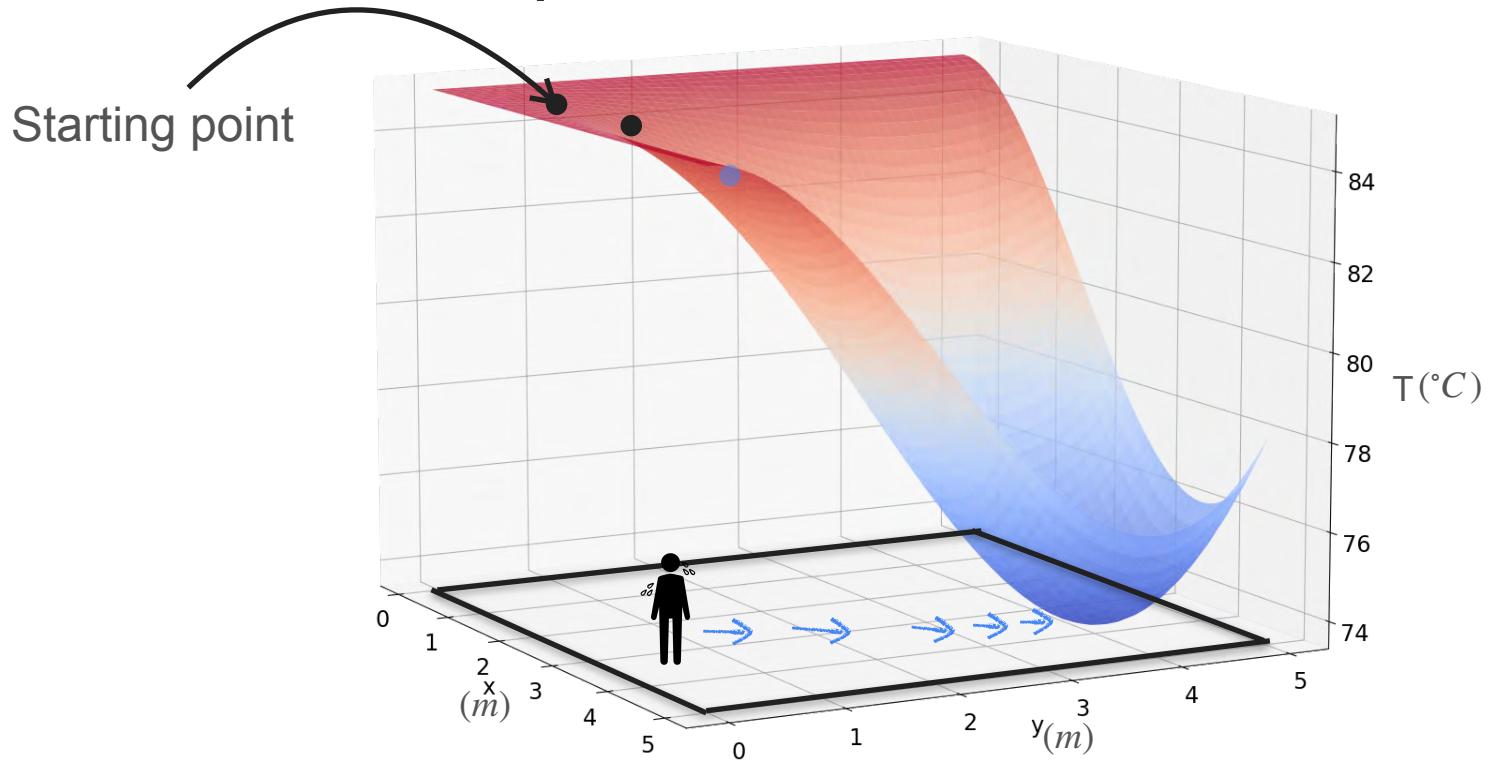
# Motivation for Optimization in Two Variables



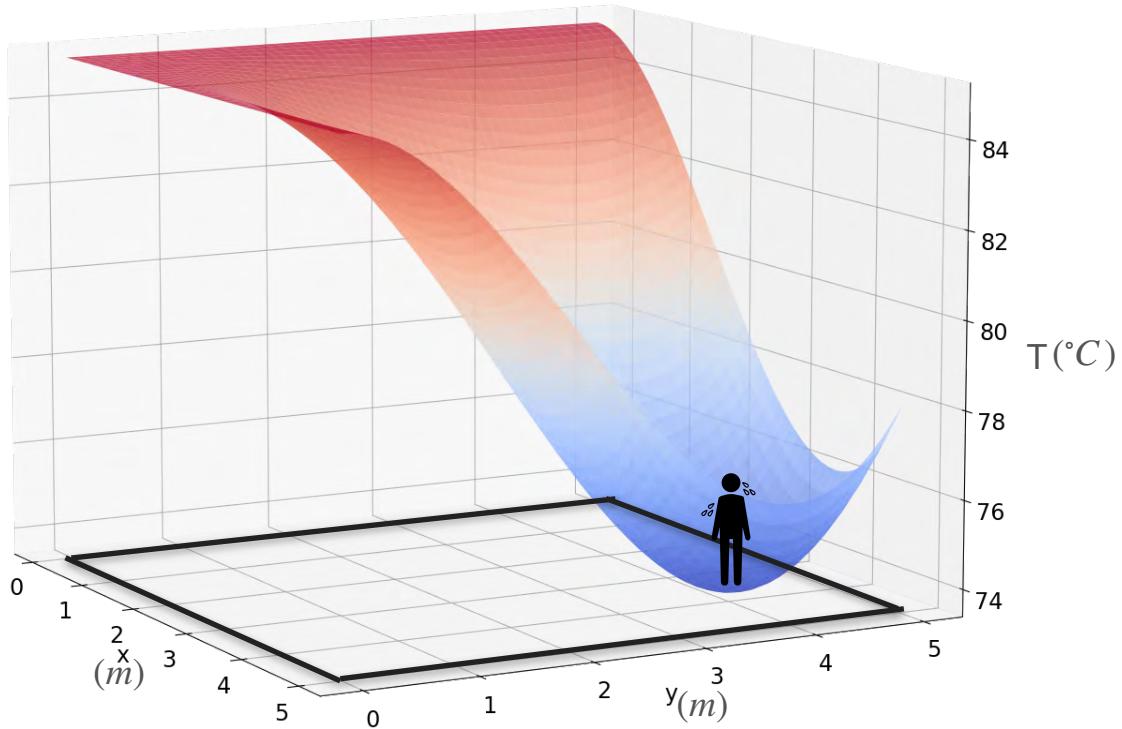
# Motivation for Optimization in Two Variables



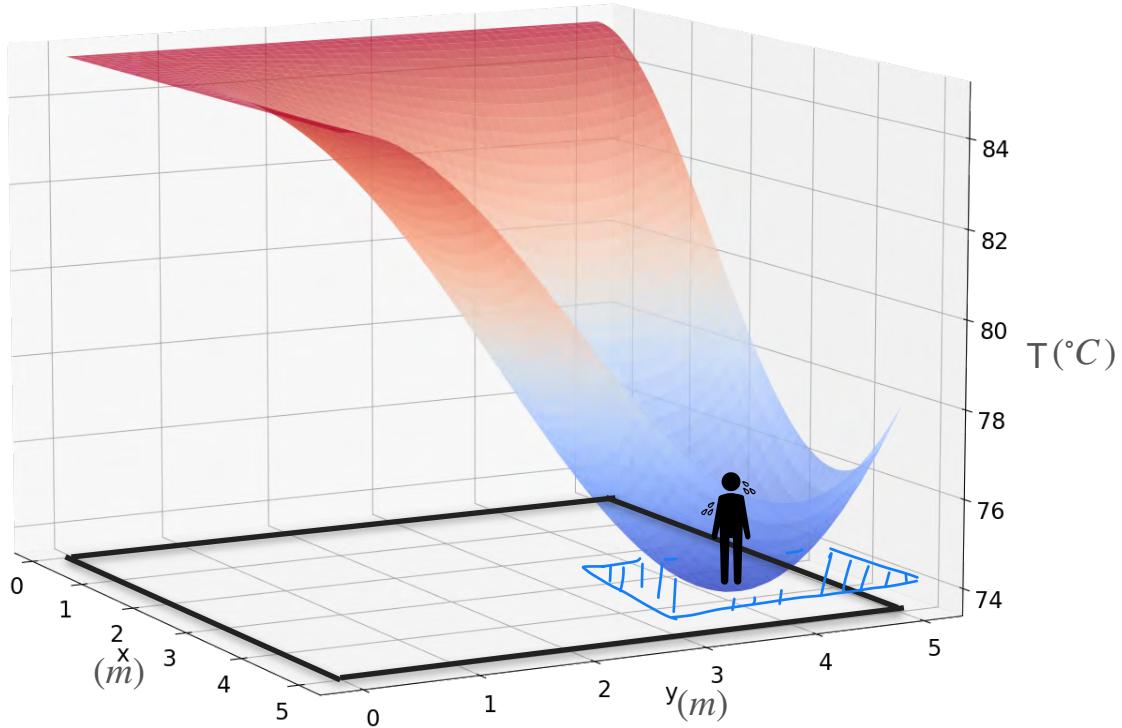
# Motivation for Optimization in Two Variables



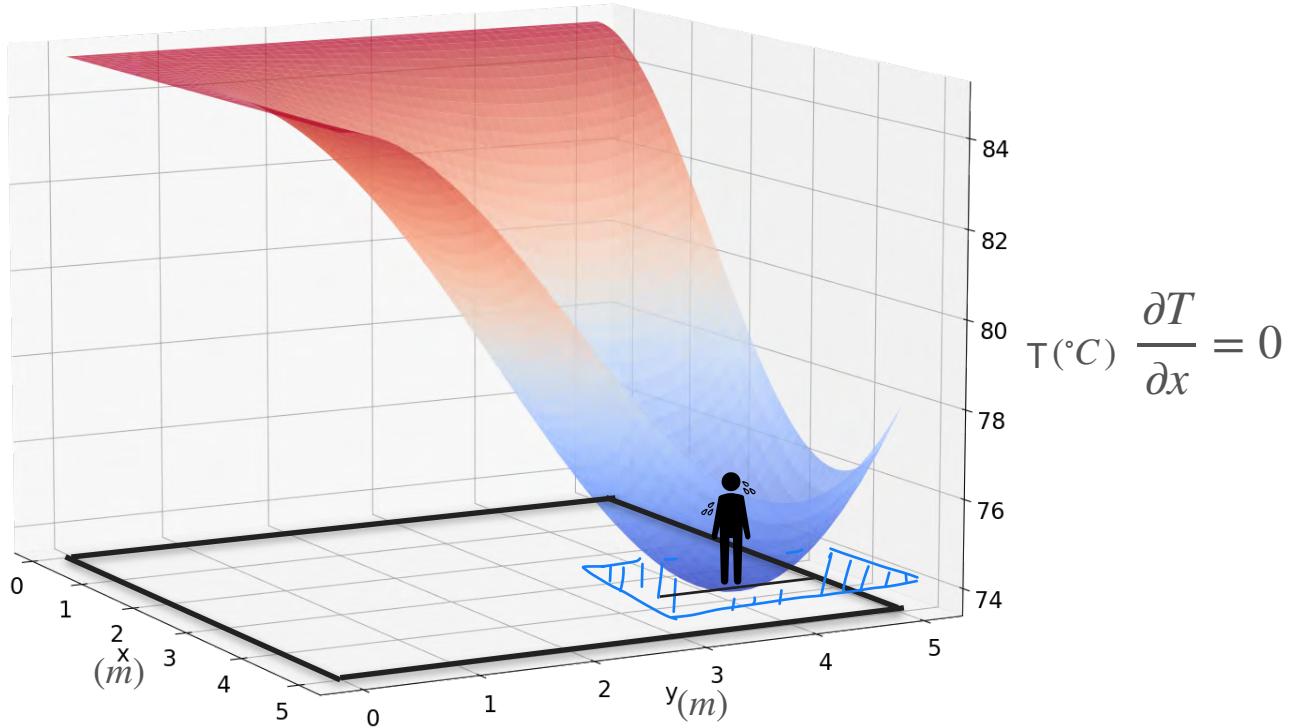
# Motivation for Optimization in Two Variables



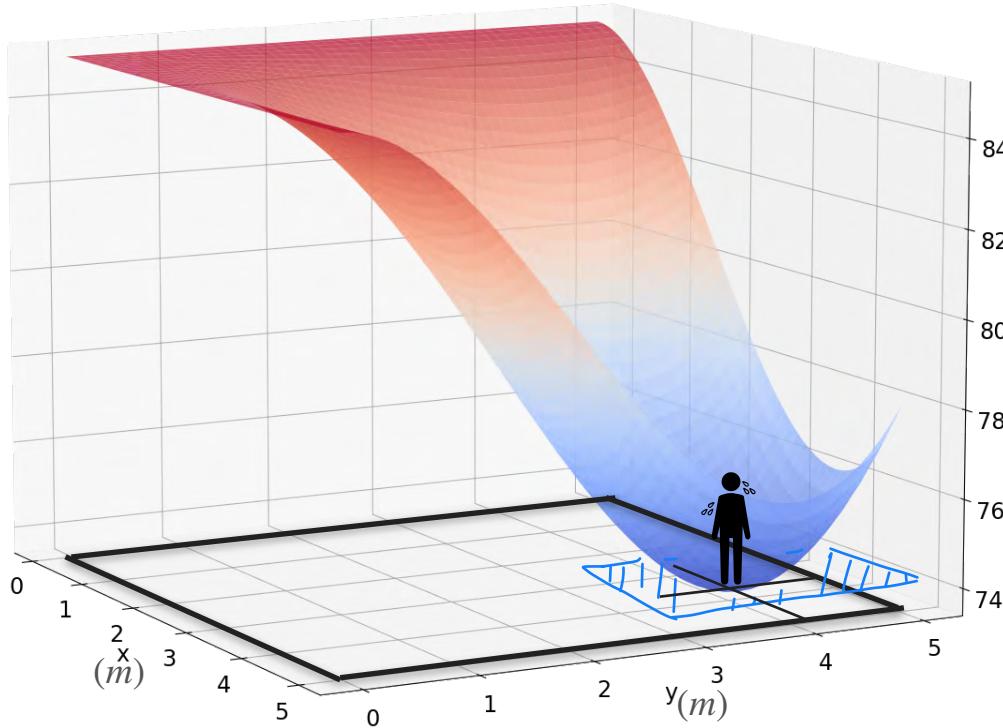
# Motivation for Optimization in Two Variables



# Motivation for Optimization in Two Variables



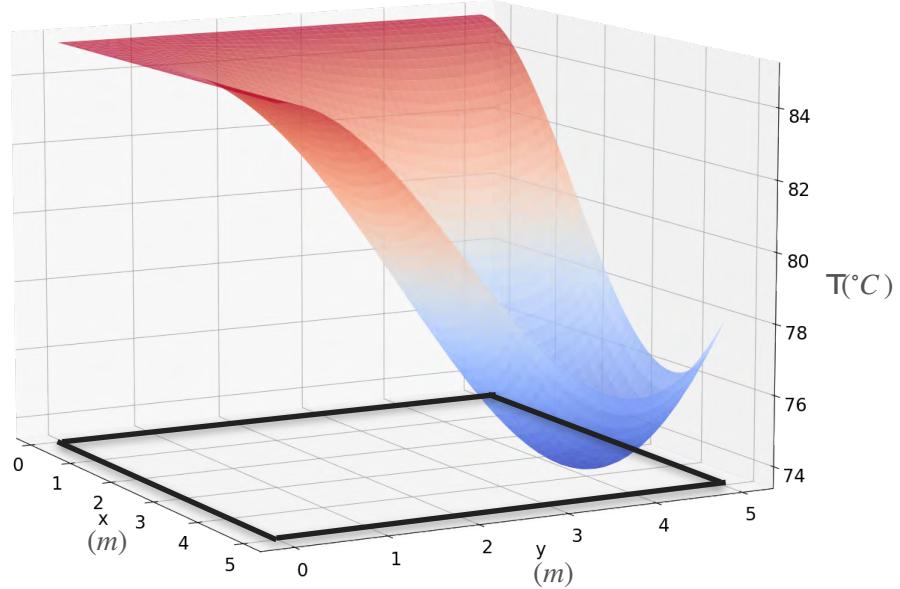
# Motivation for Optimization in Two Variables



$$T(^{\circ}\text{C}) \frac{\partial T}{\partial x} = 0$$

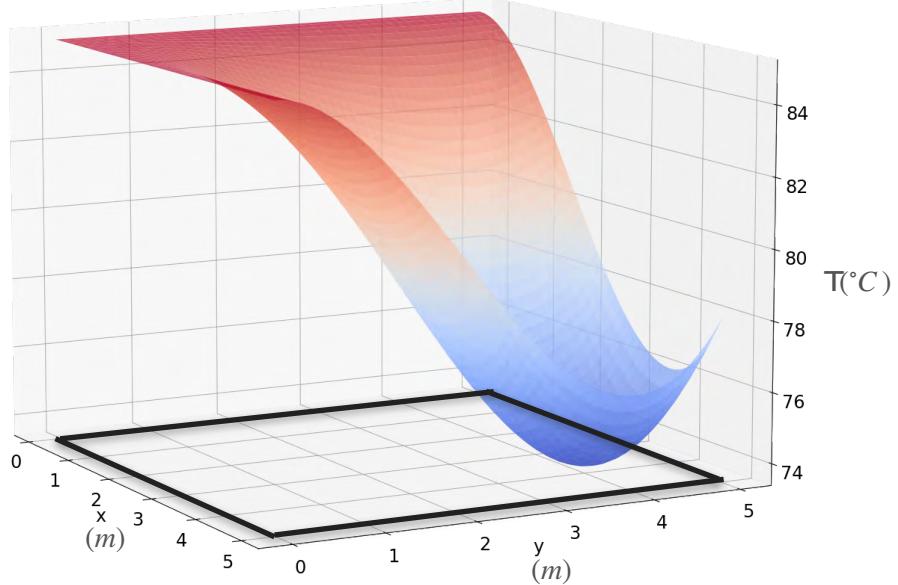
$$\frac{\partial T}{\partial y} = 0$$

# Exercise



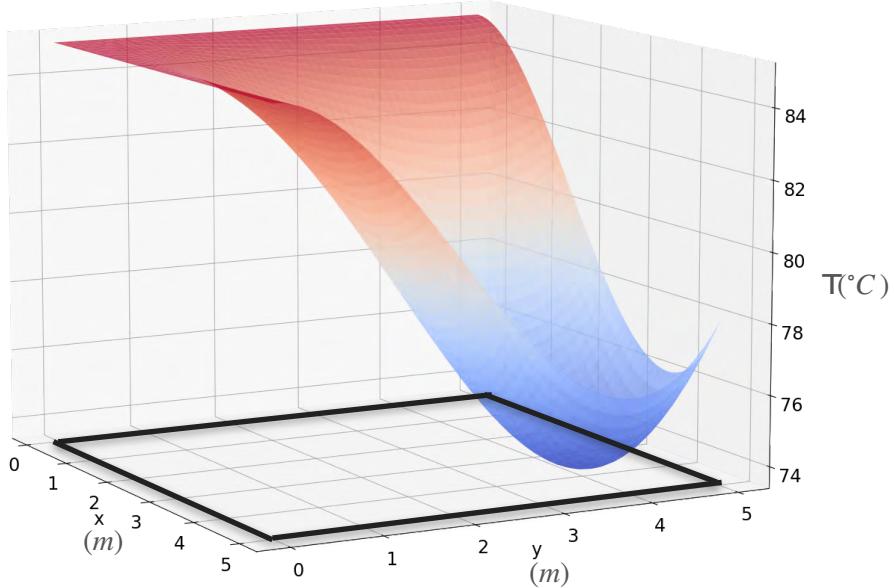
# Exercise

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



# Exercise

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

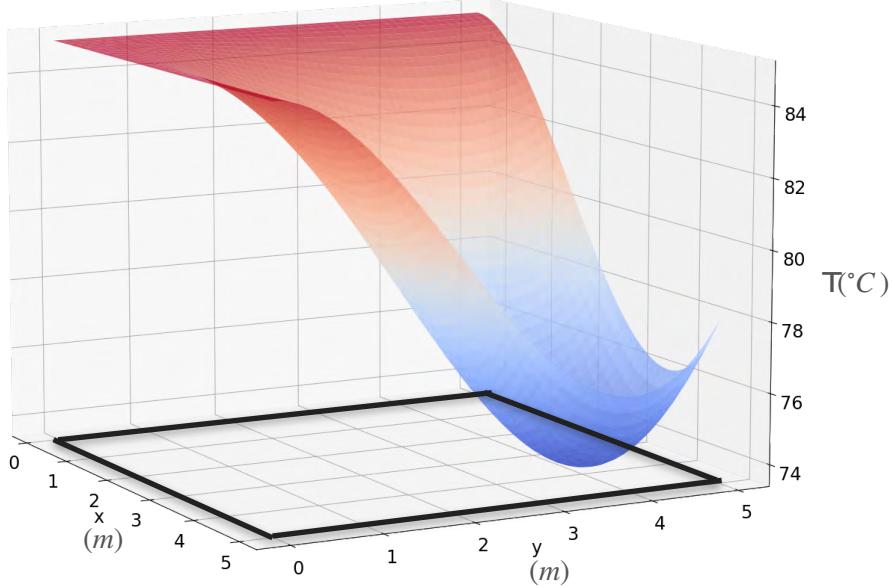


Try and calculate

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

# Exercise

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



Try and calculate

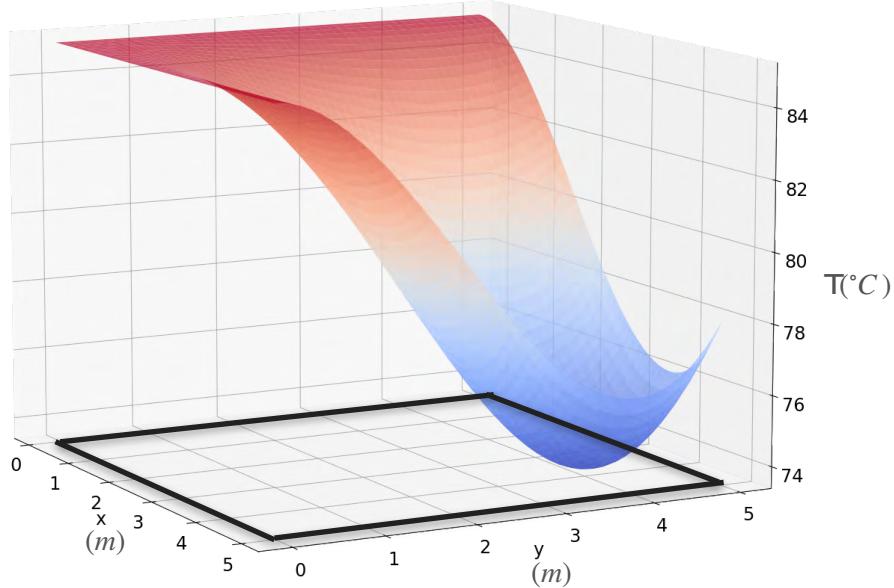
$$\frac{\partial f}{\partial x}$$

and

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

# Motivation for Optimization in Two Variables

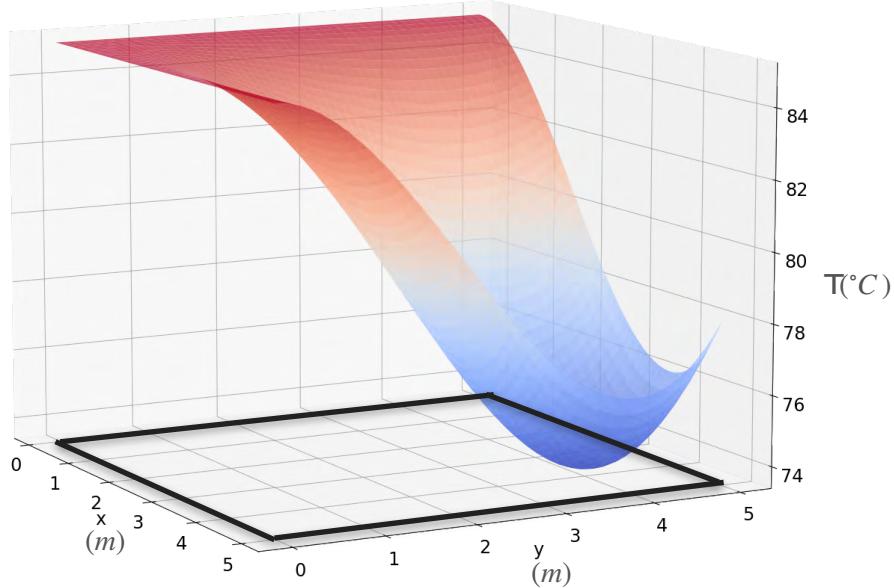
$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

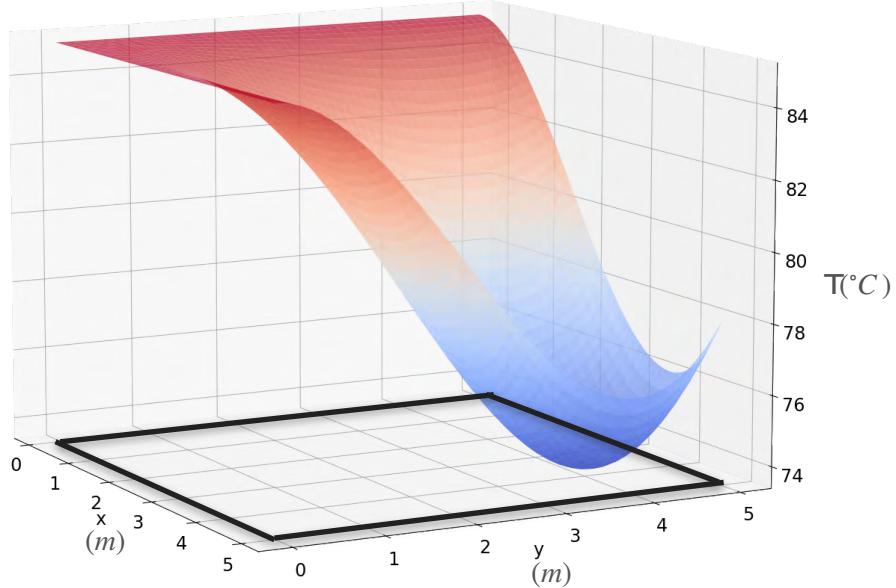
$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6)$$



# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6)$$

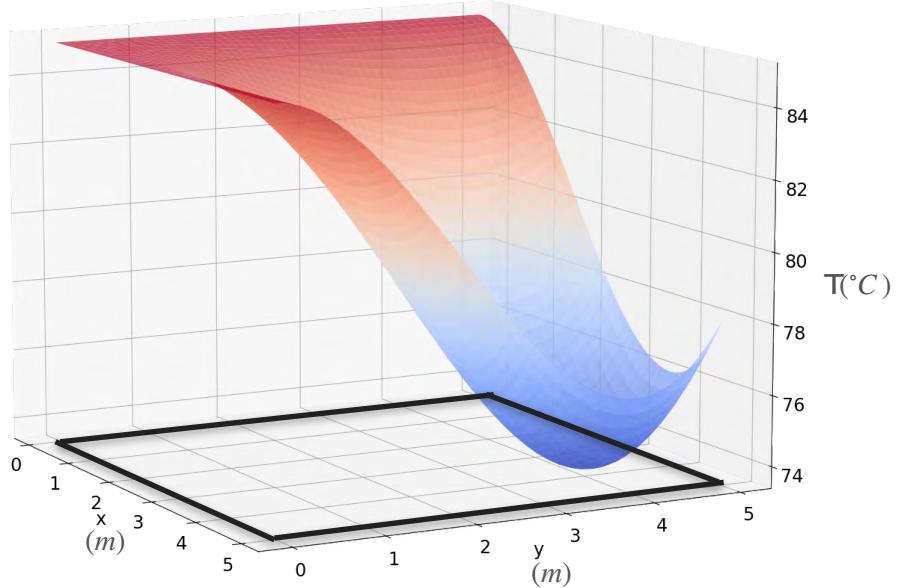


$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12)$$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

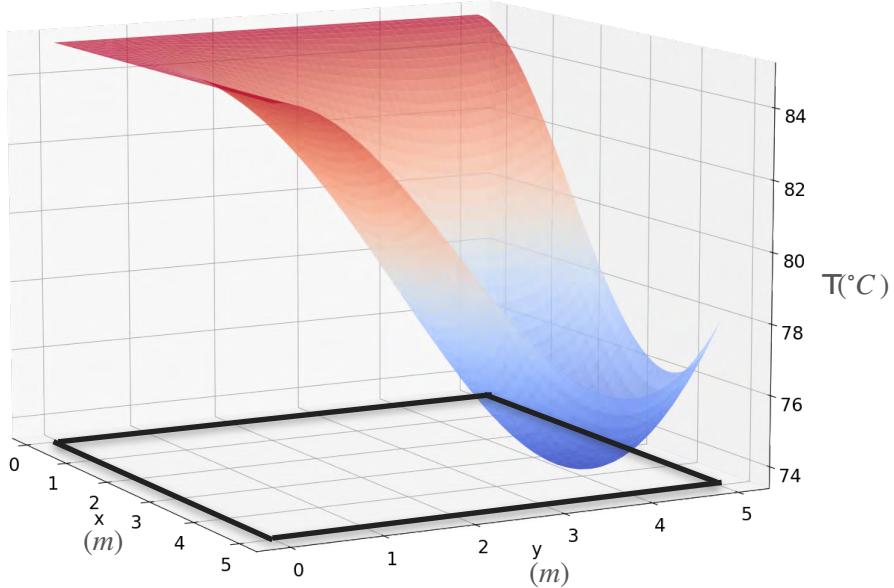
$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$



$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



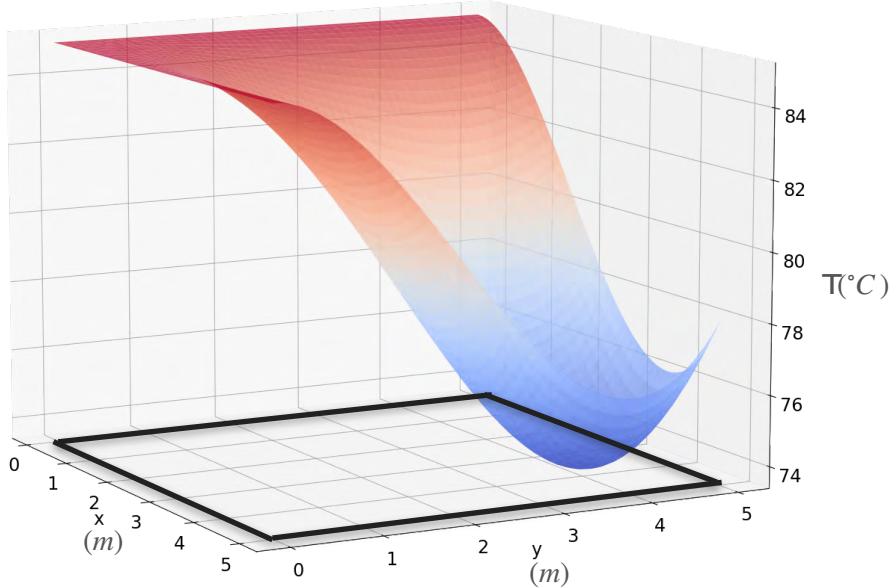
$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$

$$x = 0$$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



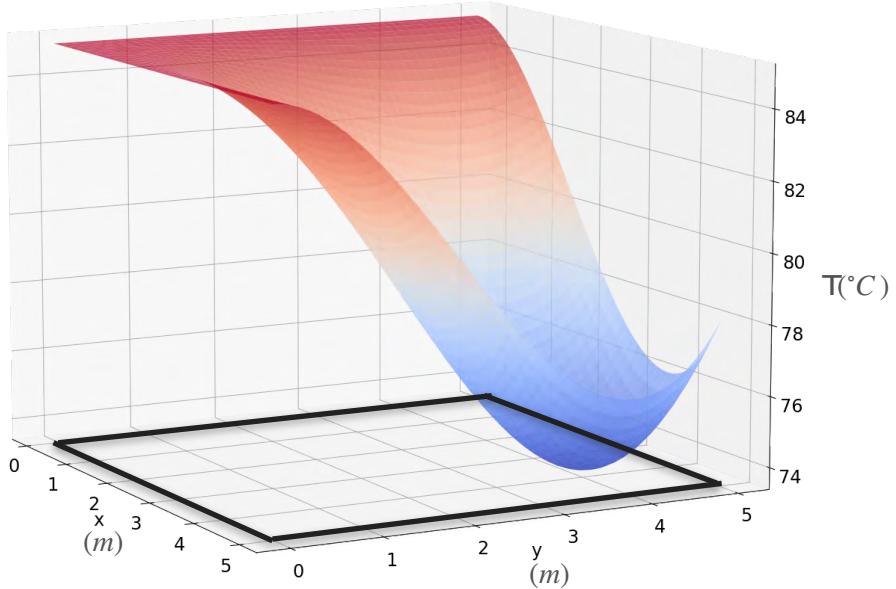
$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$

$$x = 0 \quad x = 4$$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



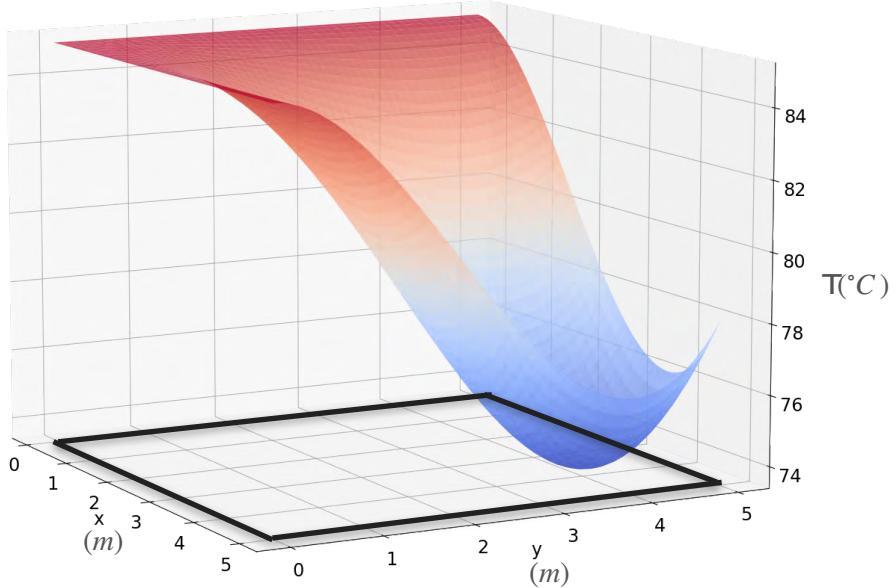
$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$

$$x = 0 \quad x = 4 \quad y = 0$$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



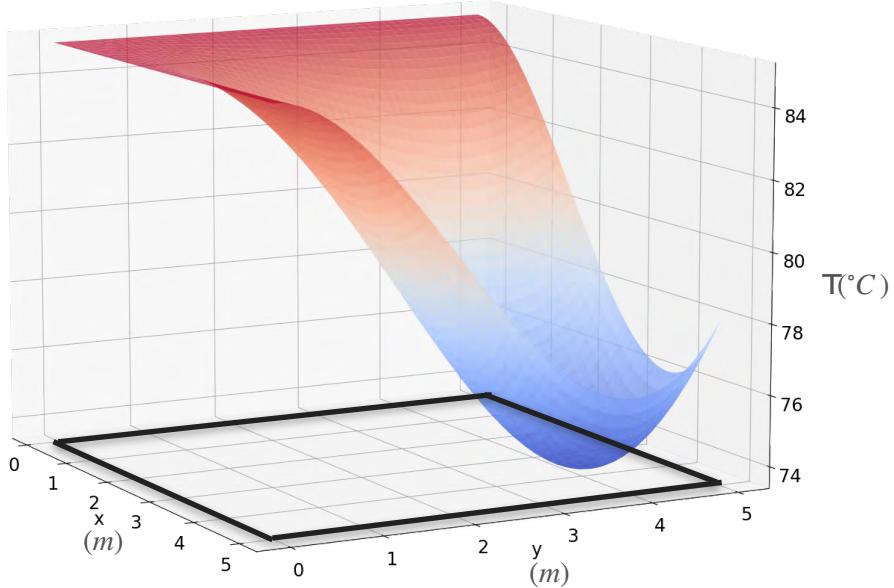
$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$

$$x = 0 \quad x = 4 \quad y = 0 \quad y = 6$$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$

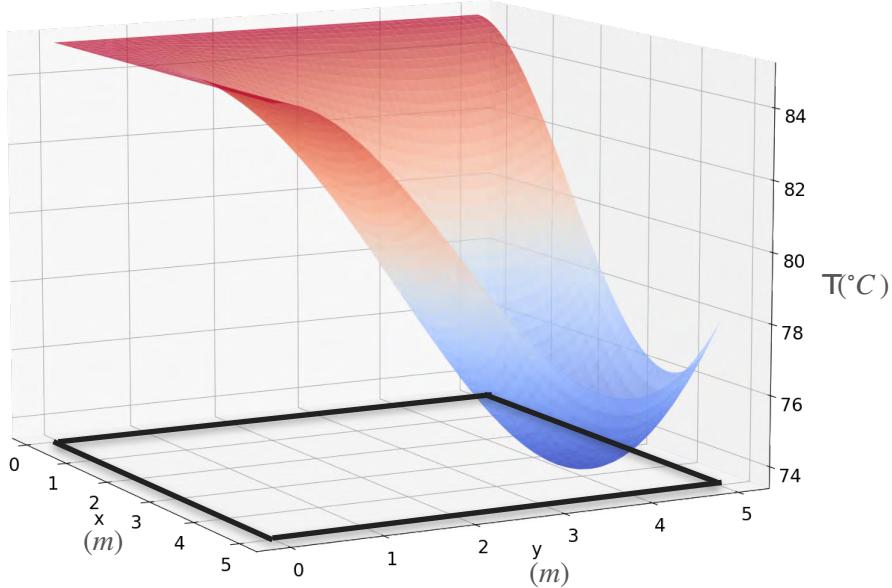
$x = 0$     $x = 4$     $y = 0$     $y = 6$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

$x = 0$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$

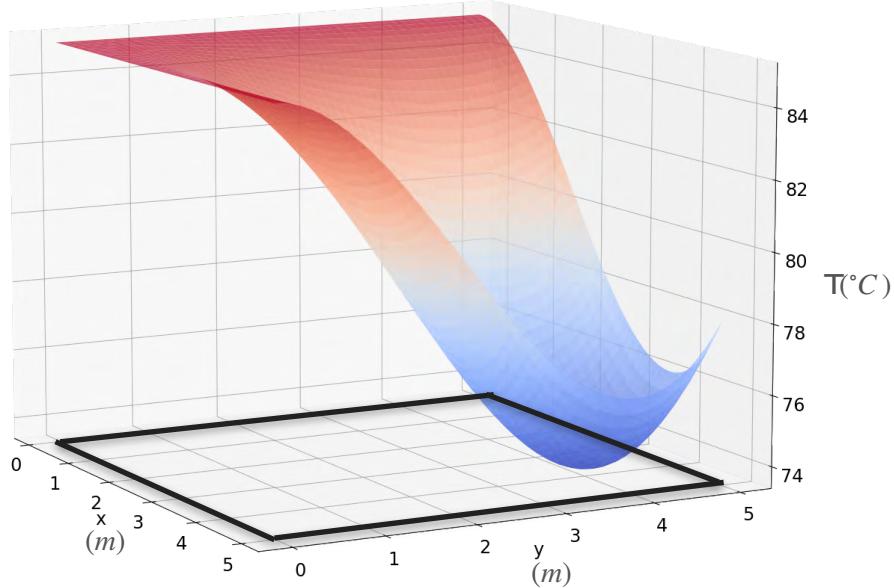
$x = 0$     $x = 4$     $y = 0$     $y = 6$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

$x = 0$     $x = 6$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$

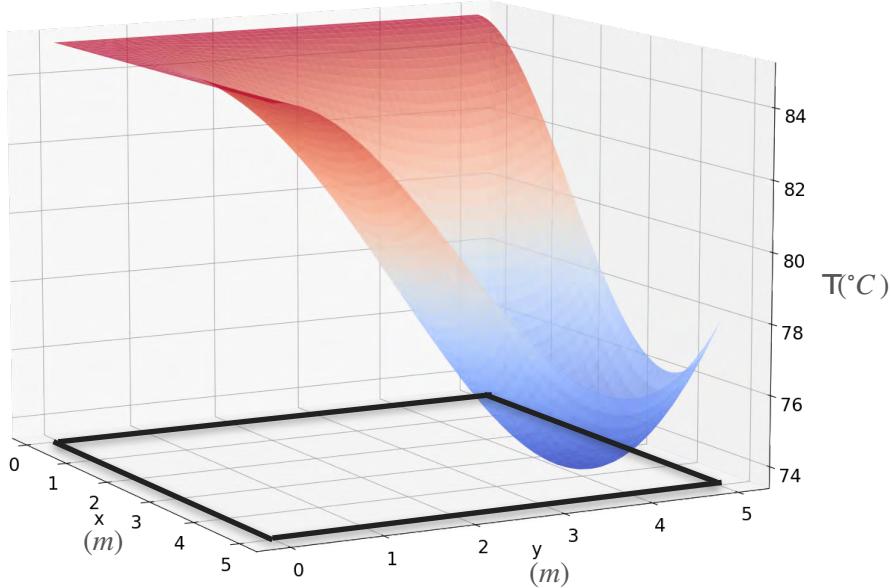
$x = 0$     $x = 4$     $y = 0$     $y = 6$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

$x = 0$     $x = 6$     $y = 0$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6) = 0$$

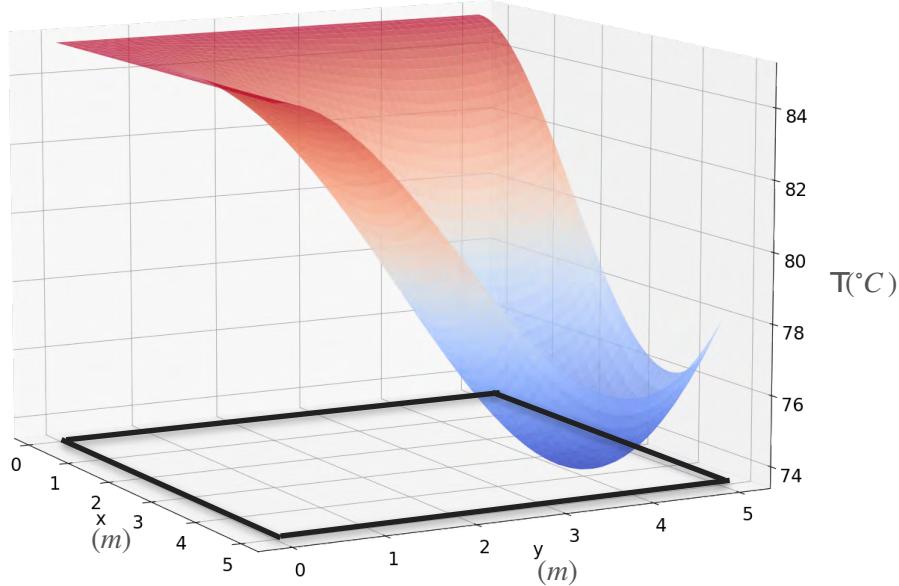
$$\begin{aligned}x &= 0 \\x &= 4 \\y &= 0 \\y &= 6\end{aligned}$$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12) = 0$$

$$\begin{aligned}x &= 0 \\x &= 6 \\y &= 0 \\y &= 4\end{aligned}$$

# Motivation for Optimization in Two Variables

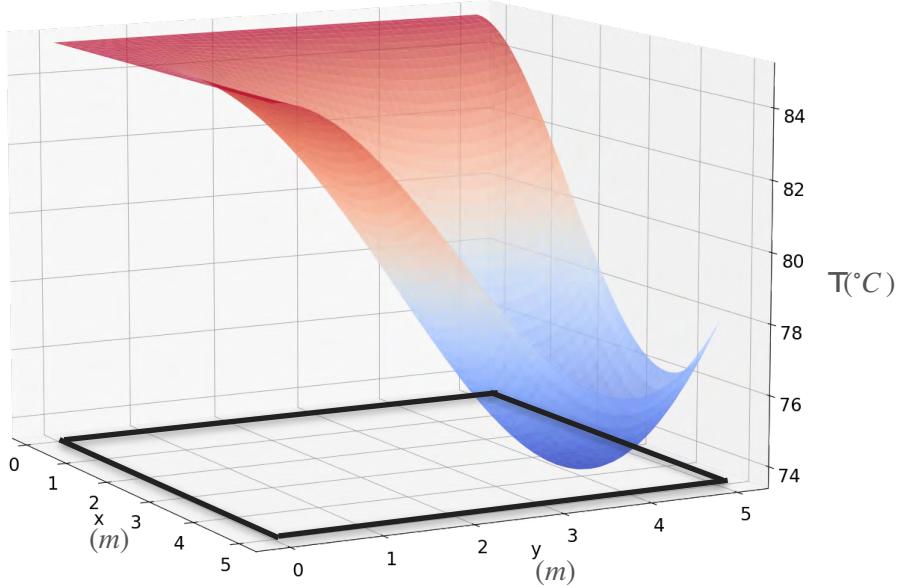
$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



# Motivation for Optimization in Two Variables

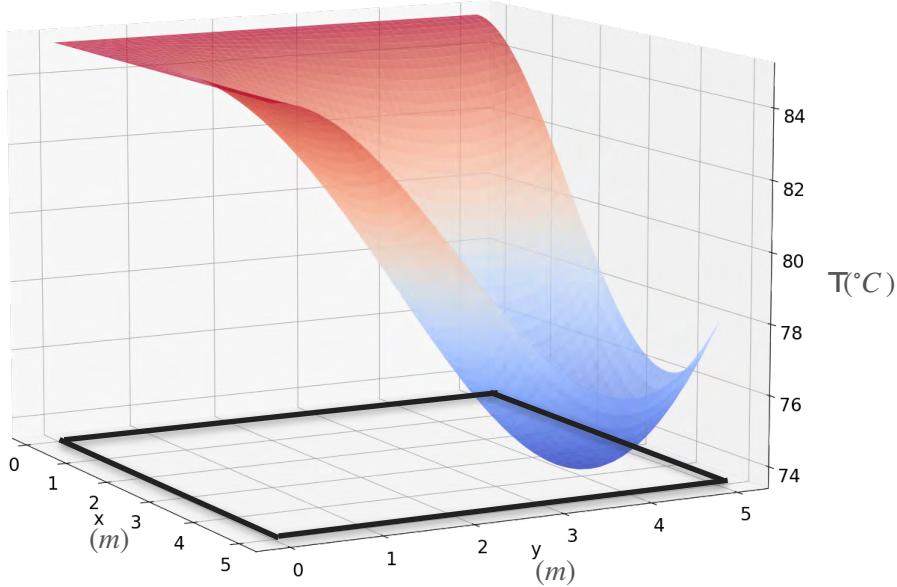
$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

Candidate points for the minima



# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



Candidate points for the minima

$$x = 0$$

$$y = 0$$

$$x = 0, y = 0$$

$$x = 0, y = 4$$

$$x = 0, y = 6$$

$$x = 4, y = 0$$

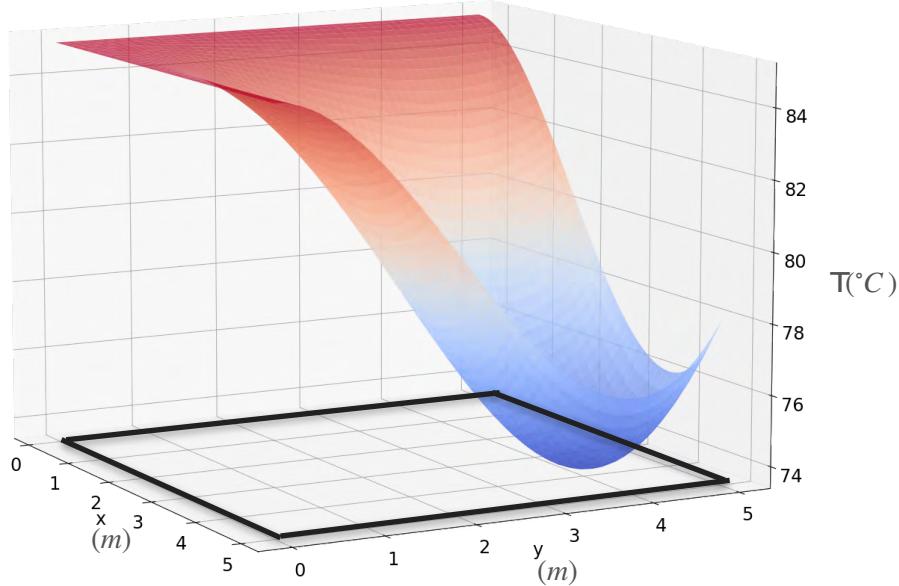
$$x = 4, y = 4$$

$$x = 6, y = 0$$

$$x = 6, y = 6$$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



Candidate points for the minima

$$x = 0$$

$$y = 0$$

$$x = 0, y = 0$$

$$x = 0, y = 4$$

$$x = 0, y = 6$$

Outside

$$x = 4, y = 0$$

$$x = 4, y = 4$$

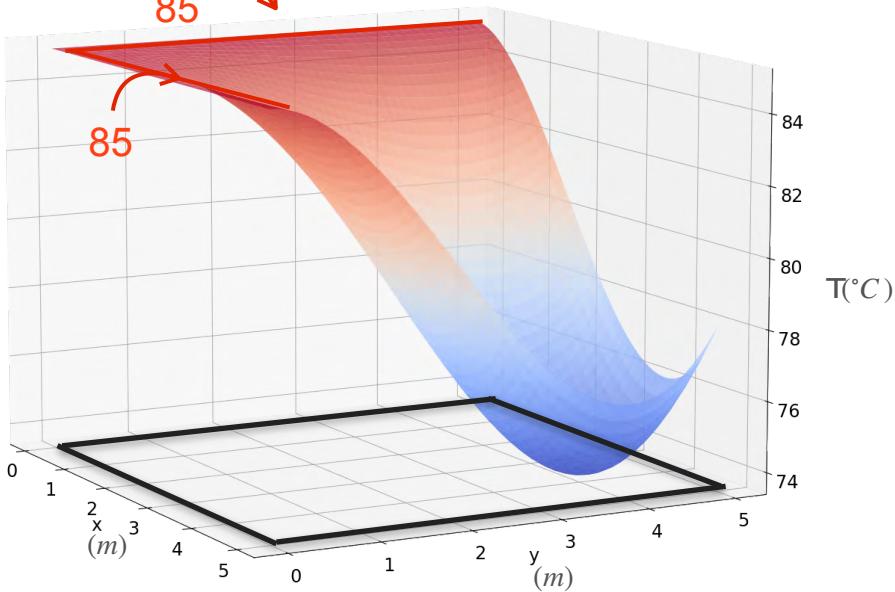
$$x = 6, y = 0$$

Outside

$$x = 6, y = 6$$

# Motivation for Optimization in Two Variables

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$



Candidate points for the minima

$$\begin{aligned}x &= 0 \\y &= 0\end{aligned}$$

Maxima

$$x = 0, y = 0$$

$$x = 0, y = 4$$

$$x = 0, y = 6$$

Outside

$$x = 4, y = 0$$

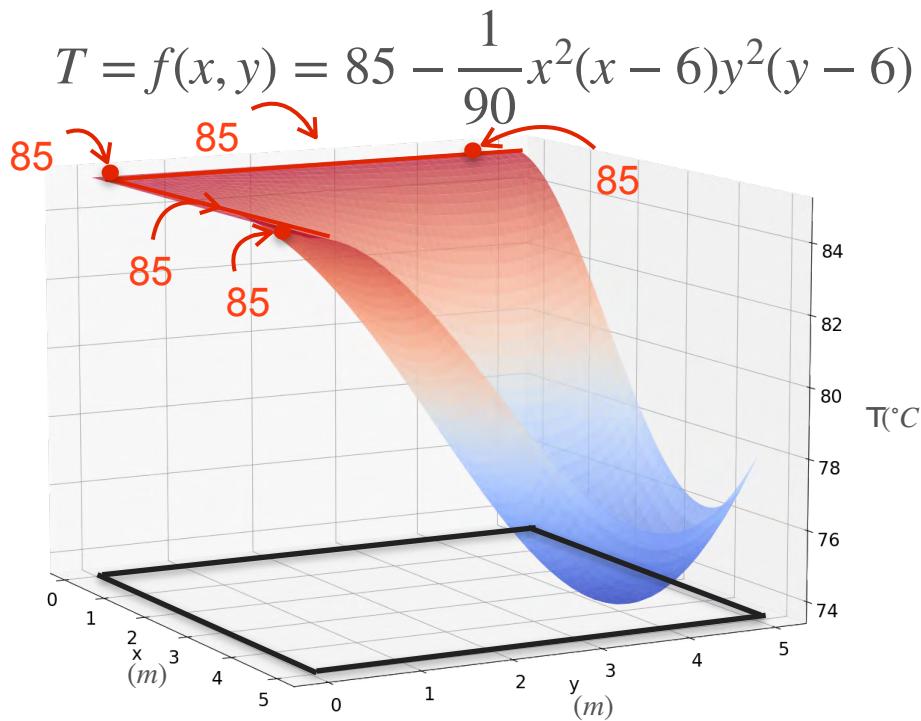
$$x = 4, y = 4$$

$$x = 6, y = 0$$

Outside

$$x = 6, y = 6$$

# Motivation for Optimization in Two Variables



Candidate points for the minima

$$\begin{aligned}x &= 0 \\y &= 0\end{aligned}$$

Maxima

$$\begin{aligned}x &= 0, y = 0 \\x &= 0, y = 4\end{aligned}$$

Maxima

$$\begin{aligned}x &= 0, y = 6\end{aligned}$$

Outside

$$\begin{aligned}x &= 4, y = 0\end{aligned}$$

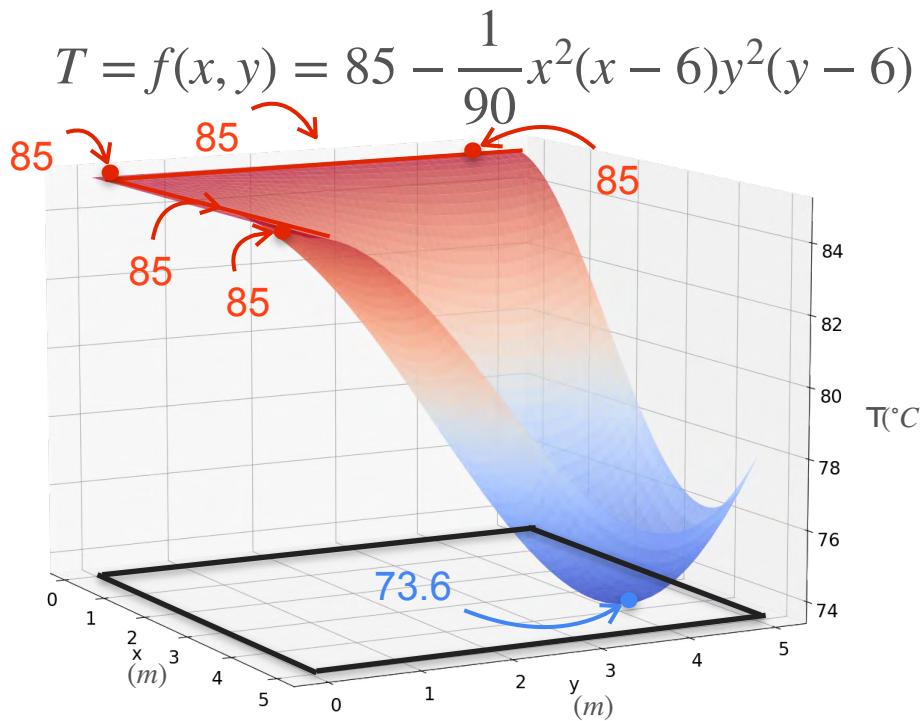
Maxima

$$\begin{aligned}x &= 4, y = 4\end{aligned}$$

$$\begin{aligned}x &= 6, y = 0 \\x &= 6, y = 6\end{aligned}$$

Outside

# Motivation for Optimization in Two Variables



Candidate points for the minima

$$\begin{aligned}x &= 0 \\y &= 0\end{aligned}$$

Maxima

$$\begin{aligned}x &= 0, y = 0 \\x &= 0, y = 4\end{aligned}$$

Maxima

$$\begin{aligned}x &= 0, y = 6\end{aligned}$$

Outside

$$\begin{aligned}x &= 4, y = 0\end{aligned}$$

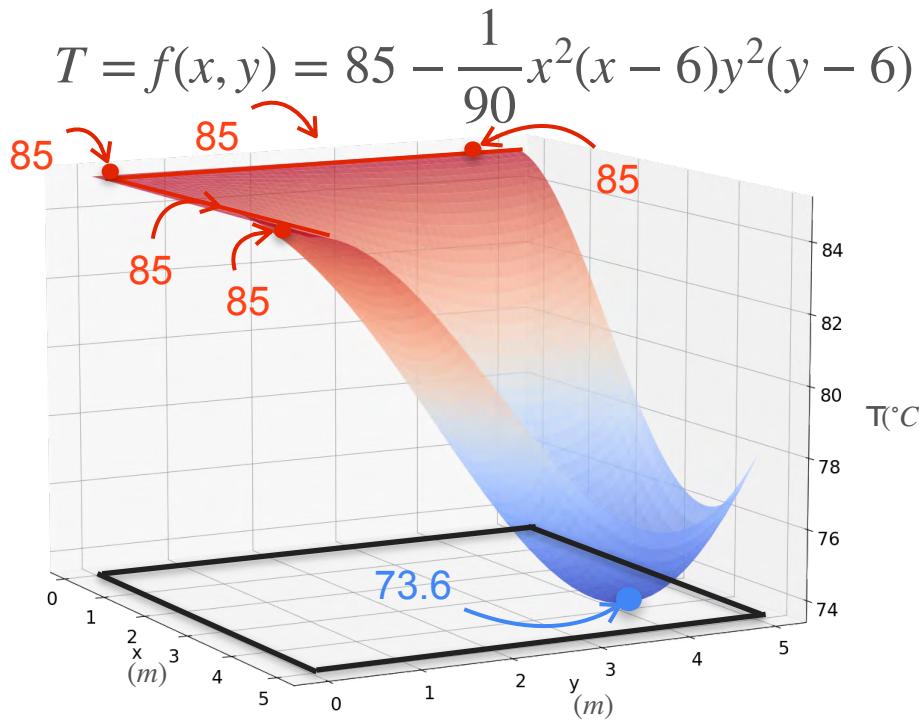
Maxima

$$\begin{aligned}x &= 4, y = 4\end{aligned}$$

Outside

$$\begin{aligned}x &= 6, y = 0 \\x &= 6, y = 6\end{aligned}$$

# Motivation for Optimization in Two Variables



Candidate points for the minima

$$\begin{aligned}x &= 0 \\y &= 0\end{aligned}$$

Maxima

$$\begin{aligned}x &= 0, y = 0 \\x &= 0, y = 4\end{aligned}$$

Maxima

$$\begin{aligned}x &= 0, y = 6\end{aligned}$$

Outside

$$\begin{aligned}x &= 4, y = 0\end{aligned}$$

Maxima

$$\begin{aligned}x &= 4, y = 4\end{aligned}$$

Minimum

$$\begin{aligned}x &= 6, y = 0 \\x &= 6, y = 6\end{aligned}$$

Outside



DeepLearning.AI

# Gradients and Gradient Descent

---

**Optimization using gradients**  
**- Analytical method**

# Linear Regression: Analytical Approach

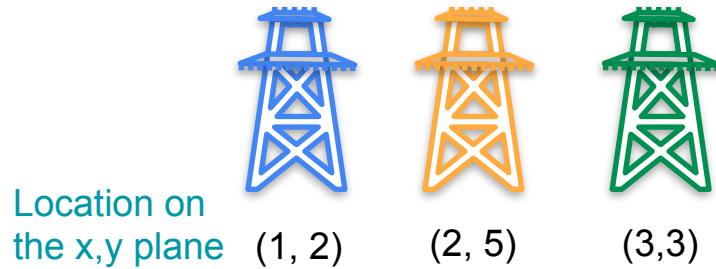
# Linear Regression: Analytical Approach



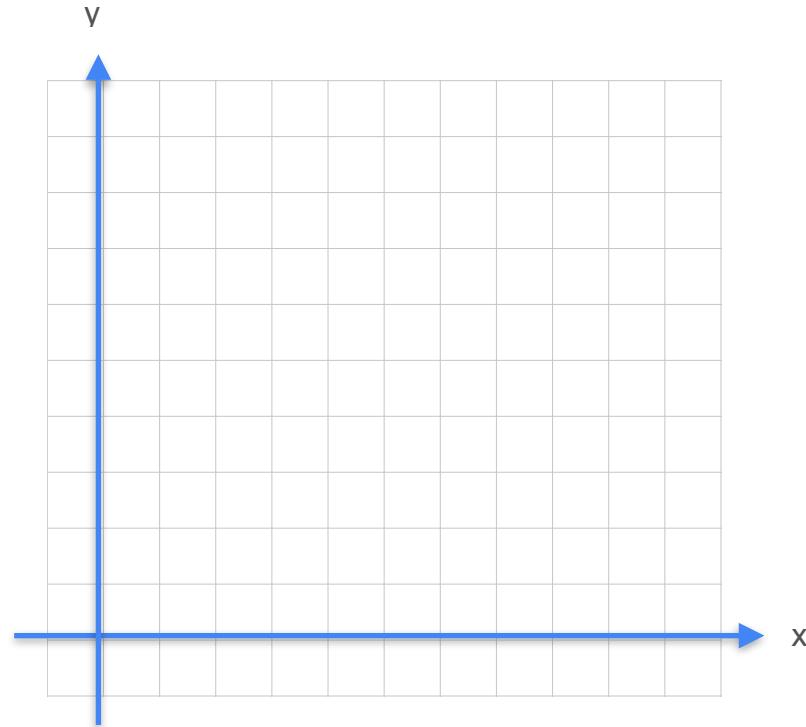
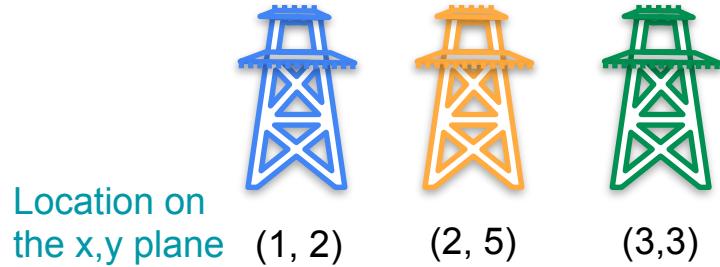
# Linear Regression: Analytical Approach



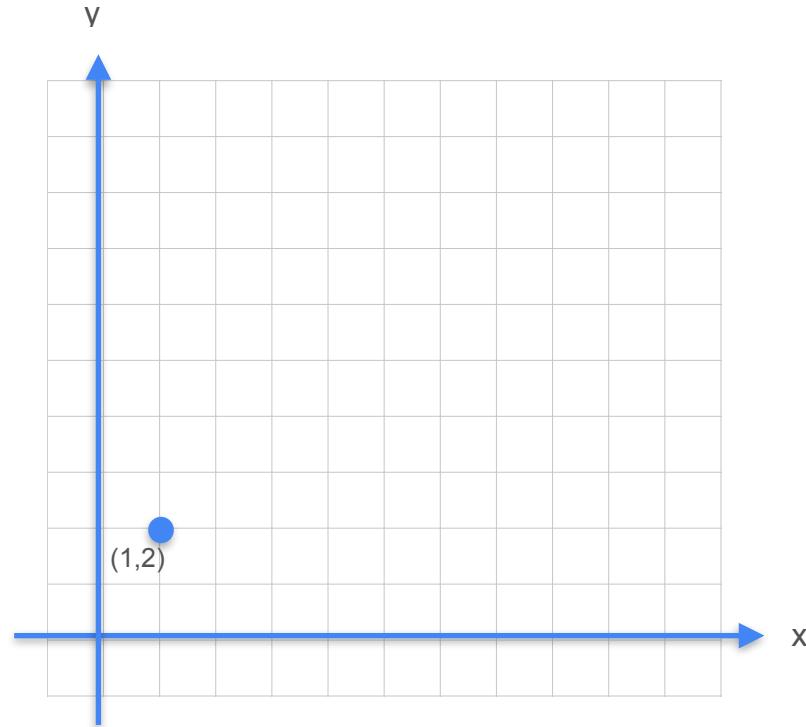
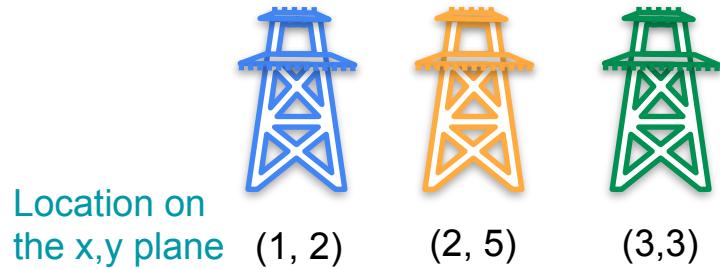
# Linear Regression: Analytical Approach



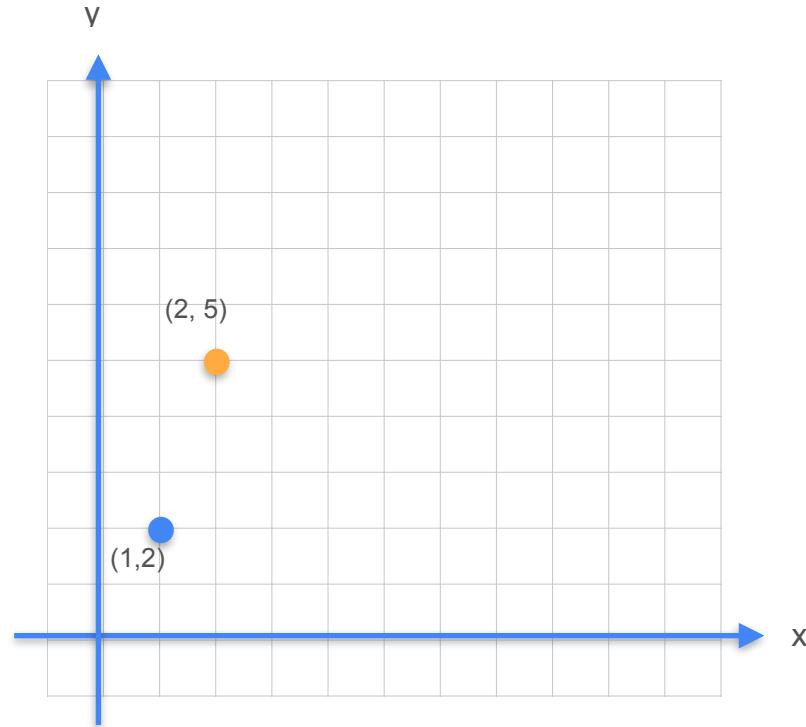
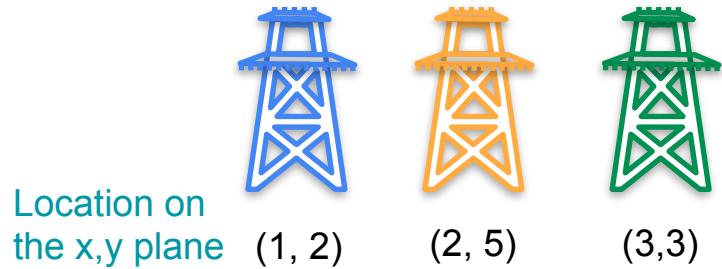
# Linear Regression: Analytical Approach



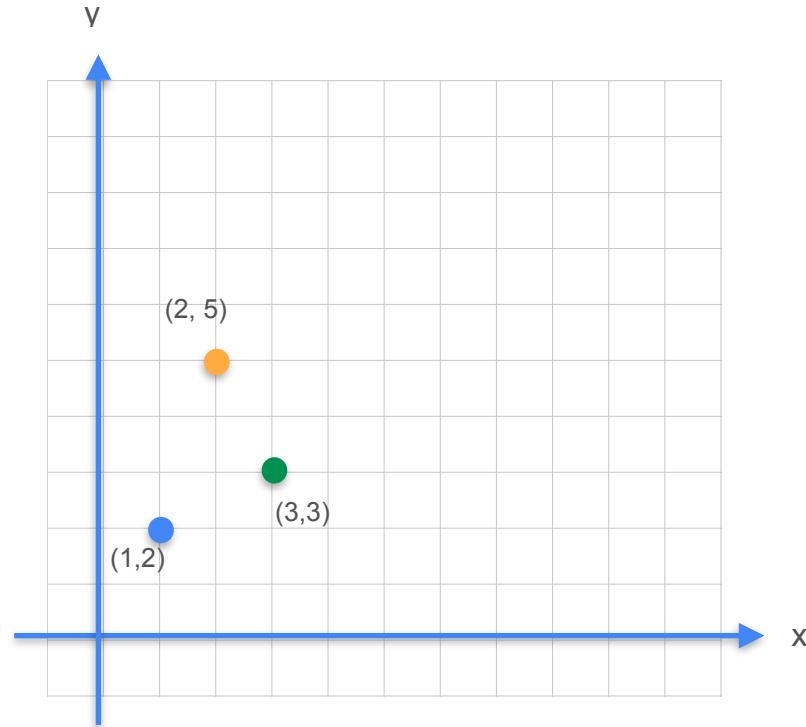
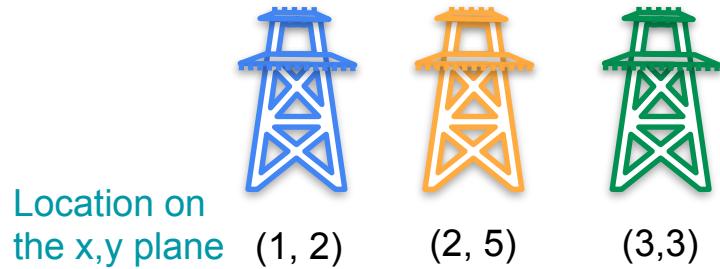
# Linear Regression: Analytical Approach



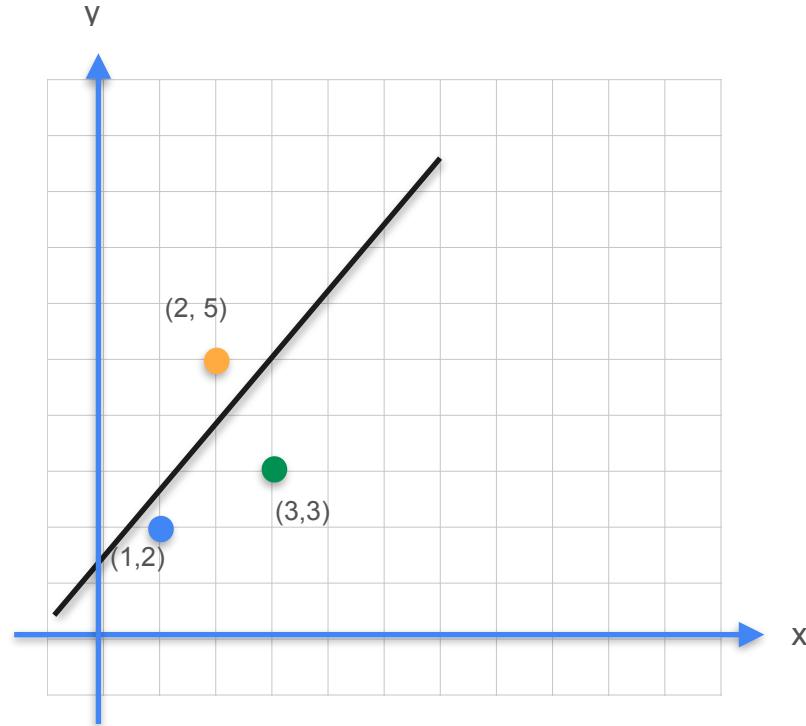
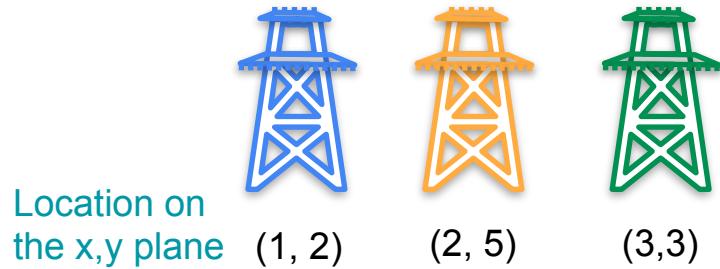
# Linear Regression: Analytical Approach



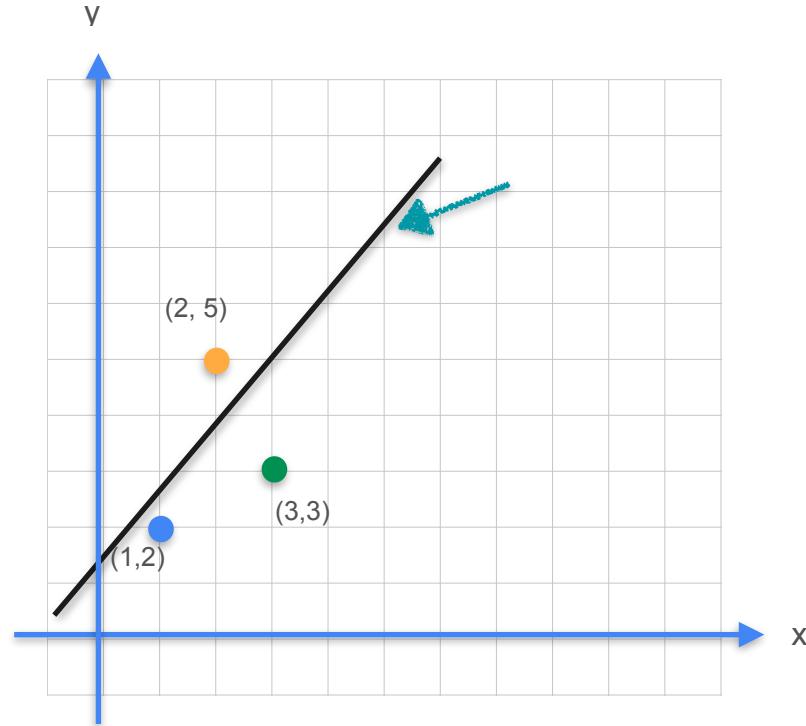
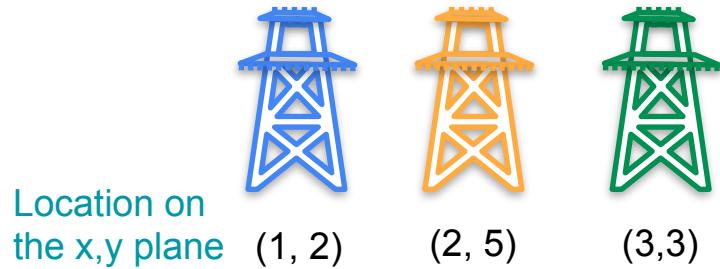
# Linear Regression: Analytical Approach



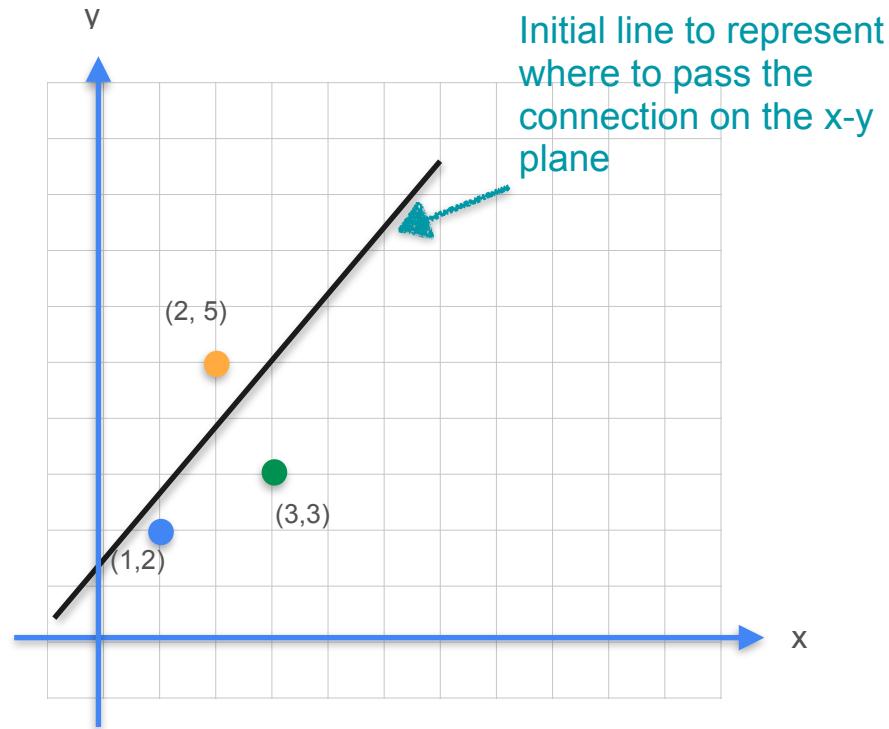
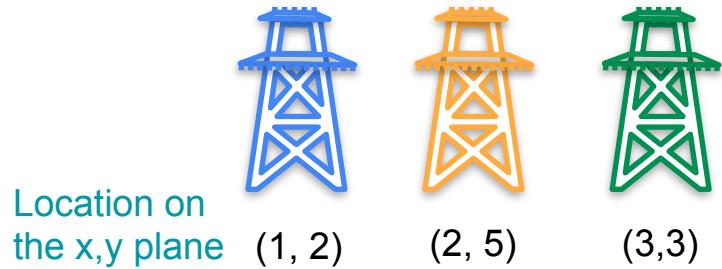
# Linear Regression: Analytical Approach



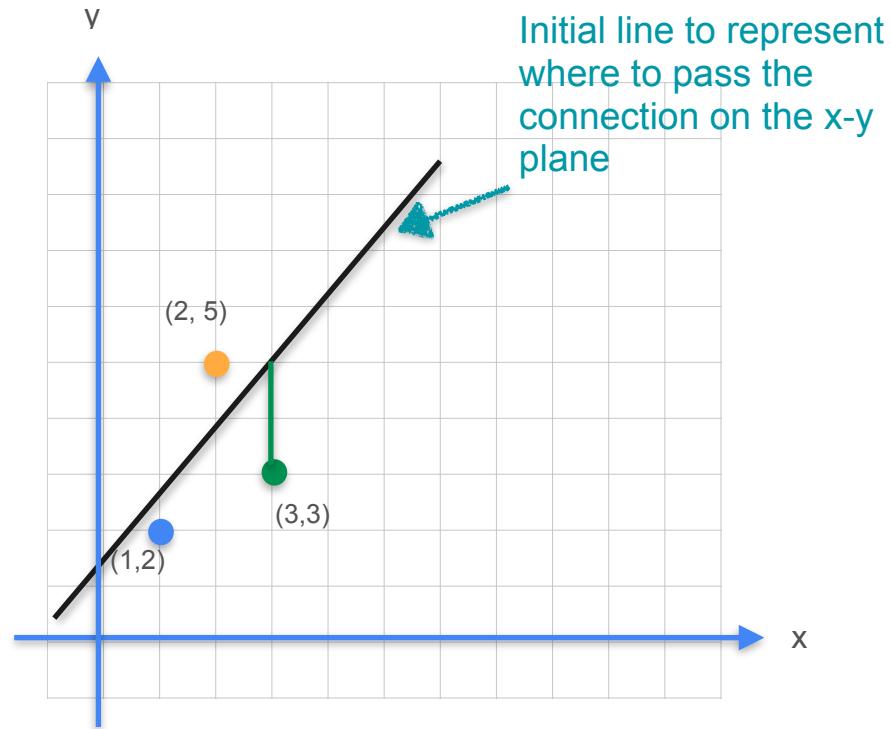
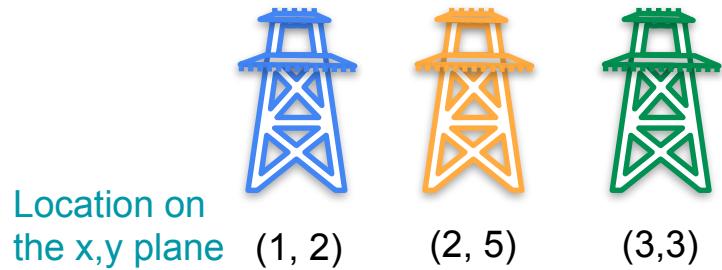
# Linear Regression: Analytical Approach



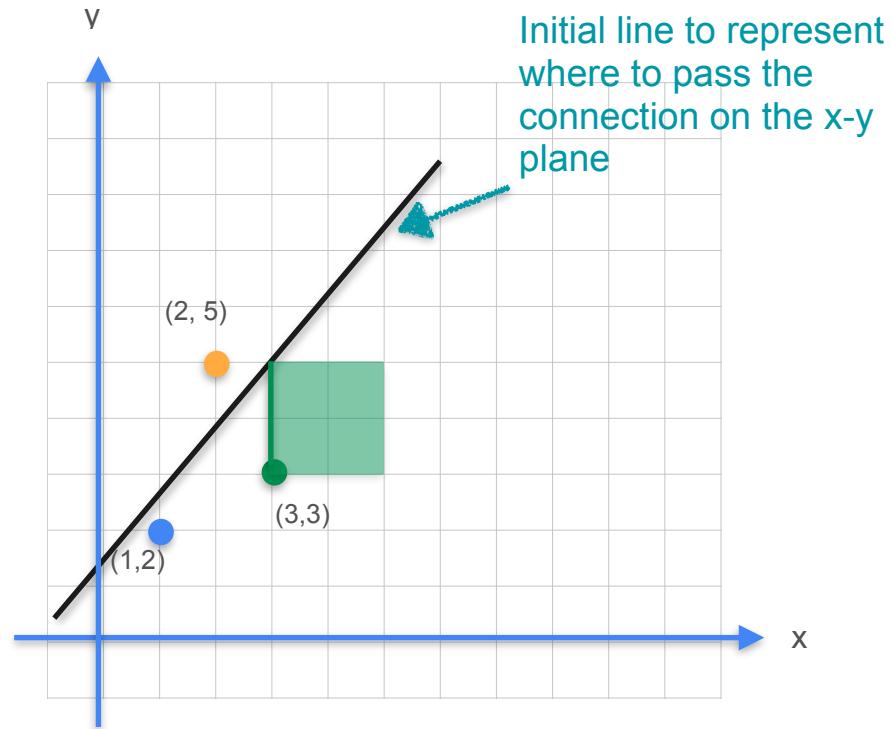
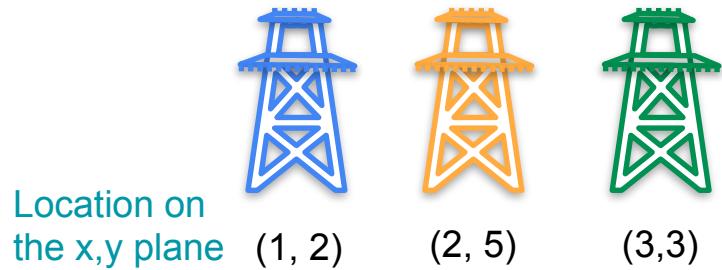
# Linear Regression: Analytical Approach



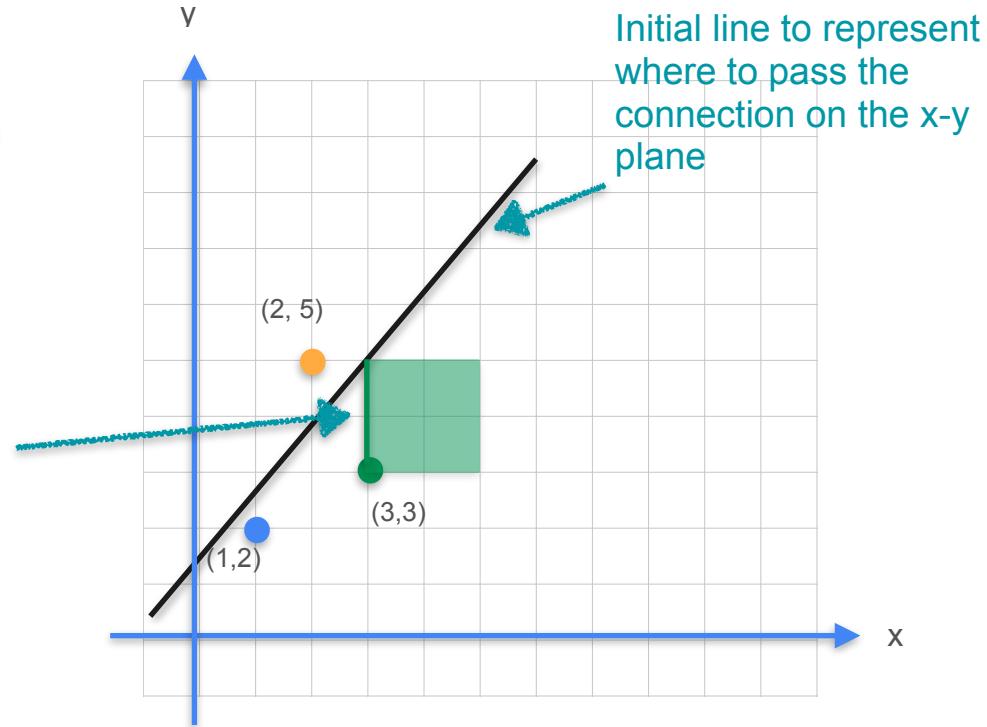
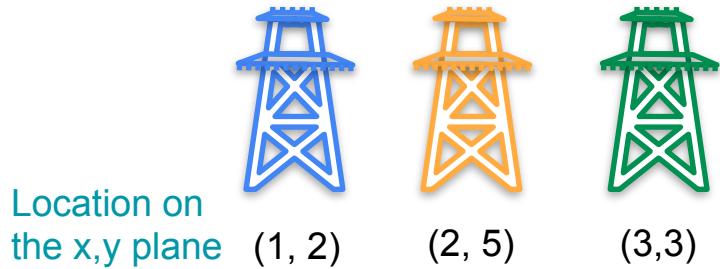
# Linear Regression: Analytical Approach



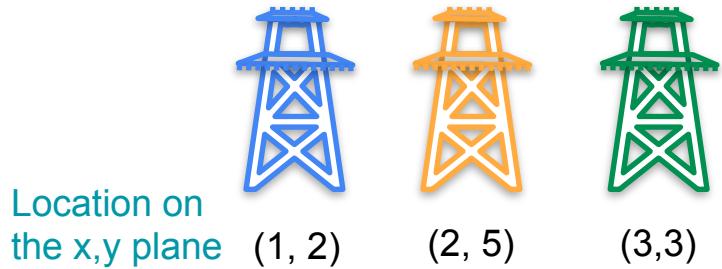
# Linear Regression: Analytical Approach



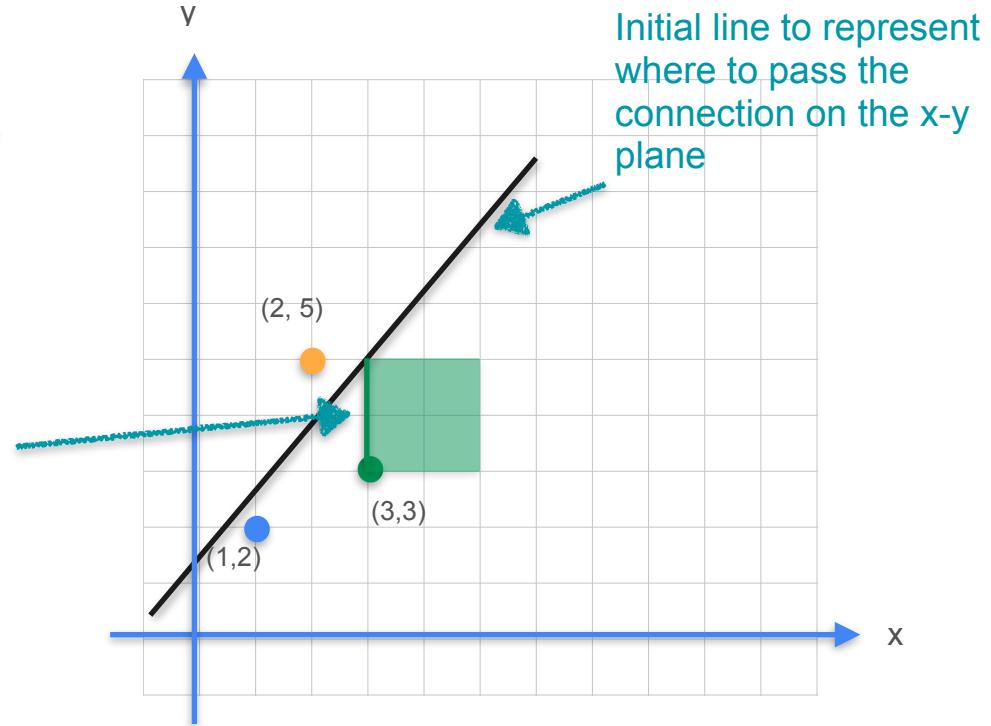
# Linear Regression: Analytical Approach



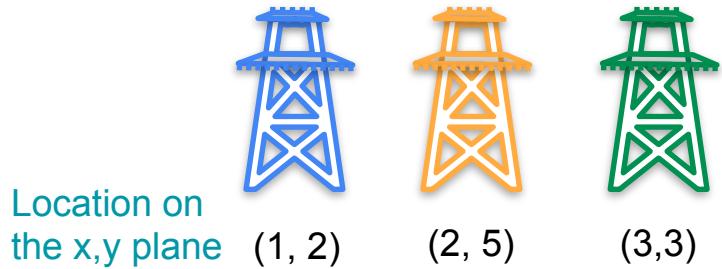
# Linear Regression: Analytical Approach



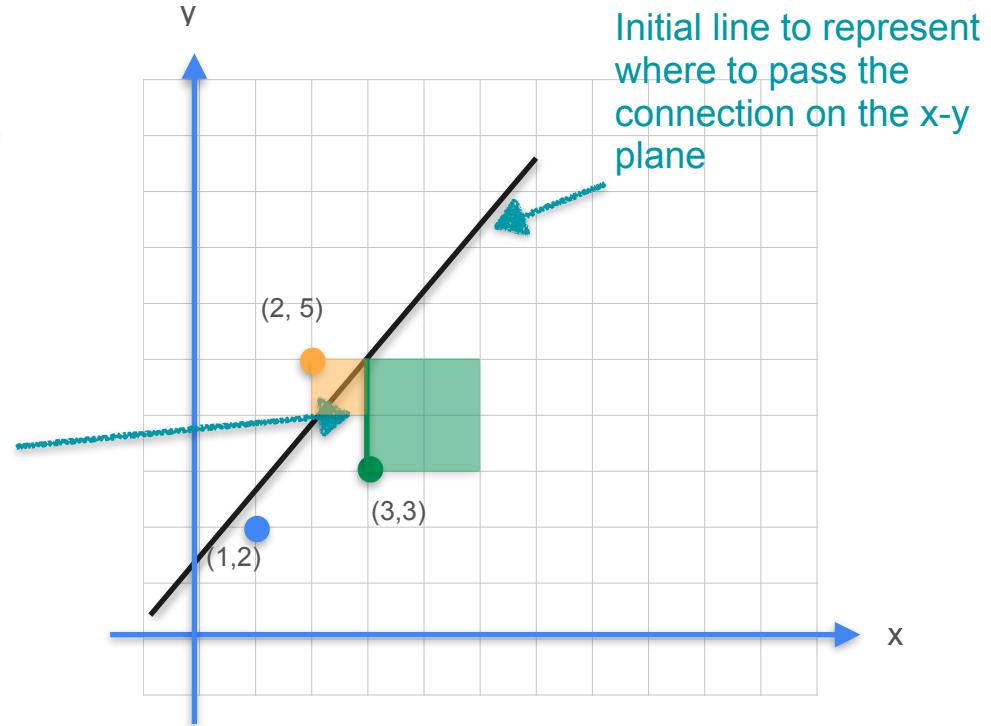
The cost of connecting connection to the powerline



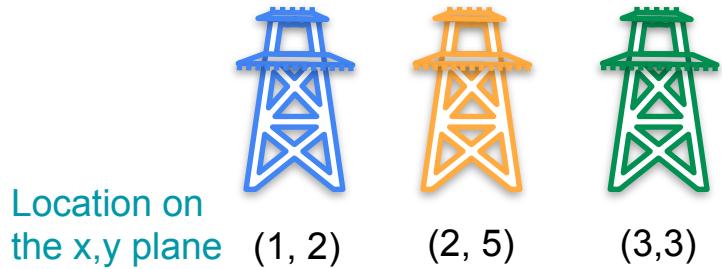
# Linear Regression: Analytical Approach



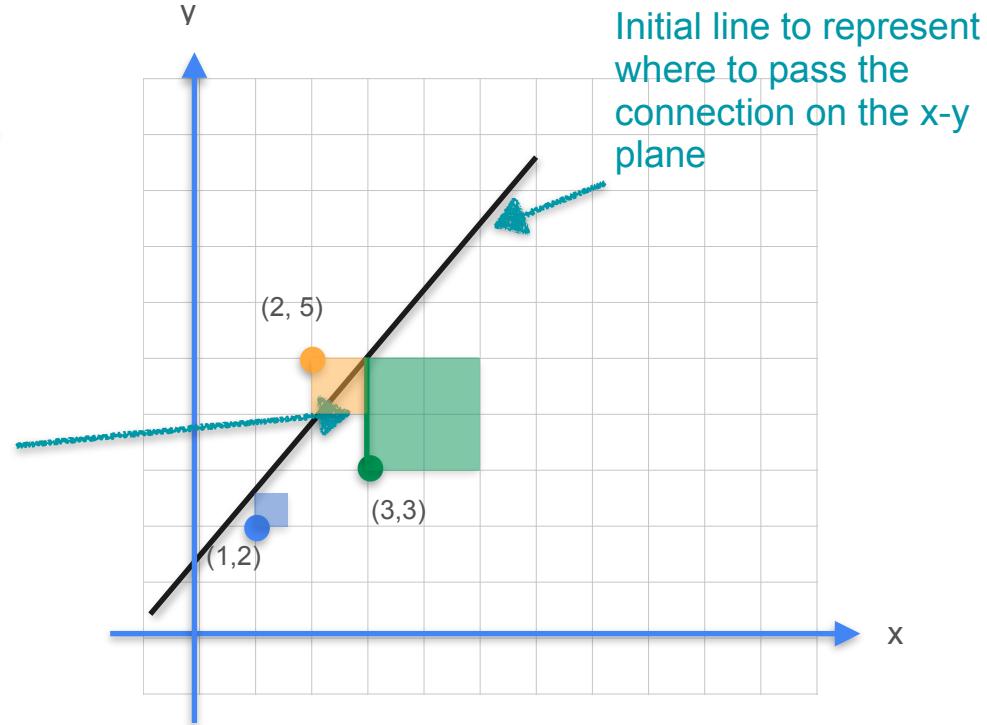
The cost of connecting connection to the powerline



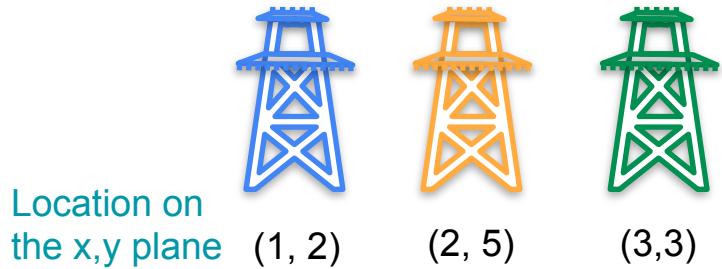
# Linear Regression: Analytical Approach



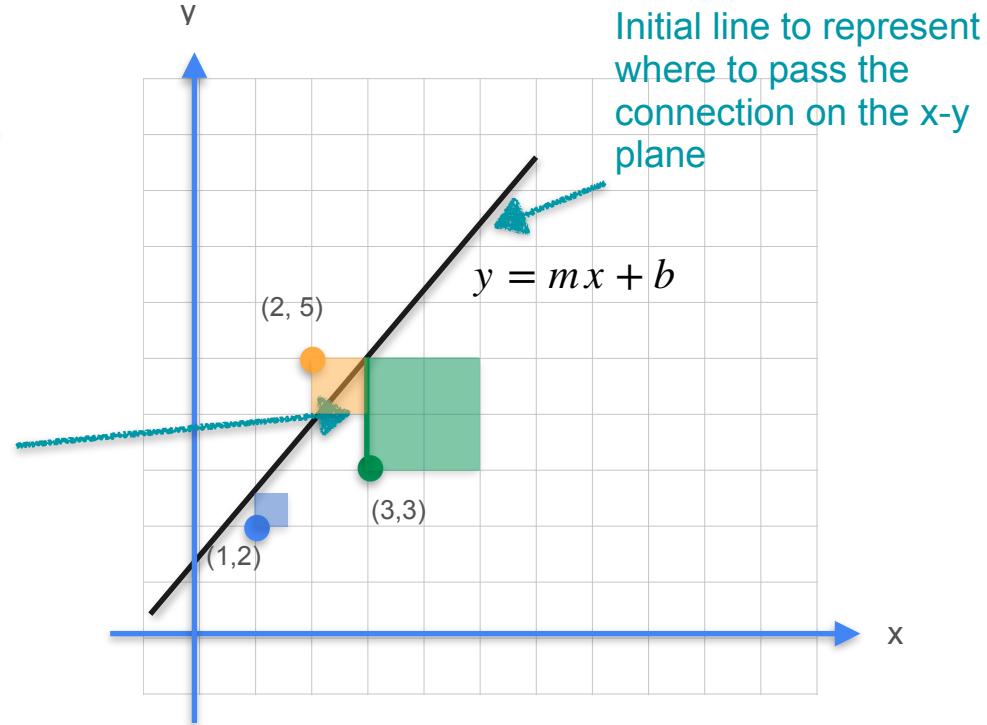
The cost of connecting connection to the powerline



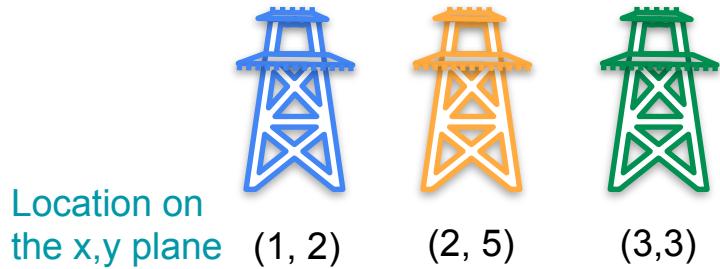
# Linear Regression: Analytical Approach



The cost of connecting connection to the powerline

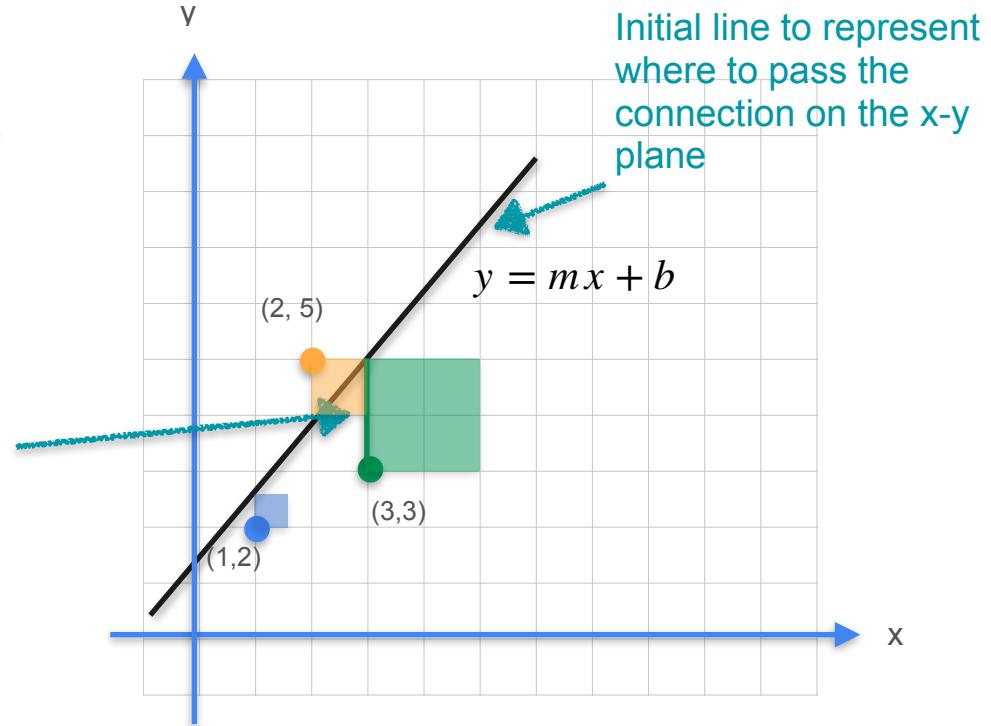


# Linear Regression: Analytical Approach

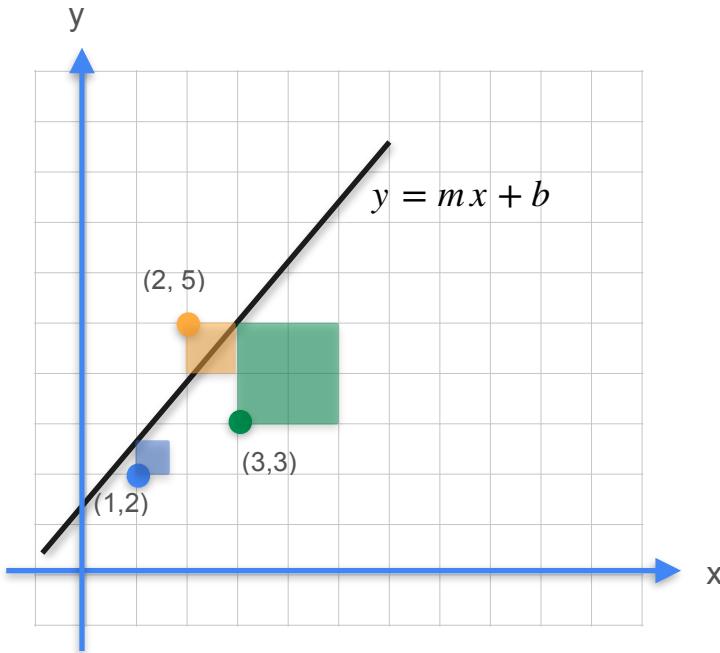


The cost of connecting connection to the powerline

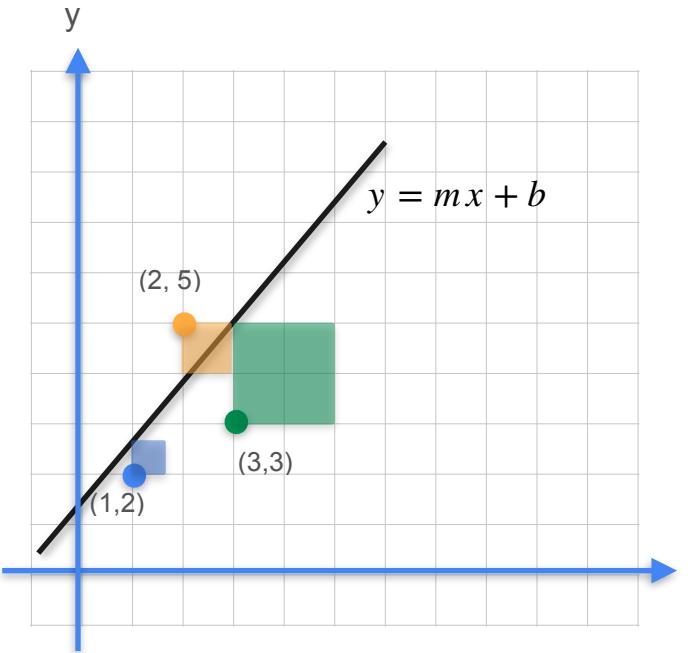
**Goal:** Find  $m, b$  such that you minimize sum of squares cost



# Linear Regression: Analytical Approach

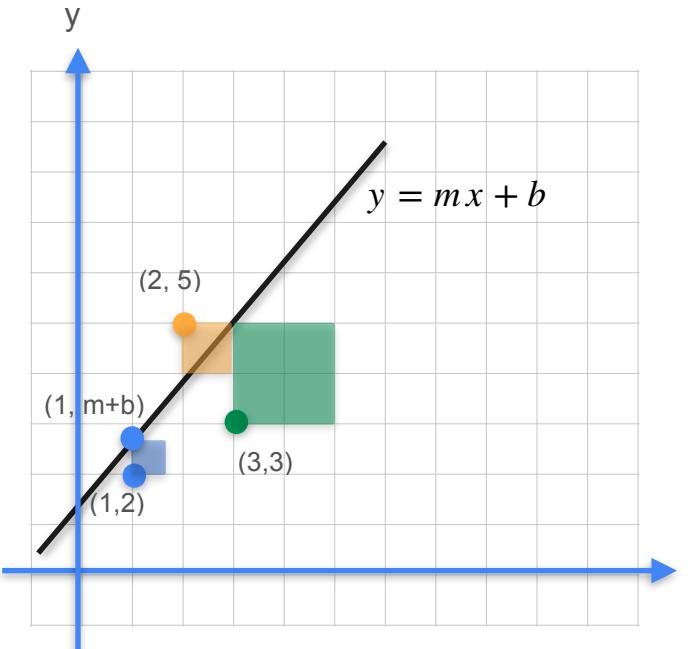


# Linear Regression: Analytical Approach



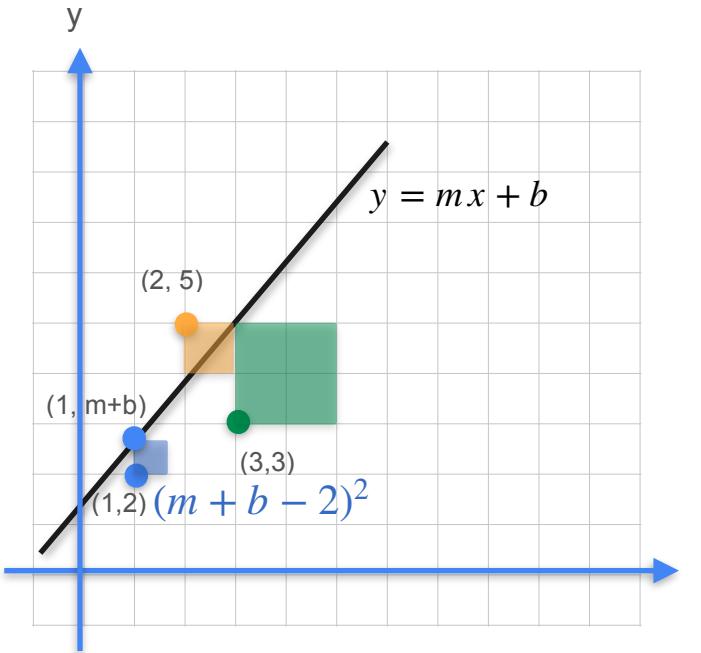
**Goal: Minimize sum of squares cost**

# Linear Regression: Analytical Approach



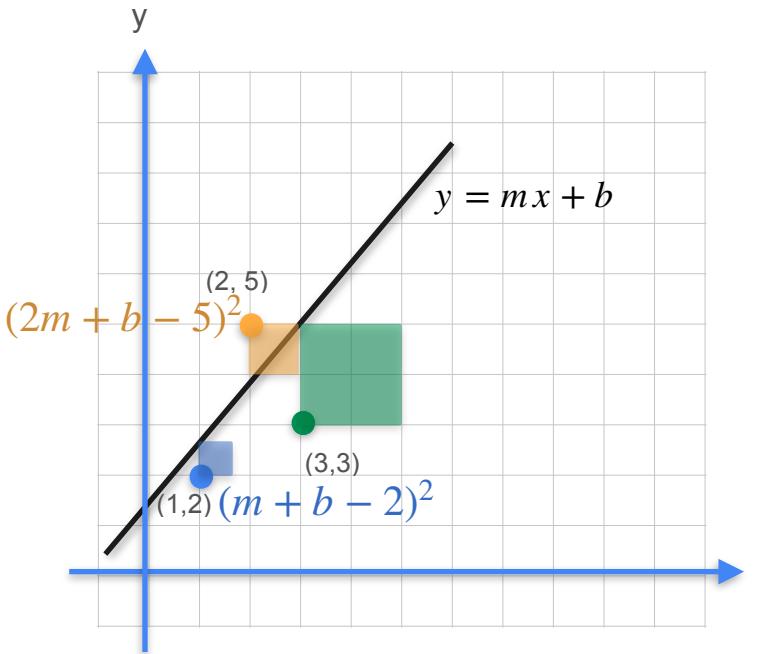
**Goal: Minimize sum of squares cost**

# Linear Regression: Analytical Approach



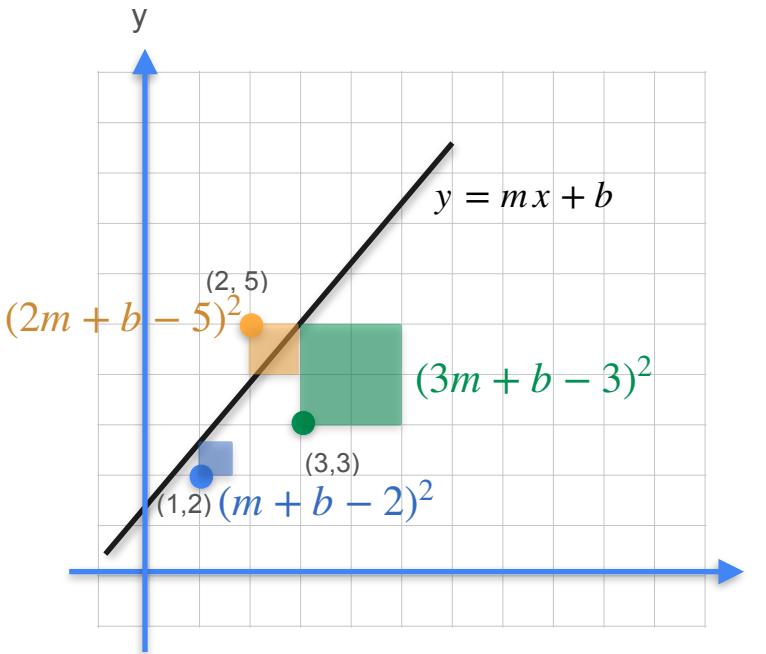
Goal: Minimize sum of squares cost

# Linear Regression: Analytical Approach



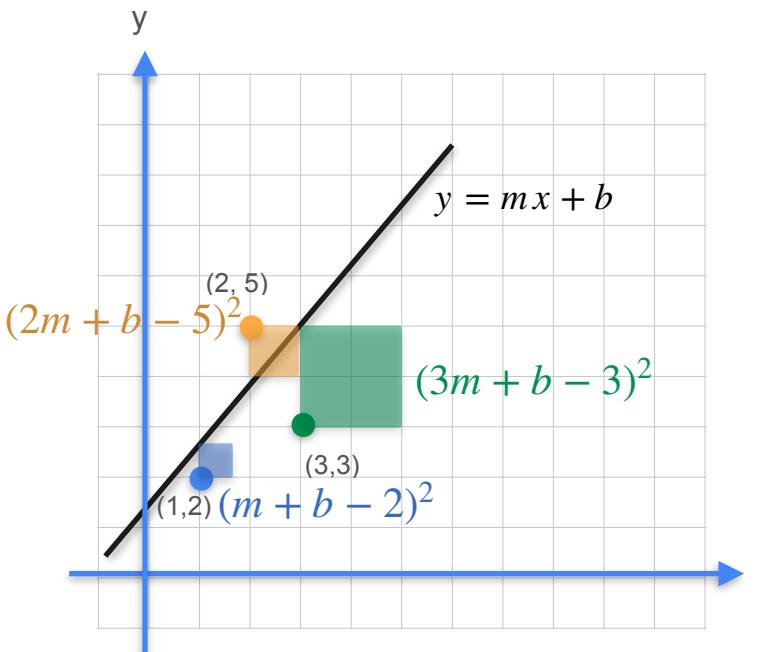
**Goal: Minimize sum of squares cost**

# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

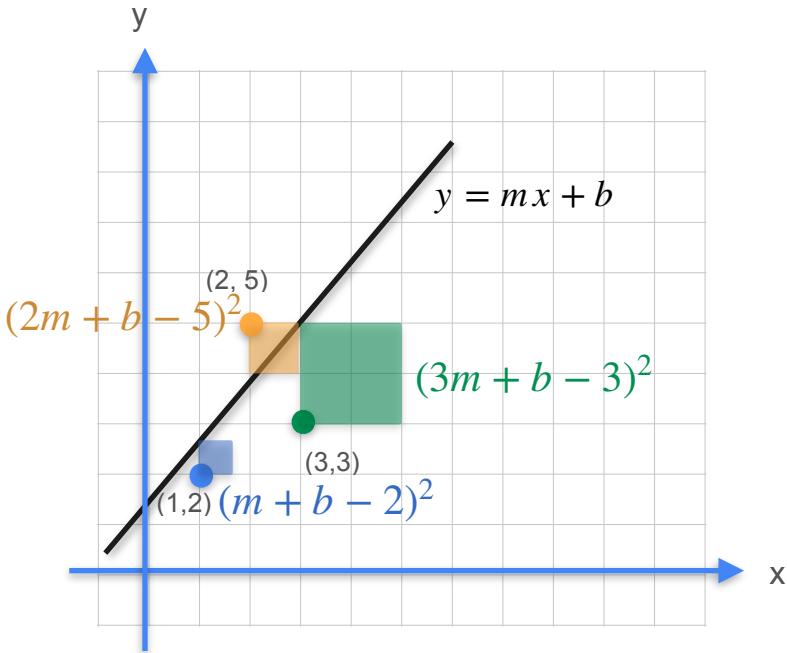
# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$(m + b - 2)^2 + (2m + b - 5)^2 + (3m + b - 3)^2$$

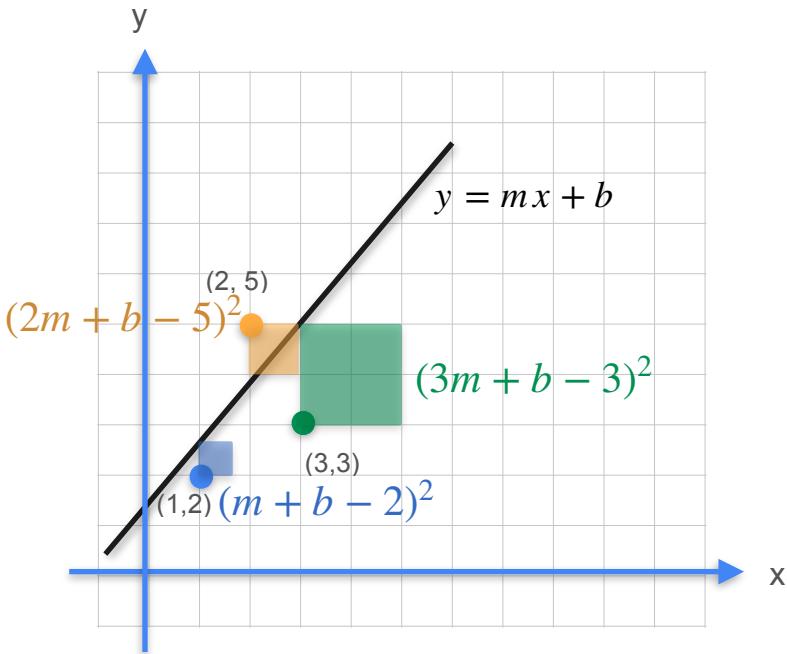
# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$(m + b - 2)^2 + (2m + b - 5)^2 + (3m + b - 3)^2$$
$$m^2 + b^2 + 4 + 2mb - 4m - 4b$$

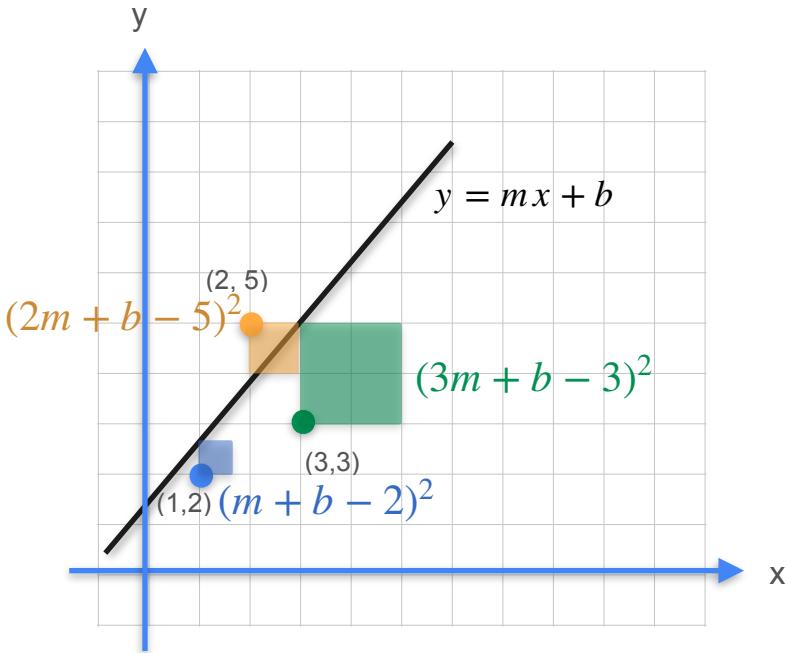
# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$\begin{aligned}(m + b - 2)^2 &+ (2m + b - 5)^2 + (3m + b - 3)^2 \\ m^2 &+ b^2 + 4 + 2mb - 4m - 4b \\ +4m^2 &+ b^2 + 25 + 4mb - 20m - 10b\end{aligned}$$

# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

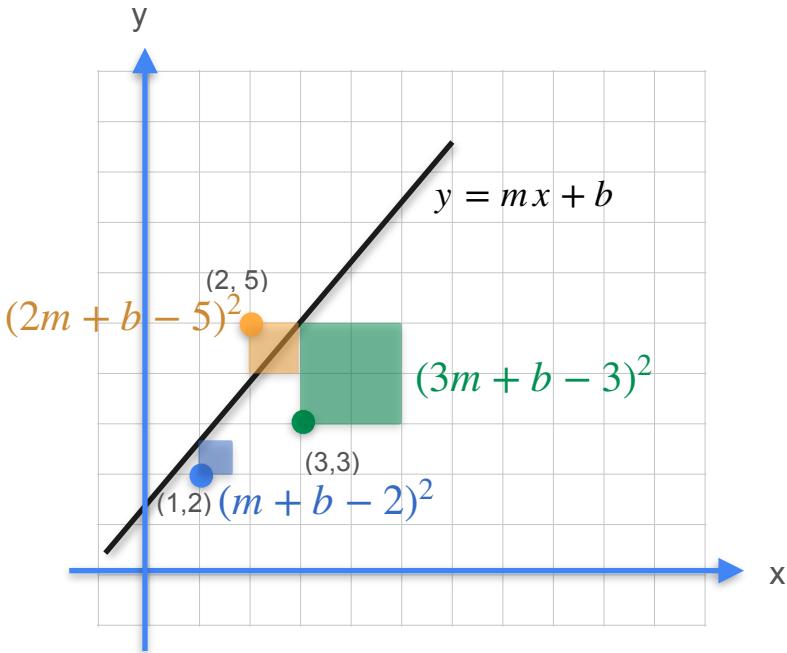
$$(m + b - 2)^2 + (2m + b - 5)^2 + (3m + b - 3)^2$$

$$m^2 + b^2 + 4 + 2mb - 4m - 4b$$

$$+4m^2 + b^2 + 25 + 4mb - 20m - 10b$$

$$+9m^2 + b^2 + 9 + 6mb - 18m - 6b$$

# Linear Regression: Analytical Approach

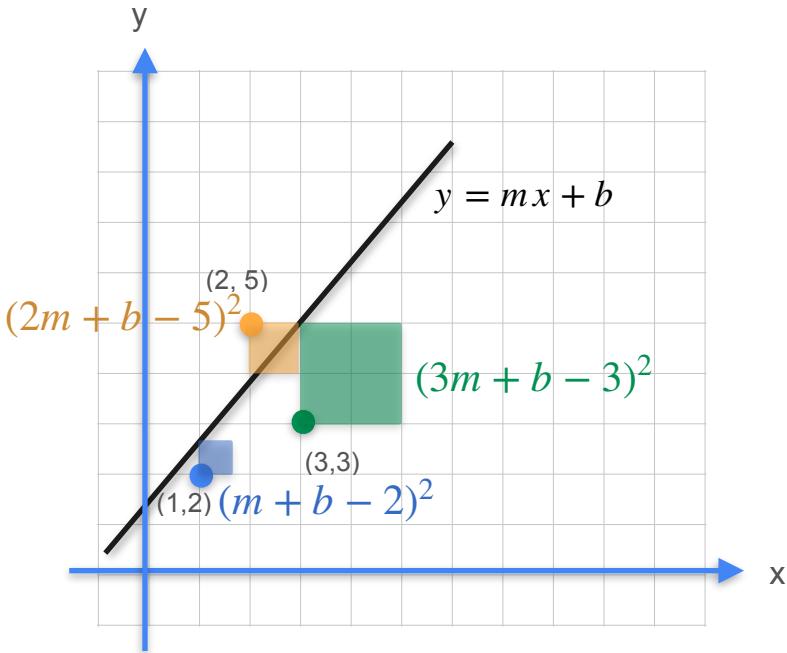


**Goal: Minimize sum of squares cost**

$$\begin{aligned}(m + b - 2)^2 &+ (2m + b - 5)^2 + (3m + b - 3)^2 \\ m^2 &+ b^2 + 4 + 2mb - 4m - 4b \\ &+ 4m^2 + b^2 + 25 + 4mb - 20m - 10b \\ &+ 9m^2 + b^2 + 9 + 6mb - 18m - 6b\end{aligned}$$

---

# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$(m + b - 2)^2 + (2m + b - 5)^2 + (3m + b - 3)^2$$

$$m^2 + b^2 + 4 + 2mb - 4m - 4b$$

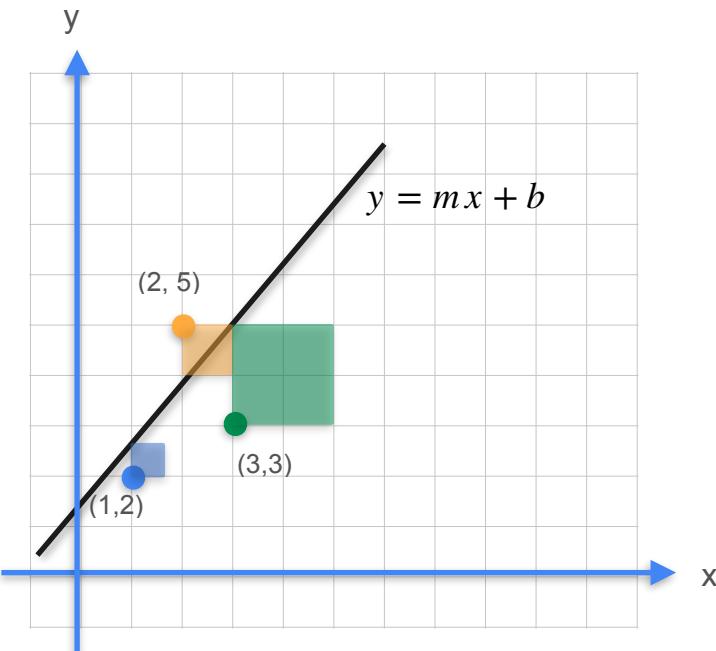
$$+4m^2 + b^2 + 25 + 4mb - 20m - 10b$$

$$+9m^2 + b^2 + 9 + 6mb - 18m - 6b$$

---

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

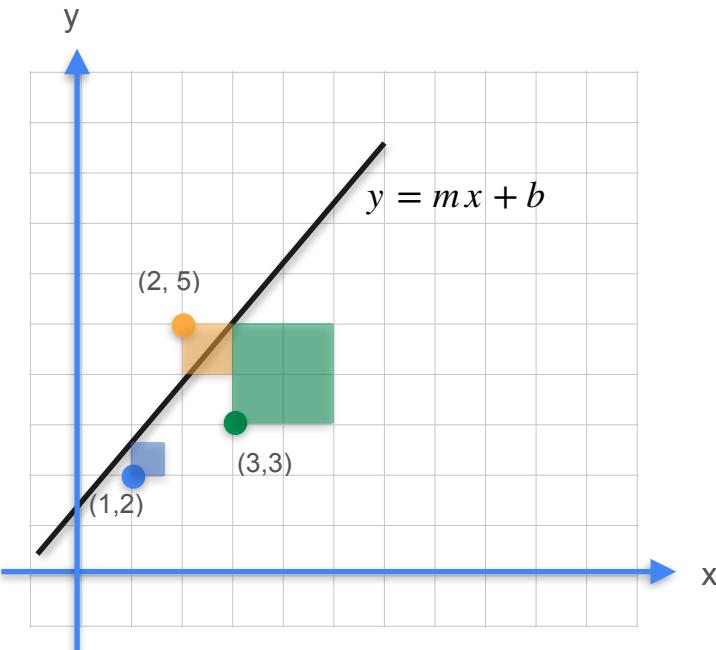
# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

# Linear Regression: Analytical Approach

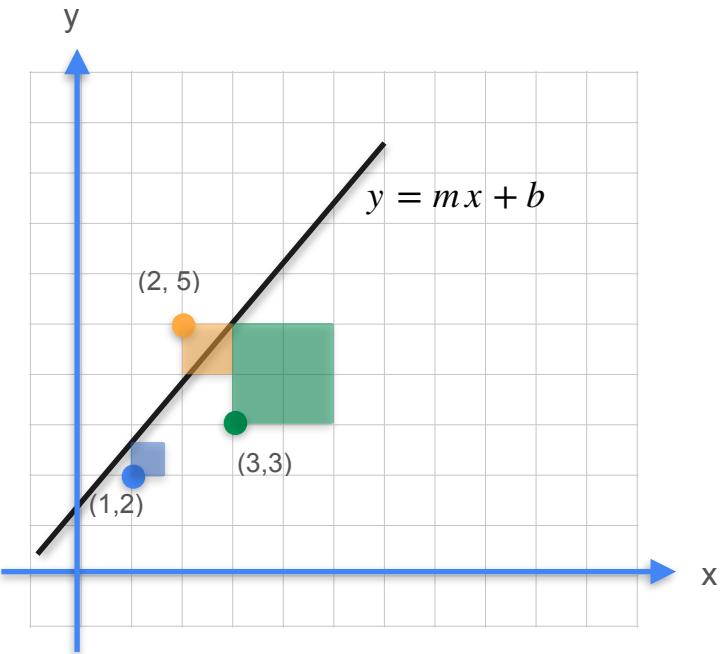


**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 0$$

# Linear Regression: Analytical Approach



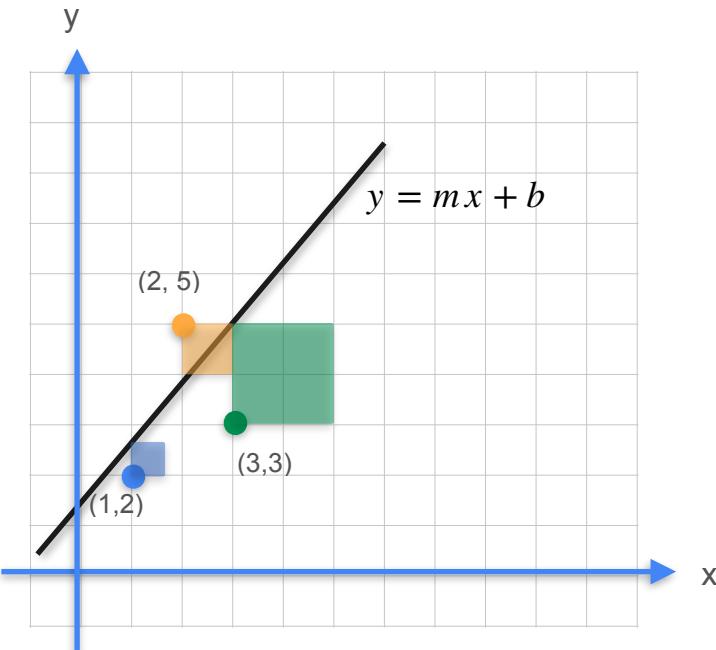
**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 0$$

$$\frac{\partial E}{\partial b} = 0$$

# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

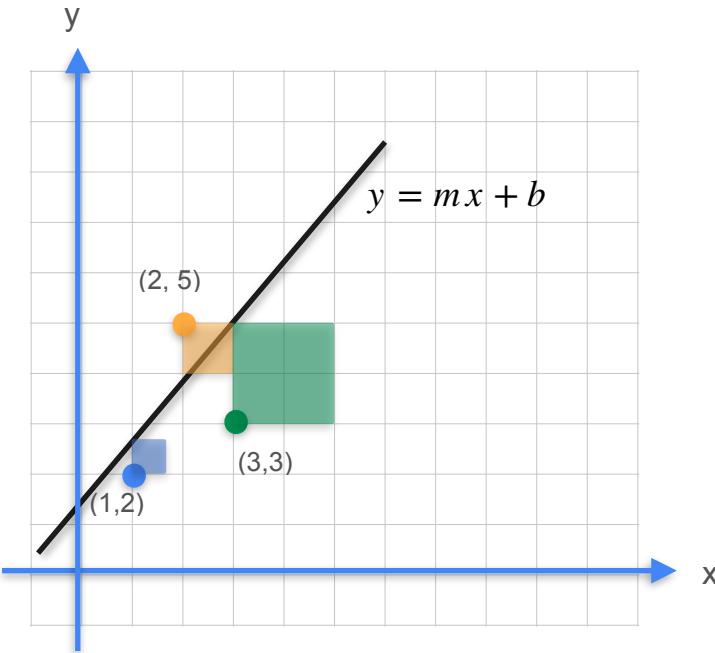
$$\frac{\partial E}{\partial m} = 0$$

**Quiz:**

$$\frac{\partial E}{\partial b} = 0$$

**Find the partial derivative of  $E$  with respect to  $m$**

# Linear Regression: Analytical Approach



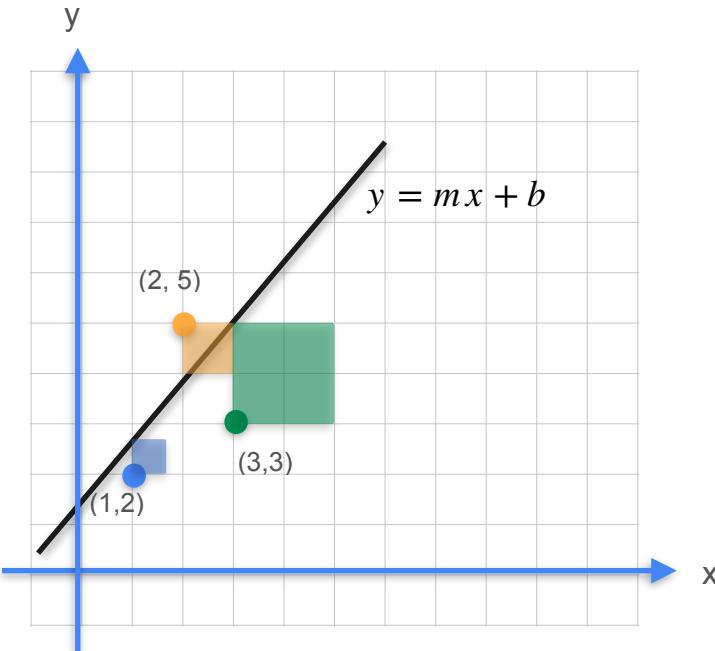
**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42$$

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

# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

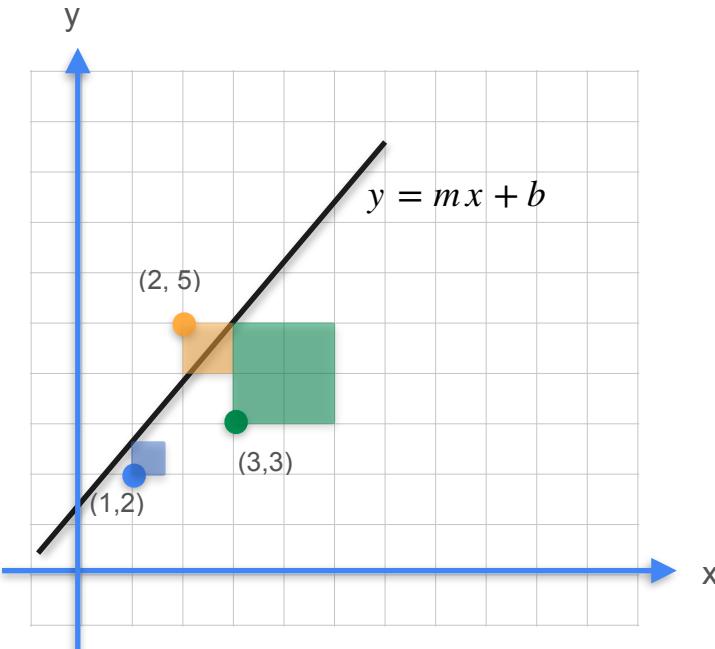
$$\frac{\partial E}{\partial m} = 28m + 12b - 42$$

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

**Quiz:**

**Find the partial derivative of  $E$  with respect to  $b$**

# Linear Regression: Analytical Approach



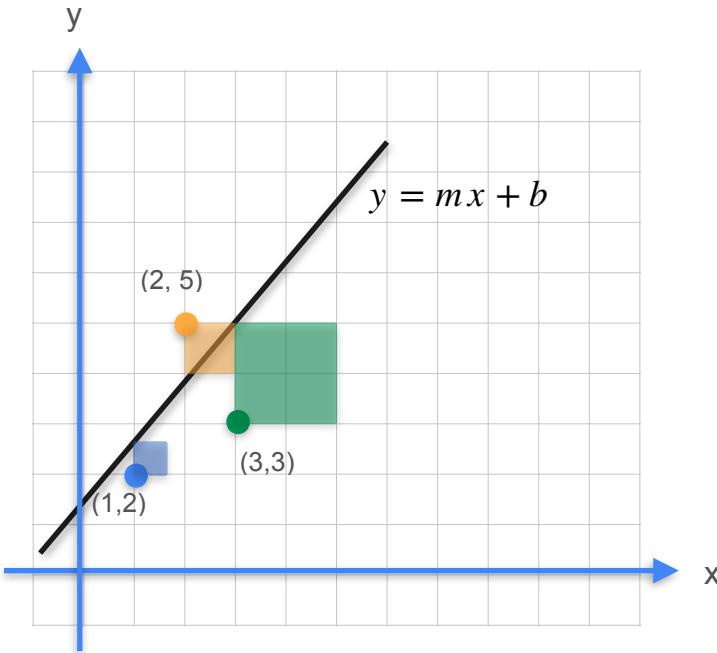
**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20$$

# Linear Regression: Analytical Approach



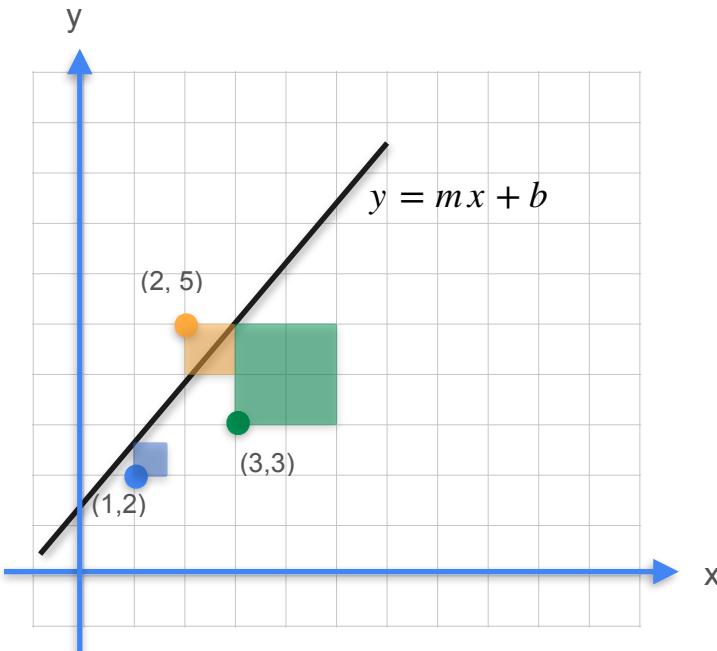
**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20$$

# Linear Regression: Analytical Approach



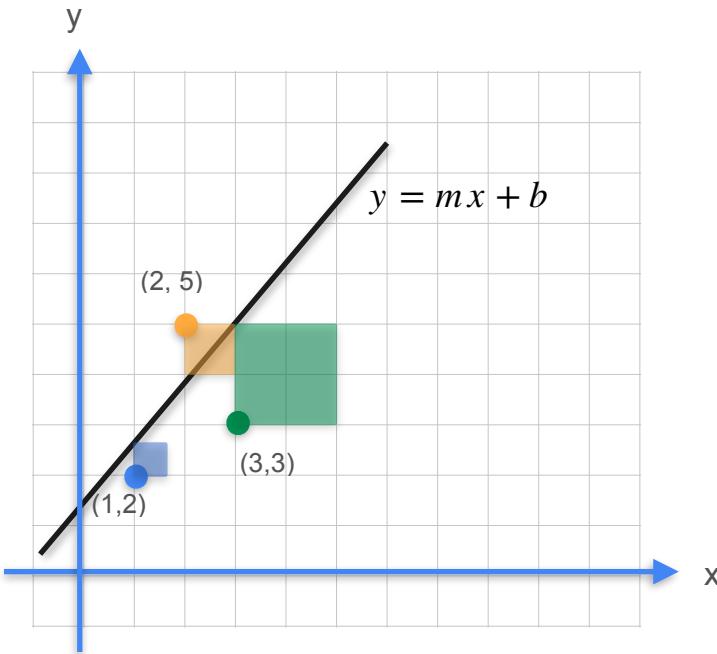
**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20$$

# Linear Regression: Analytical Approach



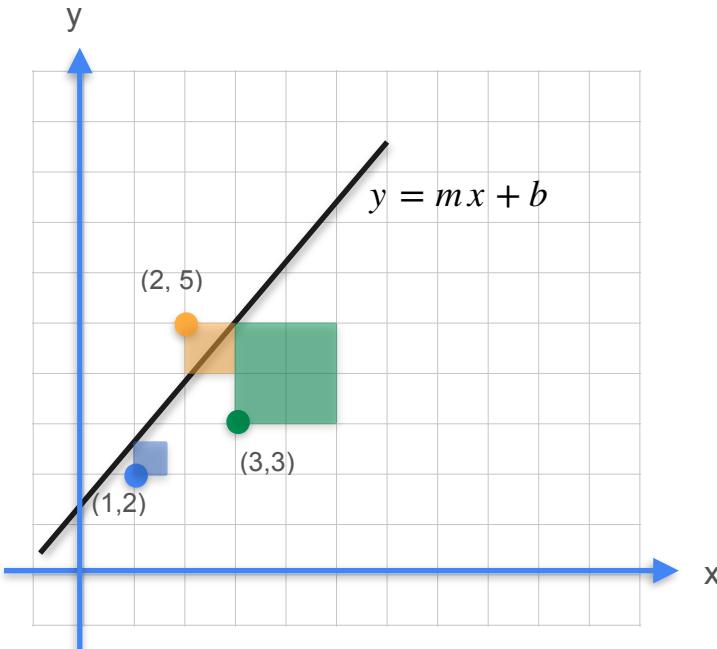
**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

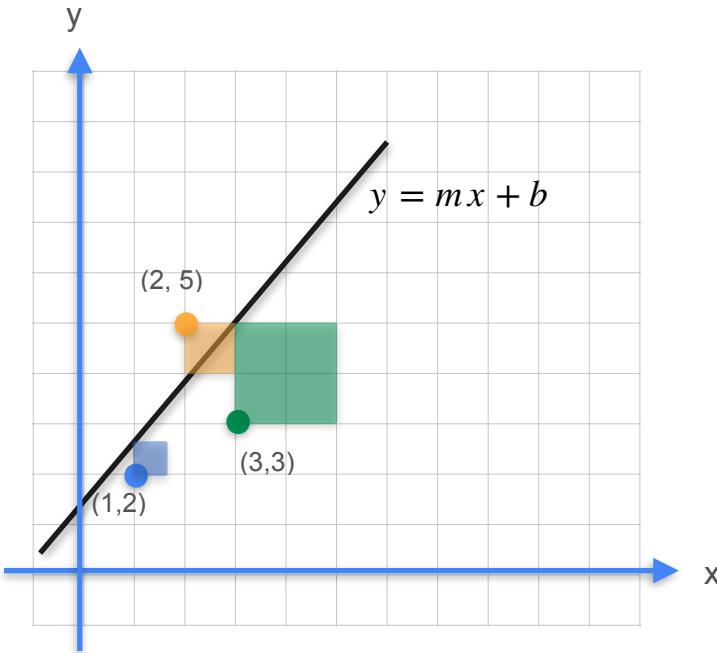
$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$m =$$

# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

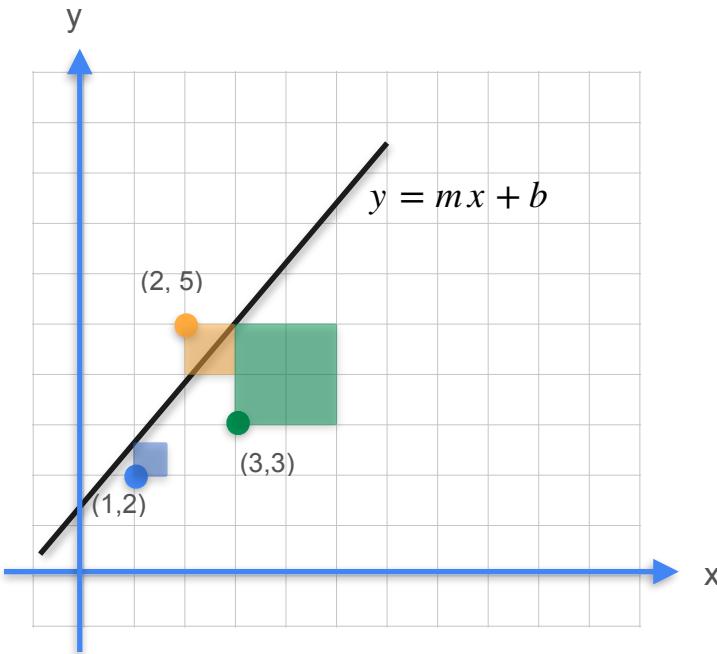
$$\frac{\partial E}{\partial m} = \boxed{28m + 12b - 42} = 0$$

$$\frac{\partial E}{\partial b} = \boxed{6b + 12m - 20} = 0$$

$$m =$$

$$b =$$

# Linear Regression: Analytical Approach



**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

# Linear Regression: Analytical Approach

**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

# Linear Regression: Analytical Approach

Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$



# Linear Regression: Analytical Approach

Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

$$12b + 24m - 40 = 0$$



# Linear Regression: Analytical Approach

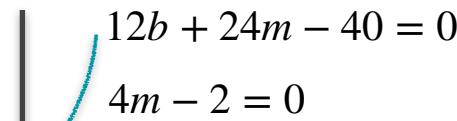
Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$


$$\begin{aligned} 12b + 24m - 40 &= 0 \\ 4m - 2 &= 0 \end{aligned}$$

# Linear Regression: Analytical Approach

Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

$$\begin{aligned} 12b + 24m - 40 &= 0 \\ 4m - 2 &= 0 \\ m &= \frac{2}{4} = 0.5 \end{aligned}$$

# Linear Regression: Analytical Approach

Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$


$$12b + 24m - 40 = 0$$

$$4m - 2 = 0$$

$$m = \frac{2}{4} = 0.5$$

$$6b + 12(0.5) - 20 = 0$$

# Linear Regression: Analytical Approach

Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$


$$12b + 24m - 40 = 0$$

$$4m - 2 = 0$$

$$m = \frac{2}{4} = 0.5$$

$$6b + 12(0.5) - 20 = 0$$

$$6b + 6 - 20 = 0$$

# Linear Regression: Analytical Approach

Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

$$12b + 24m - 40 = 0$$

$$4m - 2 = 0$$

$$m = \frac{2}{4} = 0.5$$

$$6b + 12(0.5) - 20 = 0$$

$$6b + 6 - 20 = 0$$

$$6b - 14 = 0$$

# Linear Regression: Analytical Approach

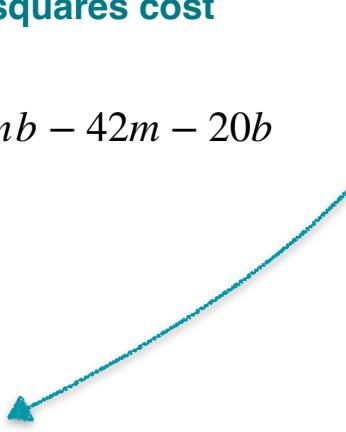
Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$



$$12b + 24m - 40 = 0$$

$$4m - 2 = 0$$

$$m = \frac{2}{4} = 0.5$$

$$6b + 12(0.5) - 20 = 0$$

$$6b + 6 - 20 = 0$$

$$6b - 14 = 0$$

$$b = \frac{14}{6} = \frac{7}{3}$$

# Linear Regression: Analytical Approach

Goal: Minimize sum of squares cost

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

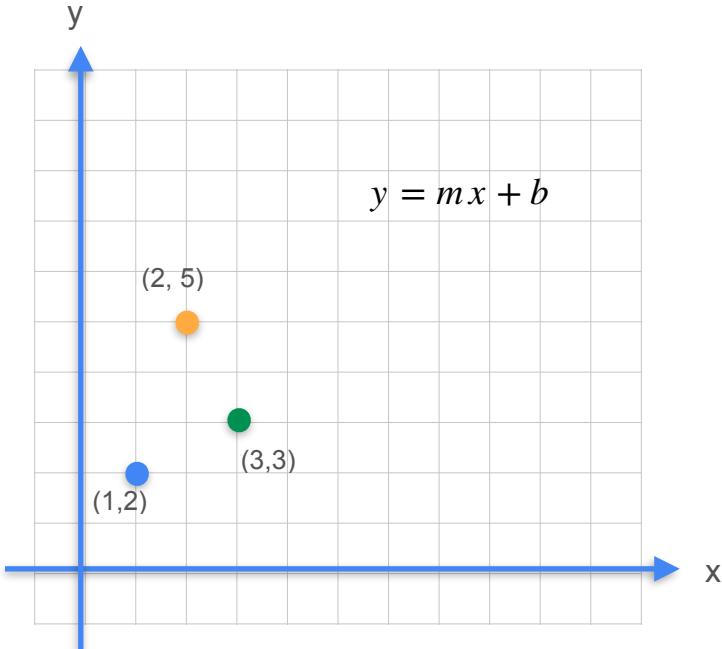
$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

$$m = \frac{1}{2}$$

$$b = \frac{7}{3}$$

$$E(m = \frac{1}{2}, b = \frac{7}{3}) \approx 4.167$$

# Linear Regression: Optimal Solution

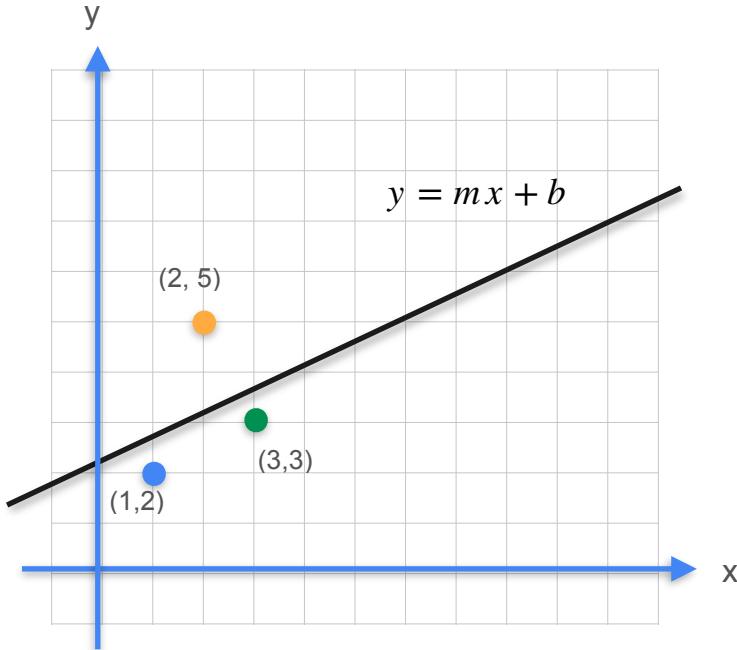


$$m = \frac{1}{2}$$

$$b = \frac{7}{3}$$

$$E(m = \frac{1}{2}, b = \frac{7}{3}) \approx 4.167$$

# Linear Regression: Optimal Solution

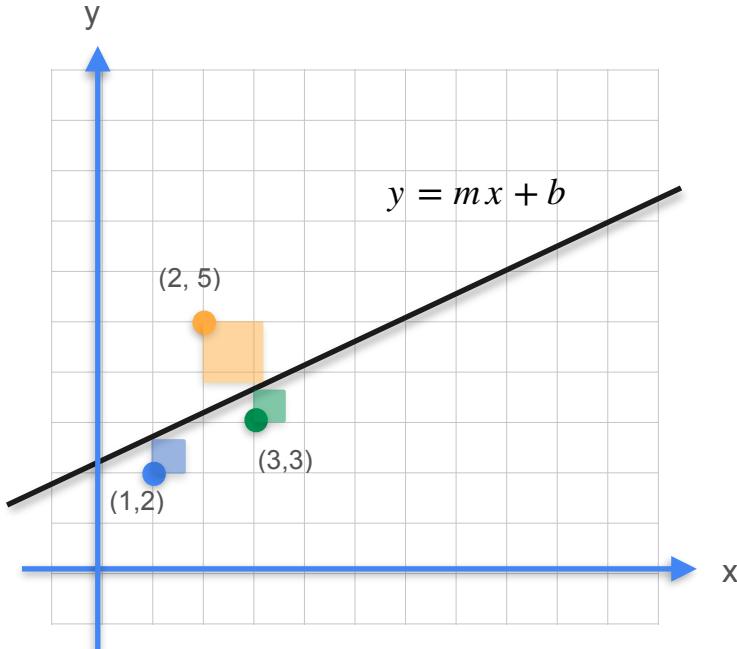


$$m = \frac{1}{2}$$

$$b = \frac{7}{3}$$

$$E(m = \frac{1}{2}, b = \frac{7}{3}) \approx 4.167$$

# Linear Regression: Optimal Solution



$$m = \frac{1}{2}$$

$$b = \frac{7}{3}$$

$$E(m = \frac{1}{2}, b = \frac{7}{3}) \approx 4.167$$

# Linear Regression: Gradient Descent

**Goal: Minimize sum of squares cost**

$$E(m, b) = 14m^2 + 3b^2 + 38 + 12mb - 42m - 20b$$

$$\frac{\partial E}{\partial m} = 28m + 12b - 42 = 0$$

$$\frac{\partial E}{\partial b} = 6b + 12m - 20 = 0$$

**Gradient Descent to the rescue**

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$



DeepLearning.AI

# Gradients and Gradient Descent

---

**Optimization using Gradient  
Descent in one variable -  
Part 1**

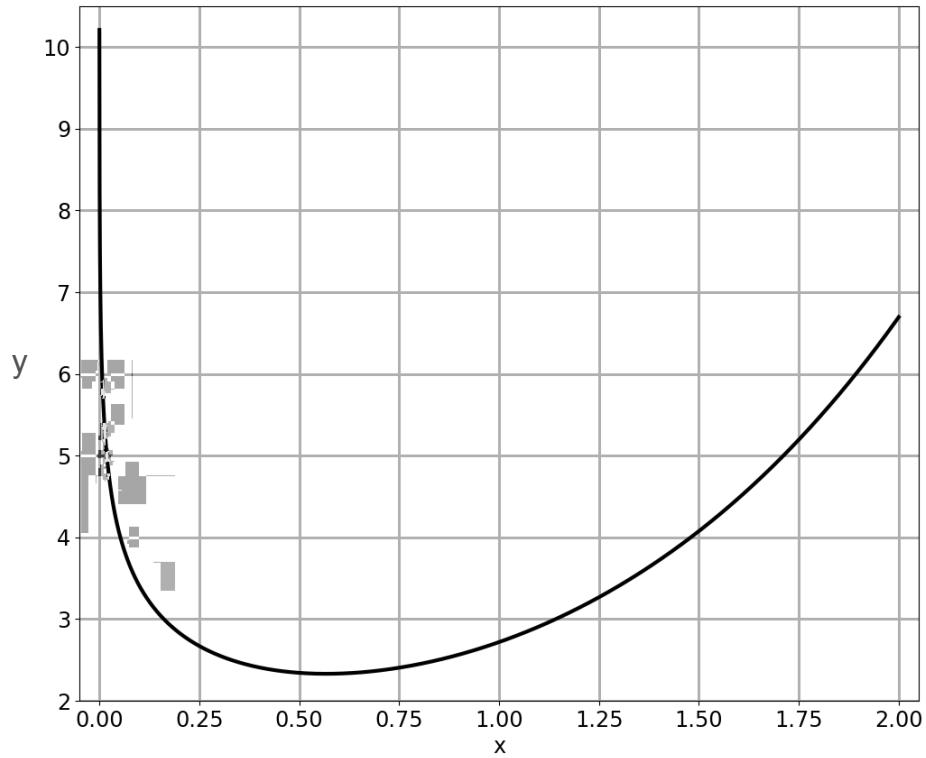
# Hard To Optimize Functions

# Hard To Optimize Functions

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

# Hard To Optimize Functions

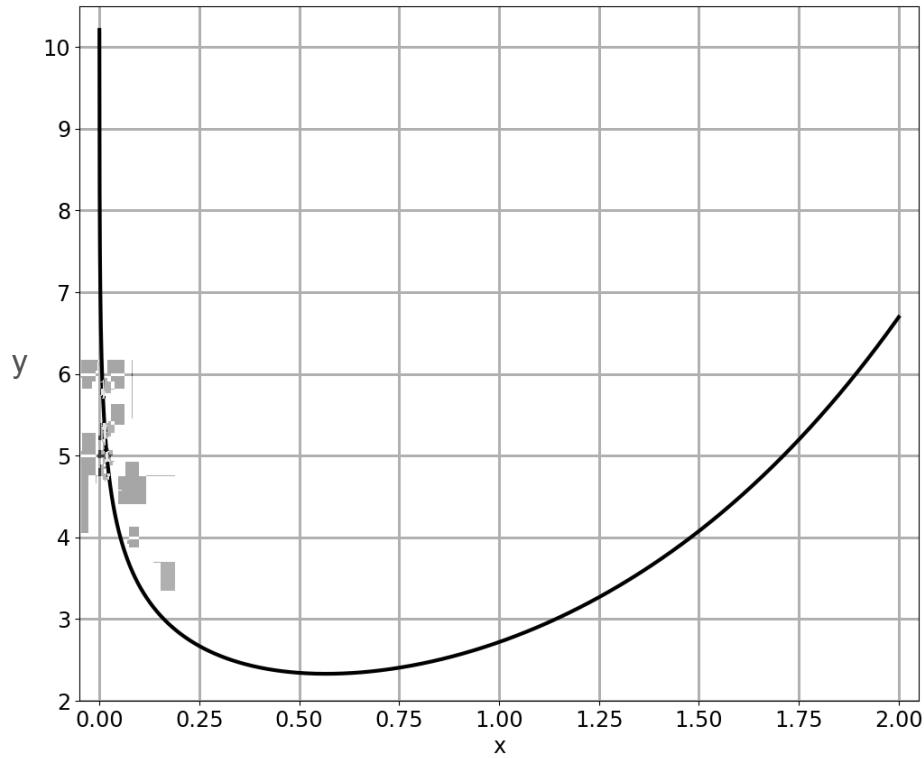
$$f(x) = e^x - \log(x)$$



# Hard To Optimize Functions

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

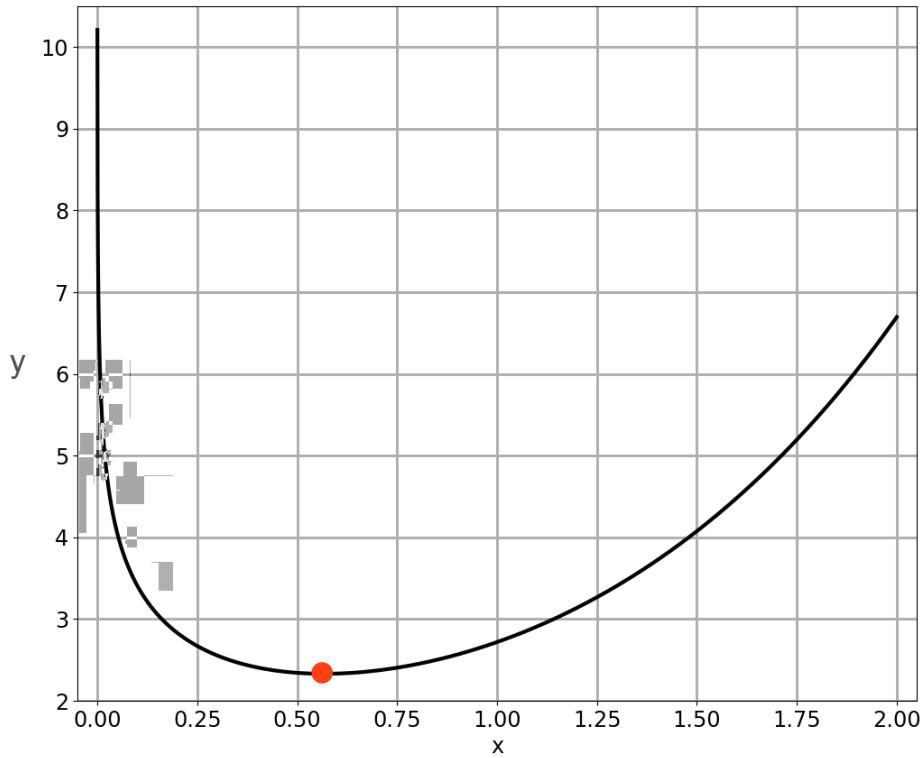
Minimum?



# Hard To Optimize Functions

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

Minimum?

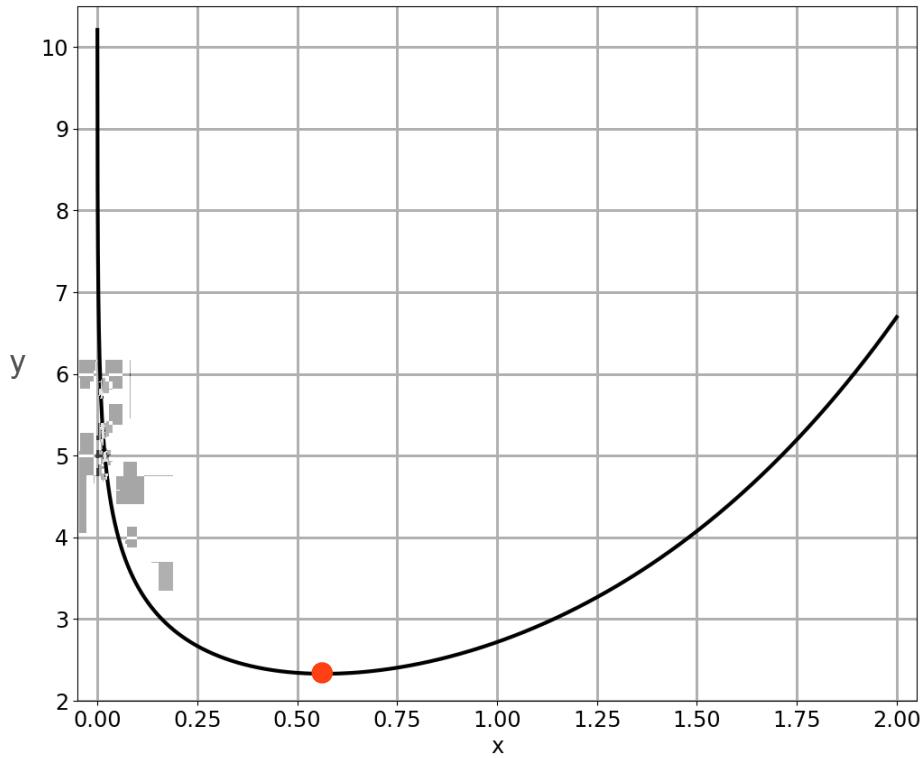


# Hard To Optimize Functions

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

Minimum?

$$f'(x)=0$$

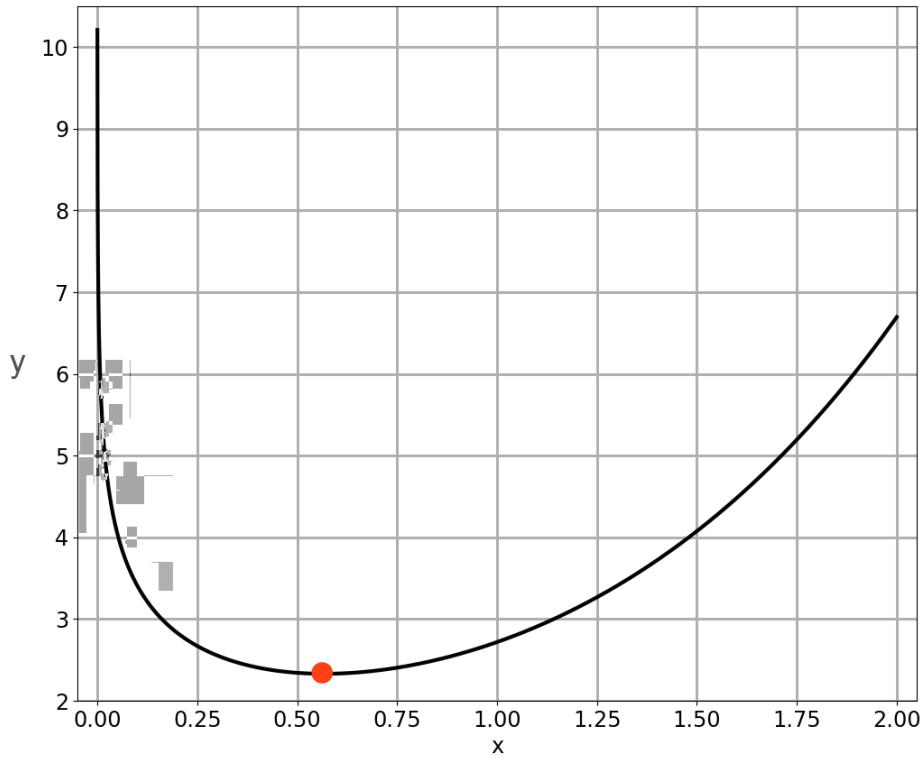


# Hard To Optimize Functions

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

Minimum?

$$f'(x) = e^x - \frac{1}{x} = 0$$

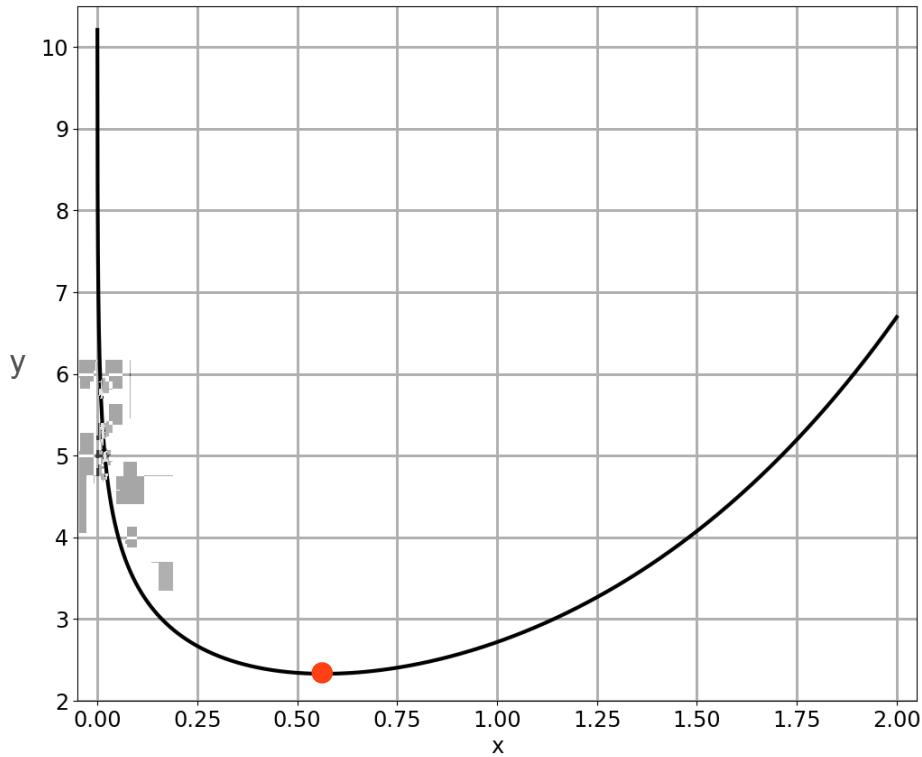


# Hard To Optimize Functions

$$f(x) = e^{\boxed{x}} - \log(x)$$

Minimum?

$$f'(x) = e^{\boxed{x}} - \frac{1}{x} = 0$$

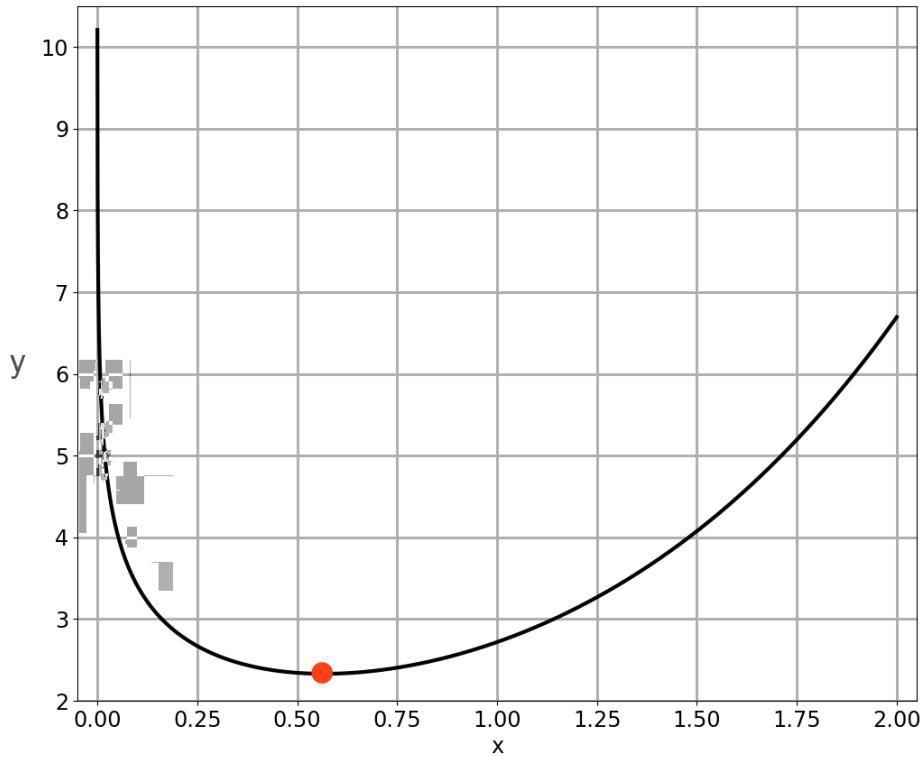


# Hard To Optimize Functions

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

Minimum?

$$f'(x) = e^x - \frac{1}{x} = 0$$

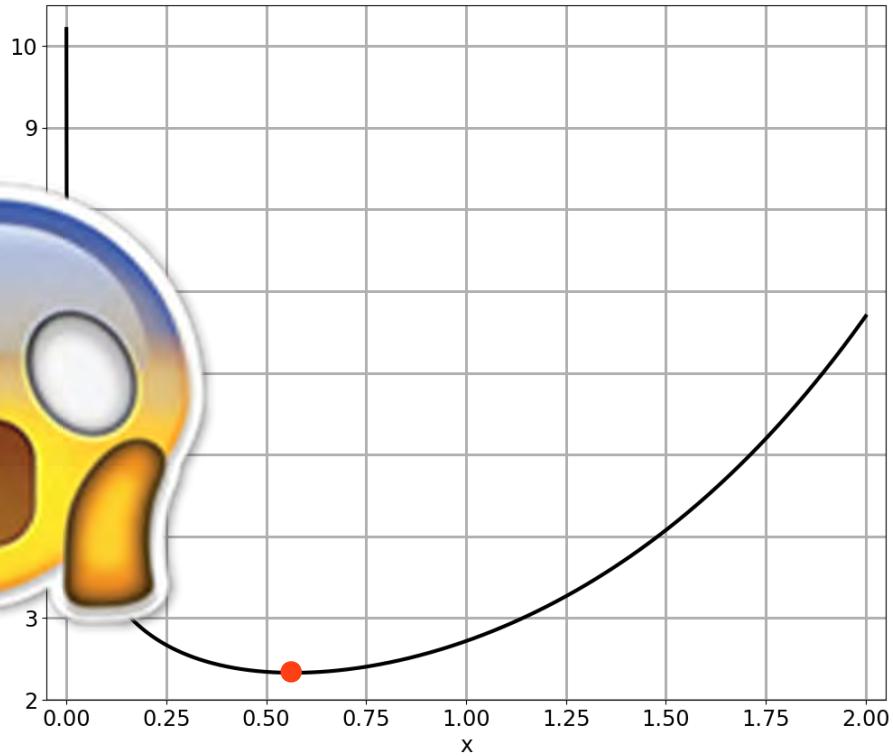


# Hard To Optimize Functions

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

Minimum?

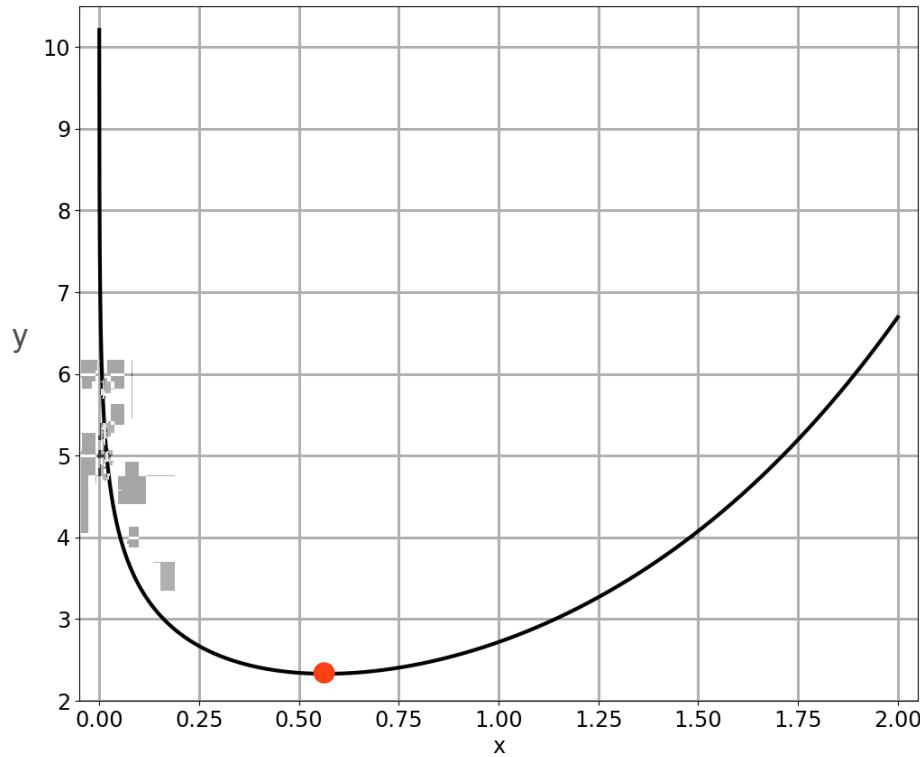
$$f'(x) = e^x - \frac{1}{x} = 0$$



# Hard To Optimize Functions

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

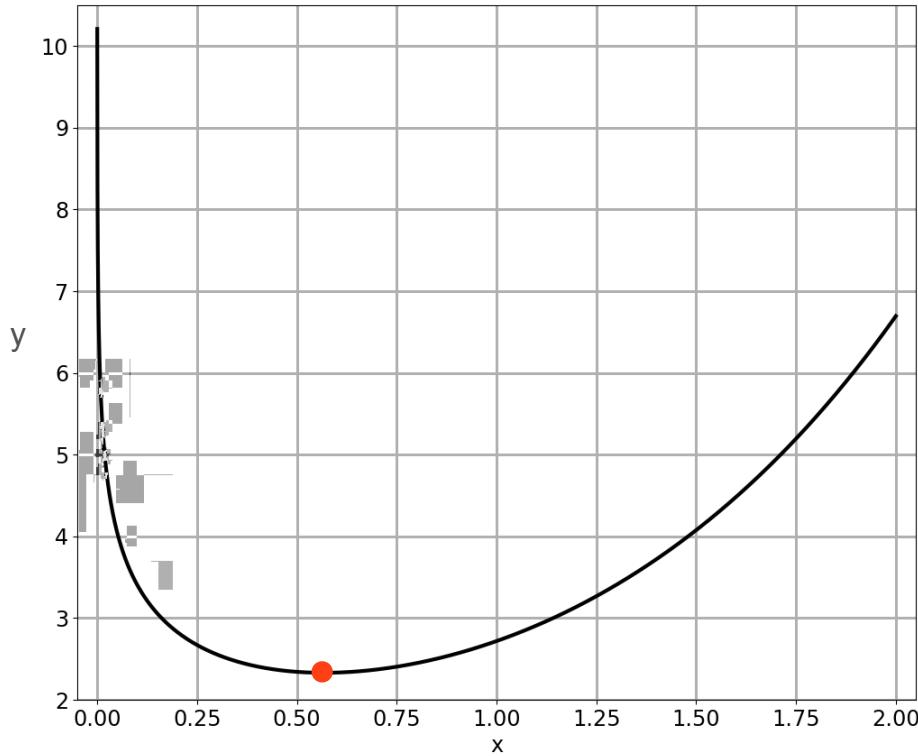
$$f'(x) = e^x - \frac{1}{x}$$



# Hard To Optimize Functions

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

$$f'(x) = e^x - \frac{1}{x} = 0$$



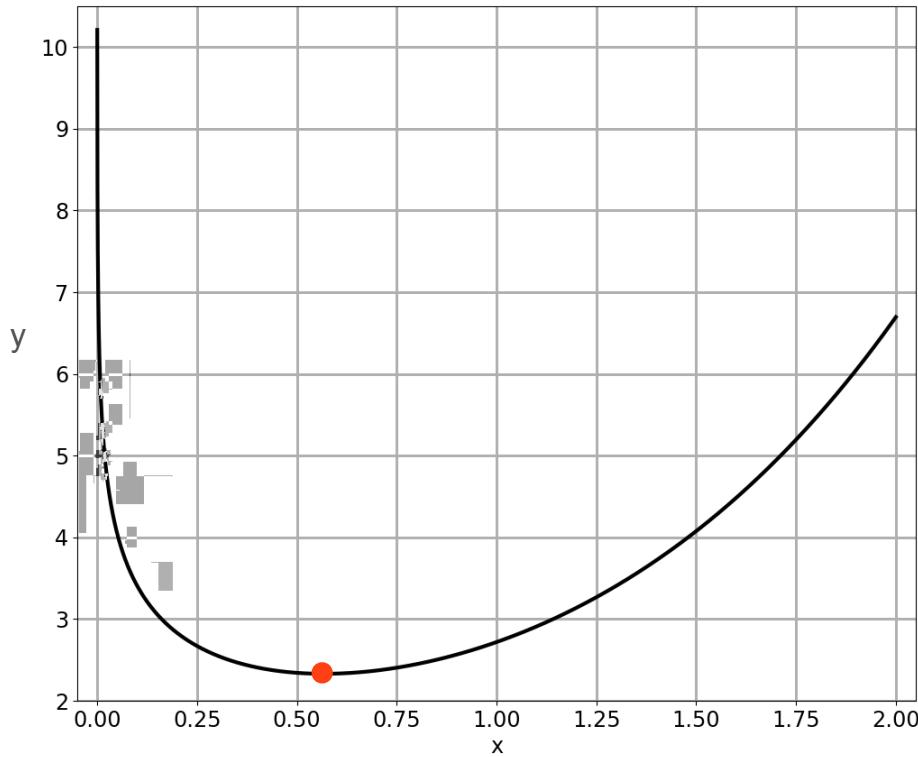
# Hard To Optimize Functions

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

$$f'(x) = e^x - \frac{1}{x} = 0$$



$$e^x = \frac{1}{x}$$



# Hard To Optimize Functions

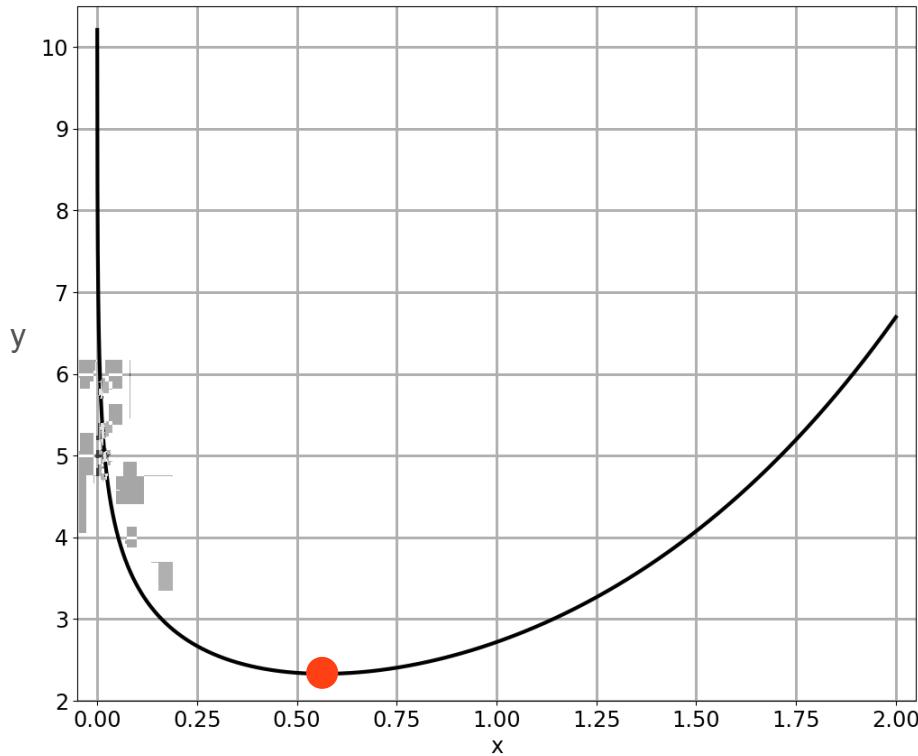
$$f(x) = e^x - \log(x)$$

$$f'(x) = e^x - \frac{1}{x} = 0$$



$$e^x = \frac{1}{x}$$

Solution:  $x = 0.5671\dots$



# Hard To Optimize Functions

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

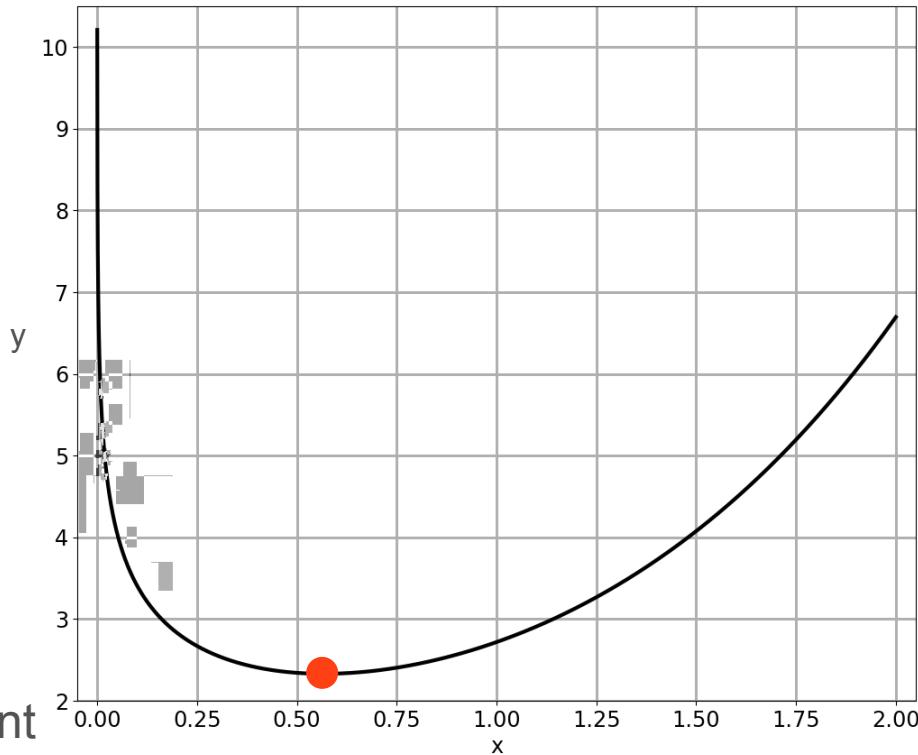
$$f'(x) = e^x - \frac{1}{x} = 0$$



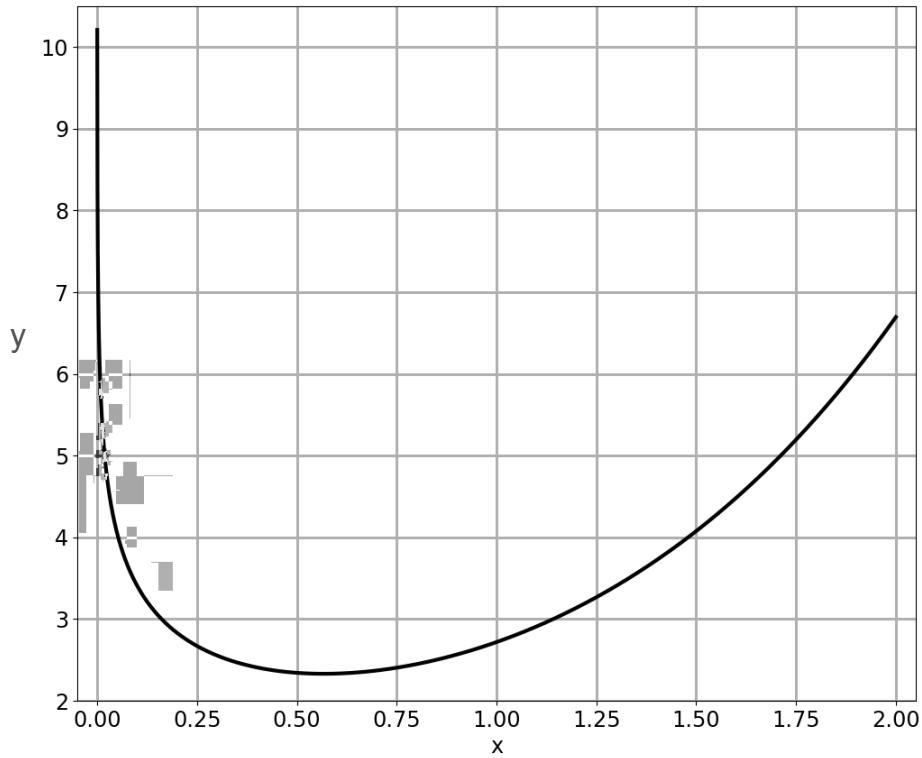
$$e^x = \frac{1}{x}$$

Solution:  $x = 0.5671\dots$

Also known as the Omega constant

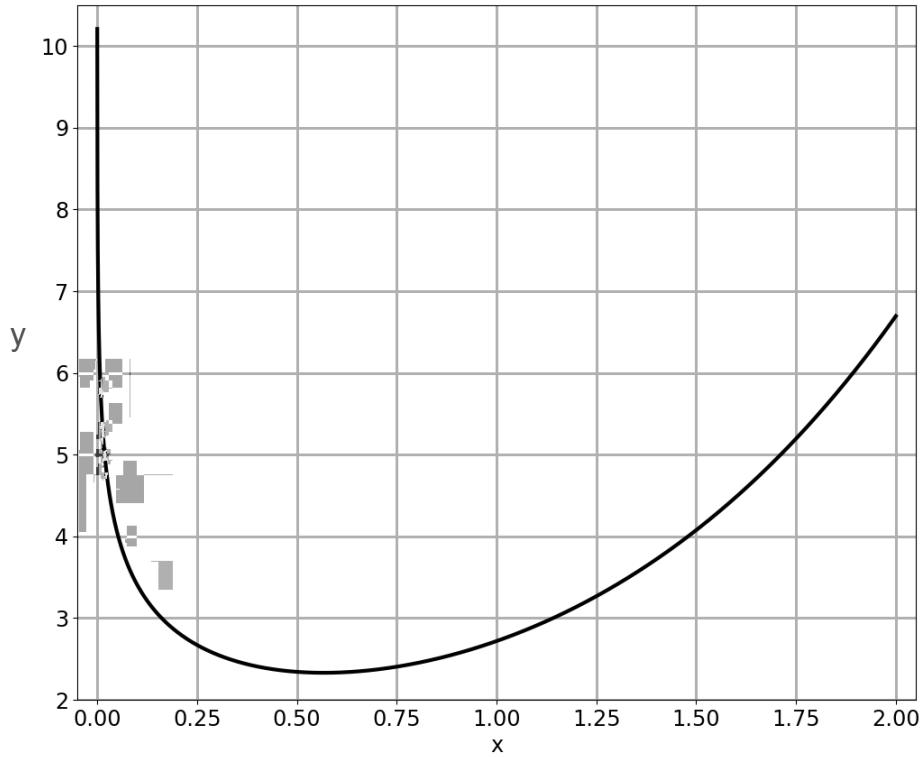


# Method 1: Try Both Directions



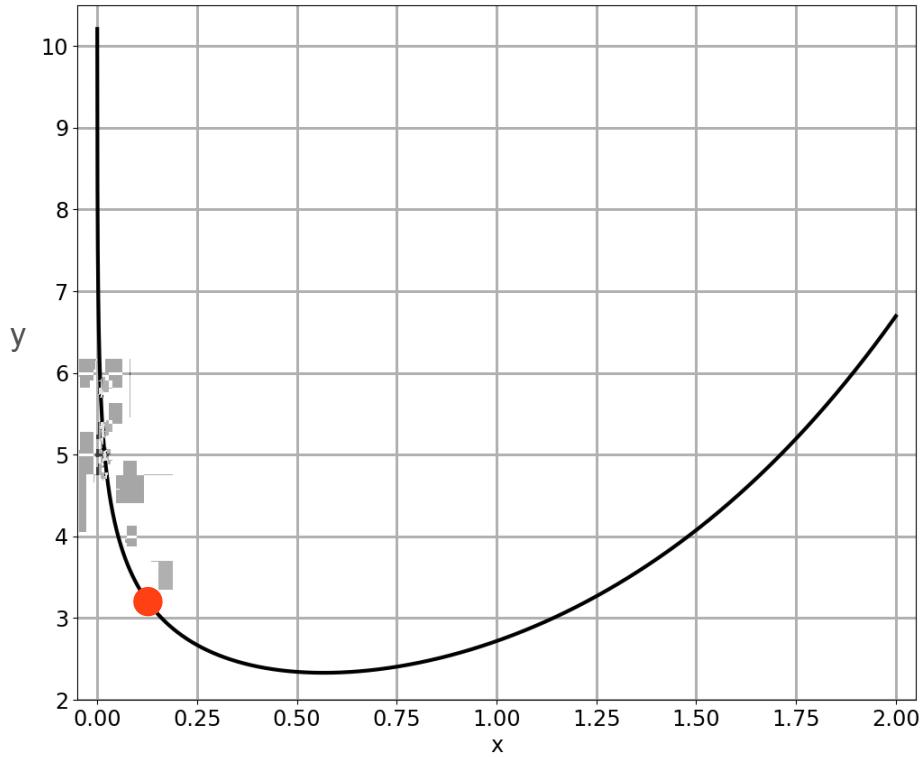
# Method 1: Try Both Directions

Is there any  
other way?



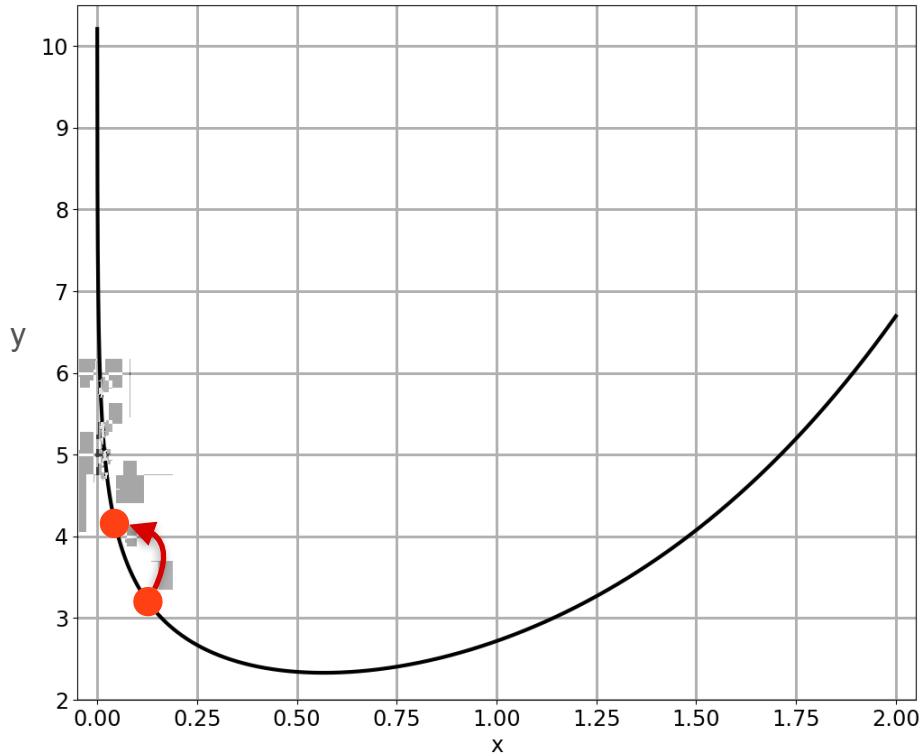
# Method 1: Try Both Directions

Is there any  
other way?



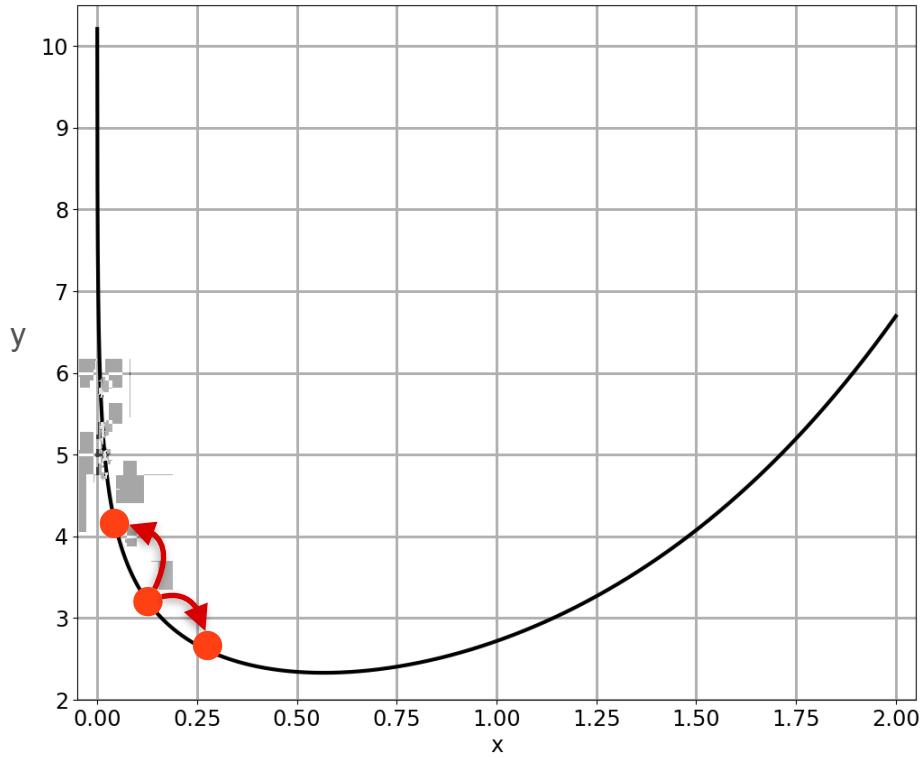
# Method 1: Try Both Directions

Is there any  
other way?



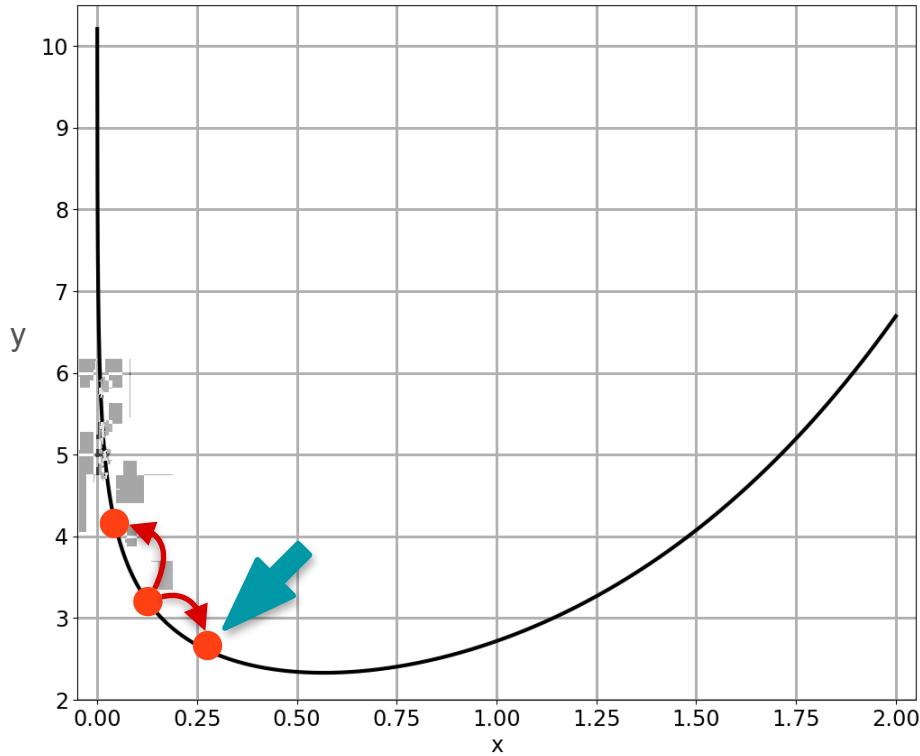
# Method 1: Try Both Directions

Is there any  
other way?



# Method 1: Try Both Directions

Is there any  
other way?

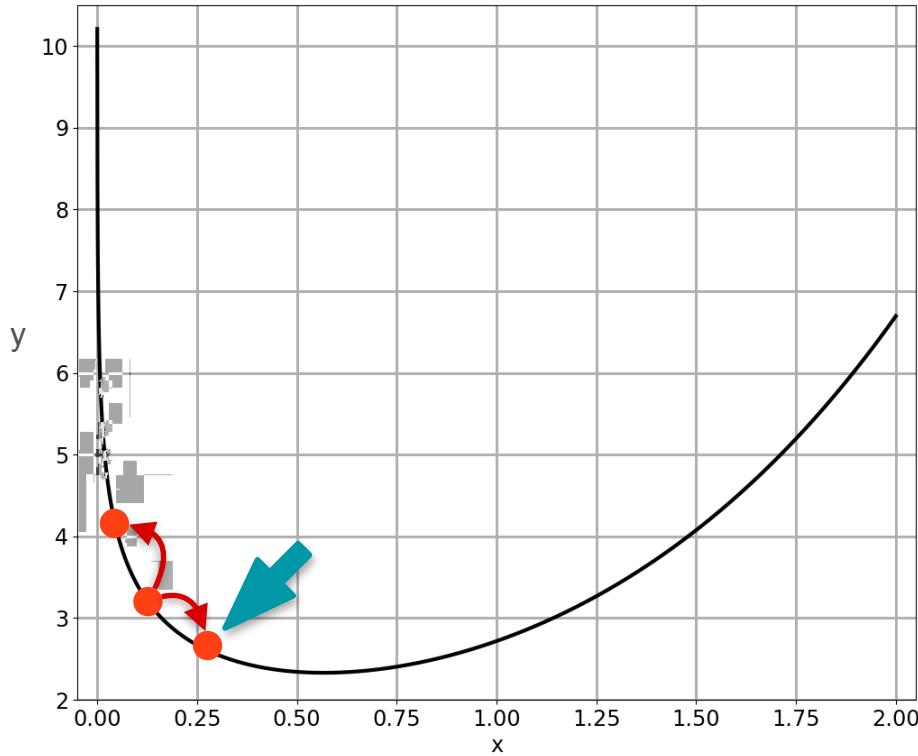


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

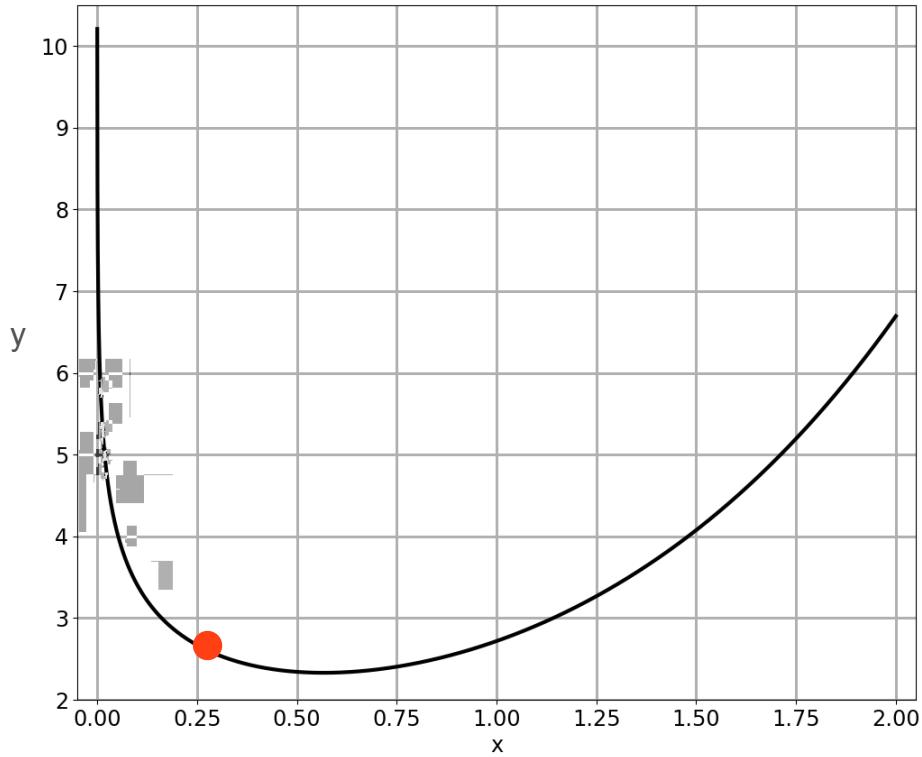


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

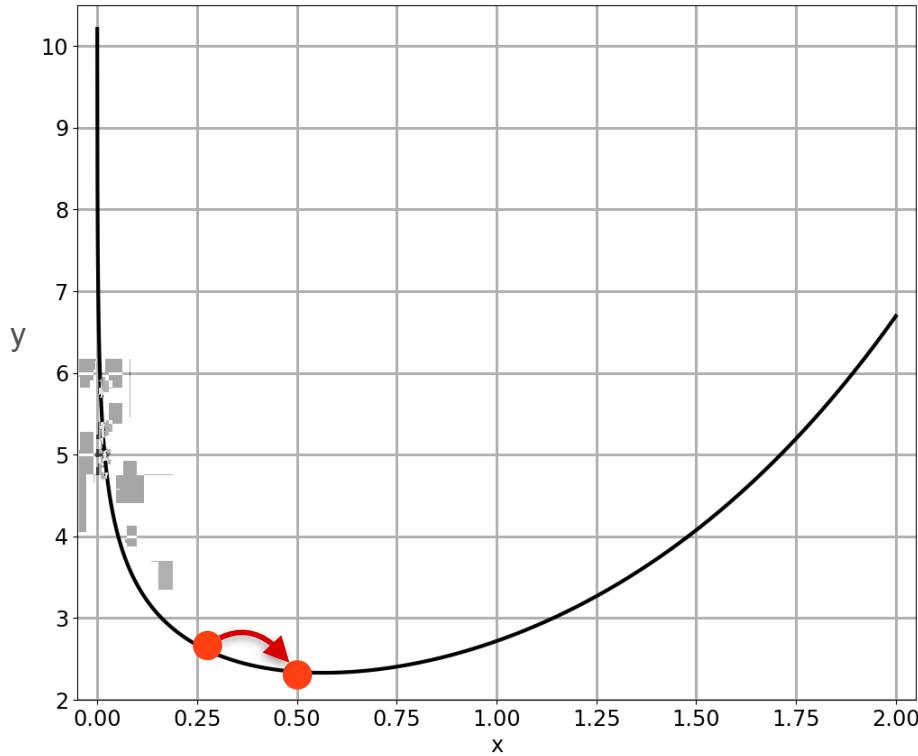


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

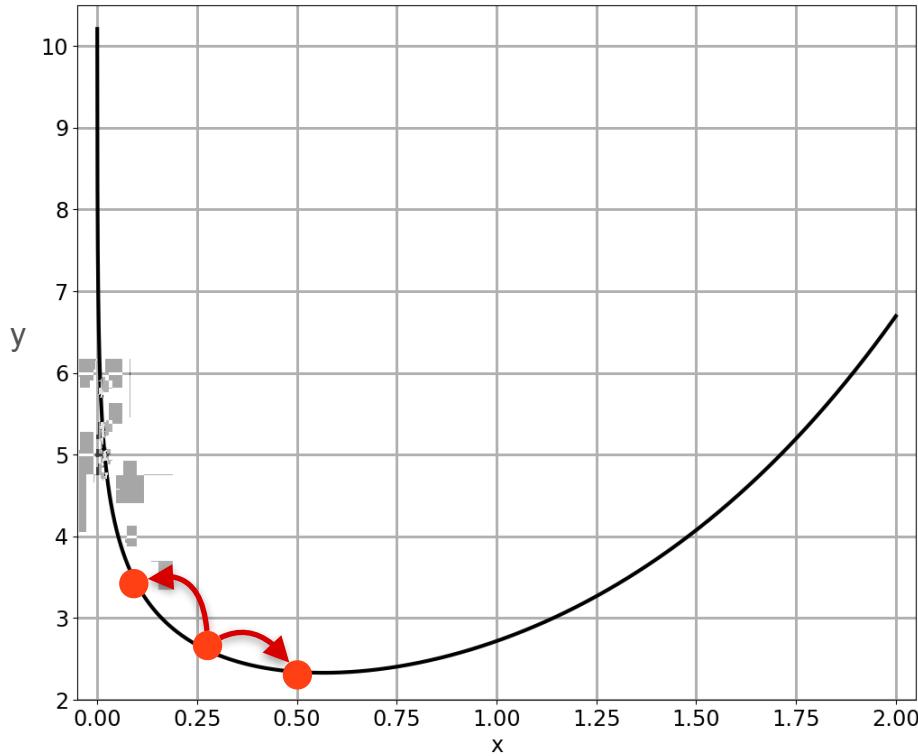


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

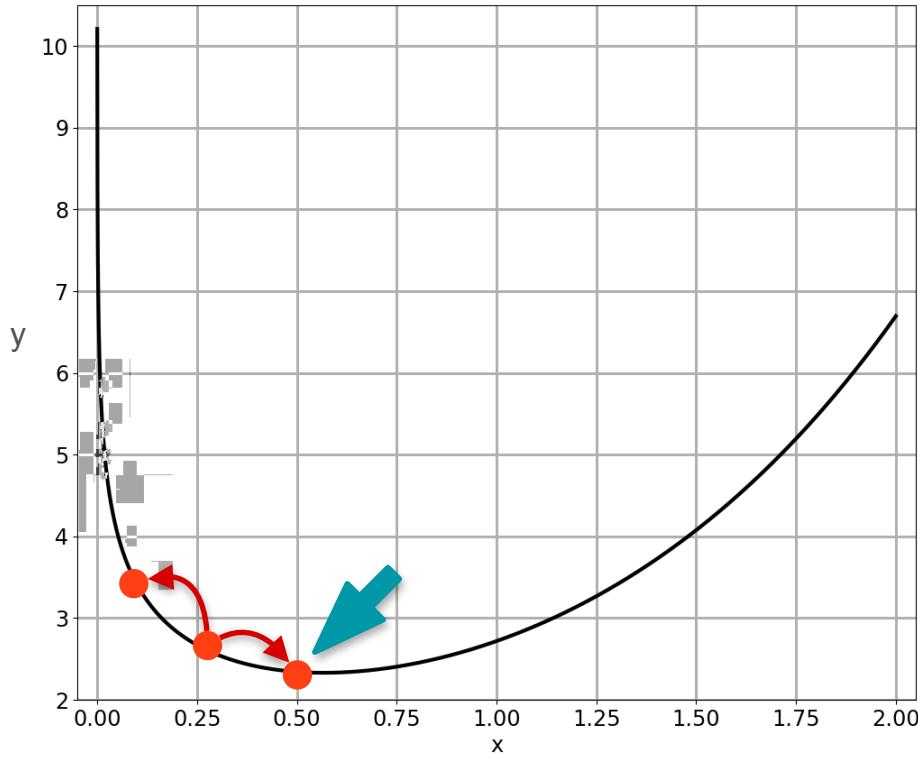


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

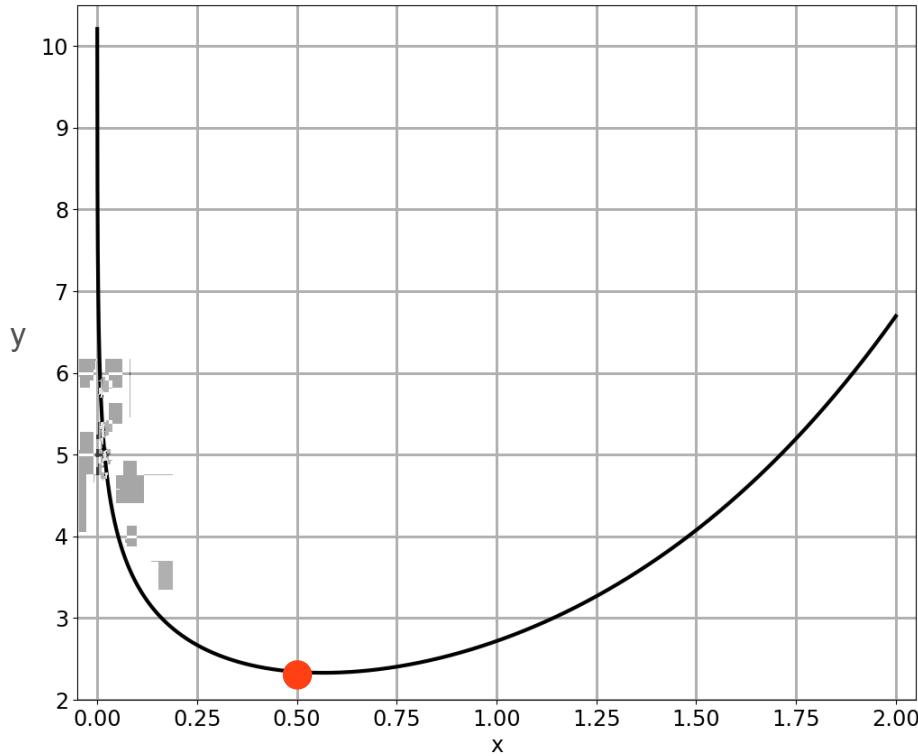


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

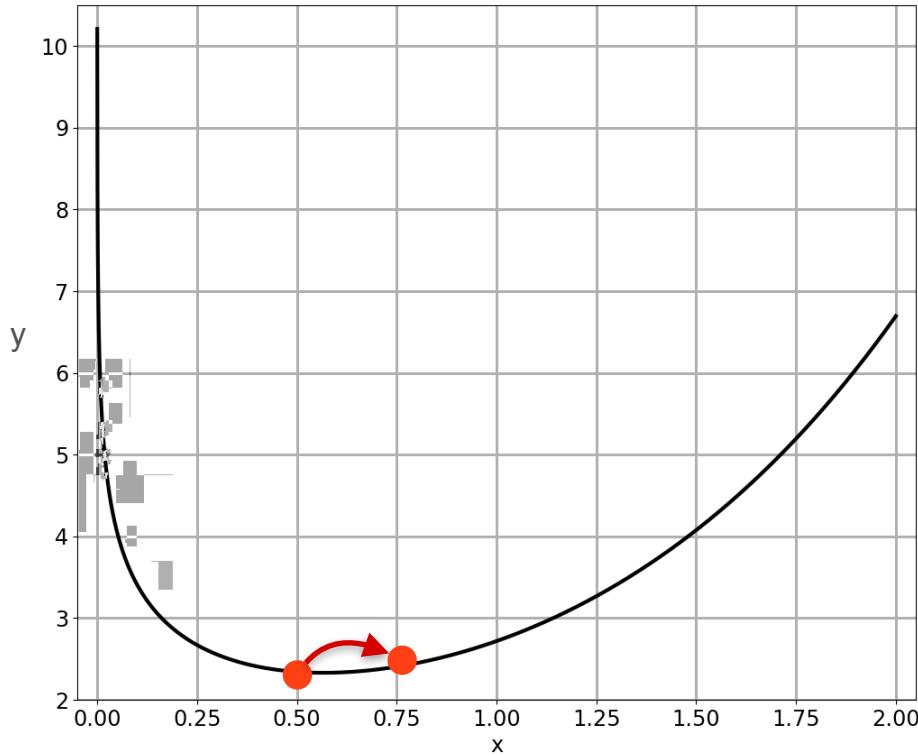


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

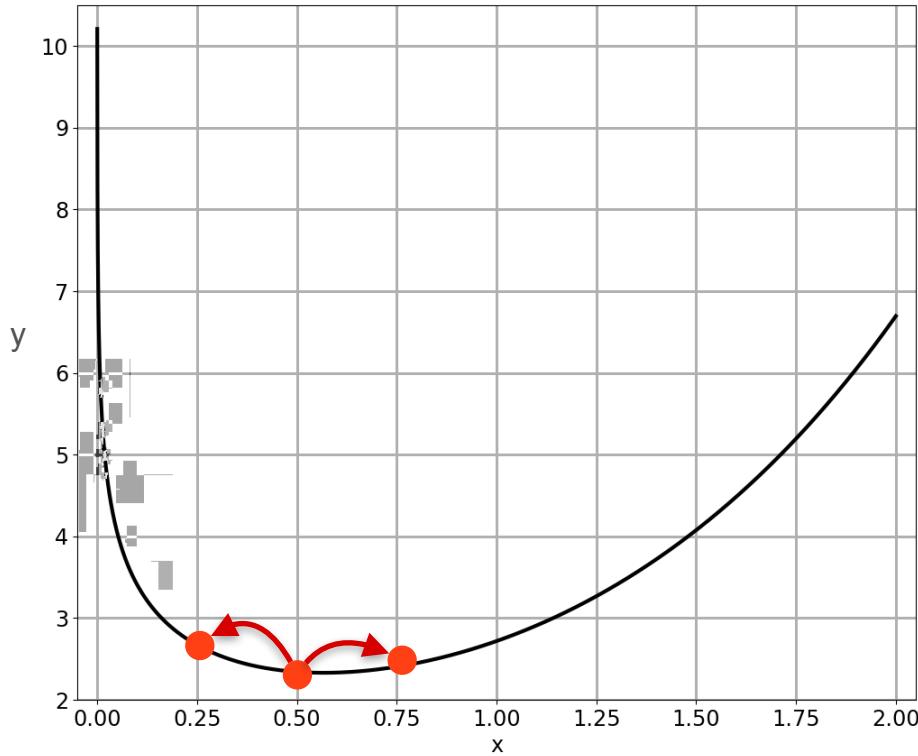


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

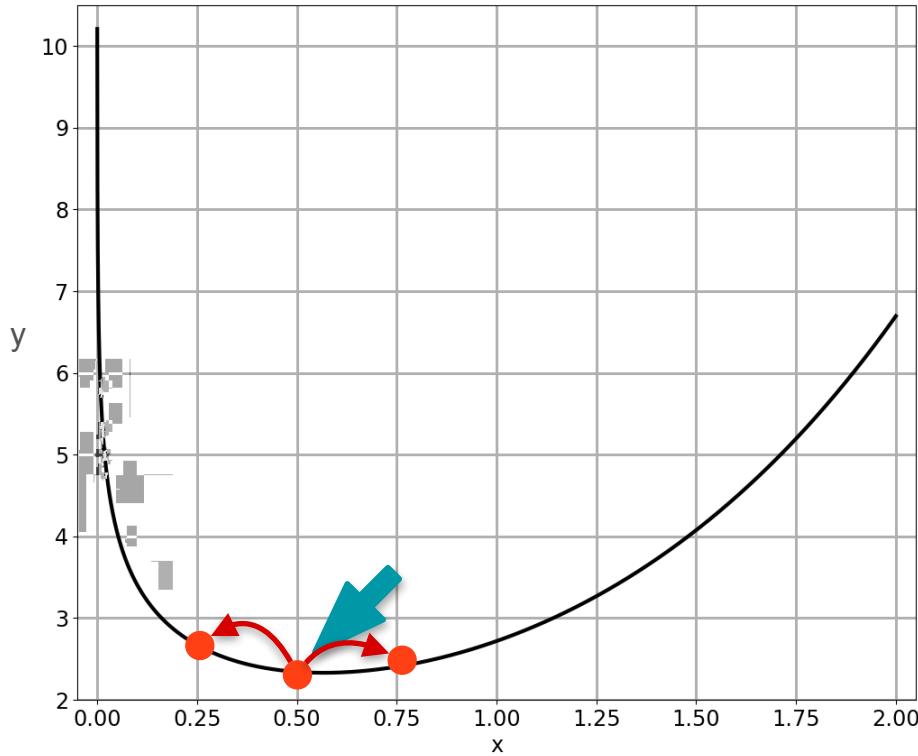


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!

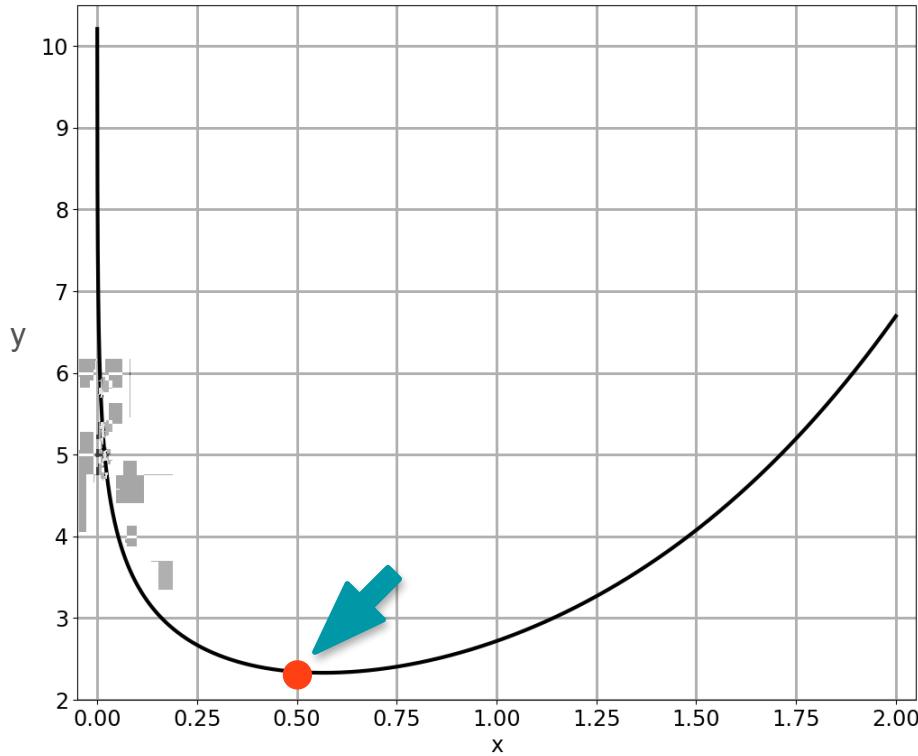


# Method 1: Try Both Directions

Is there any  
other way?



Repeat!





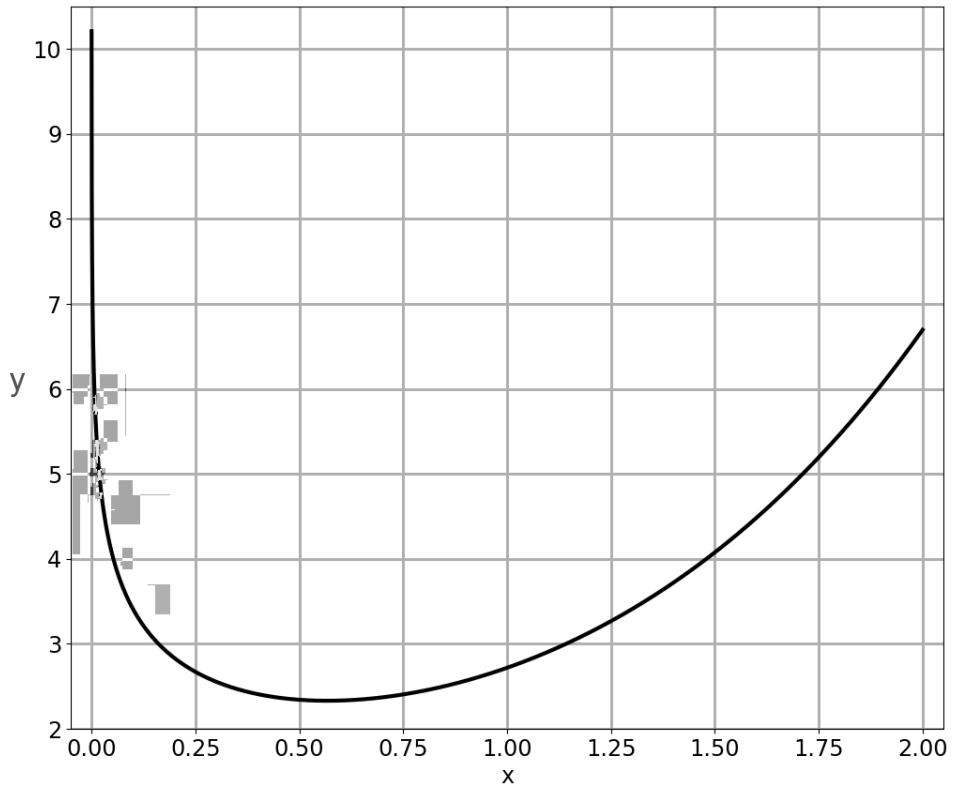
DeepLearning.AI

# Gradients and Gradient Descent

---

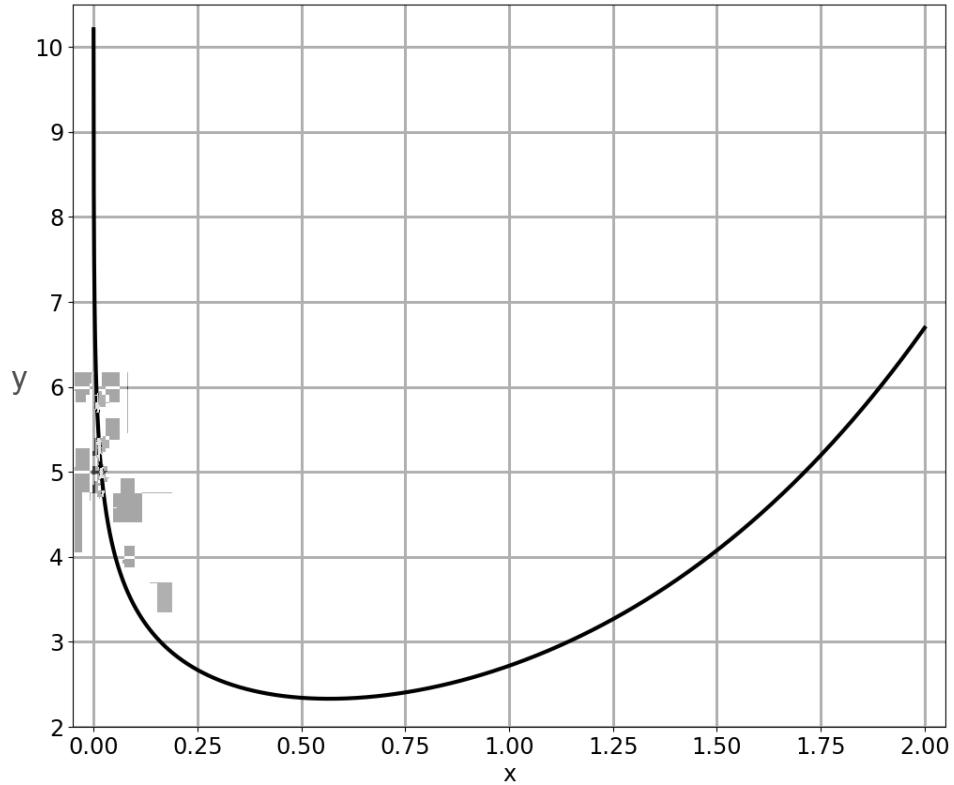
**Optimization using Gradient  
Descent in one variable -  
Part 2**

# Method 2: Be Clever



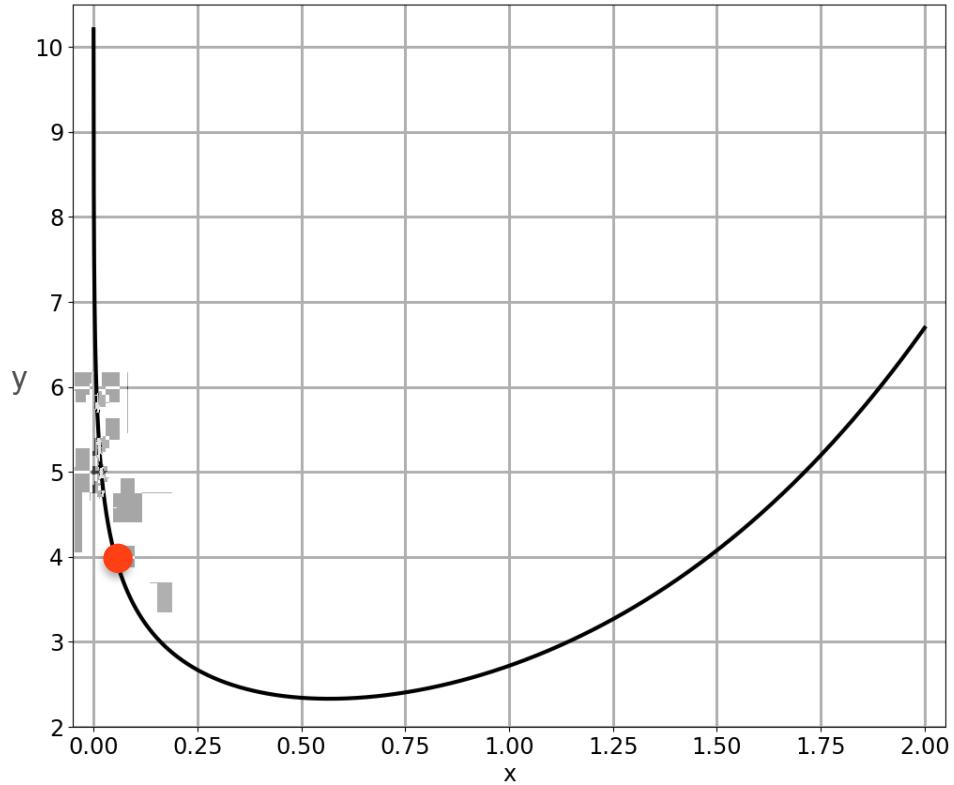
# Method 2: Be Clever

Try something  
smarter...



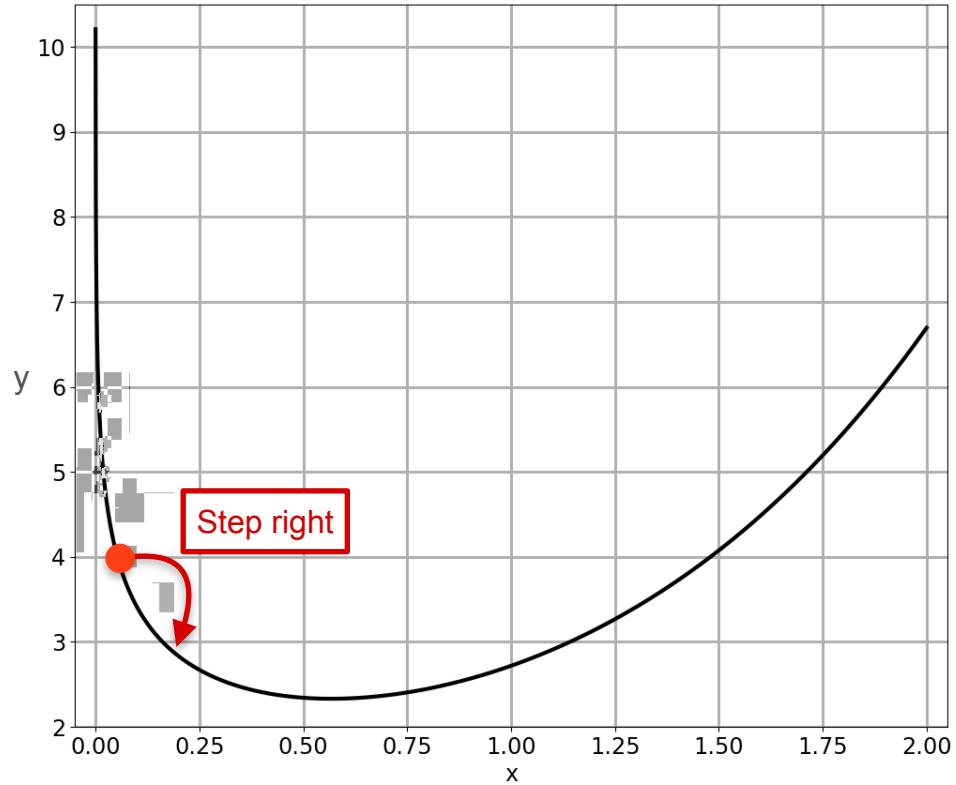
# Method 2: Be Clever

Try something  
smarter...



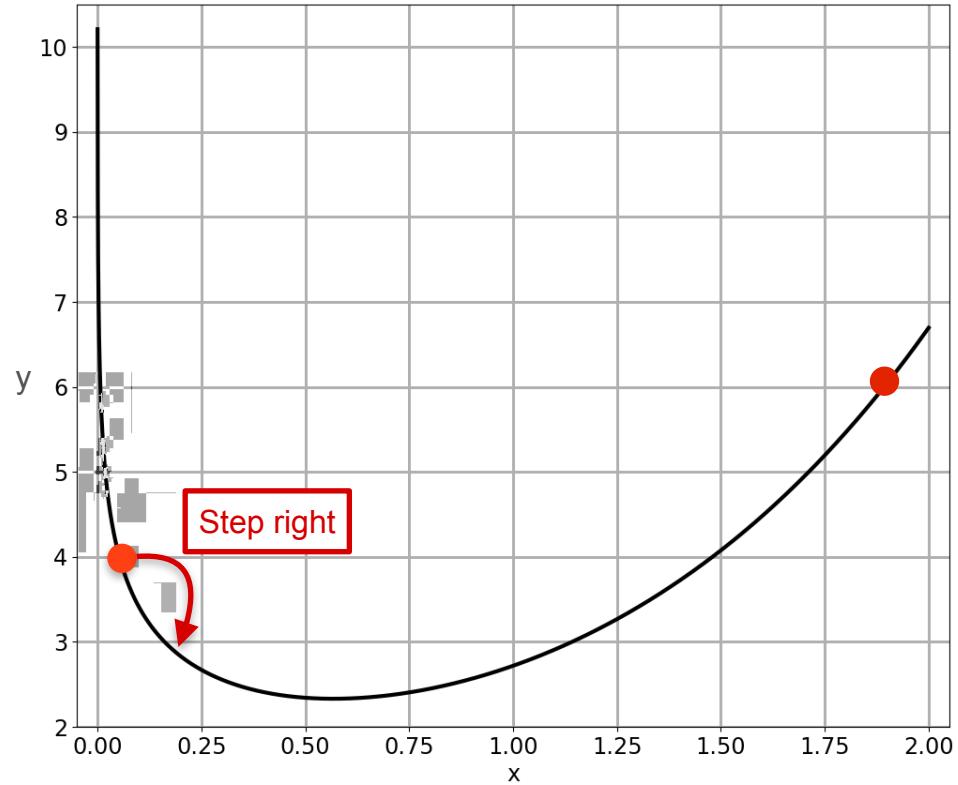
# Method 2: Be Clever

Try something  
smarter...



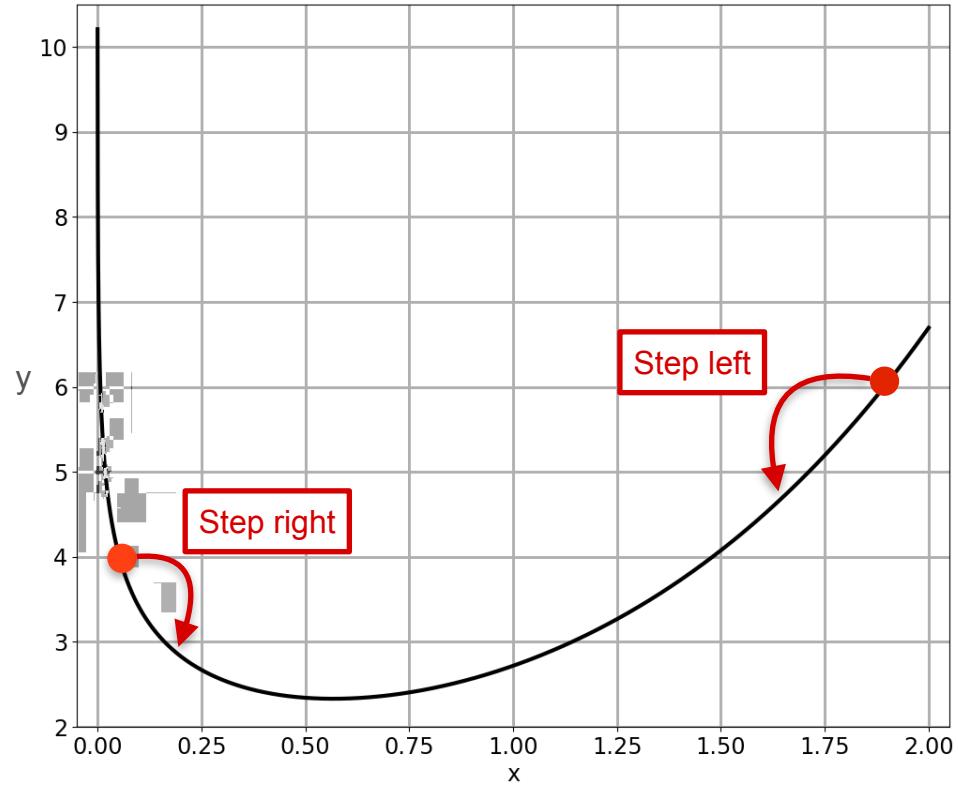
# Method 2: Be Clever

Try something  
smarter...



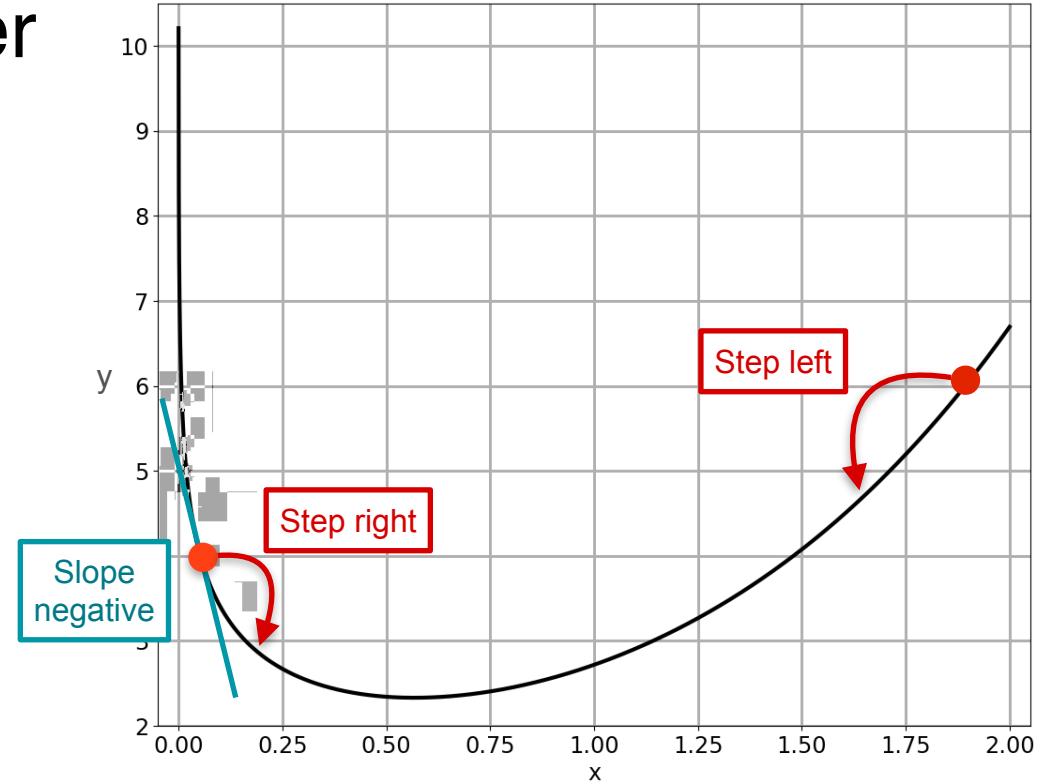
# Method 2: Be Clever

Try something  
smarter...



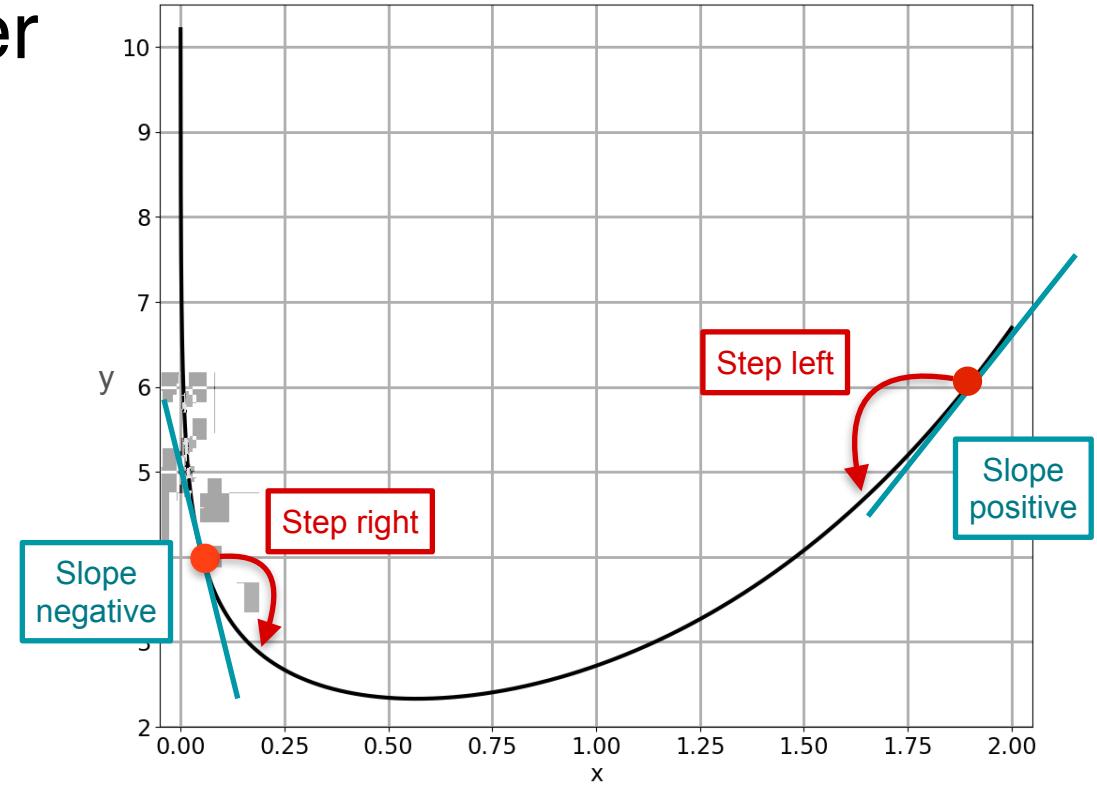
# Method 2: Be Clever

Try something  
smarter...



# Method 2: Be Clever

Try something  
smarter...

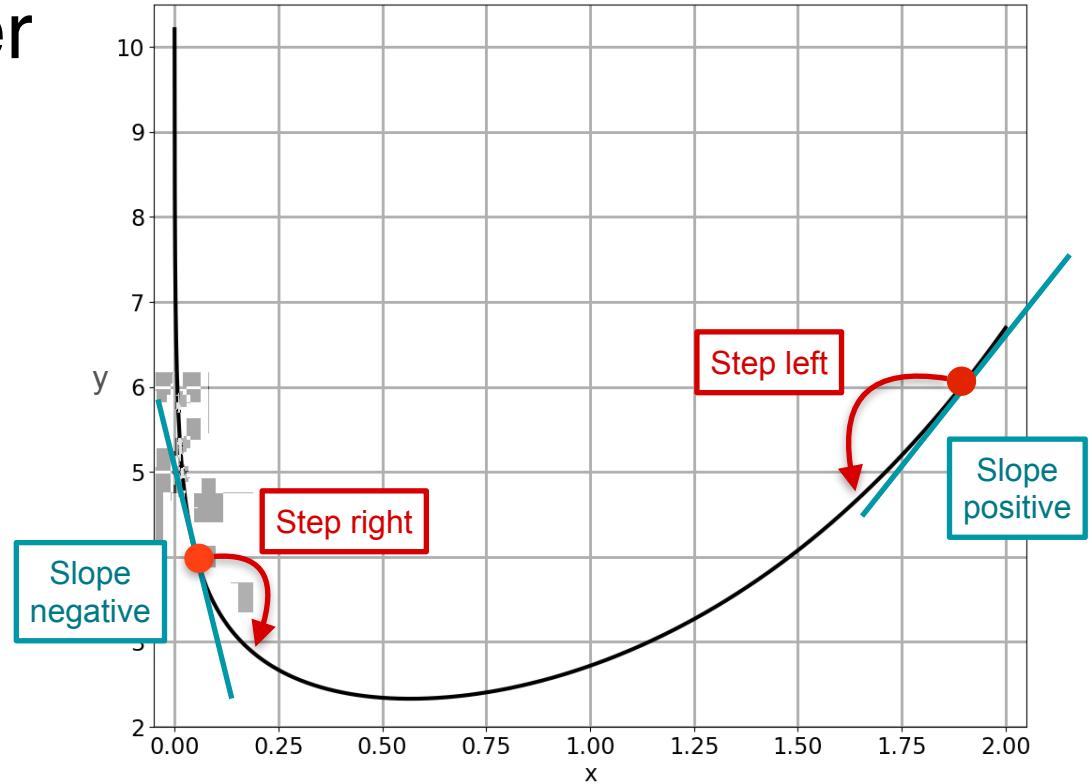


# Method 2: Be Clever

Try something  
smarter...



new point

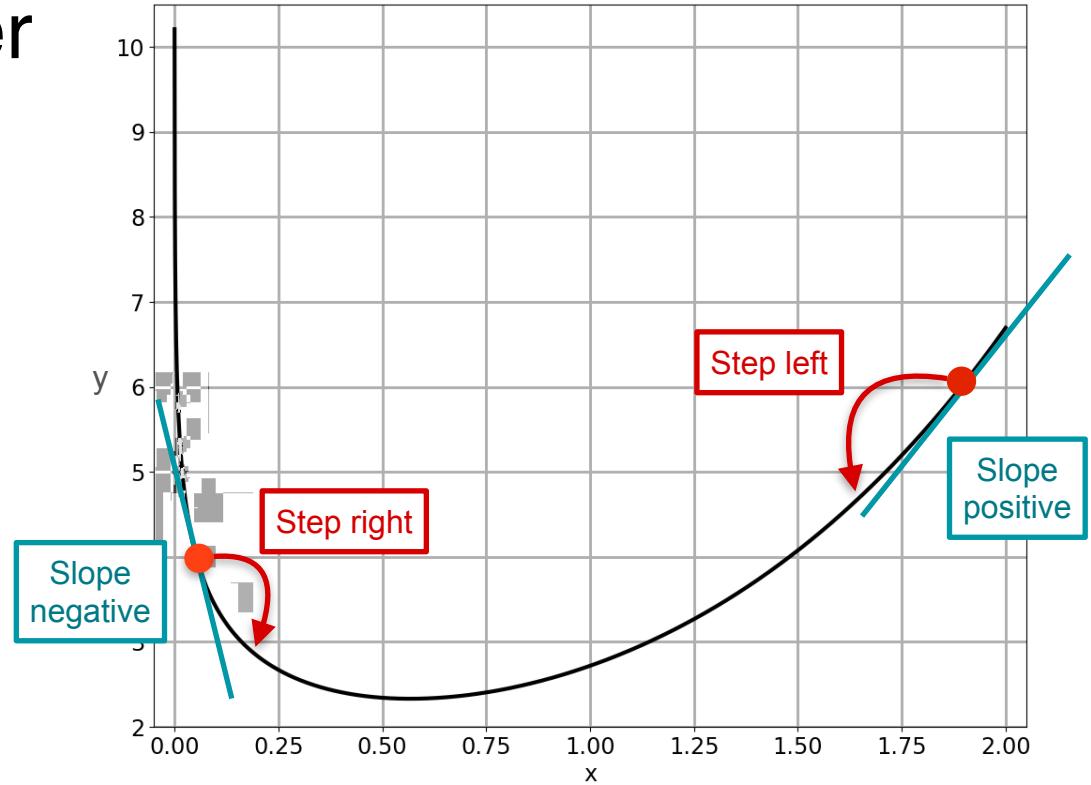


# Method 2: Be Clever

Try something  
smarter...



new point = old point

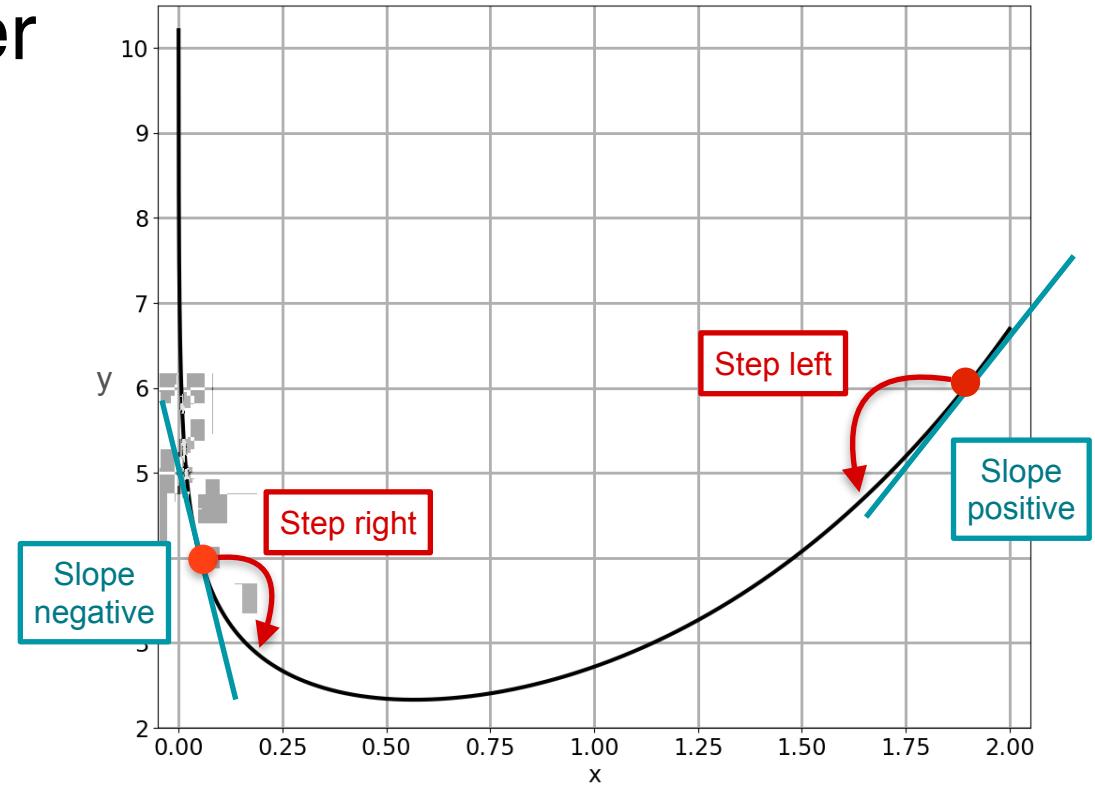


# Method 2: Be Clever

Try something  
smarter...



new point = old point - slope



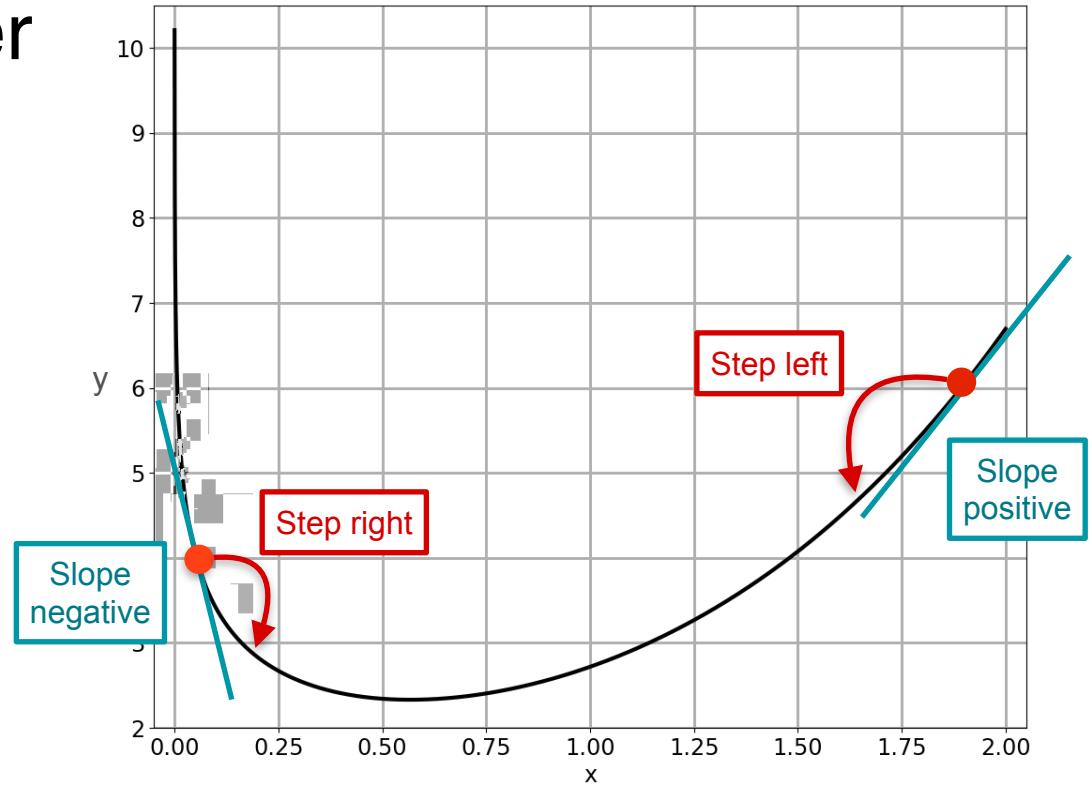
# Method 2: Be Clever

Try something  
smarter...



new point = old point - slope

$x_1$



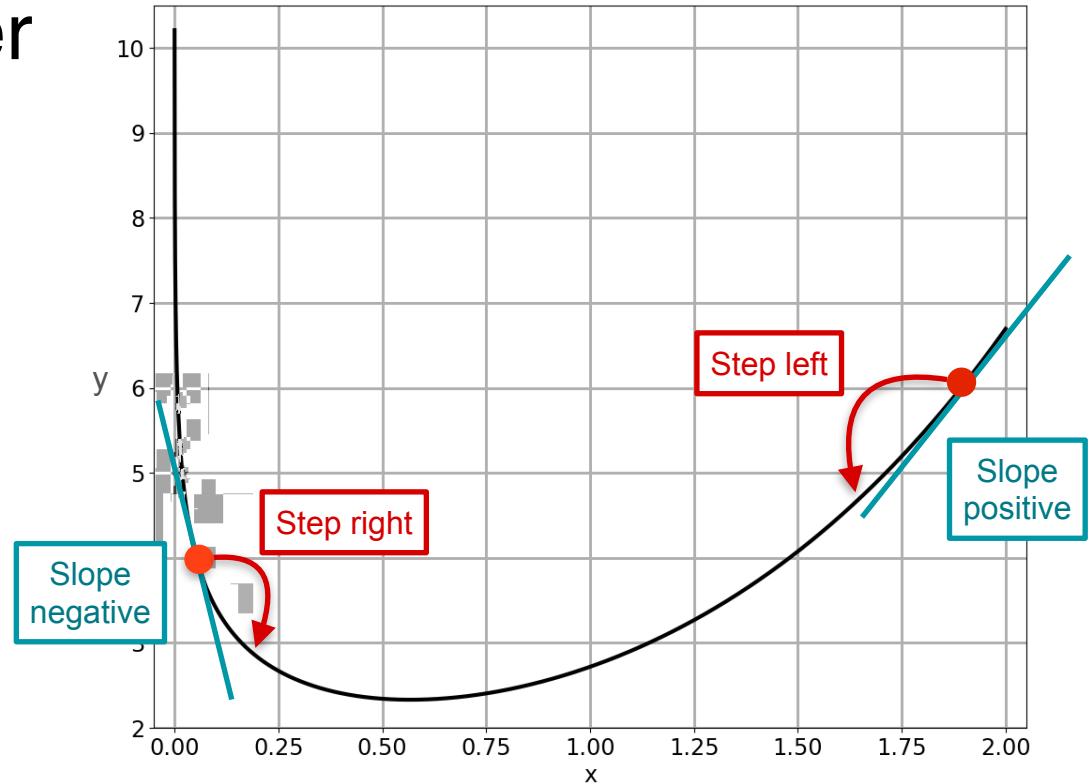
# Method 2: Be Clever

Try something  
smarter...



new point = old point - slope

$$x_1 = x_0$$



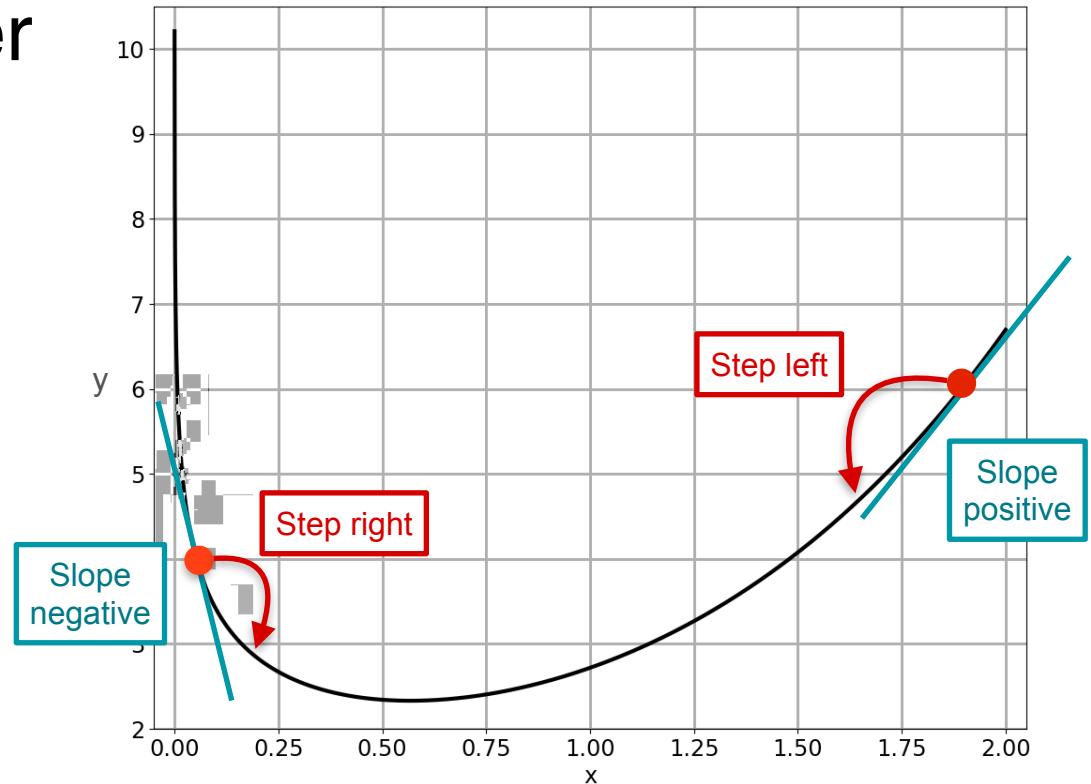
# Method 2: Be Clever

Try something  
smarter...



new point = old point - slope

$$x_1 = x_0 - f'(x_0)$$



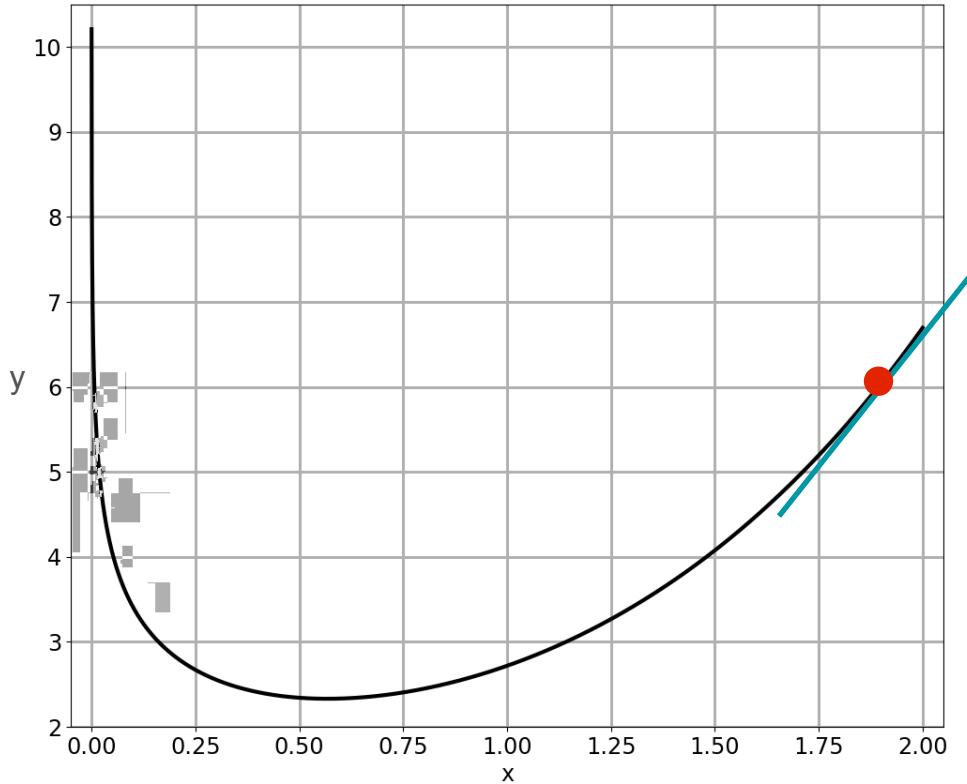
# Method 2: Be Clever

Try something  
smarter...



new point = old point - slope

$$x_1 = x_0 - f'(x_0)$$



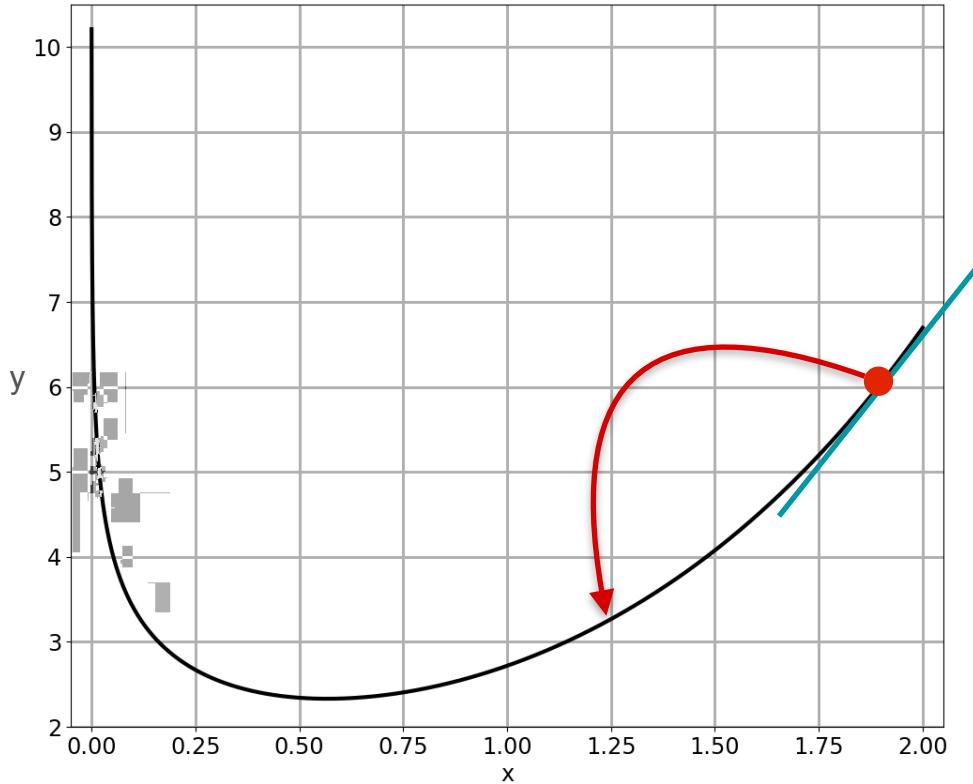
# Method 2: Be Clever

Try something  
smarter...



new point = old point - slope

$$x_1 = x_0 - f'(x_0)$$



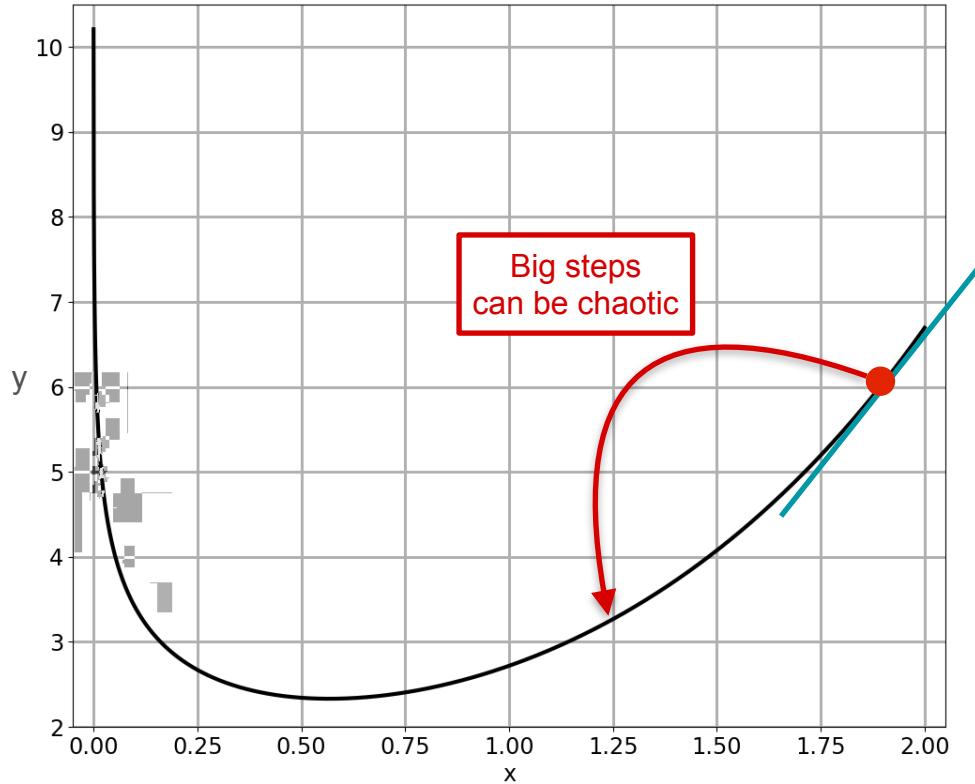
# Method 2: Be Clever

Try something  
smarter...



new point = old point - slope

$$x_1 = x_0 - f'(x_0)$$



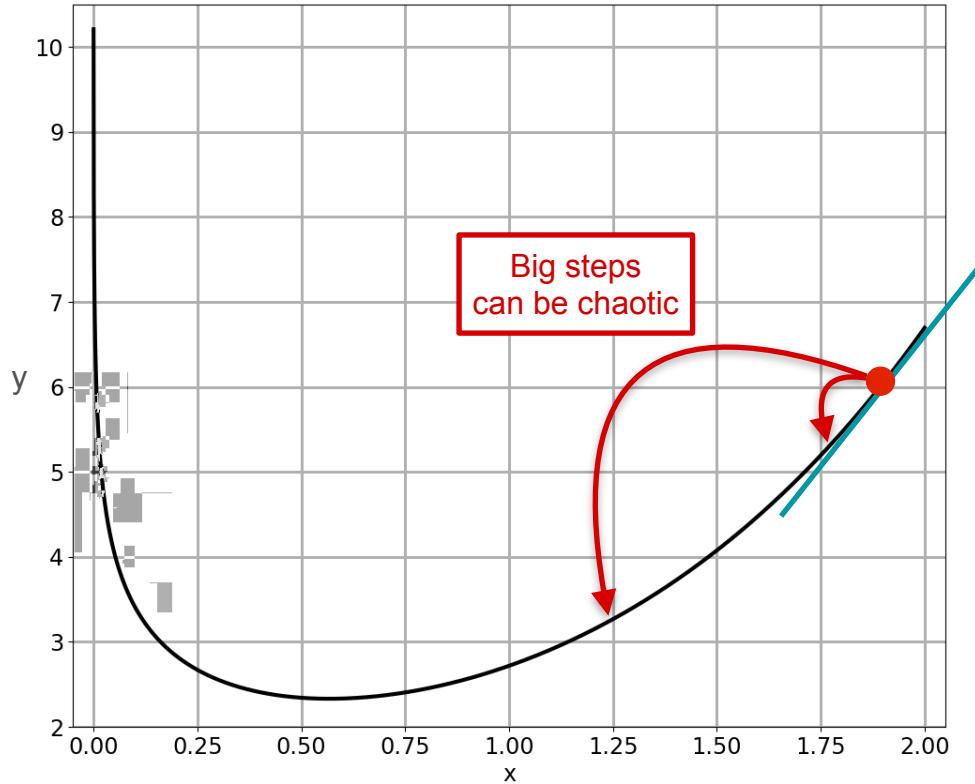
# Method 2: Be Clever

Try something  
smarter...



new point = old point - slope

$$x_1 = x_0 - f'(x_0)$$



# Method 2: Be Clever

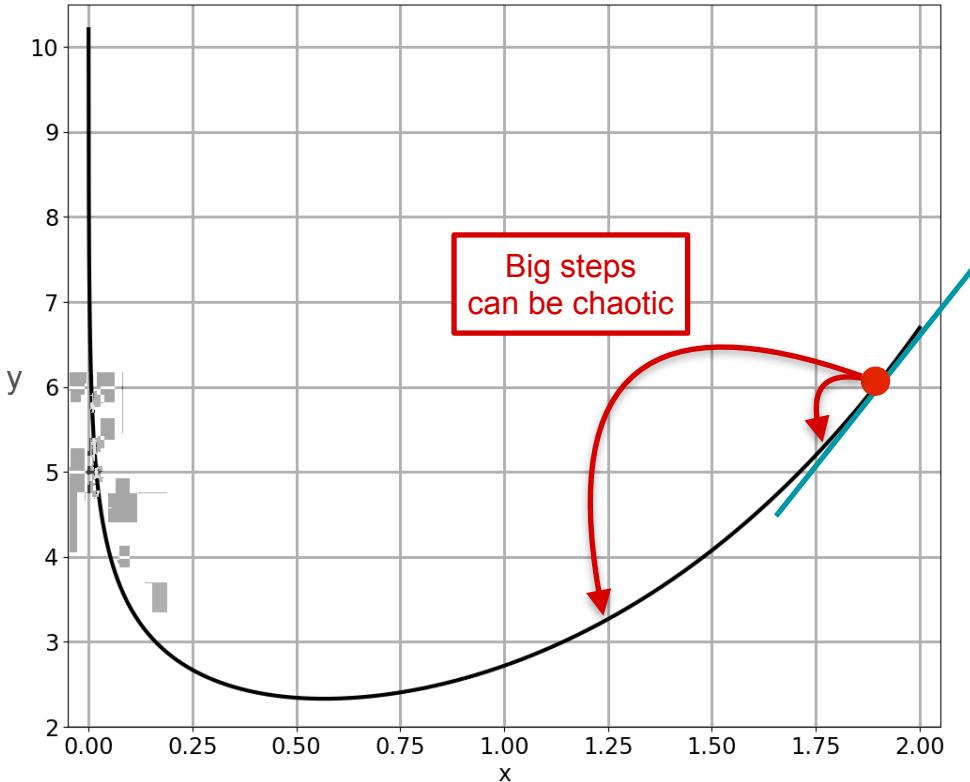
Try something  
smarter...



new point = old point - slope

$$x_1 = x_0 - f'(x_0)$$

$$x_1 = x_0 - 0.01 f'(x_0)$$



# Method 2: Be Clever

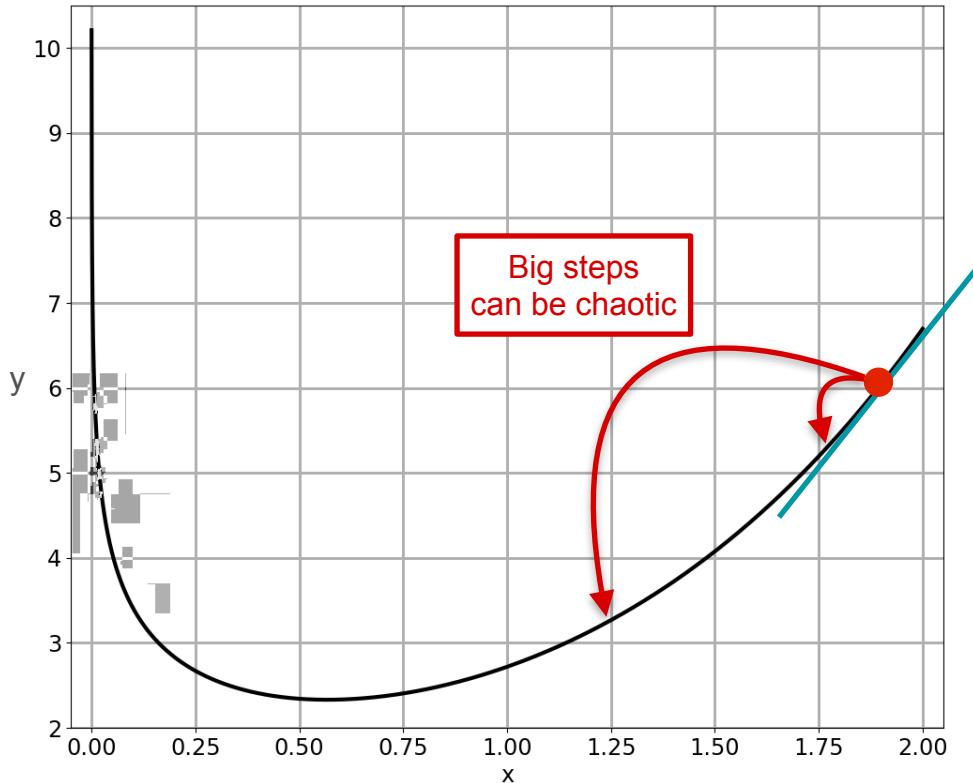
Try something  
smarter...



new point = old point - slope

$$x_1 = x_0 - f'(x_0)$$

$$x_1 = x_0 - \alpha f'(x_0)$$



# Method 2: Be Clever

Try something  
smarter...



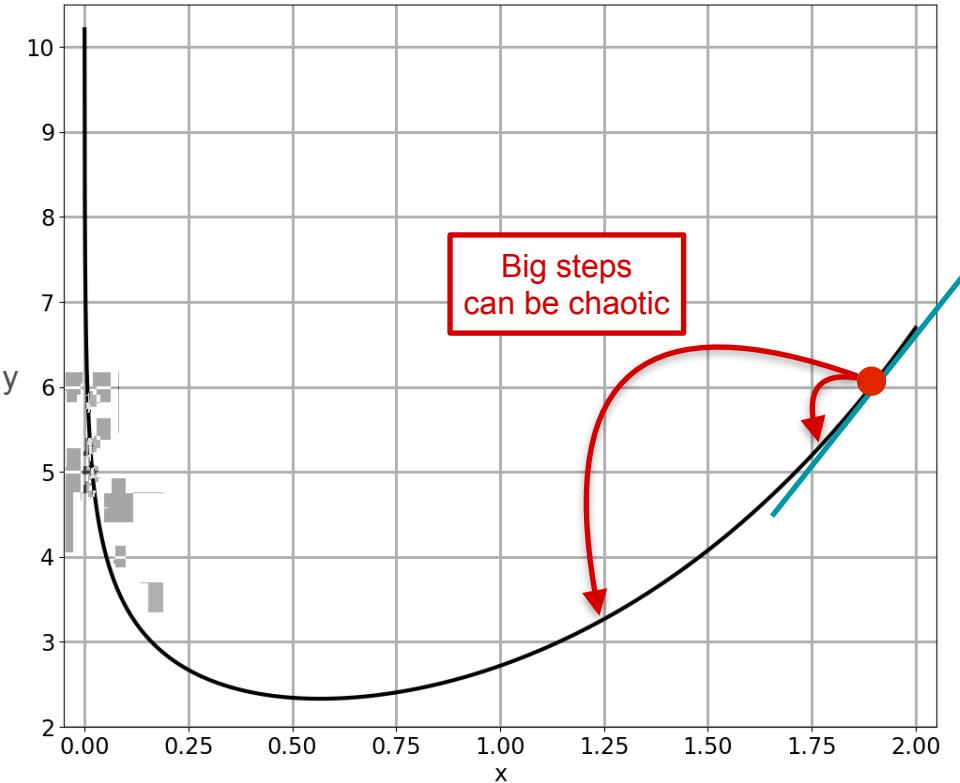
new point = old point - slope

$$x_1 = x_0 - f'(x_0)$$

$$x_1 = x_0 - \alpha f'(x_0)$$



Learning rate

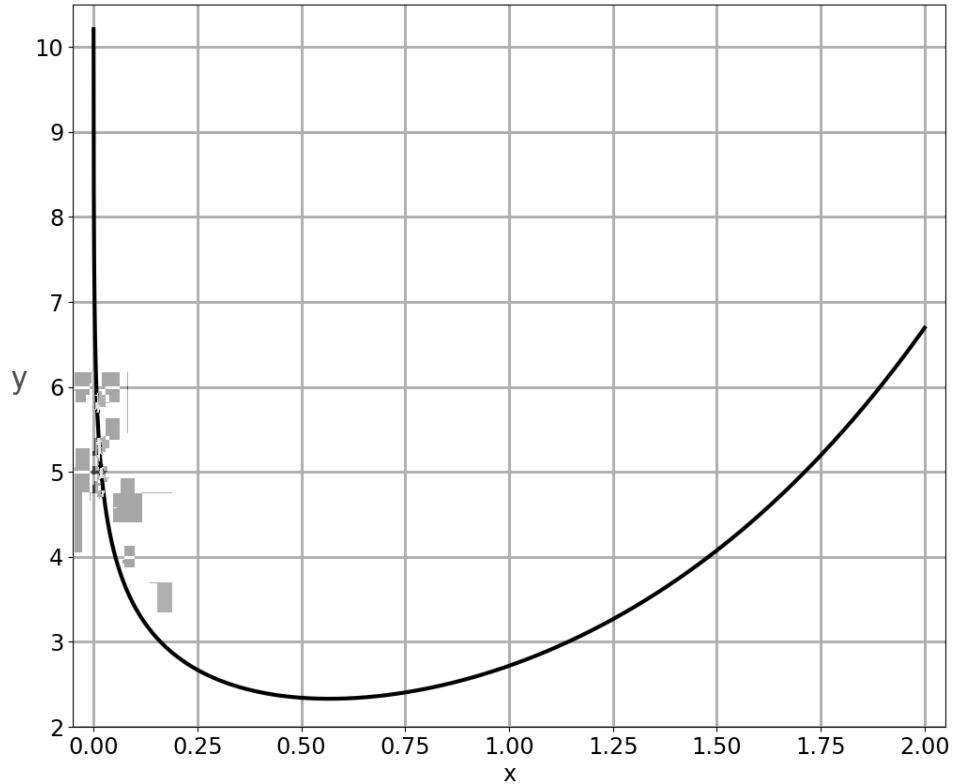


# Method 2: Be Clever

Try something  
smarter...



$$x_1 = x_0 - \alpha f'(x_0)$$

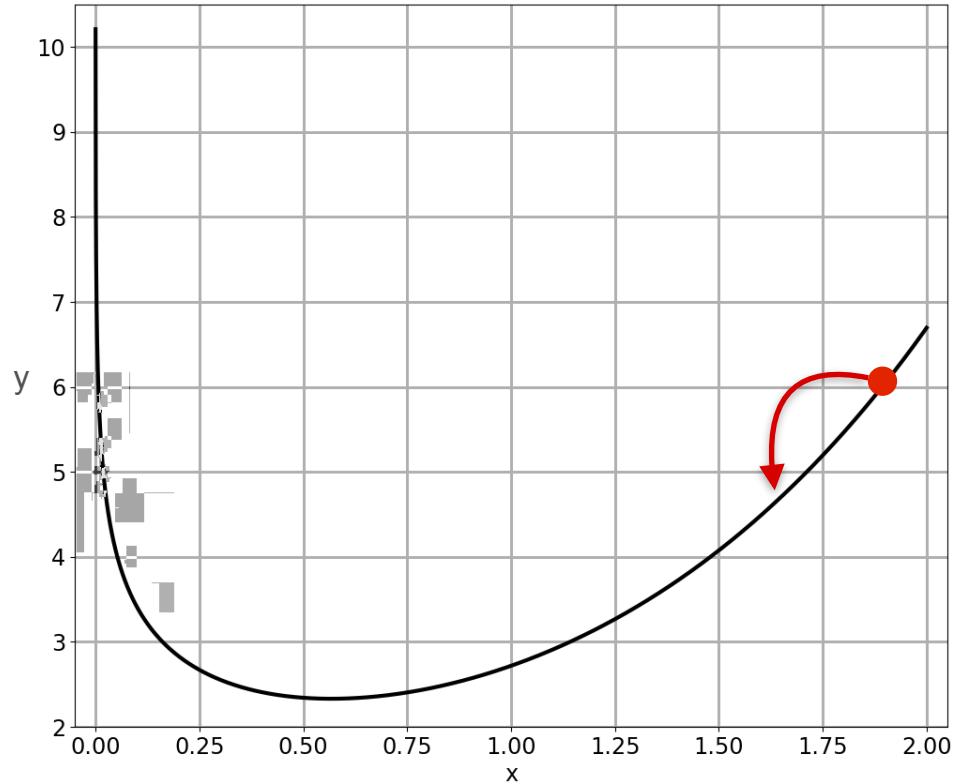


# Method 2: Be Clever

Try something  
smarter...



$$x_1 = x_0 - \alpha f'(x_0)$$

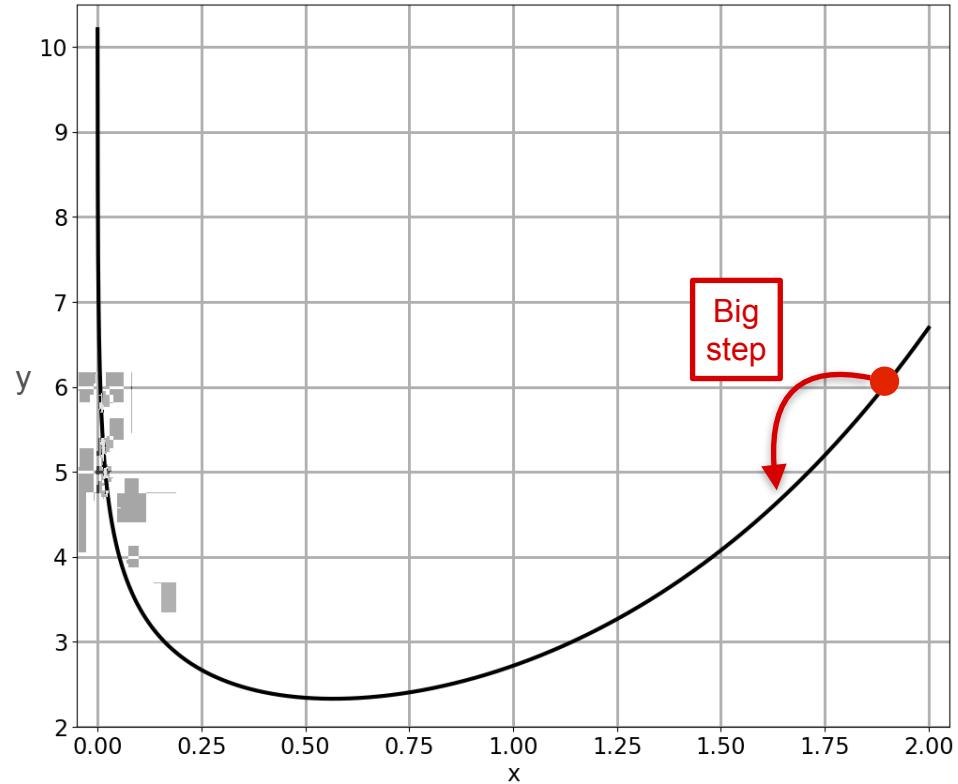


# Method 2: Be Clever

Try something  
smarter...



$$x_1 = x_0 - \alpha f'(x_0)$$

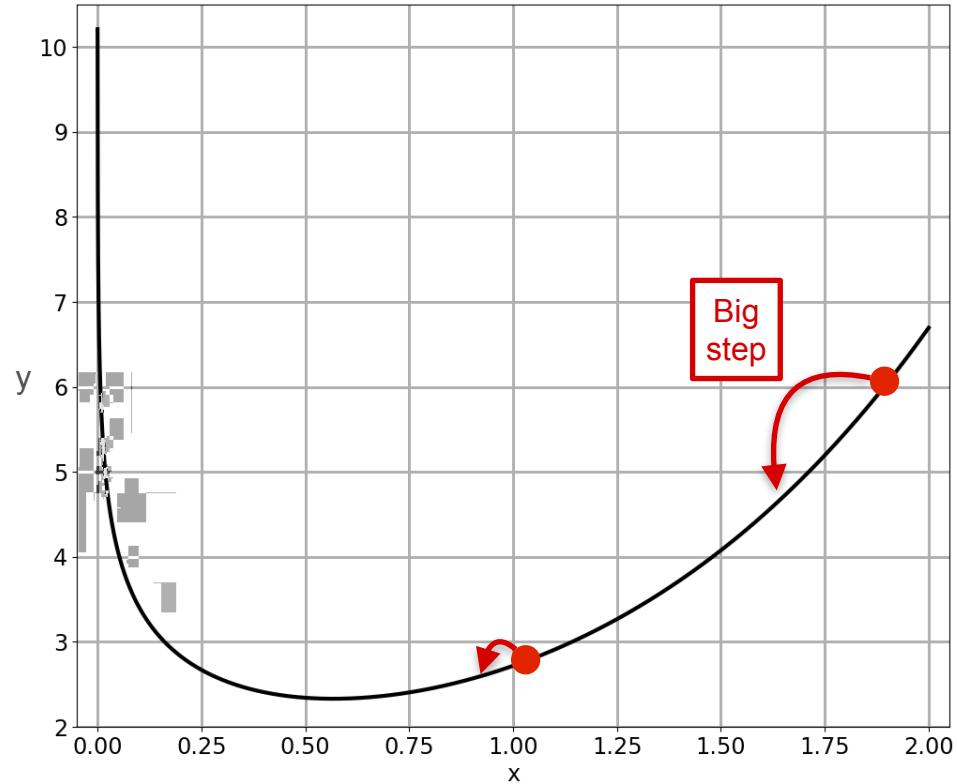


# Method 2: Be Clever

Try something  
smarter...



$$x_1 = x_0 - \alpha f'(x_0)$$

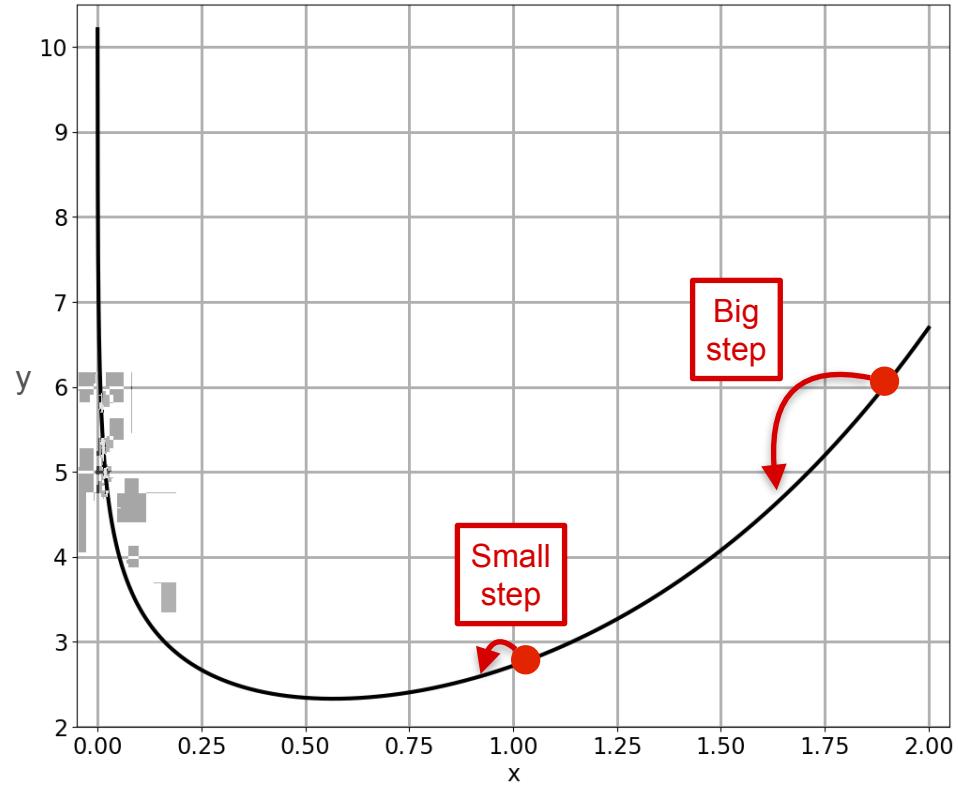


# Method 2: Be Clever

Try something  
smarter...



$$x_1 = x_0 - \alpha f'(x_0)$$

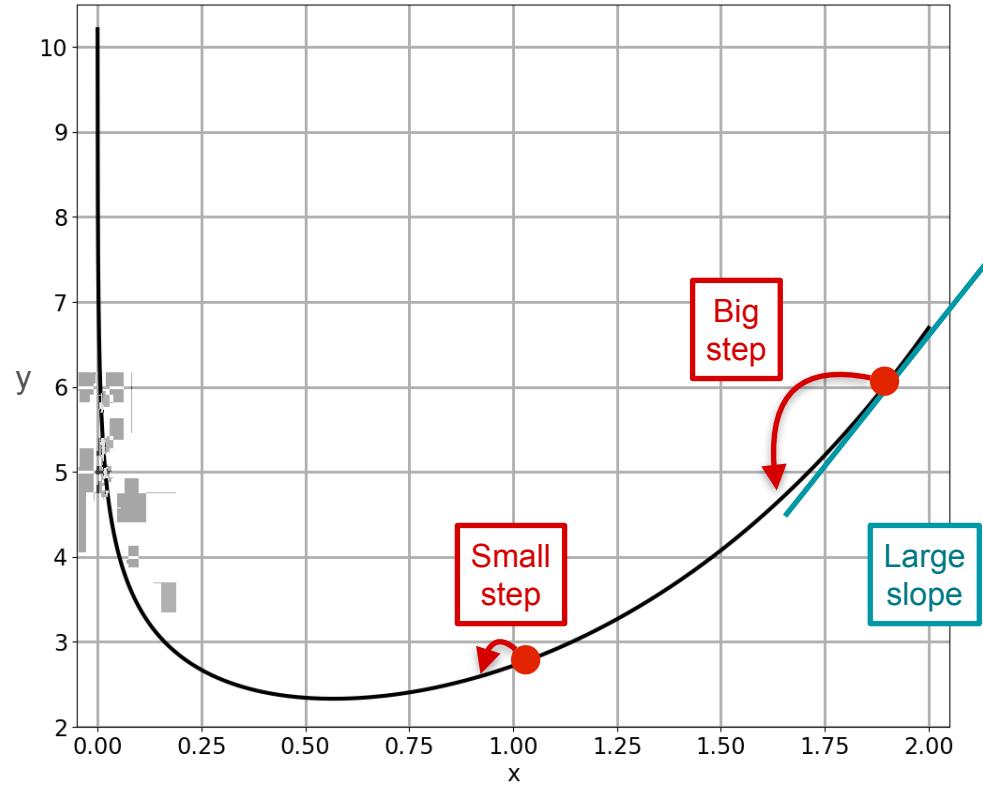


# Method 2: Be Clever

Try something  
smarter...



$$x_1 = x_0 - \alpha f'(x_0)$$

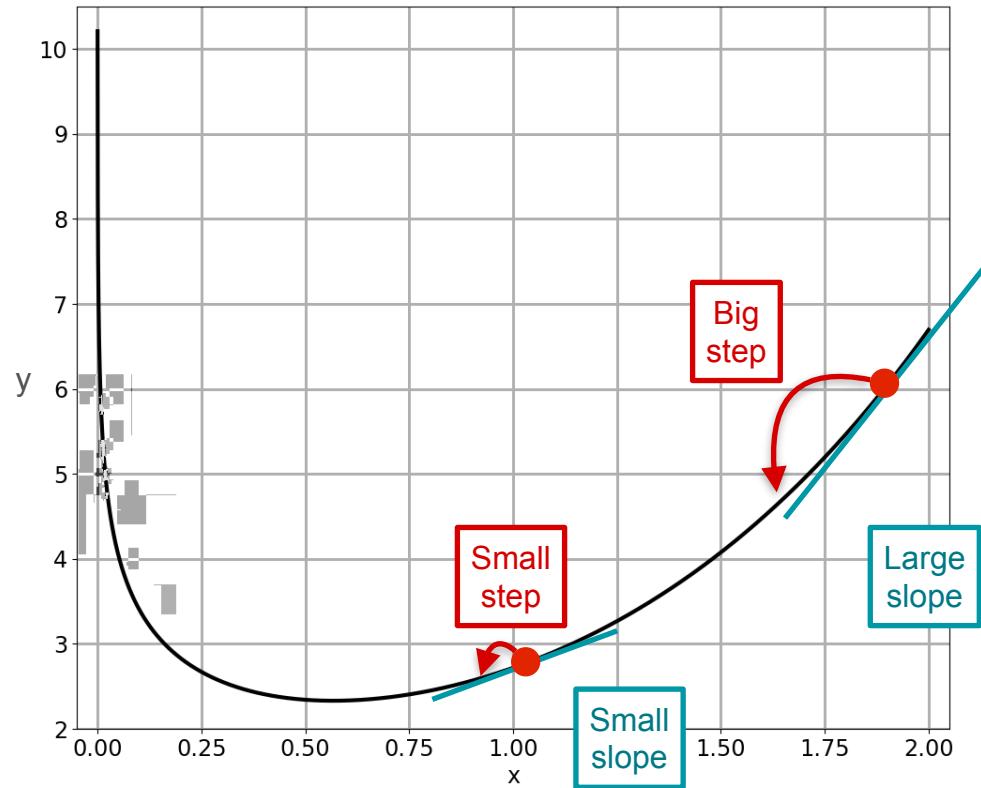


# Method 2: Be Clever

Try something  
smarter...



$$x_1 = x_0 - \alpha f'(x_0)$$



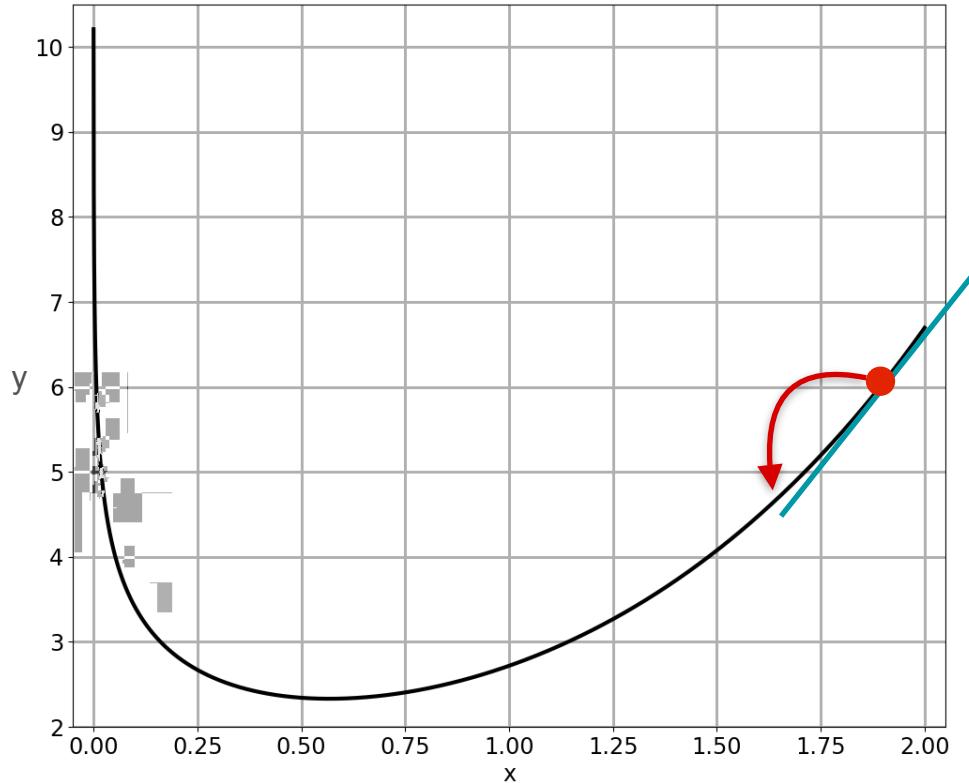
# Method 2: Be Clever

Try something  
smarter...



$$x_1 = x_0 - \alpha f'(x_0)$$

## Gradient descent



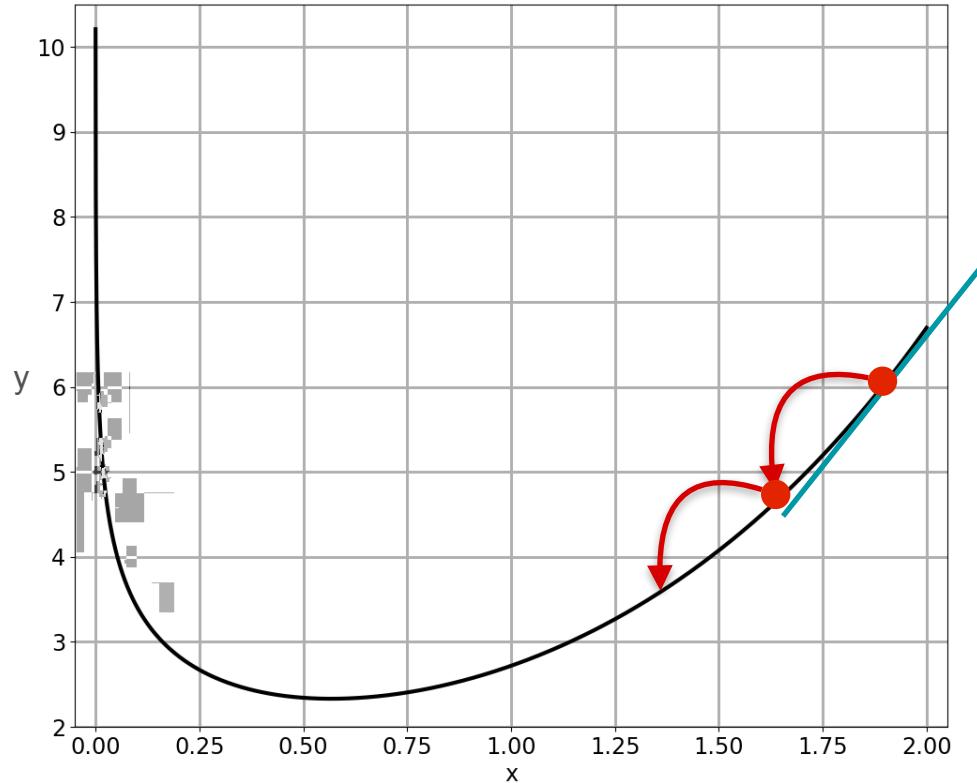
# Method 2: Be Clever

Try something  
smarter...



$$x_2 = x_1 - \alpha f'(x_1)$$

## Gradient descent



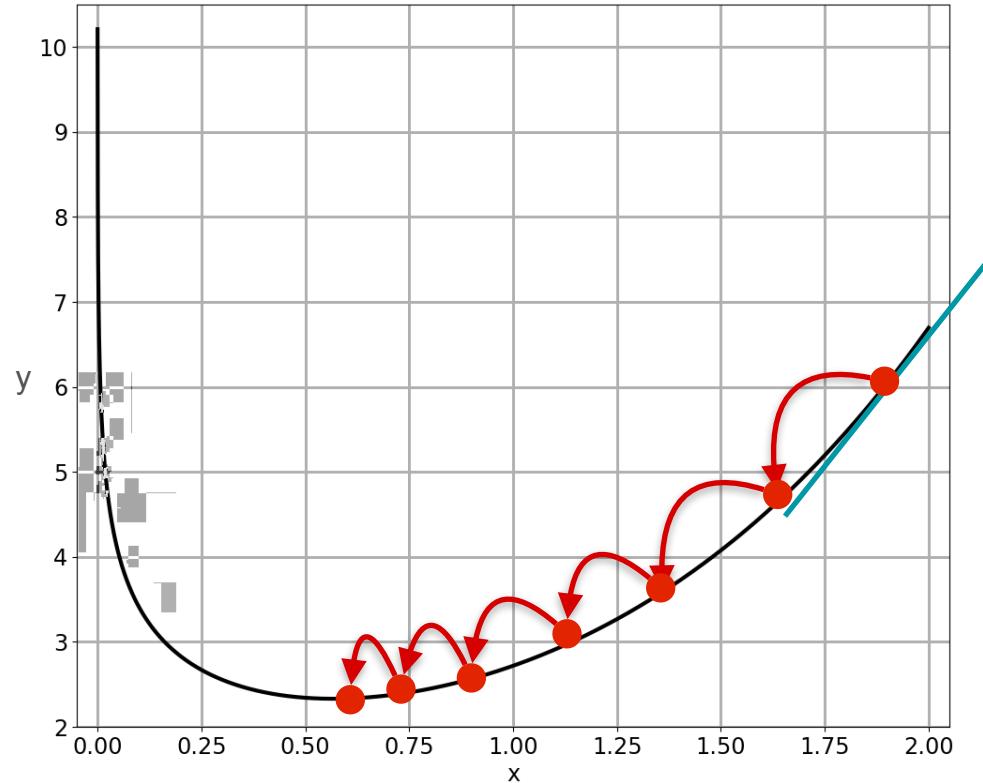
# Method 2: Be Clever

Try something  
smarter...



$$x_{20} = x_{19} - \alpha f'(x_{19})$$

## Gradient descent



# Gradient Descent

Function:  $f(x)$

Goal: find minimum of  $f(x)$

Step 1:

Define a learning rate  $\alpha$

Choose a starting point  $x_0$

Step 2:

Update:  $x_k = x_{k-1} - \alpha f'(x_{k-1})$

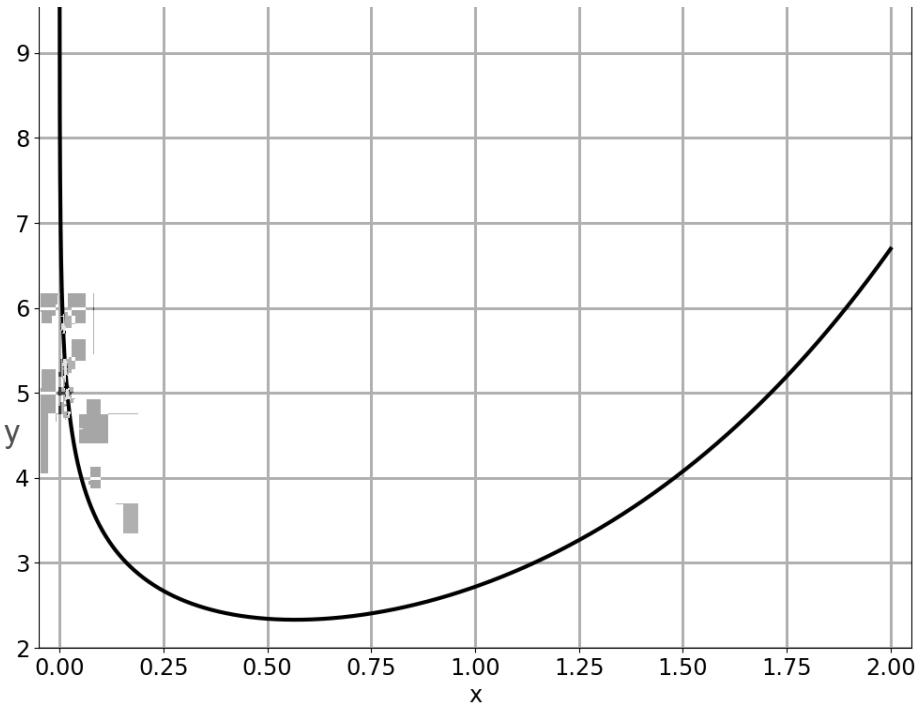
Step 3:

Repeat Step 2 until you are close enough to  
the true minimum  $x^*$

# Gradient Descent

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

$$f'(x) = e^x - \frac{1}{x}$$



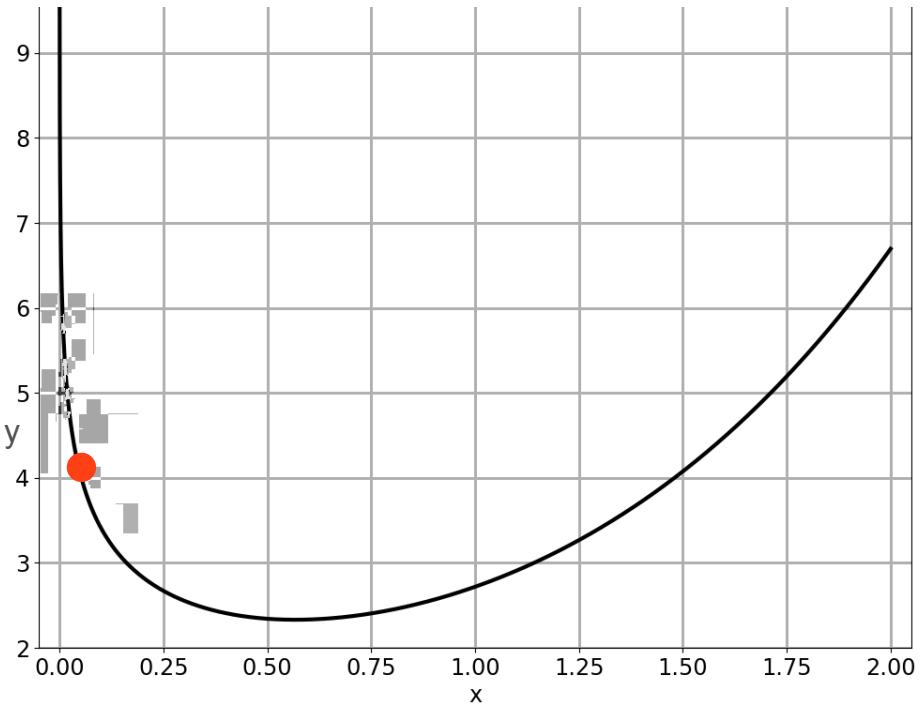
# Gradient Descent

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

$$f'(x) = e^x - \frac{1}{x}$$

Start:  $x = 0.05$

Rate:  $\alpha = 0.005$



# Gradient Descent

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

$$f'(x) = e^x - \frac{1}{x}$$

Start:  $x = 0.05$

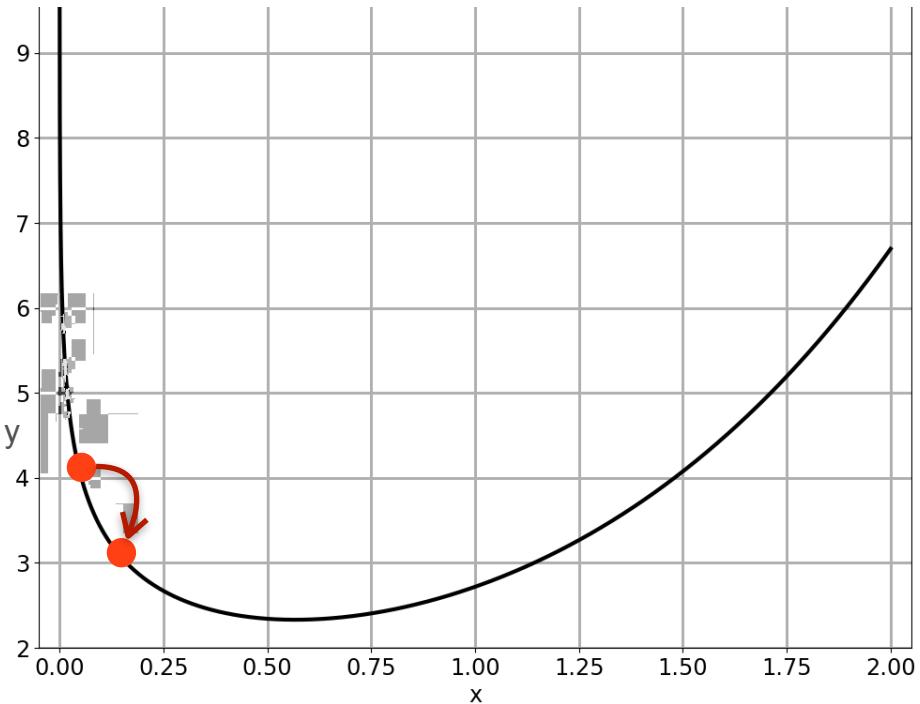
Rate:  $\alpha = 0.005$

Find:

$$f'(0.05) = -18.9$$

Move by  $-0.005f'(0.05)$

$$x \mapsto 0.1447$$



# Gradient Descent

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

$$f'(x) = e^x - \frac{1}{x}$$

Start:  $x = 0.05$

Rate:  $\alpha = 0.005$

Find:

$$f'(0.05) = -18.9$$

Move by  $-0.005f'(0.05)$

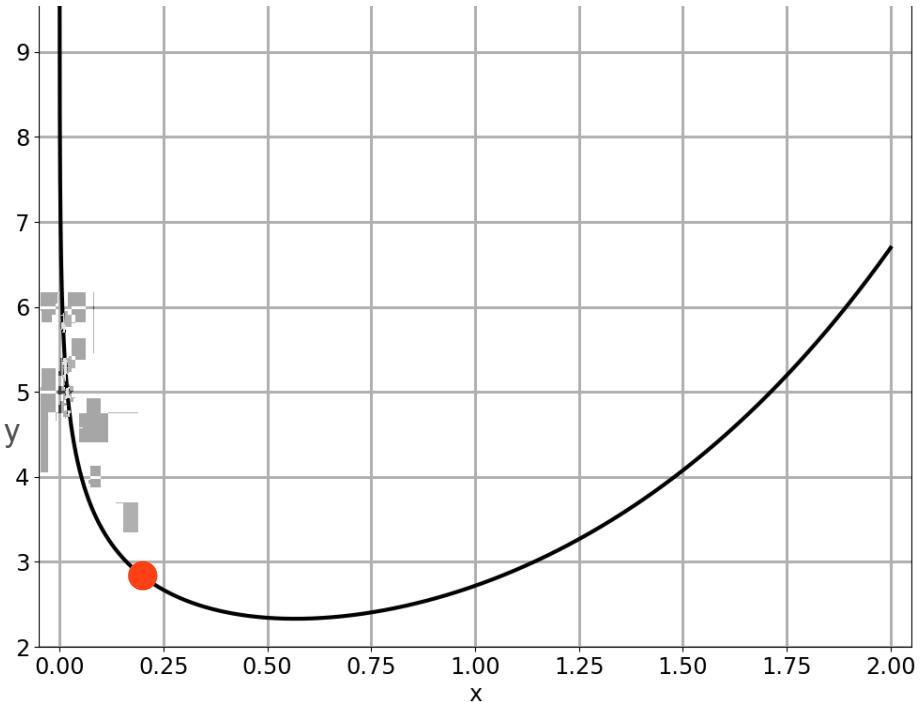
$$x \mapsto 0.1447$$

Find:

$$f'(0.1447) = -5.7552$$

Move by  $-0.005f'(0.05)$

$$x \mapsto 0.1735$$



# Gradient Descent

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

$$f'(x) = e^x - \frac{1}{x}$$

Start:  $x = 0.05$

Rate:  $\alpha = 0.005$

Find:

$$f'(0.05) = -18.9$$

Move by  $-0.005f'(0.05)$

$$x \mapsto 0.1447$$

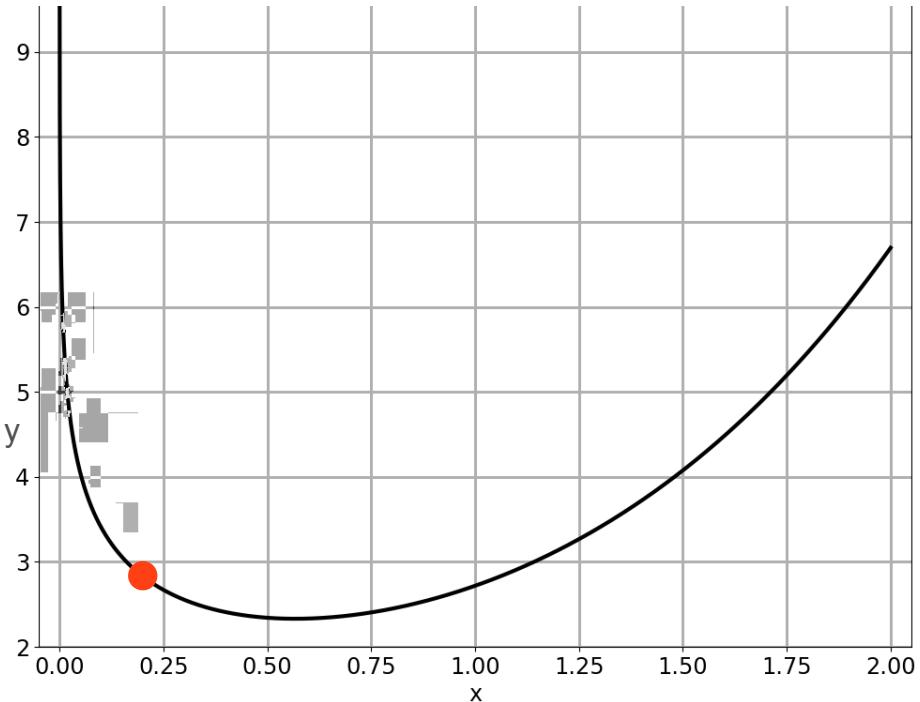
Find:

$$f'(0.1447) = -5.7552$$

Move by  $-0.005f'(0.05)$

$$x \mapsto 0.1735$$

**Repeat!**



# Gradient Descent

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

$$f'(x) = e^x - \frac{1}{x}$$

Start:  $x = 0.05$

Rate:  $\alpha = 0.005$

Find:

$$f'(0.05) = -18.9$$

Move by  $-0.005f'(0.05)$

$$x \mapsto 0.1447$$

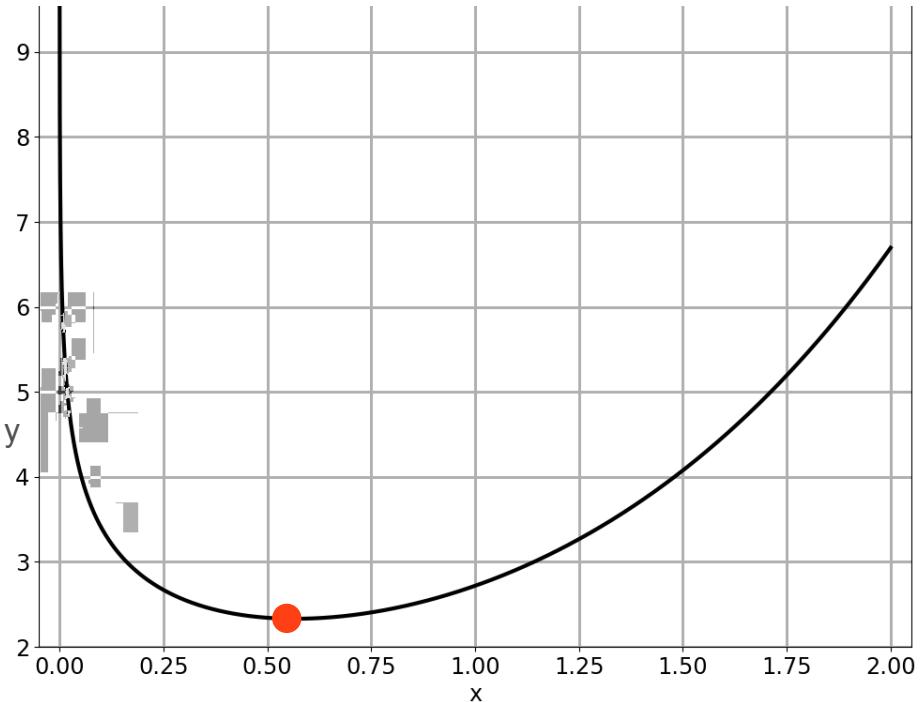
Find:

$$f'(0.1447) = -5.7552$$

Move by  $-0.005f'(0.05)$

$$x \mapsto 0.1735$$

**Repeat!**





DeepLearning.AI

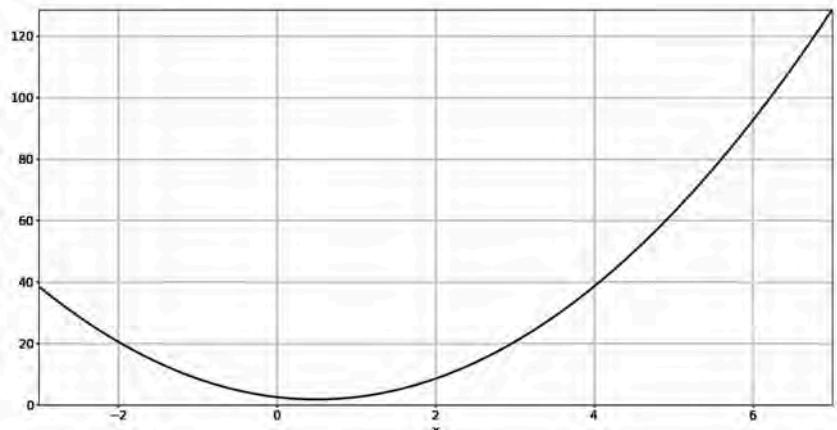
# Gradients and Gradient Descent

---

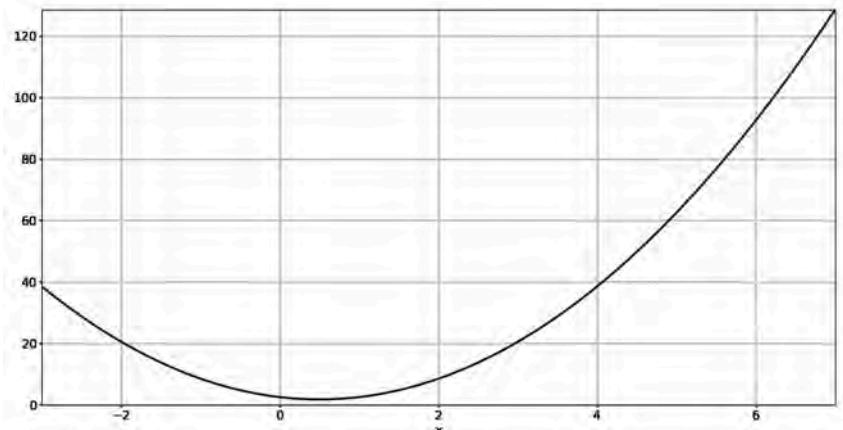
**Optimization using Gradient  
Descent in one variable -  
Part 3**

# What Is a Good Learning Rate?

Too large

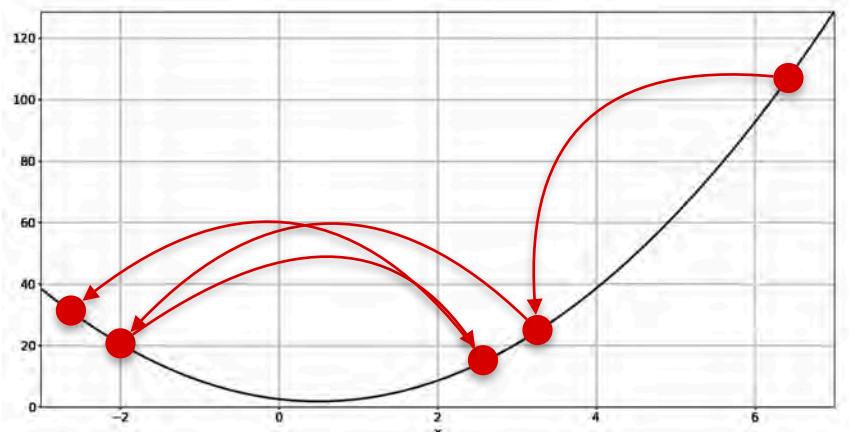


Too small

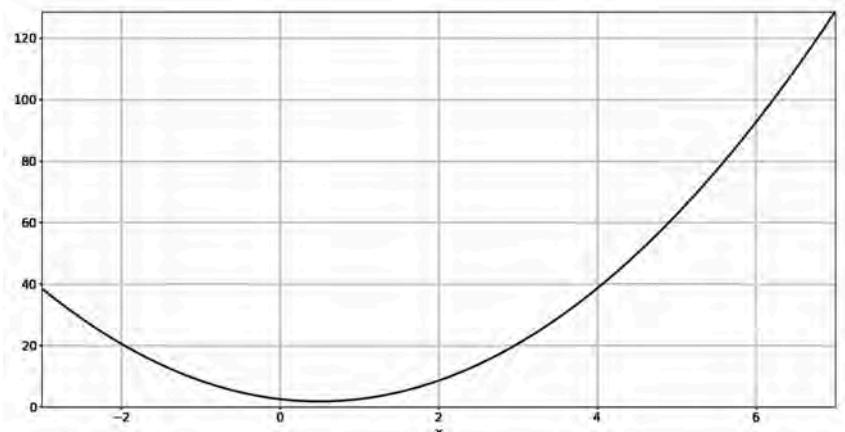


# What Is a Good Learning Rate?

Too large

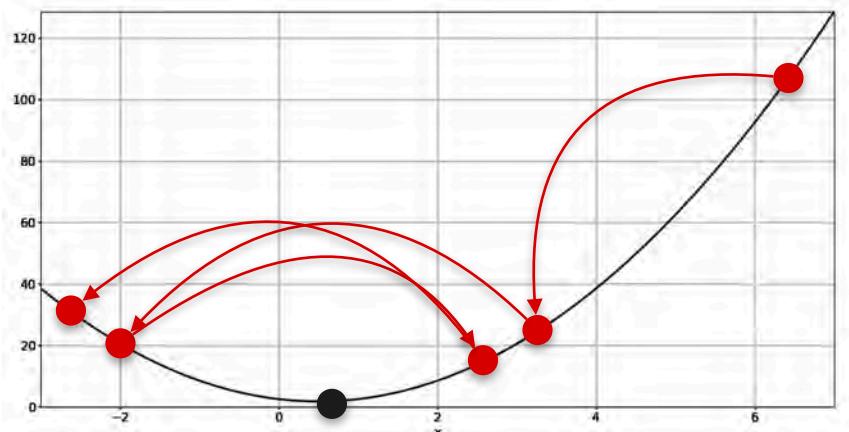


Too small

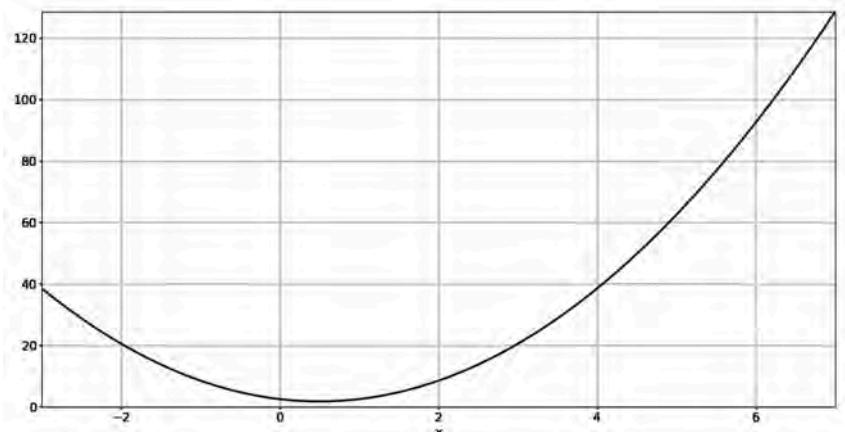


# What Is a Good Learning Rate?

Too large

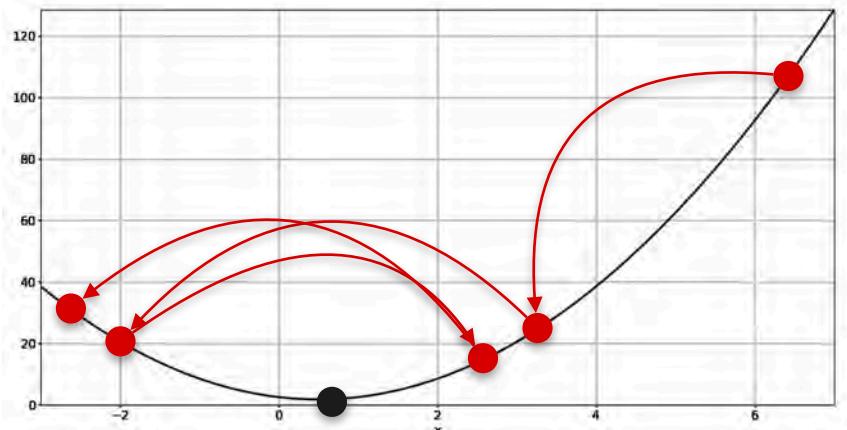


Too small

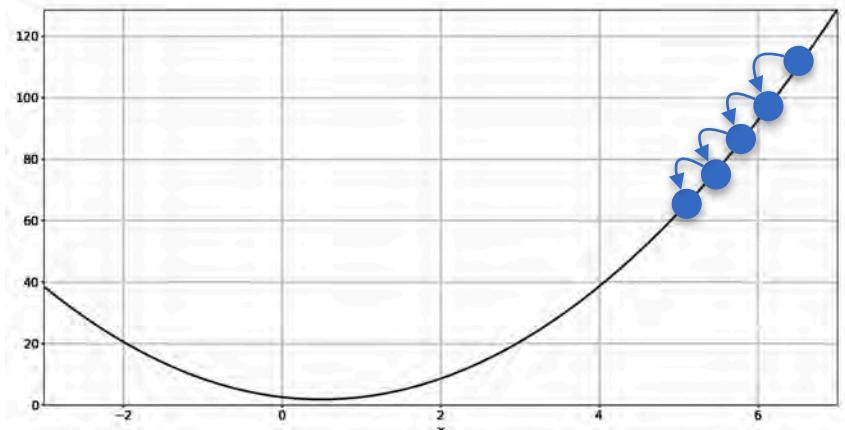


# What Is a Good Learning Rate?

Too large

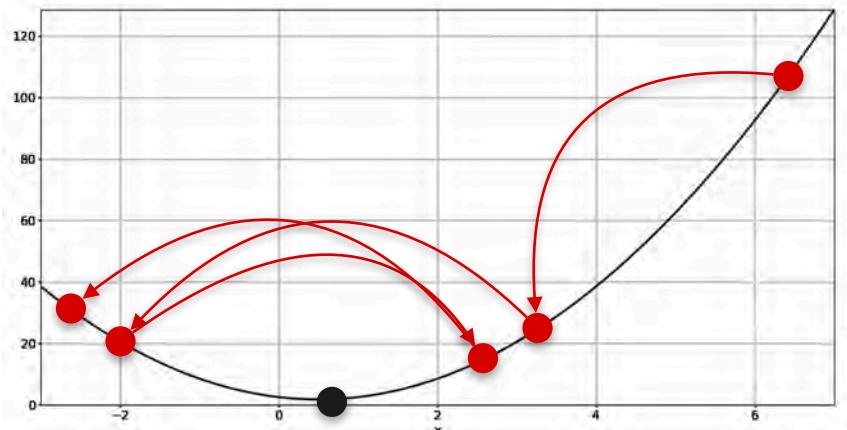


Too small

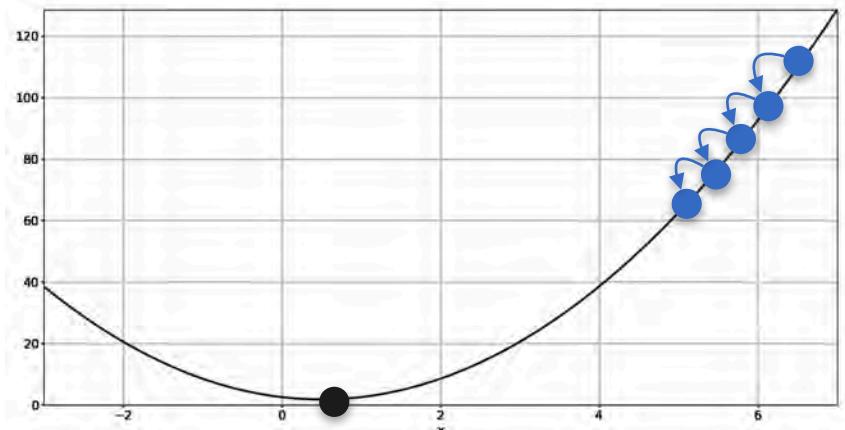


# What Is a Good Learning Rate?

Too large

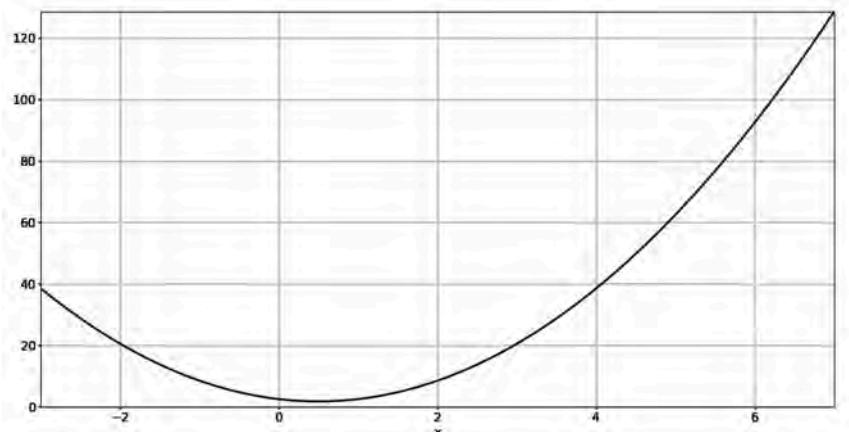


Too small



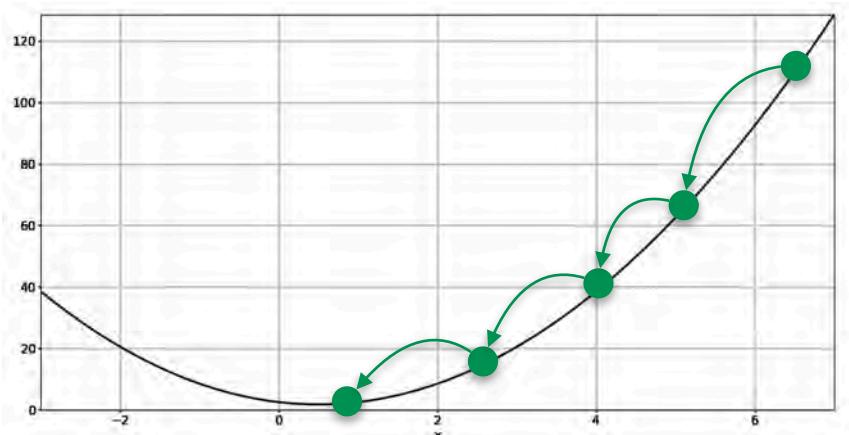
# What Is a Good Learning Rate?

Just right



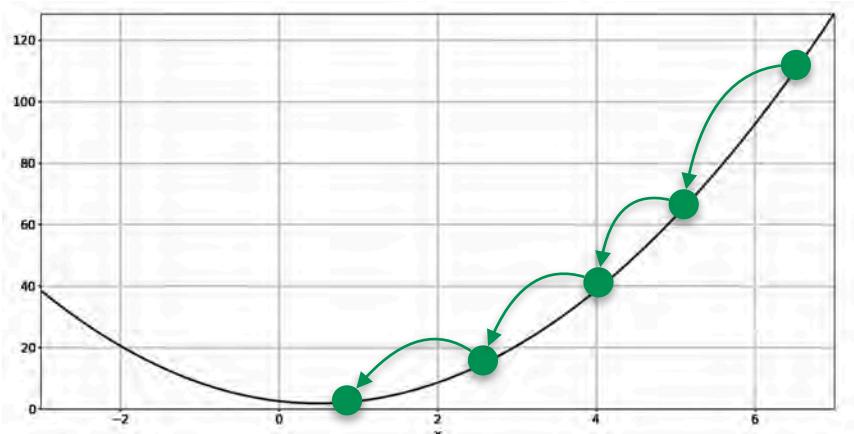
# What Is a Good Learning Rate?

Just right



# What Is a Good Learning Rate?

Just right

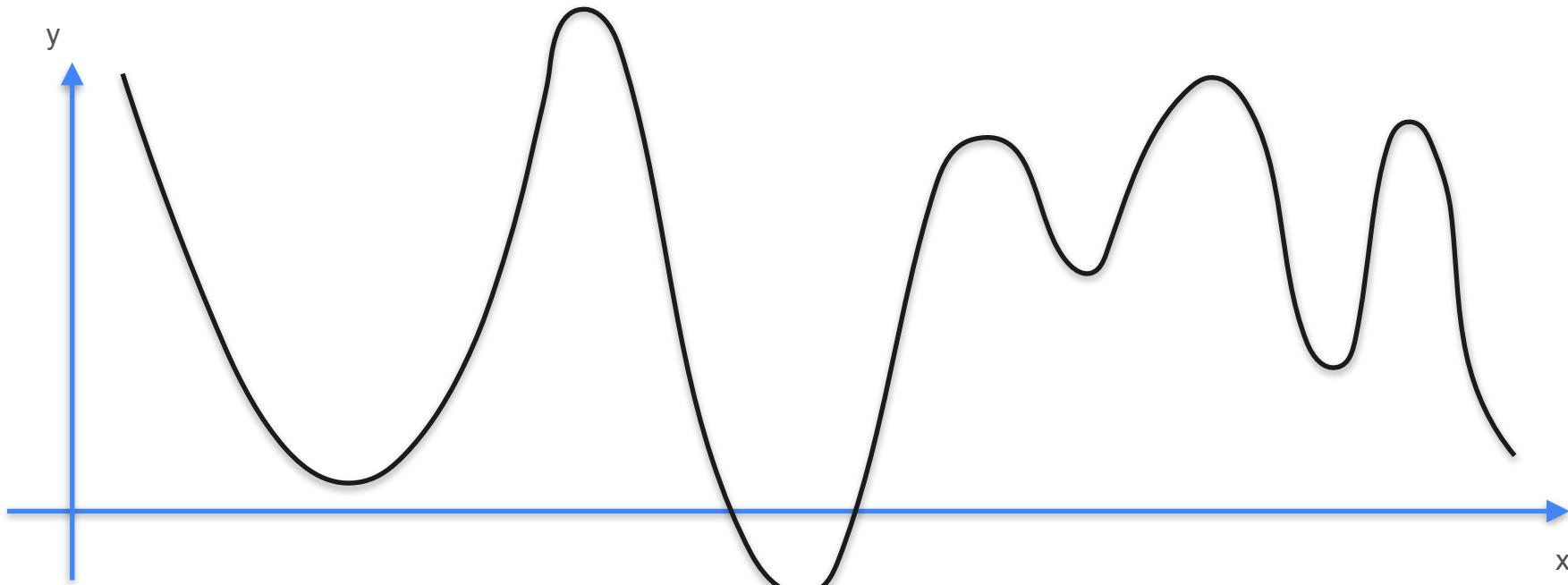


Unfortunately, there is no rule to give the best learning rate  $\alpha$

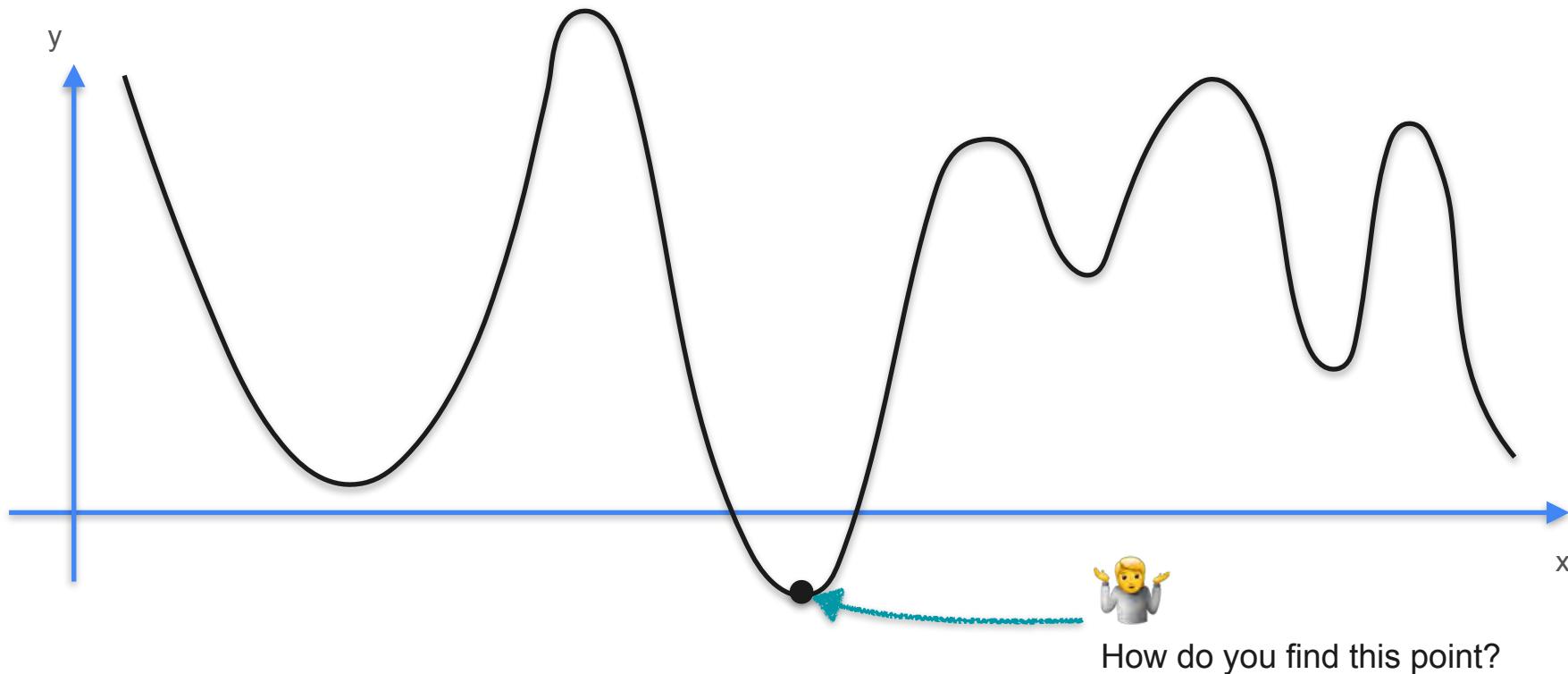


# Drawbacks of Gradient Descent

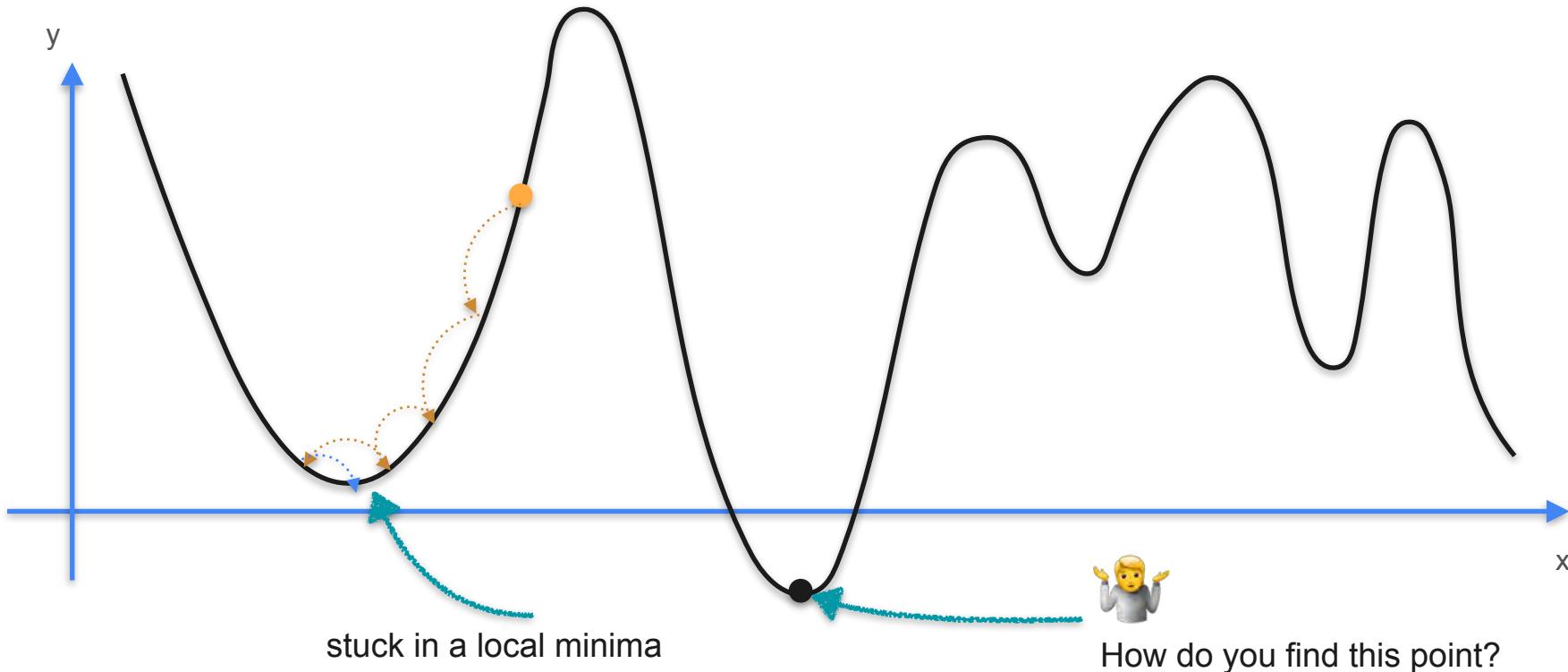
# Drawbacks of Gradient Descent



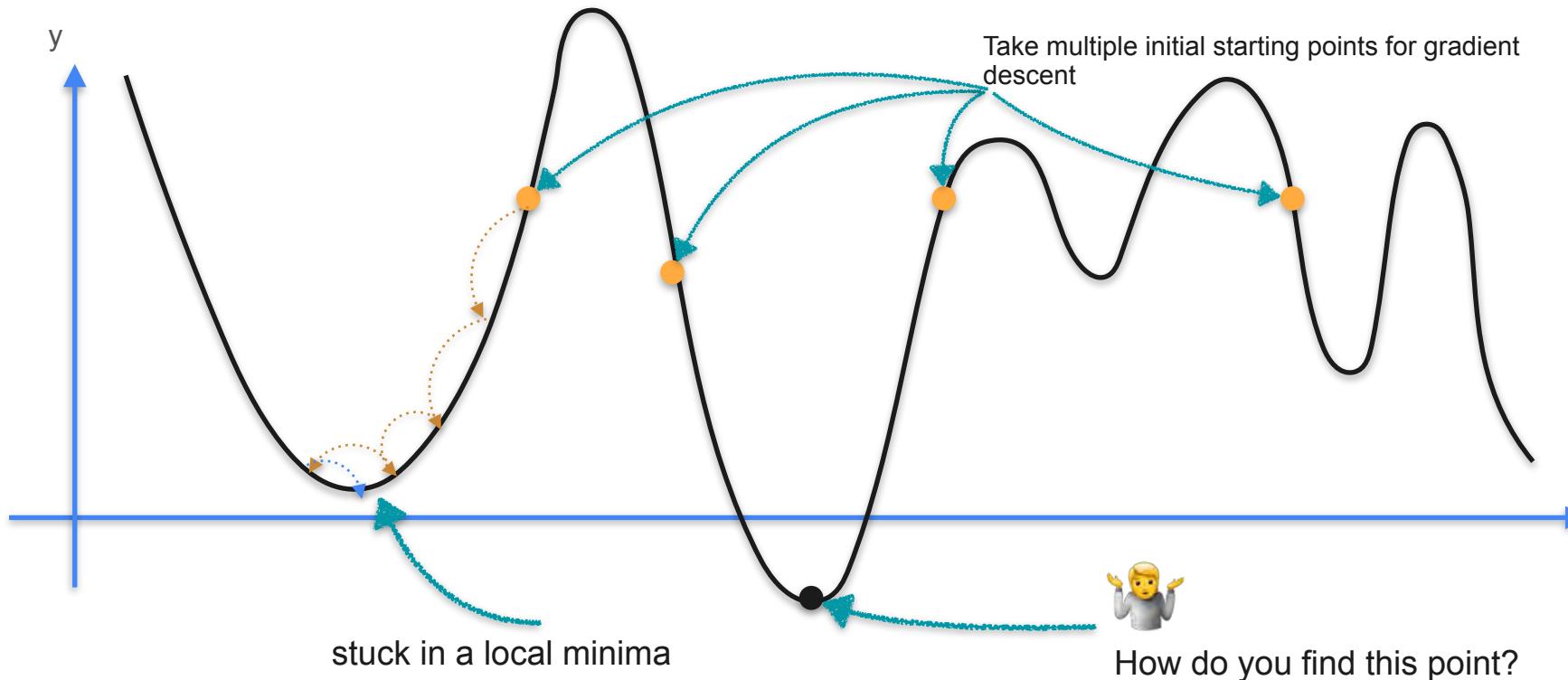
# Drawbacks of Gradient Descent



# Drawbacks of Gradient Descent



# Drawbacks of Gradient Descent





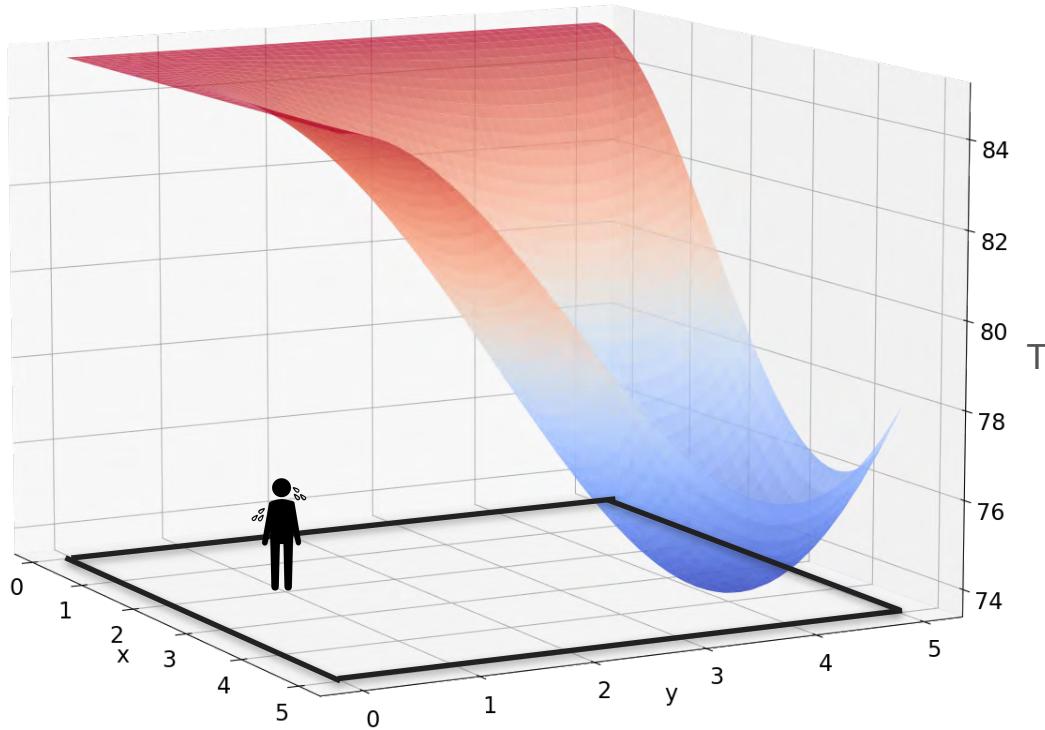
DeepLearning.AI

# Gradients and Gradient Descent

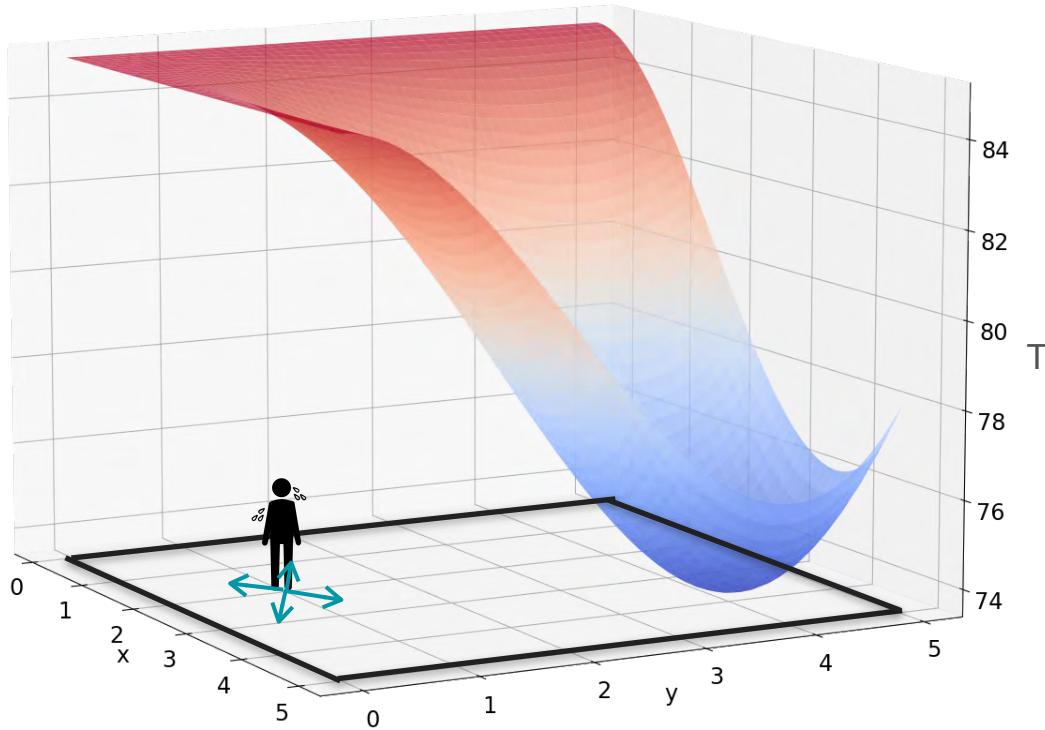
---

**Optimization using Gradient  
Descent in two variables -  
Part 1**

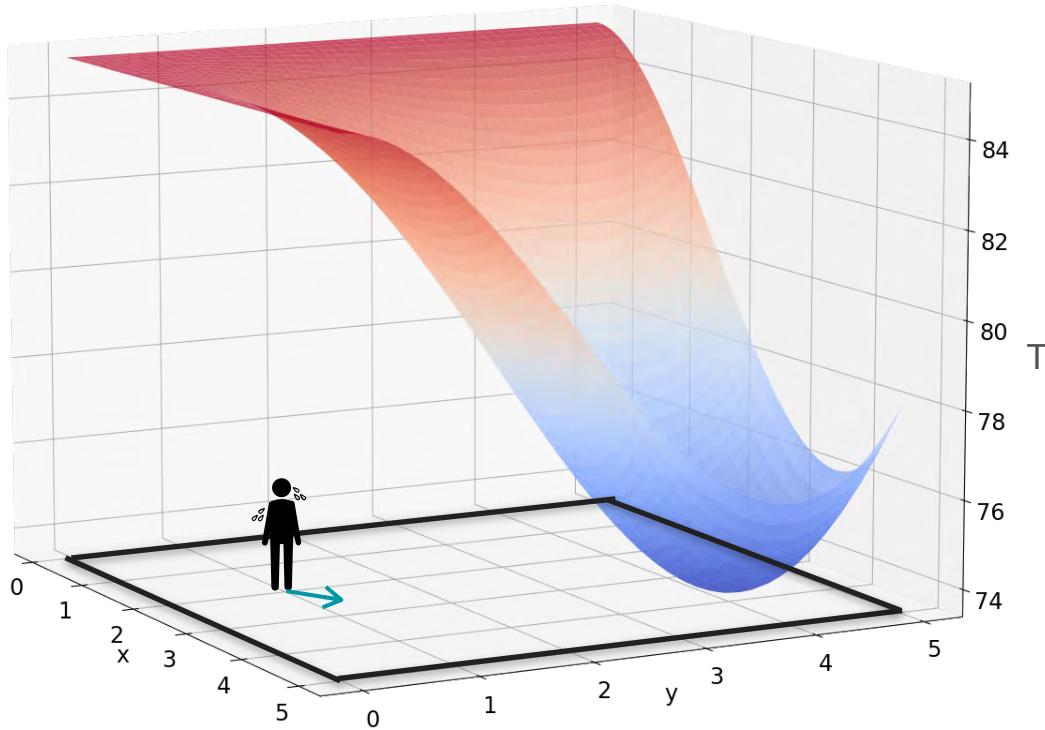
# Gradient Descent With Heat Example



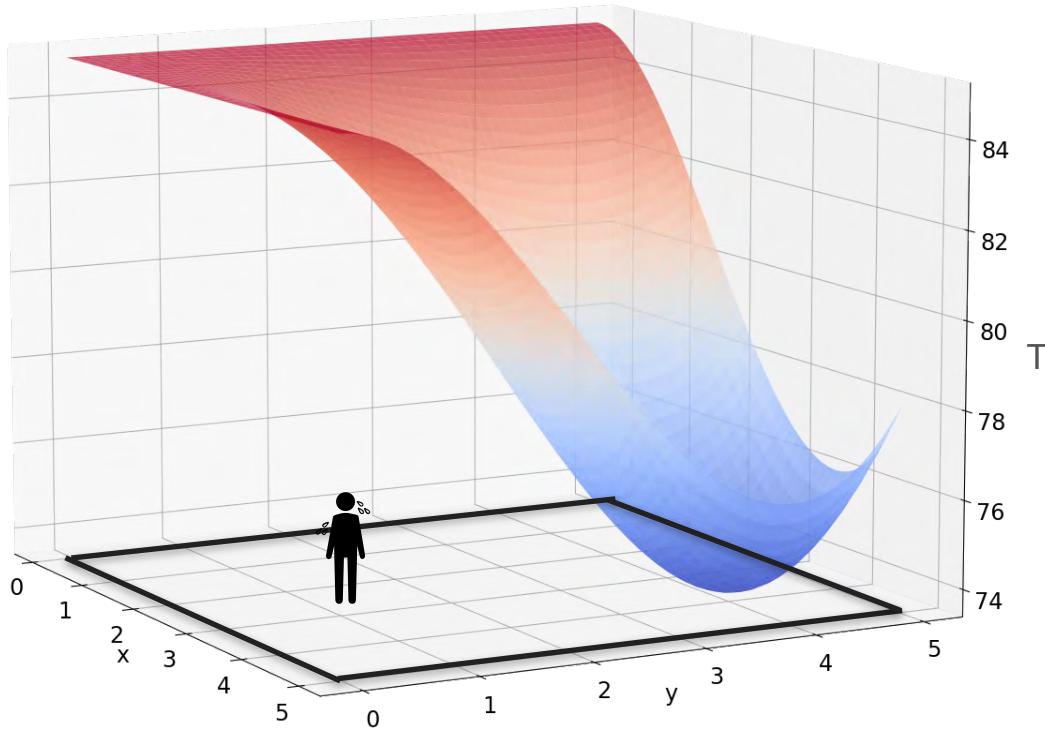
# Gradient Descent With Heat Example



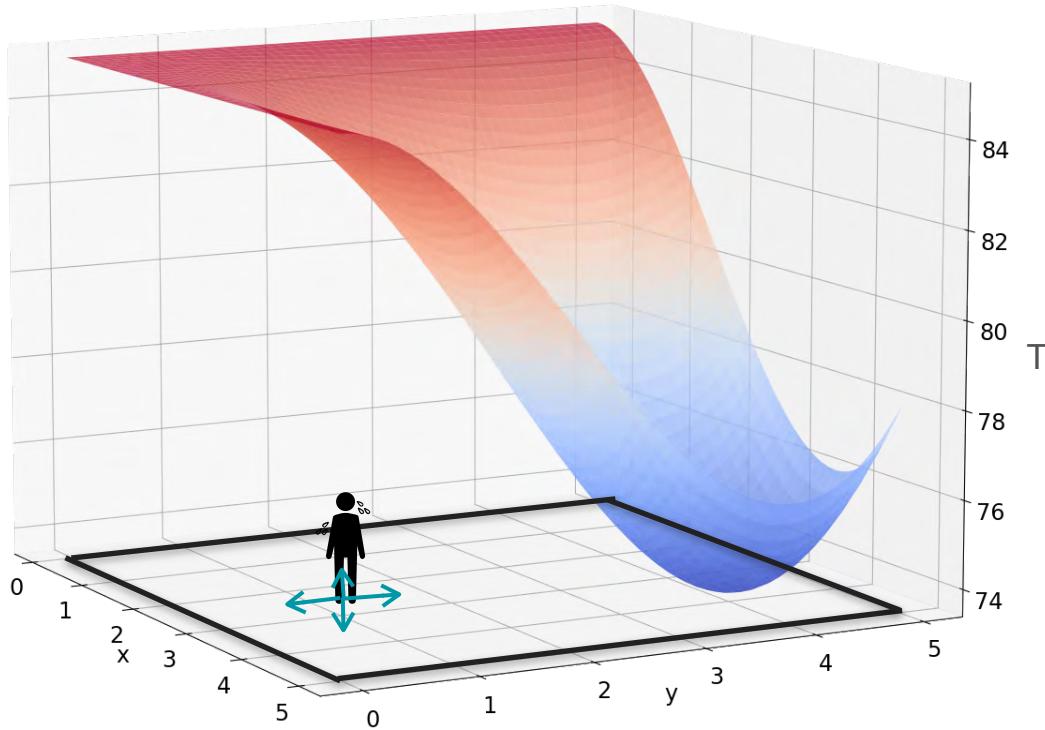
# Gradient Descent With Heat Example



# Gradient Descent With Heat Example

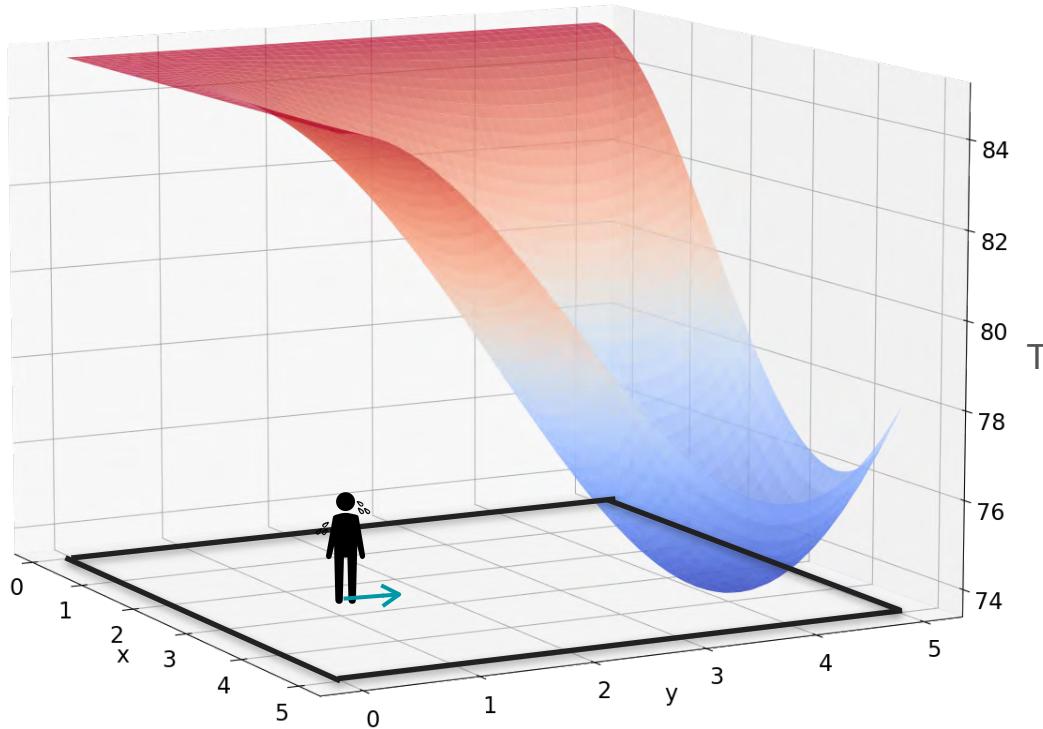


# Gradient Descent With Heat Example

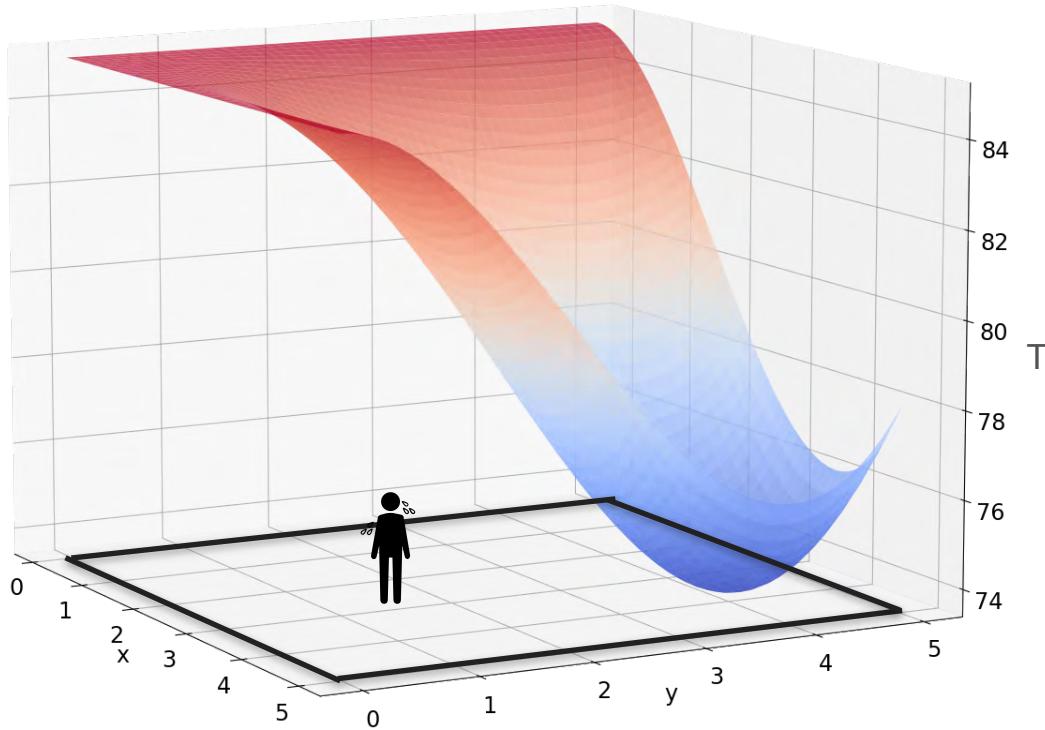


# Gradient Descent With Heat Example

Repeat!

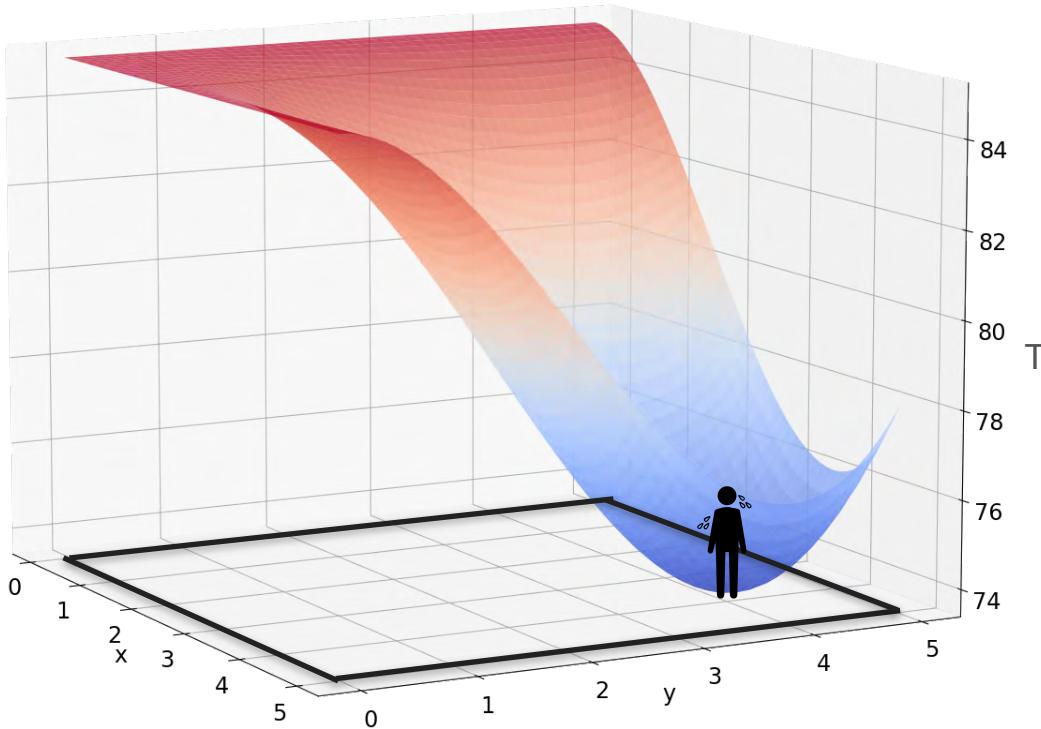


# Gradient Descent With Heat Example



# Gradient Descent With Heat Example

Repeat!





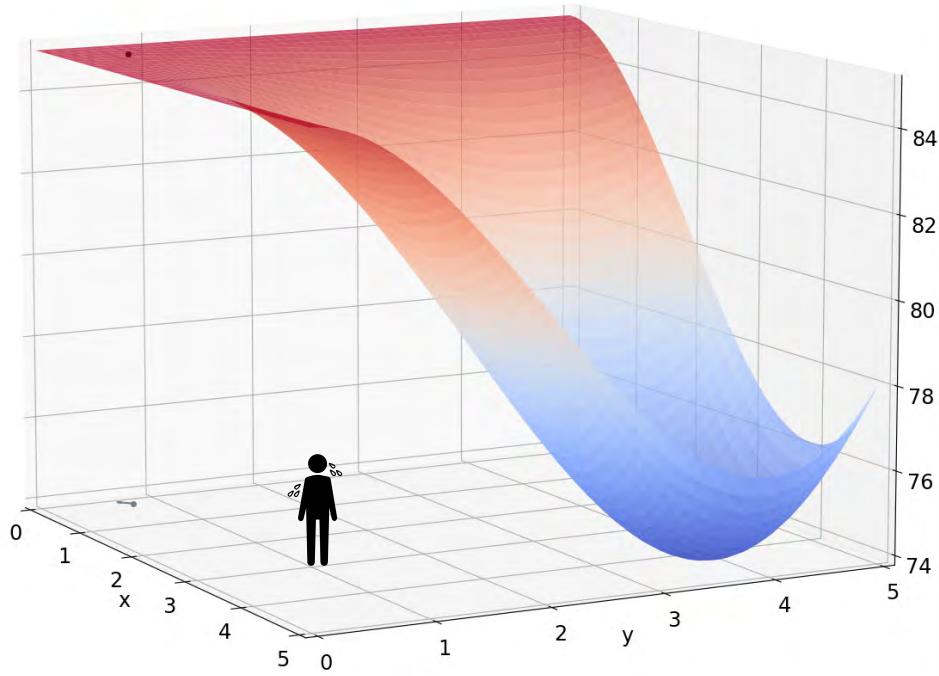
DeepLearning.AI

# Gradients and Gradient Descent

---

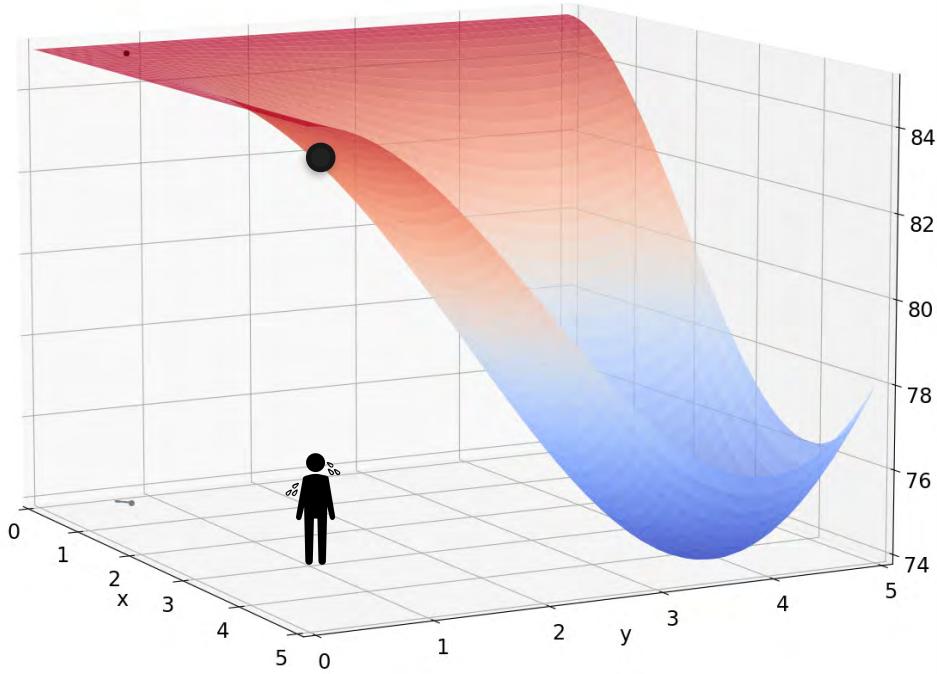
**Optimization using Gradient  
Descent in two variables -  
Part 2**

# Idea for Gradient Descent



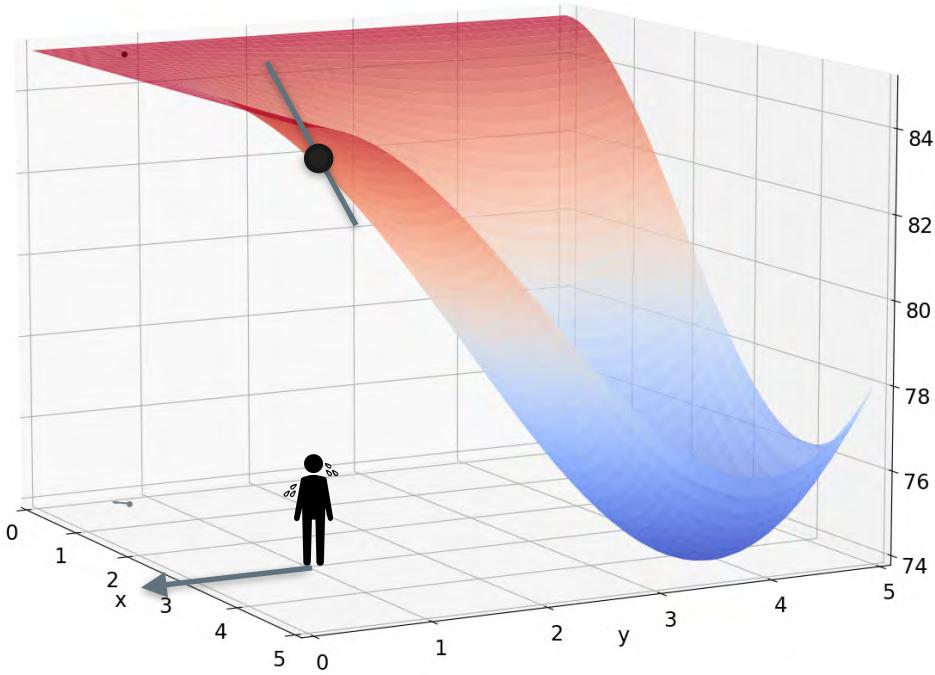
# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$



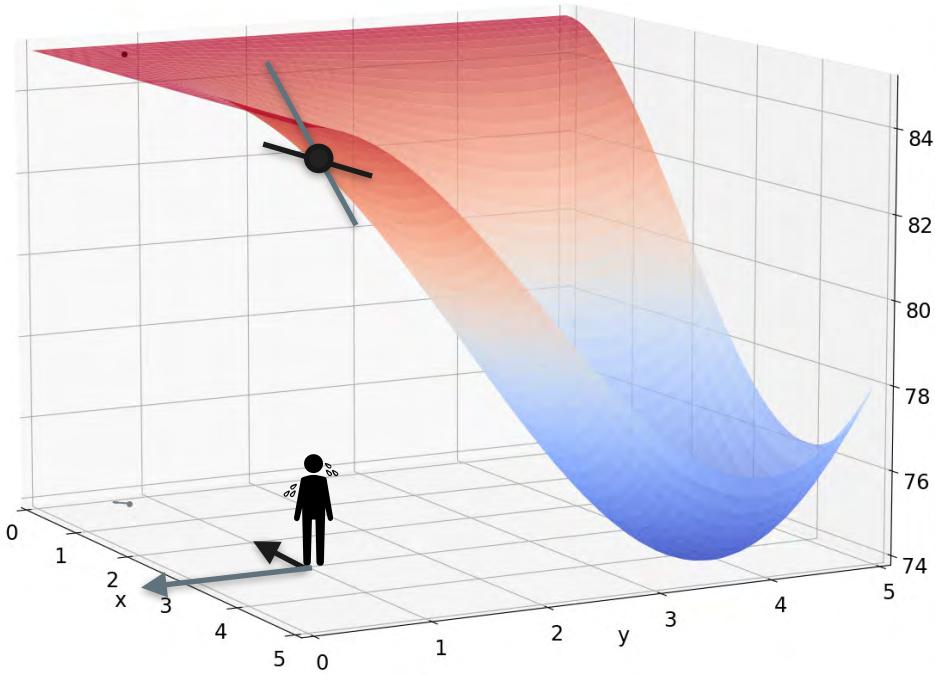
# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$



# Idea for Gradient Descent

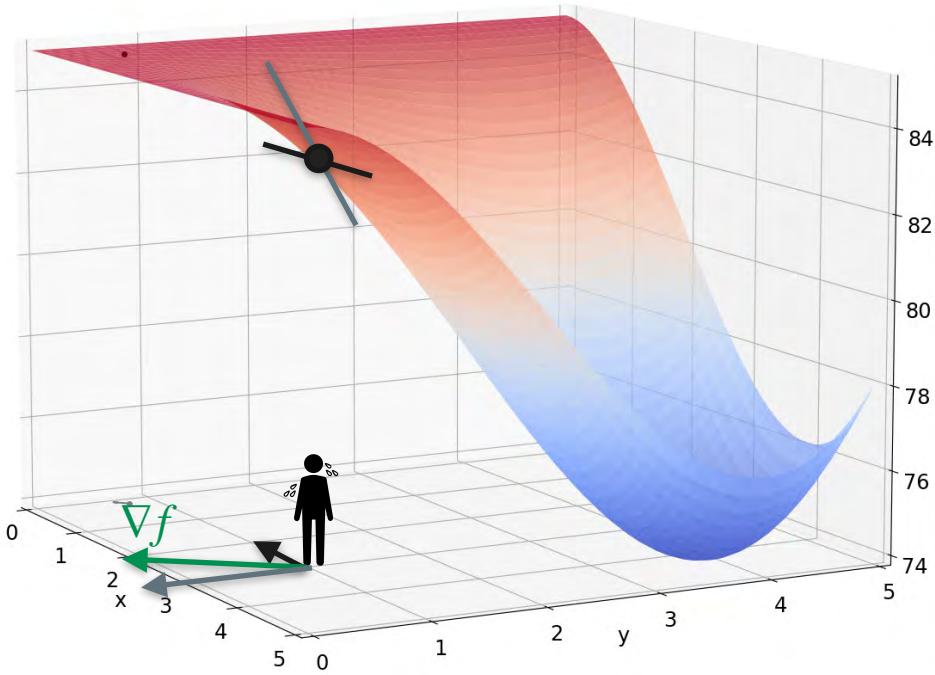
Initial position:  $(x_0, y_0)$



# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$

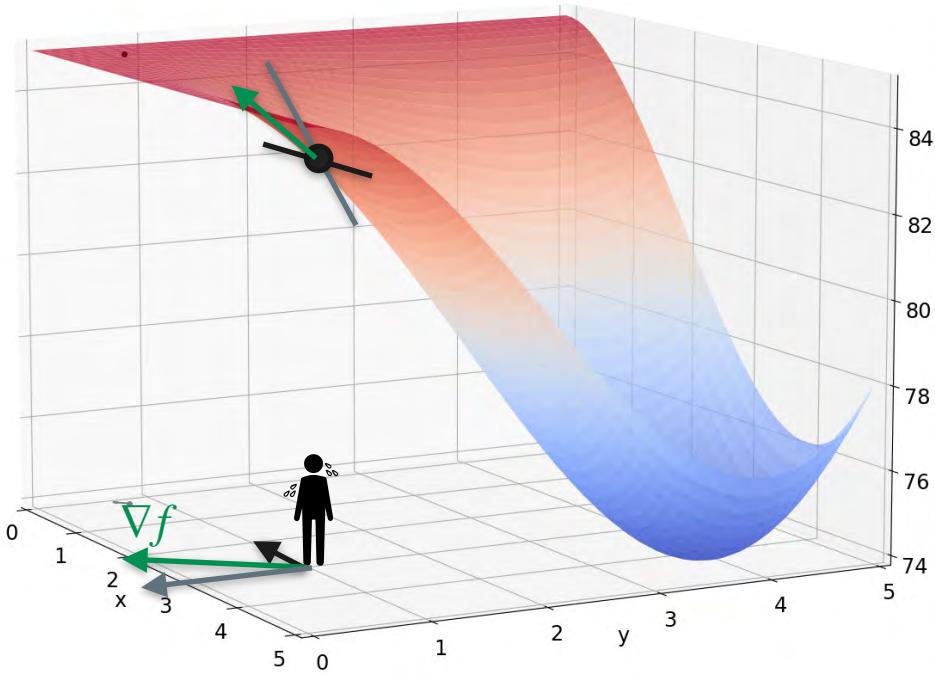
Direction of greatest ascent:  $\nabla f$



# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$

Direction of greatest ascent:  $\nabla f$

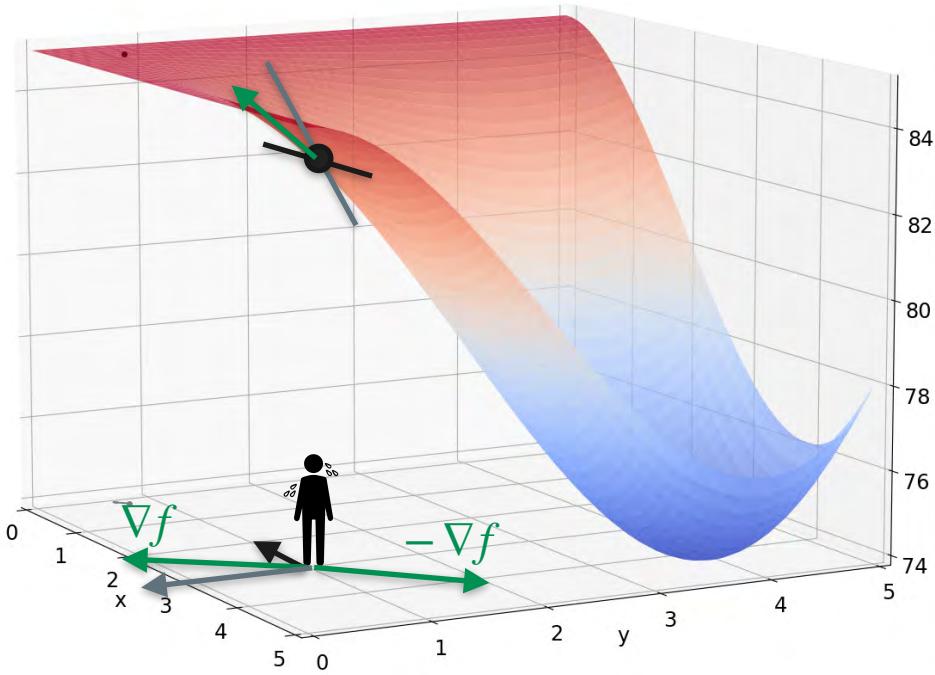


# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$

Direction of greatest ascent:  $\nabla f$

Direction of greatest descent:  $-\nabla f$

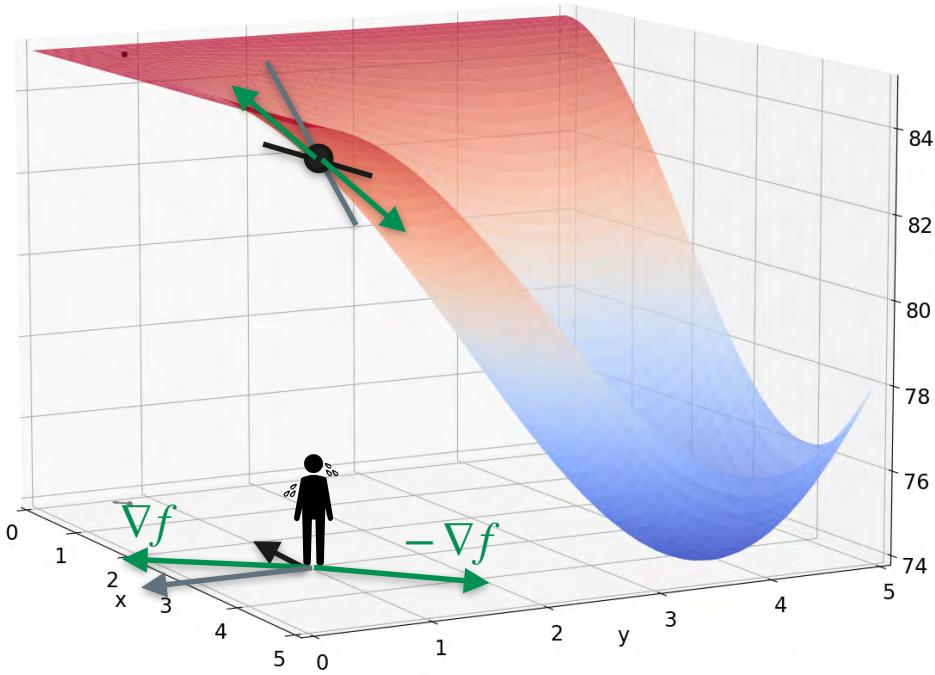


# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$

Direction of greatest ascent:  $\nabla f$

Direction of greatest descent:  $-\nabla f$

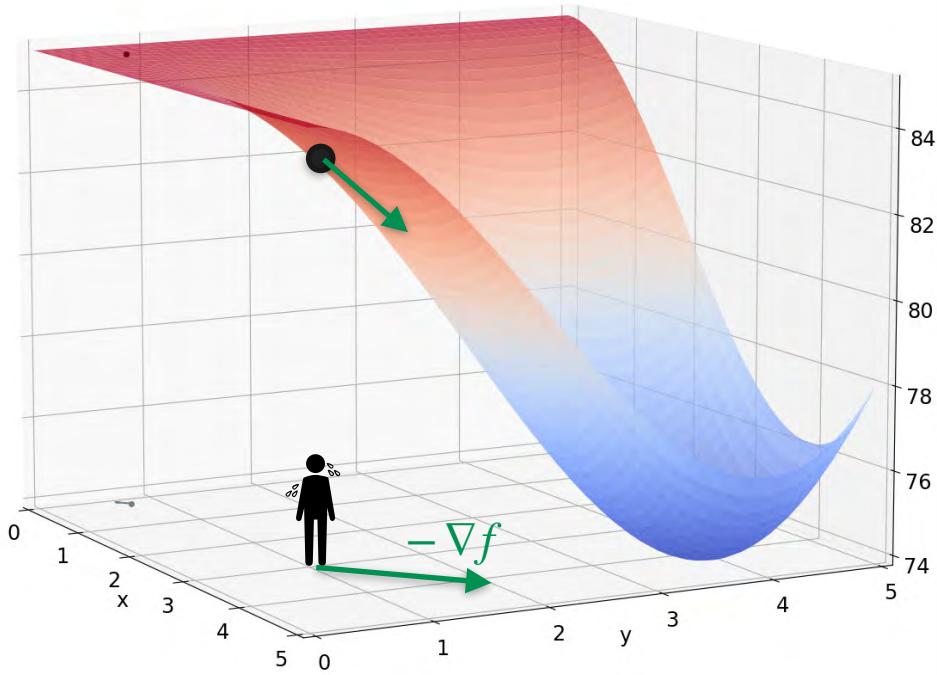


# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$

Direction of greatest ascent:  $\nabla f$

Direction of greatest descent:  $-\nabla f$



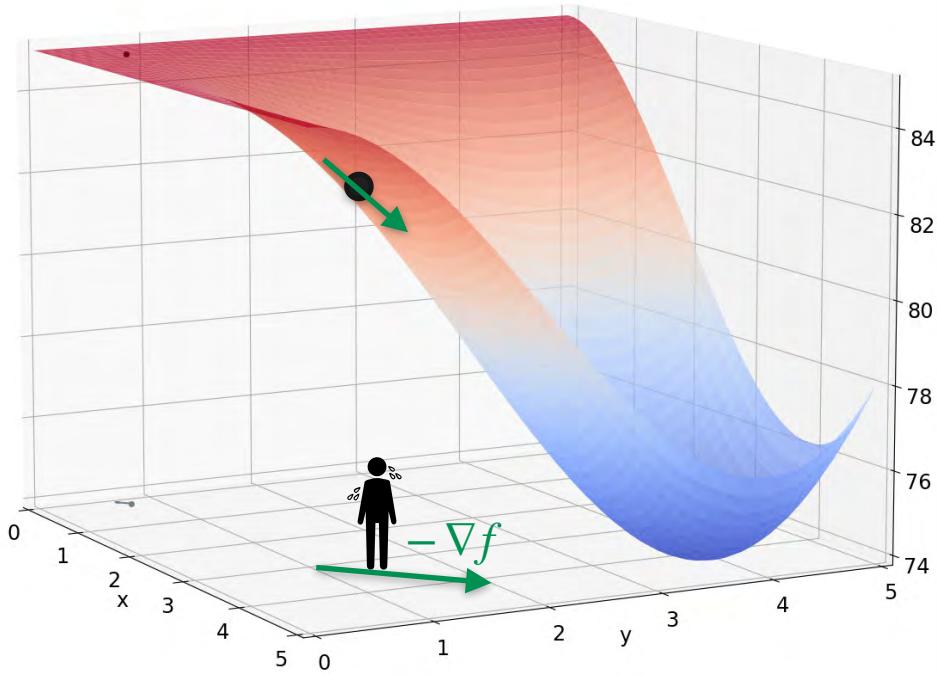
# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$

Direction of greatest ascent:  $\nabla f$

Direction of greatest descent:  $-\nabla f$

Updated position:  $(x_0, y_0) - \alpha \nabla f$



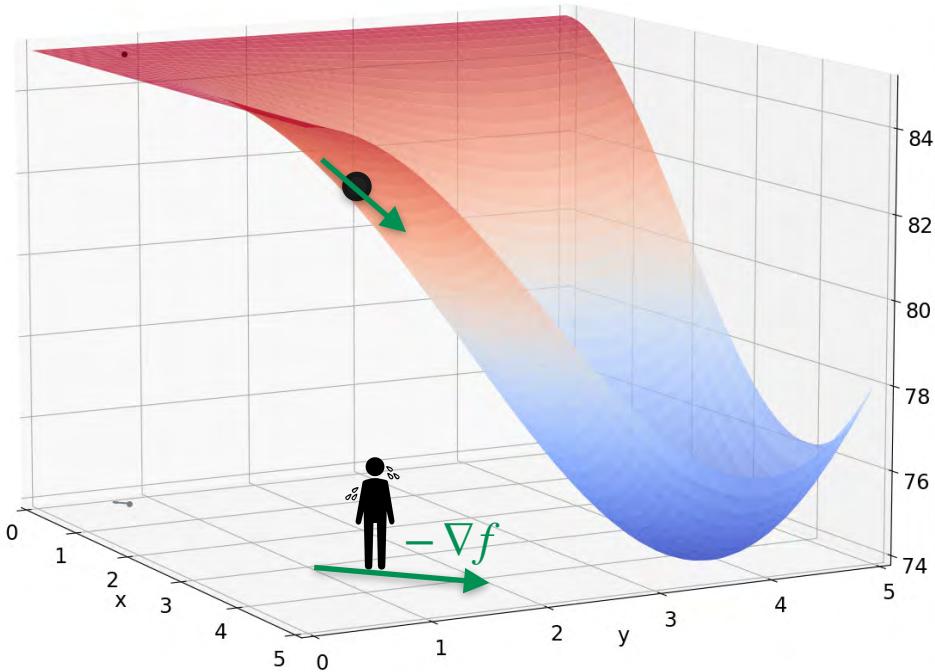
# Idea for Gradient Descent

Initial position:  $(x_0, y_0)$

Direction of greatest ascent:  $\nabla f$

Direction of greatest descent:  $-\nabla f$

Updated position:  $\underbrace{(x_0, y_0) - \alpha \nabla f}_{(x_1, y_1)}$



# Idea for Gradient Descent

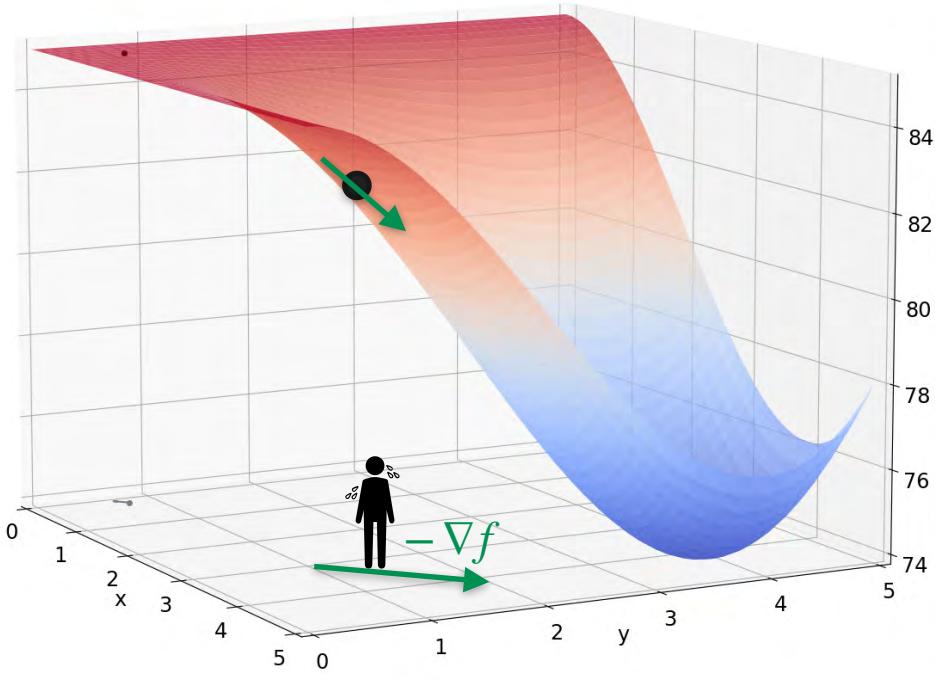
Initial position:  $(x_0, y_0)$

Direction of greatest ascent:  $\nabla f$

Direction of greatest descent:  $-\nabla f$

Updated position:  $\underbrace{(x_0, y_0) - \alpha \nabla f}_{(x_1, y_1)}$

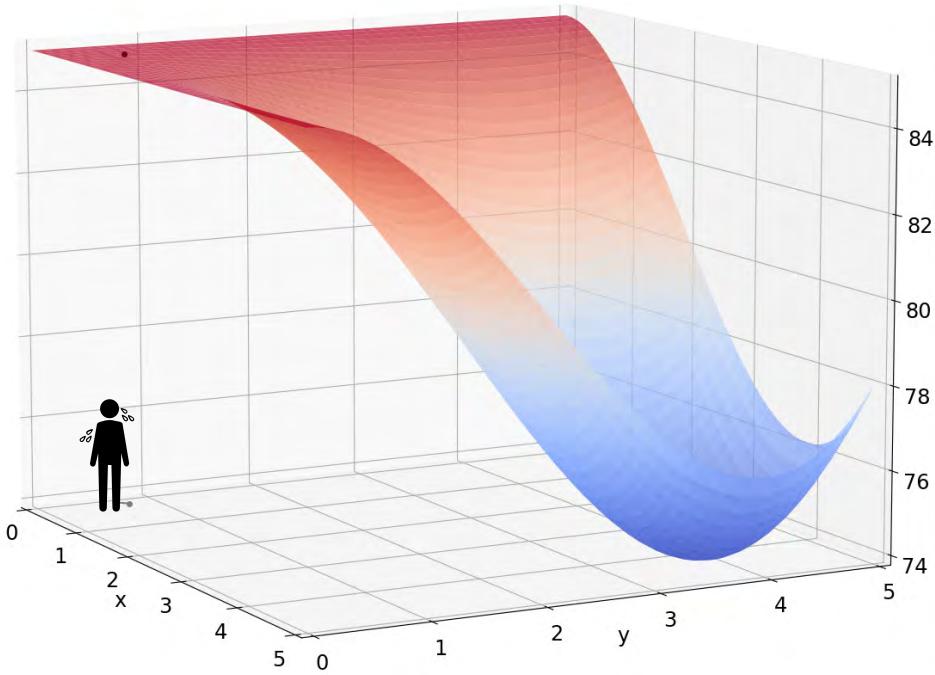
Better point!



# Method 2: Gradient Descent

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

Start:  $x = 0.5, y = 0.6$

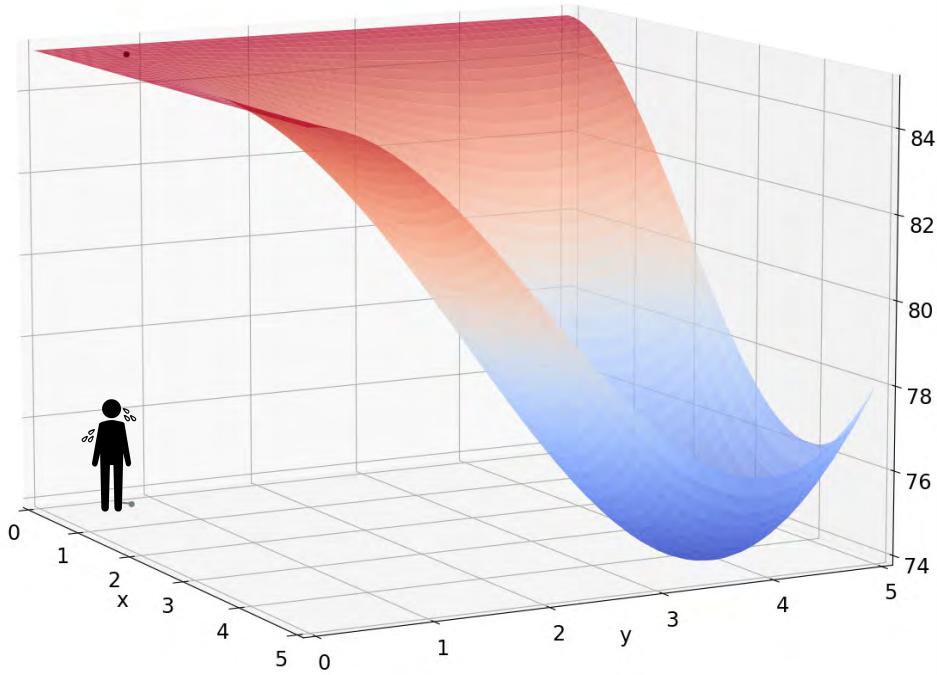


# Method 2: Gradient Descent

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

Start:  $x = 0.5, y = 0.6$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$



# Method 2: Gradient Descent

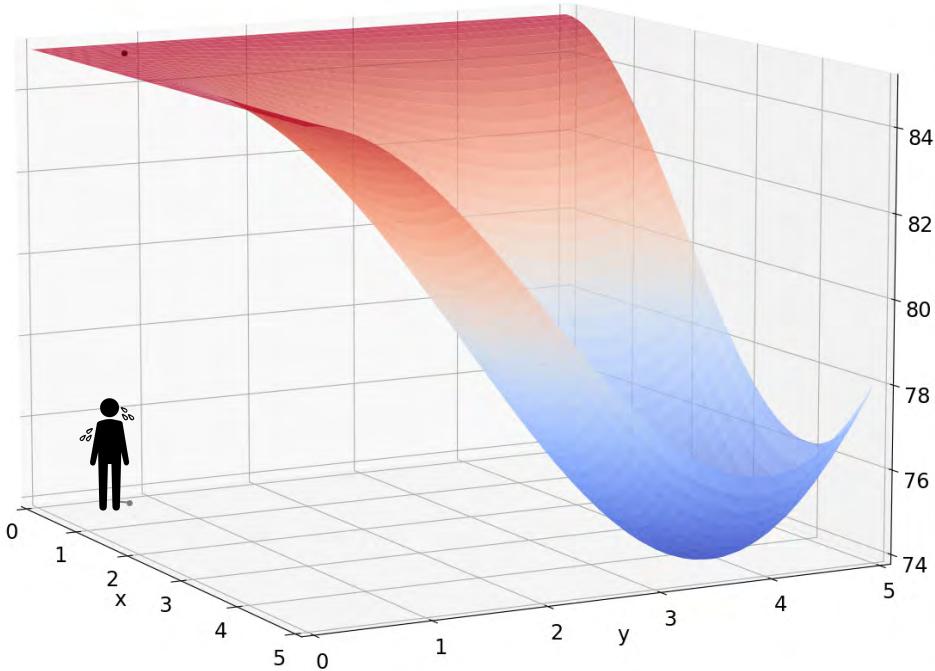
$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

Start:  $x = 0.5, y = 0.6$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

$$\frac{\partial f}{\partial x} = -\frac{1}{90}x(3x - 12)y^2(y - 6)$$

$$\frac{\partial f}{\partial y} = -\frac{1}{90}x^2(x - 6)y(3y - 12)$$



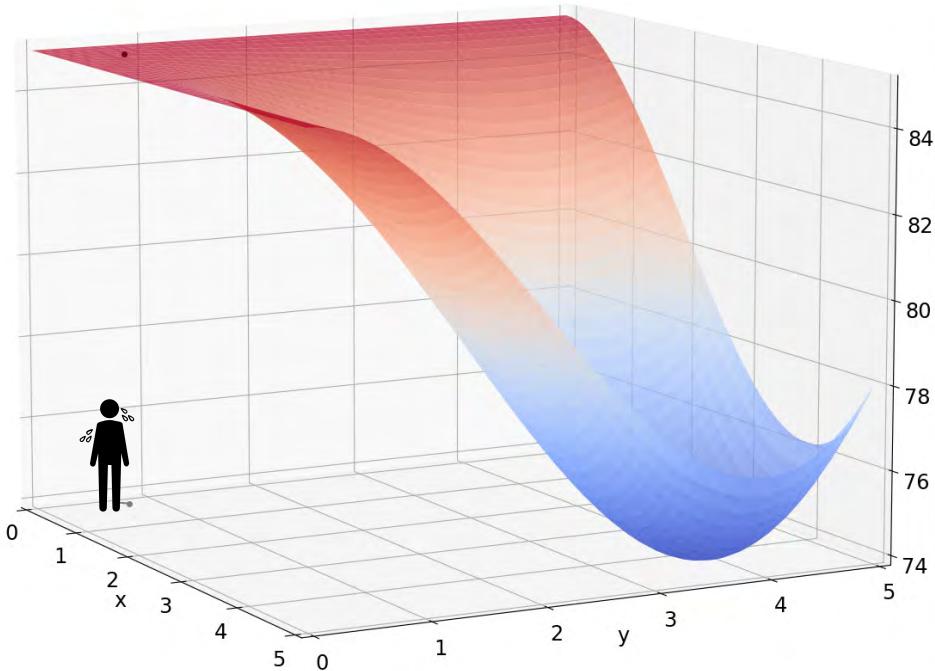
# Method 2: Gradient Descent

$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

Start:  $x = 0.5, y = 0.6$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

$$\nabla f = \begin{bmatrix} -\frac{1}{90}x(3x - 12)y^2(y - 6) \\ -\frac{1}{90}x^2(x - 6)y(3y - 12) \end{bmatrix}$$



# Method 2: Gradient Descent

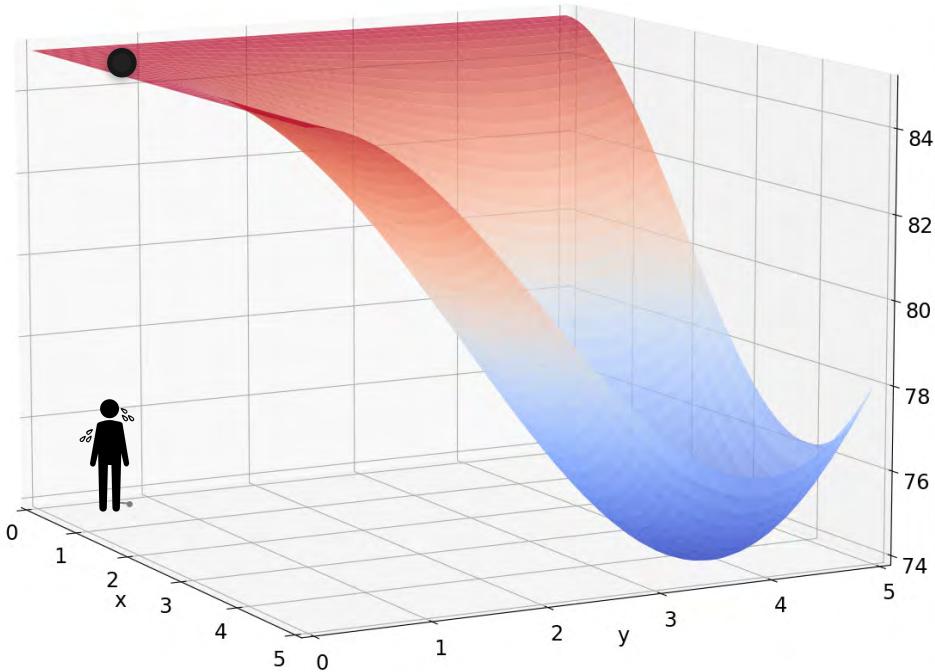
$$T = f(x, y) = 85 - \frac{1}{90}x^2(x - 6)y^2(y - 6)$$

Start:  $x = 0.5, y = 0.6$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

$$\nabla f = \begin{bmatrix} -\frac{1}{90}x(3x - 12)y^2(y - 6) \\ -\frac{1}{90}x^2(x - 6)y(3y - 12) \end{bmatrix}$$

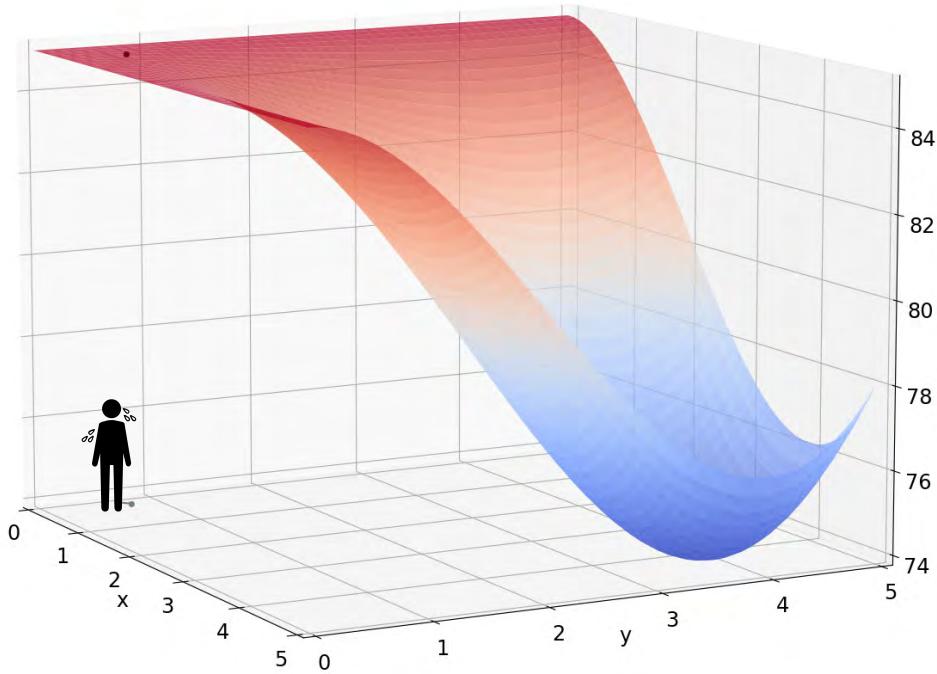
$$\nabla f(0.5, 0.6) = \begin{bmatrix} 0.1134 \\ -0.0935 \end{bmatrix}$$



# Method 2: Gradient Descent

Start:  $x = 0.5, y = 0.6$

$$\nabla f(0.5, 0.6) = \begin{bmatrix} 0.1134 \\ -0.0935 \end{bmatrix}$$

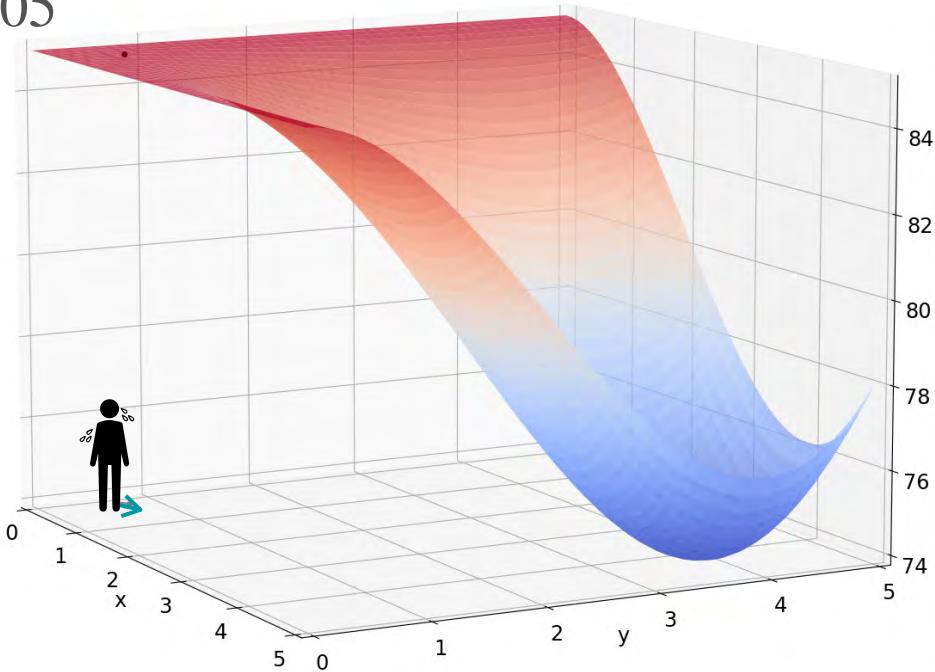


# Method 2: Gradient Descent

Start:  $x = 0.5, y = 0.6$     Rate:  $\alpha = 0.05$

$$\nabla f(0.5, 0.6) = \begin{bmatrix} 0.1134 \\ -0.0935 \end{bmatrix}$$

Move by  
 $-0.05 \nabla f(0.5, 0.6)$



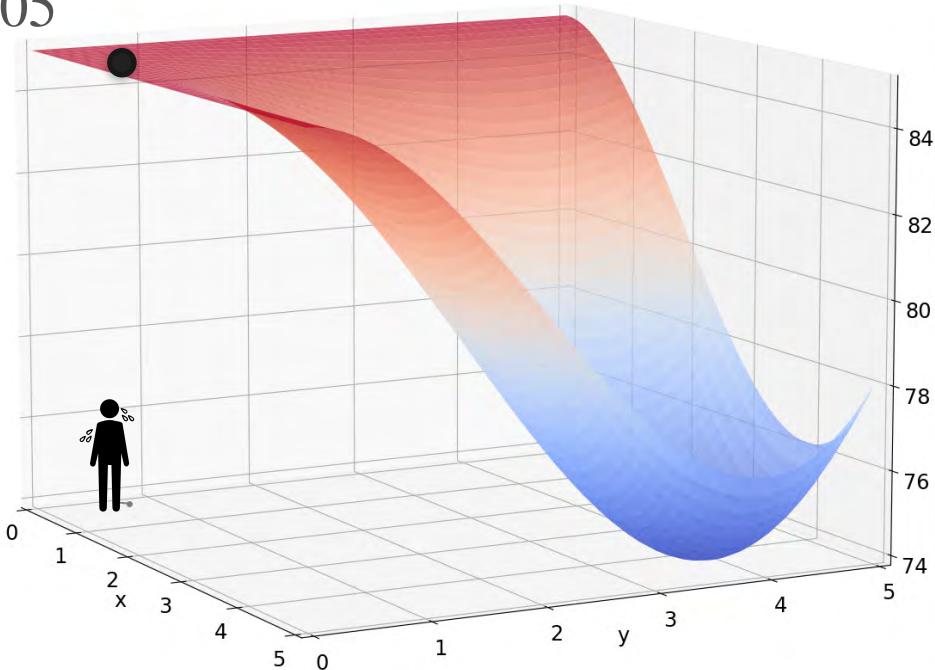
# Method 2: Gradient Descent

Start:  $x = 0.5$ ,  $y = 0.6$     Rate:  $\alpha = 0.05$

$$\nabla f(0.5, 0.6) = \begin{bmatrix} 0.1134 \\ -0.0935 \end{bmatrix}$$

Move by  
 $-0.05 \nabla f(0.5, 0.6)$

$$\begin{aligned} x &\mapsto 0.5057 \\ y &\mapsto 0.6047 \end{aligned}$$



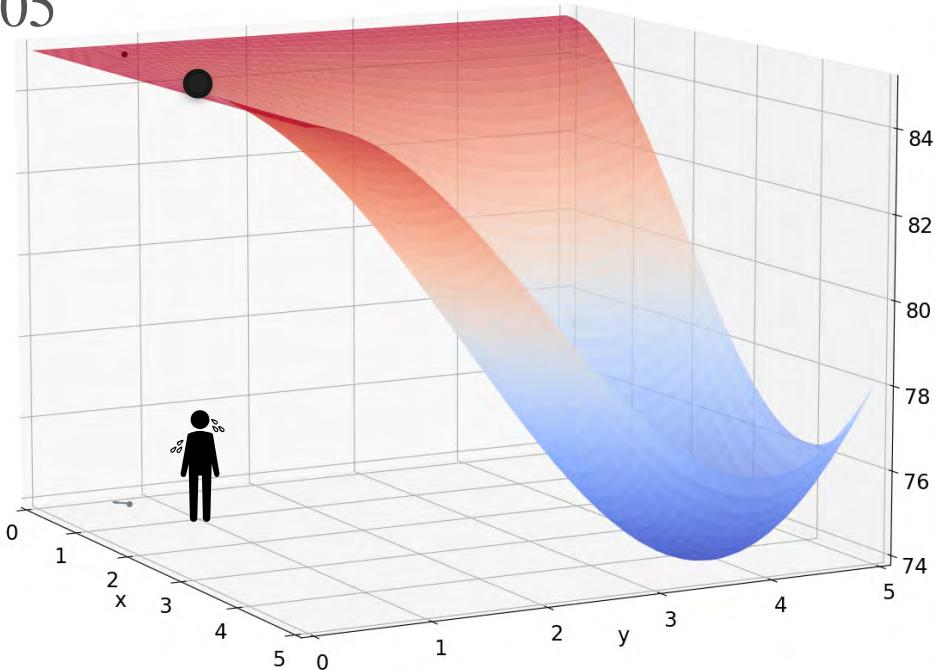
# Method 2: Gradient Descent

Start:  $x = 0.5$ ,  $y = 0.6$     Rate:  $\alpha = 0.05$

$$\nabla f(0.5, 0.6) = \begin{bmatrix} 0.1134 \\ -0.0935 \end{bmatrix}$$

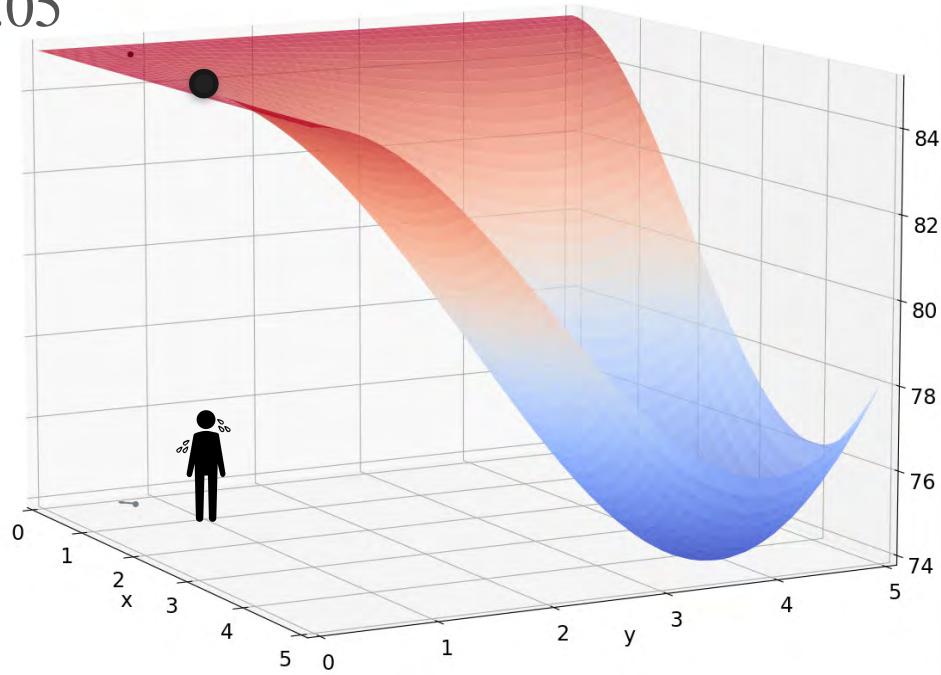
Move by  
 $-0.05 \nabla f(0.5, 0.6)$

$$\begin{aligned} x &\mapsto 0.5057 \\ y &\mapsto 0.6047 \end{aligned}$$



# Method 2

Start:  $x = 0.5, y = 0.6$     Rate:  $\alpha = 0.05$



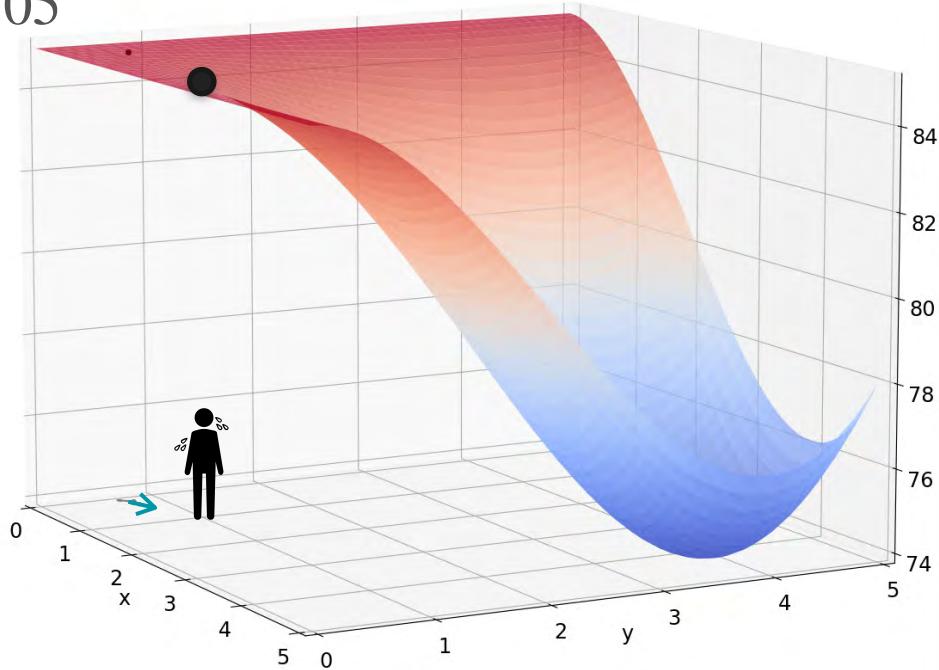
# Method 2

Start:  $x = 0.5, y = 0.6$     Rate:  $\alpha = 0.05$

Find:

$$\nabla f(0.5057, 0.6047) = \begin{bmatrix} -0.1162 \\ -0.0961 \end{bmatrix}$$

Move by  
 $-0.05 \nabla f(0.5057, 0.6047)$



# Method 2

Start:  $x = 0.5$ ,  $y = 0.6$     Rate:  $\alpha = 0.05$

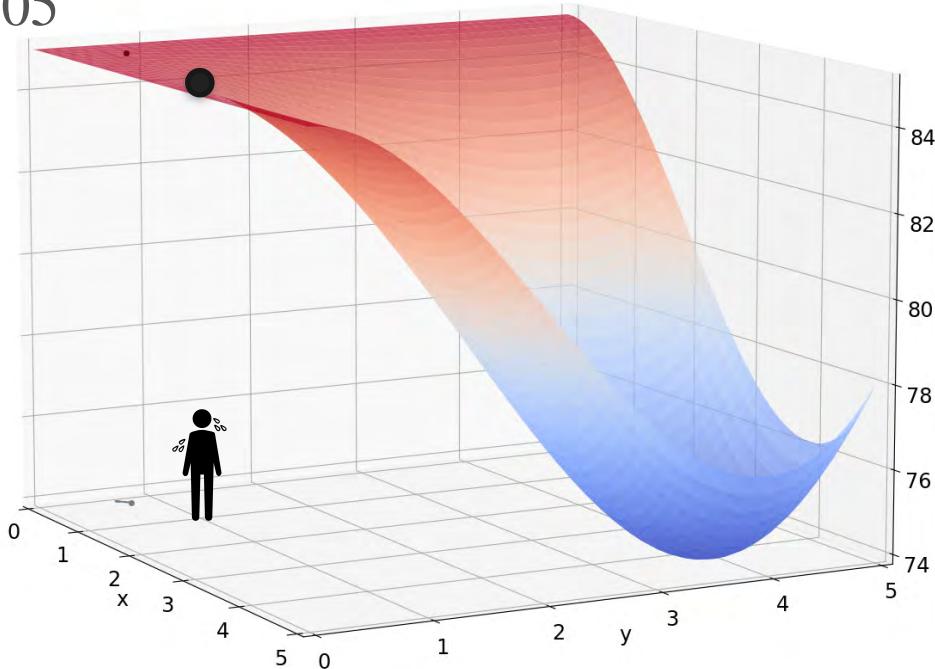
Find:

$$\nabla f(0.5057, 0.6047) = \begin{bmatrix} -0.1162 \\ -0.0961 \end{bmatrix}$$

Move by  
 $-0.05 \nabla f(0.5057, 0.6047)$

$$\begin{aligned} x &\mapsto 0.5115 \\ y &\mapsto 0.6095 \end{aligned}$$

**Repeat!**



# Gradient Descent

Start:  $x = 0.5$ ,  $y = 0.6$     Rate:  $\alpha = 0.05$

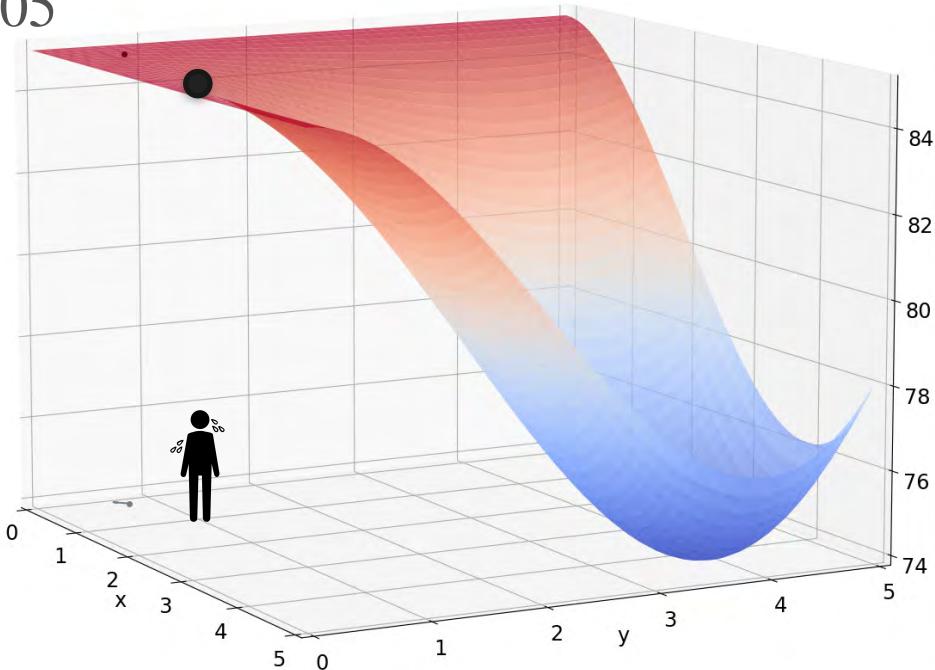
Find:

$$\nabla f(0.5057, 0.6047) = \begin{bmatrix} -0.1162 \\ -0.0961 \end{bmatrix}$$

Move by  
 $-0.05 \nabla f(0.5057, 0.6047)$

$$\begin{aligned} x &\mapsto 0.5115 \\ y &\mapsto 0.6095 \end{aligned}$$

**Repeat!**



# Gradient Descent

Start:  $x = 0.5$ ,  $y = 0.6$     Rate:  $\alpha = 0.05$

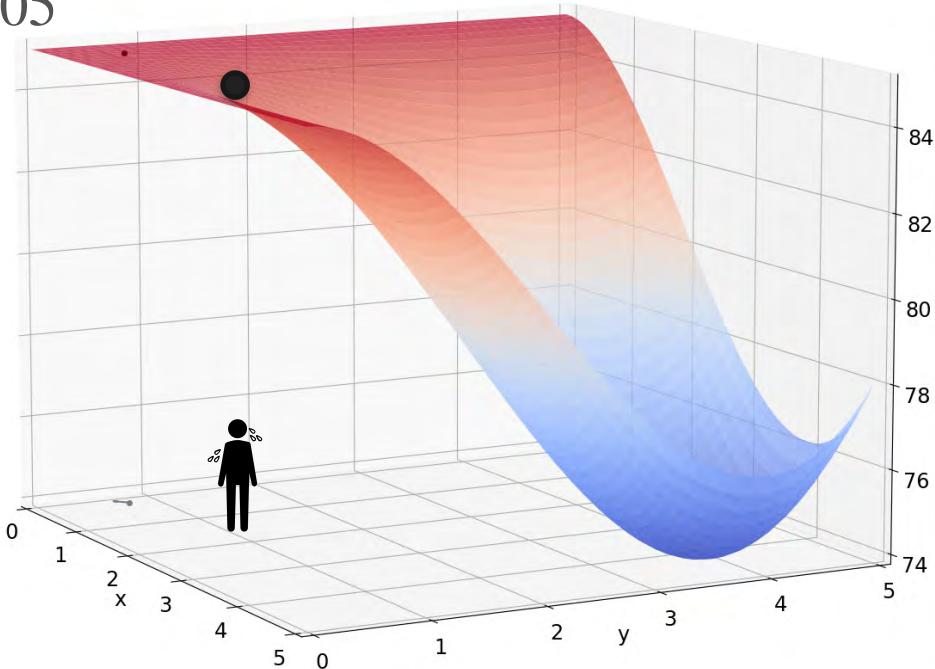
Find:

$$\nabla f(0.5057, 0.6047) = \begin{bmatrix} -0.1162 \\ -0.0961 \end{bmatrix}$$

Move by  
 $-0.05 \nabla f(0.5057, 0.6047)$

$$\begin{aligned} x &\mapsto 0.5115 \\ y &\mapsto 0.6095 \end{aligned}$$

**Repeat!**



# Gradient Descent

Start:  $x = 0.5$ ,  $y = 0.6$     Rate:  $\alpha = 0.05$

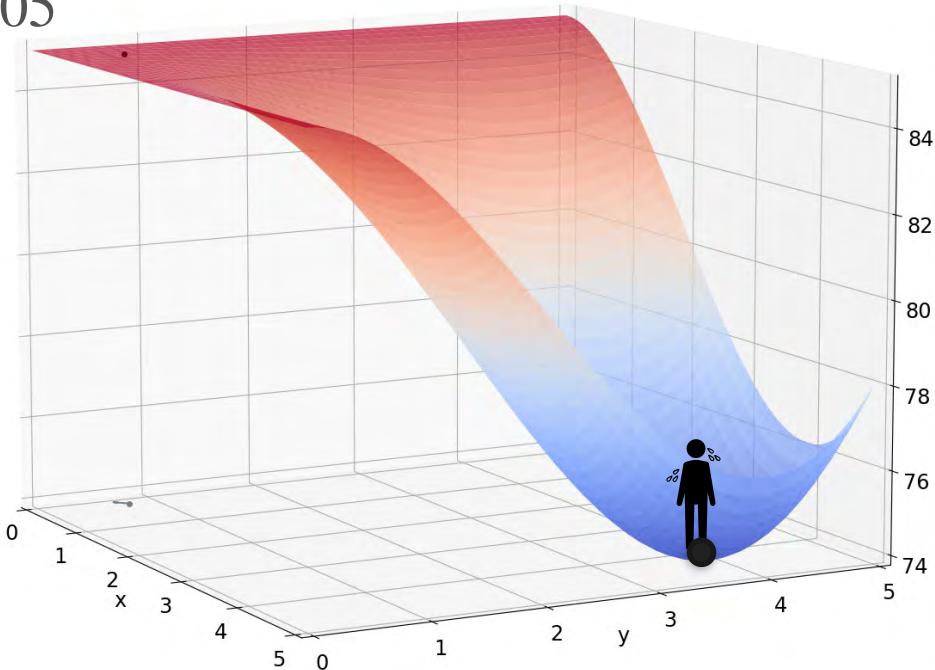
Find:

$$\nabla f(0.5057, 0.6047) = \begin{bmatrix} -0.1162 \\ -0.0961 \end{bmatrix}$$

Move by  
 $-0.05 \nabla f(0.5057, 0.6047)$

$$\begin{aligned} x &\mapsto 0.5115 \\ y &\mapsto 0.6095 \end{aligned}$$

**Repeat!**



# Gradient Descent

# Gradient Descent

Function:  $f(x, y)$

# Gradient Descent

Function:  $f(x, y)$

Goal: find minimum of  $f(x, y)$

# Gradient Descent

Function:  $f(x, y)$

Goal: find minimum of  $f(x, y)$

## Step 1:

Define a learning rate  $\alpha$

Choose a starting point  $(x_0, y_0)$

# Gradient Descent

Function:  $f(x, y)$

Goal: find minimum of  $f(x, y)$

## Step 1:

Define a learning rate  $\alpha$

Choose a starting point  $(x_0, y_0)$

## Step 2:

Update: 
$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix} - \alpha \nabla f(x_{k-1}, y_{k-1})$$

# Gradient Descent

Function:  $f(x, y)$

Goal: find minimum of  $f(x, y)$

## Step 1:

Define a learning rate  $\alpha$

Choose a starting point  $(x_0, y_0)$

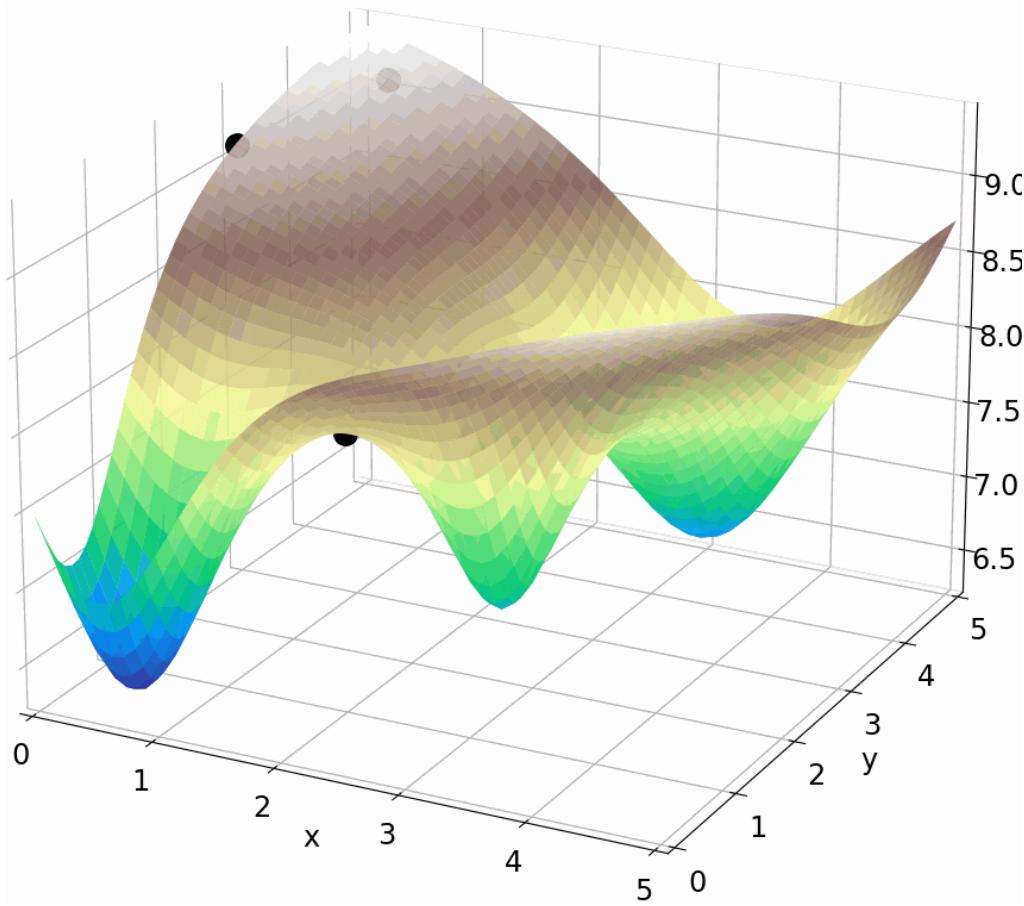
## Step 2:

Update:  $\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix} - \alpha \nabla f(x_{k-1}, y_{k-1})$

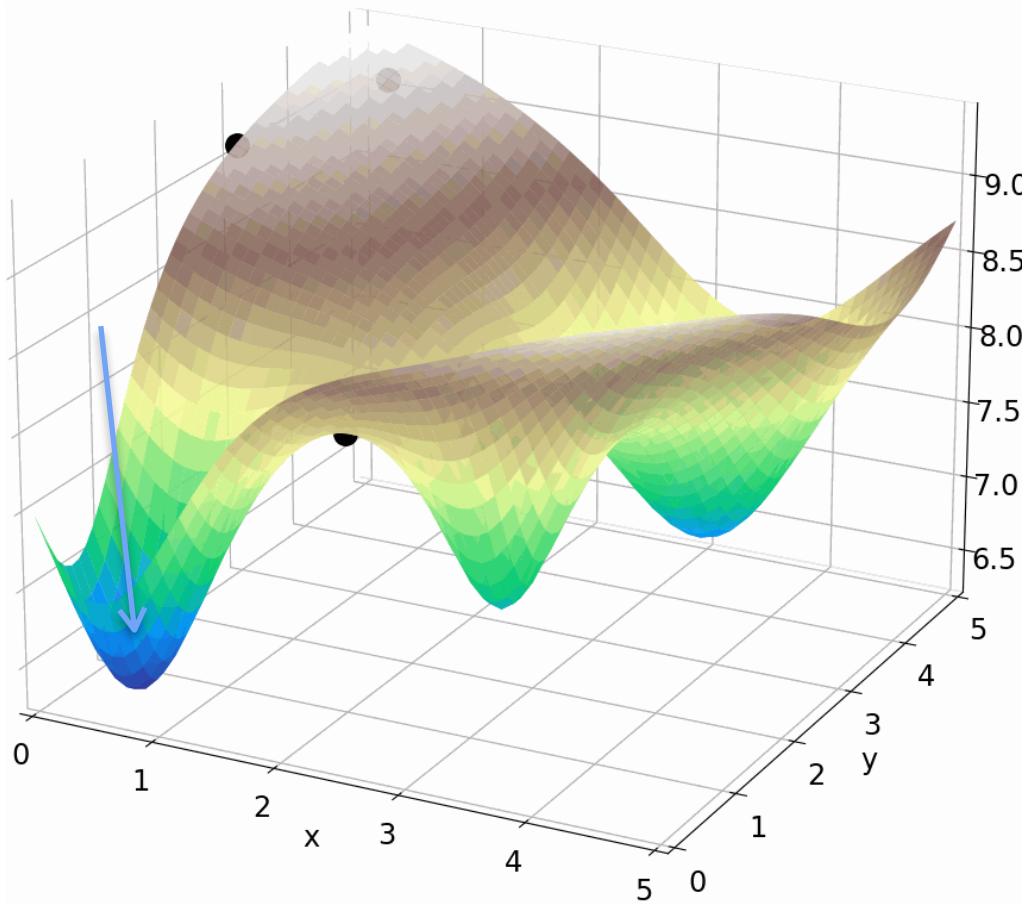
## Step 3:

Repeat Step 2 until you are close enough to  
the true minimum  $(x^*, y^*)$

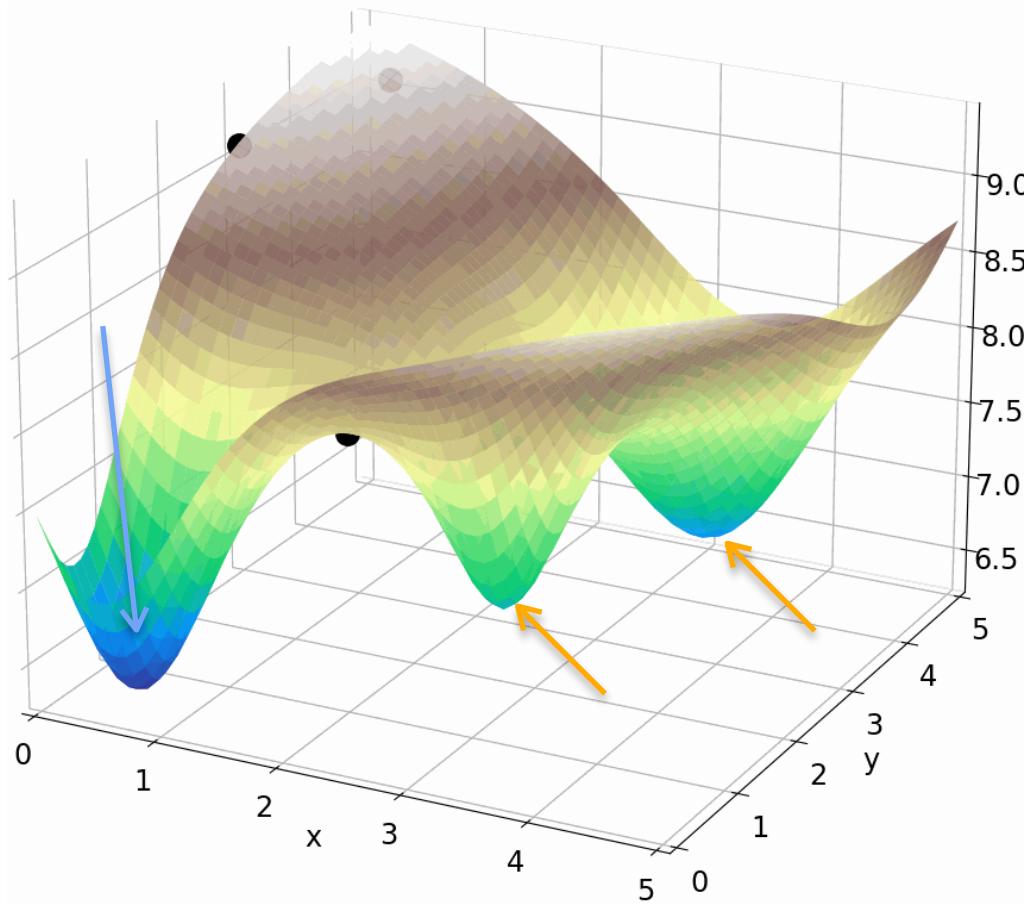
# Gradient Descent With Local Minima



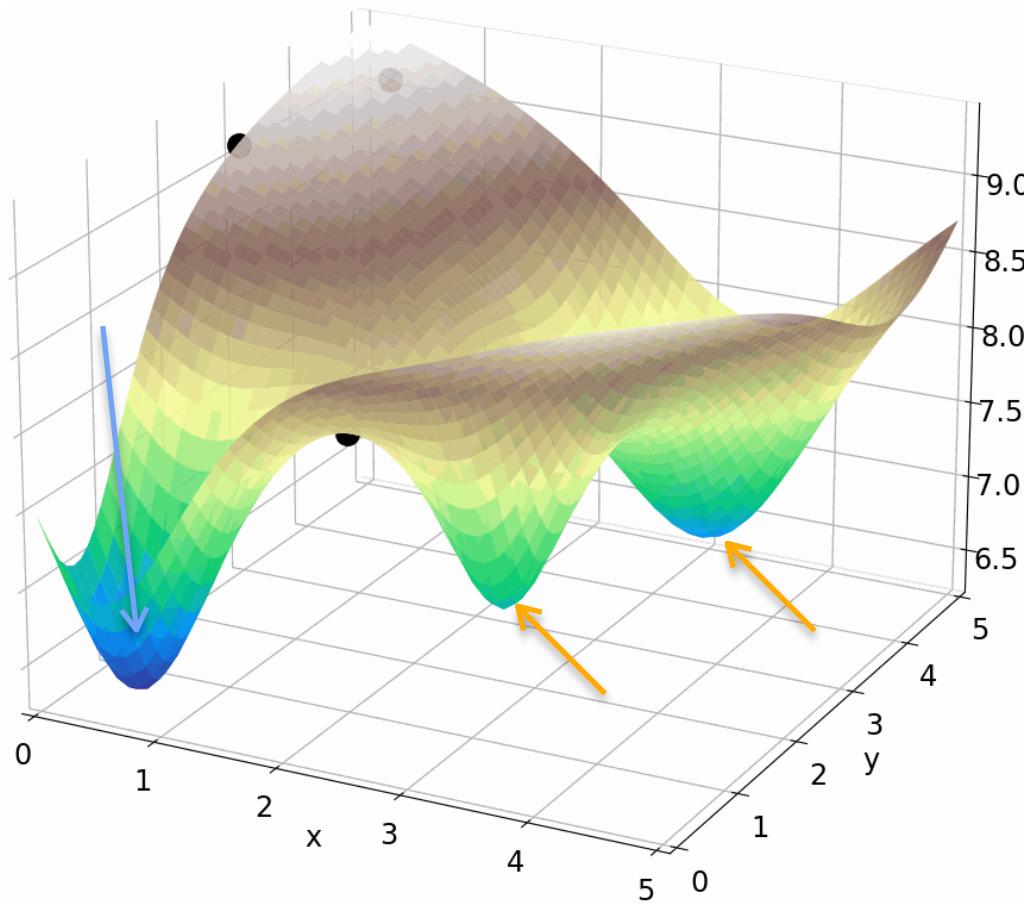
# Gradient Descent With Local Minima



# Gradient Descent With Local Minima



# Gradient Descent With Local Minima





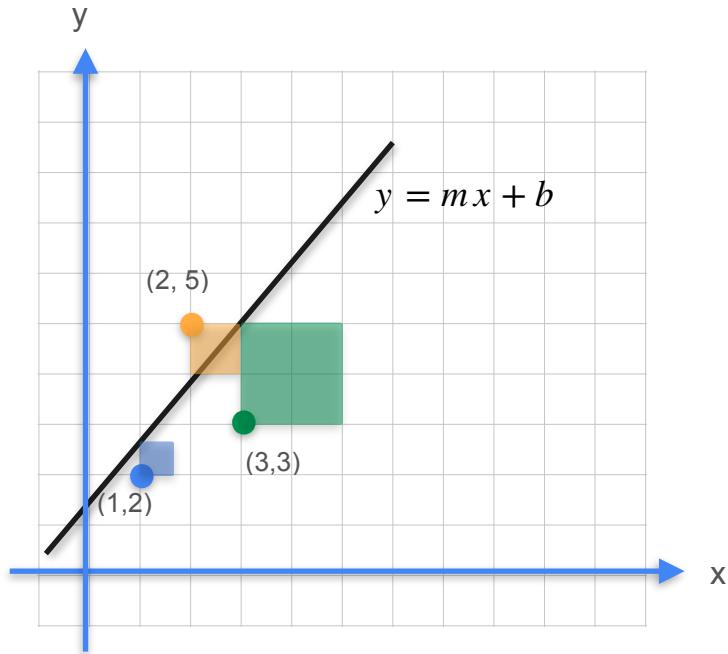
DeepLearning.AI

# Gradients and Gradient Descent

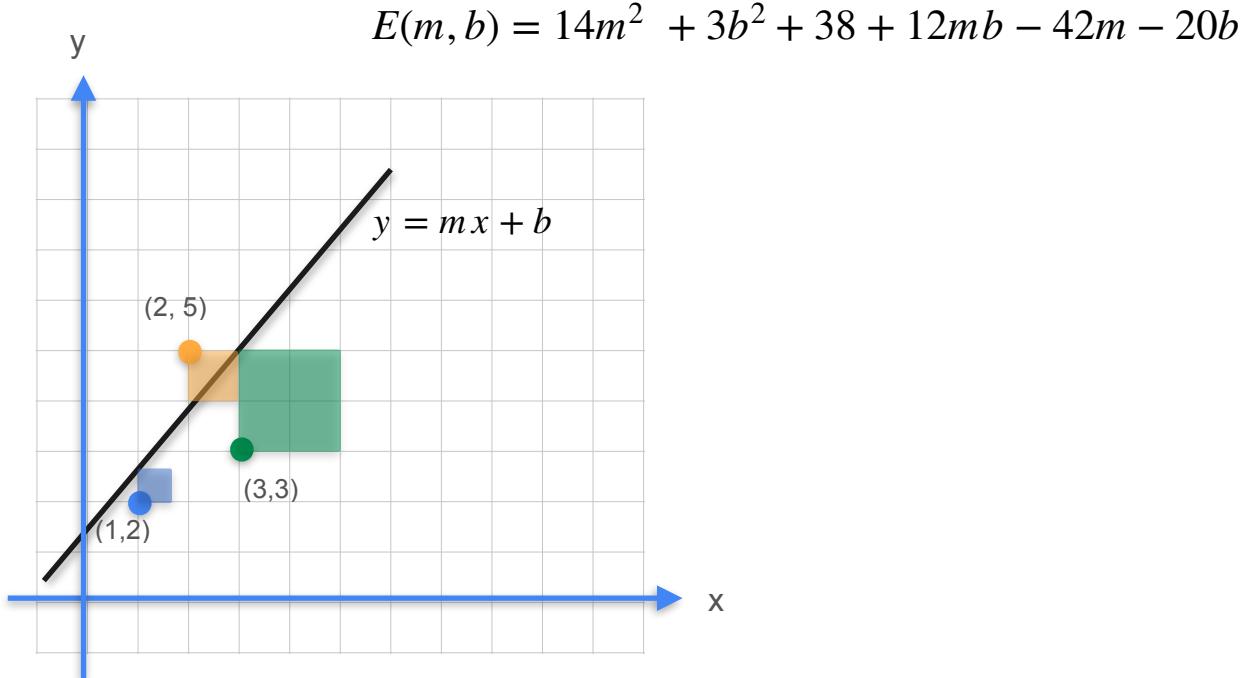
---

**Optimization using Gradient  
Descent - Least squares**

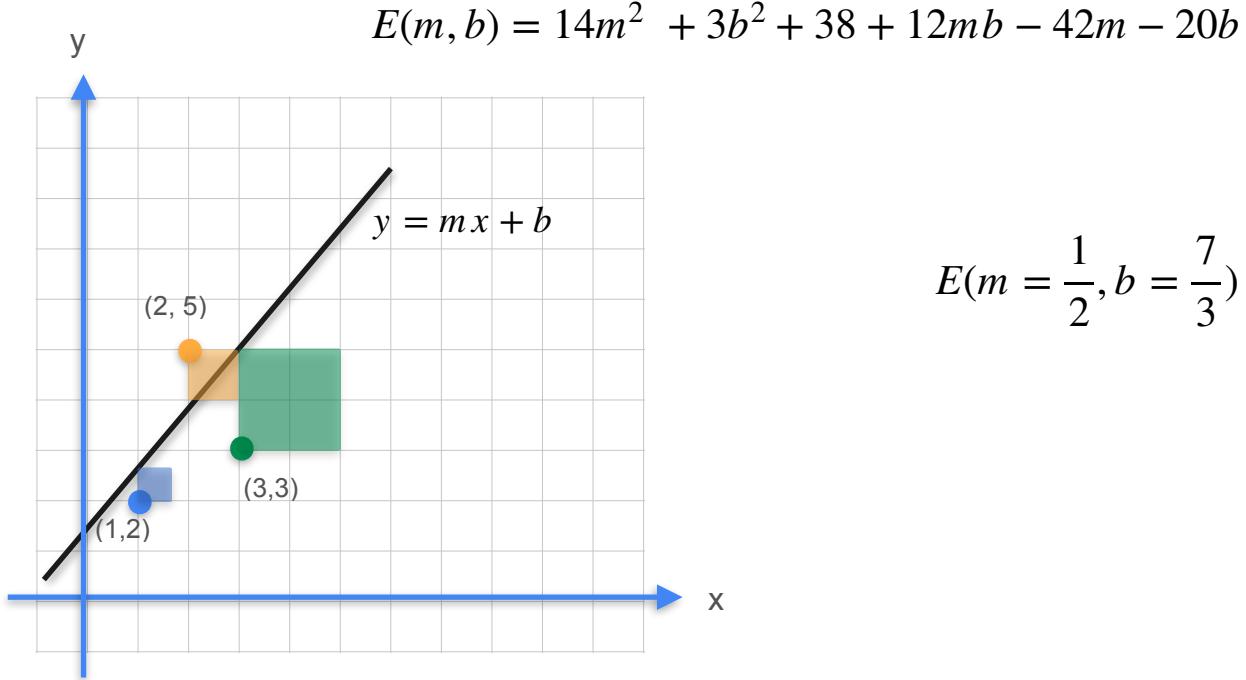
# Gradient Descent With Power Lines Example



# Gradient Descent With Power Lines Example

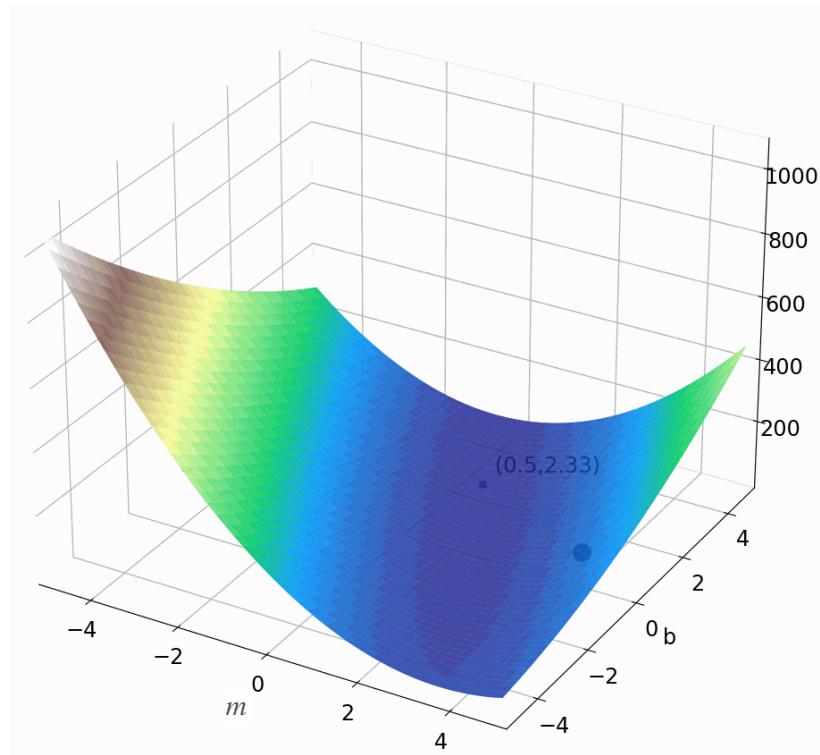


# Gradient Descent With Power Lines Example



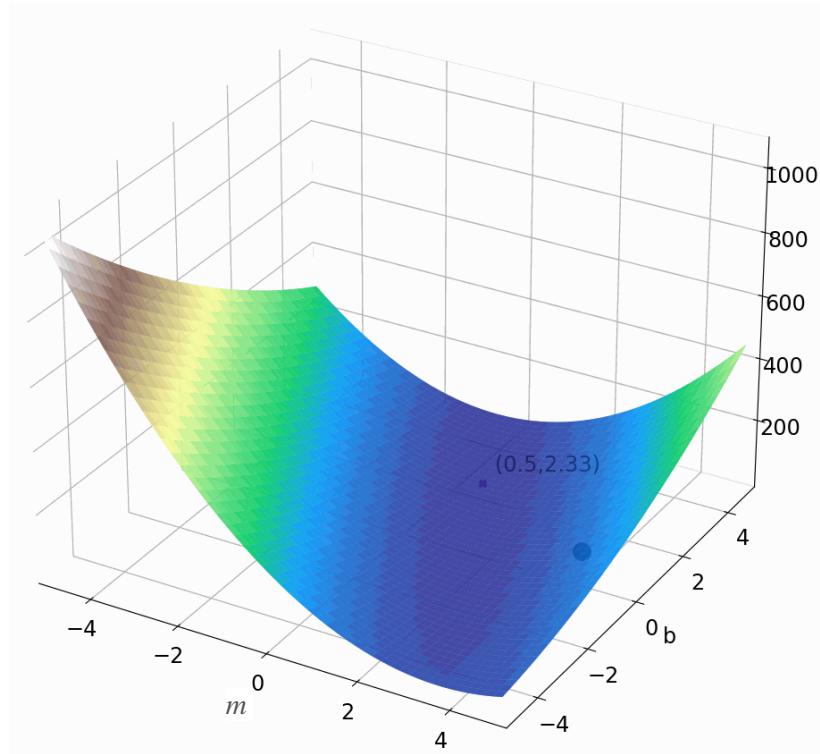
$$E\left(m = \frac{1}{2}, b = \frac{7}{3}\right) \approx 4.167$$

# Linear Regression: Gradient Descent



# Linear Regression: Gradient Descent

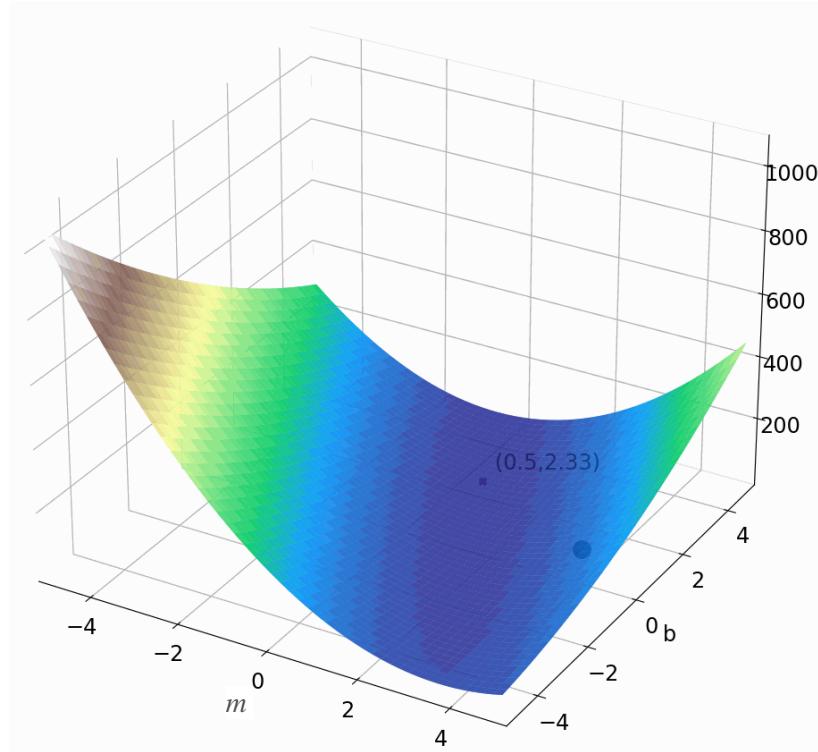
Goal: Minimize sum of squares cost



# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$



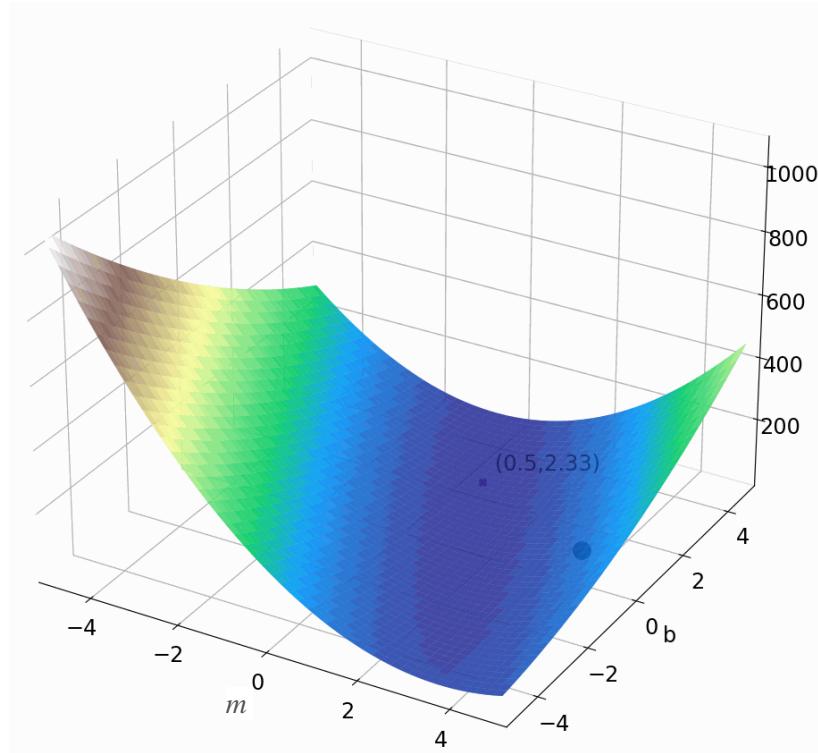
# Linear Regression: Gradient Descent

**Goal: Minimize sum of squares cost**

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$m =$$

$$b =$$

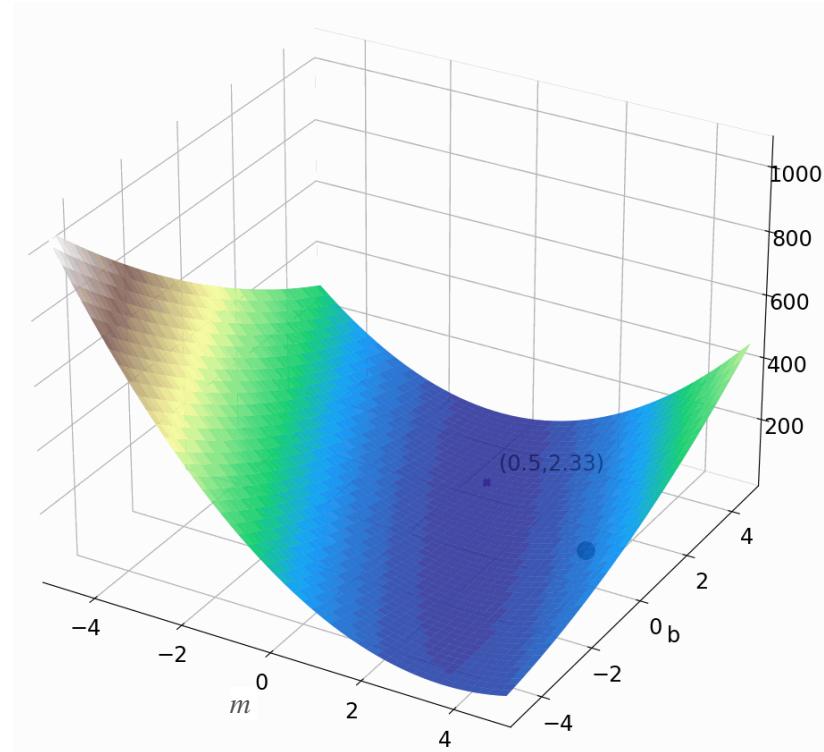


# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

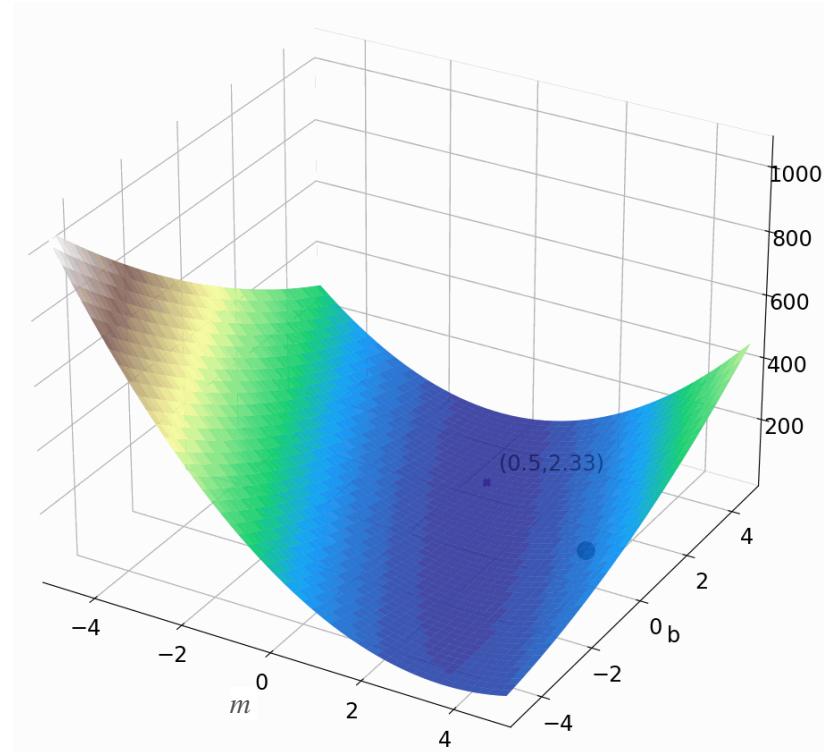


# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

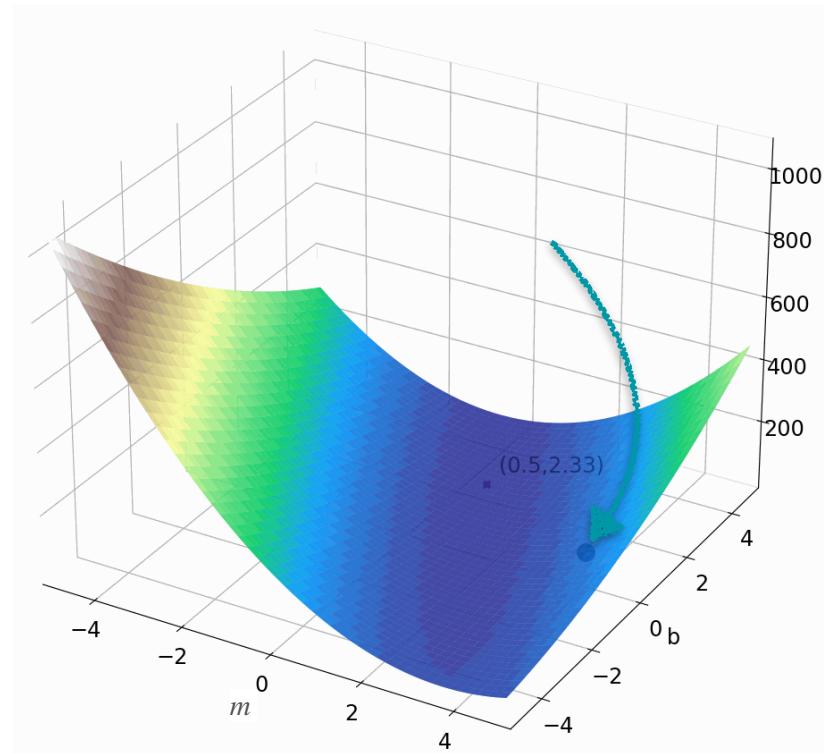


# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

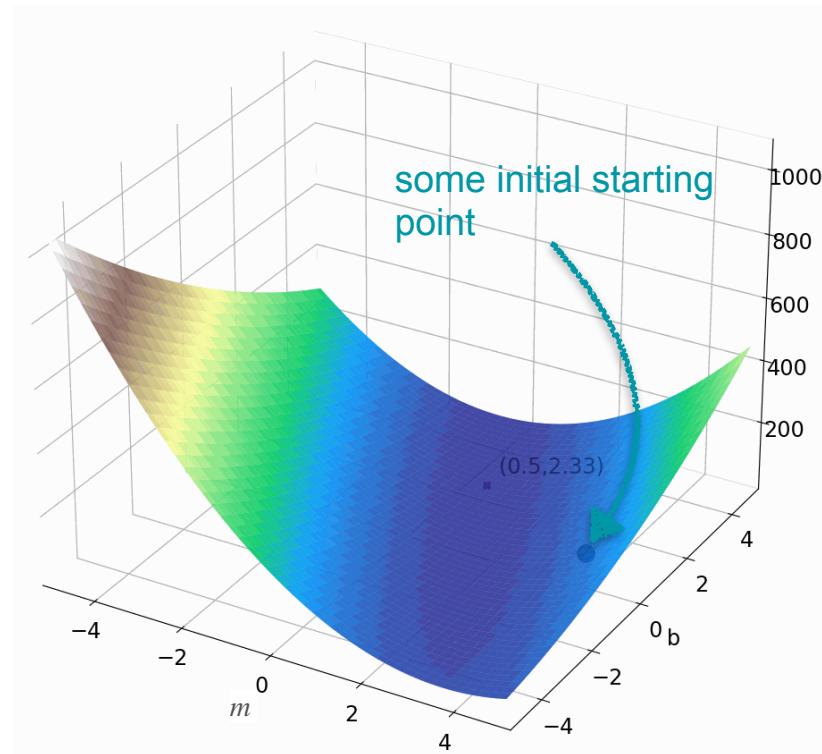


# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

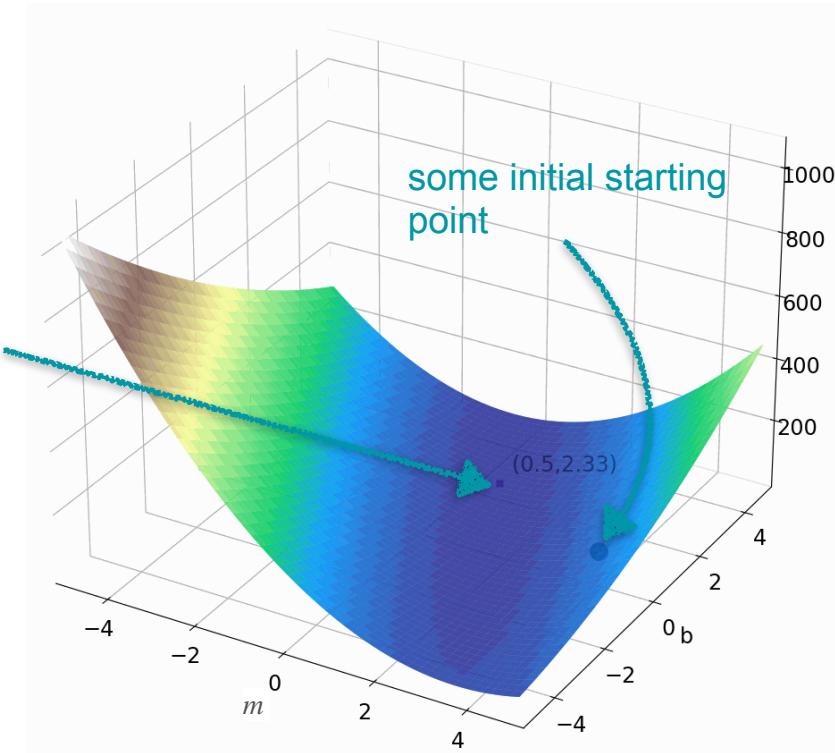


# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$



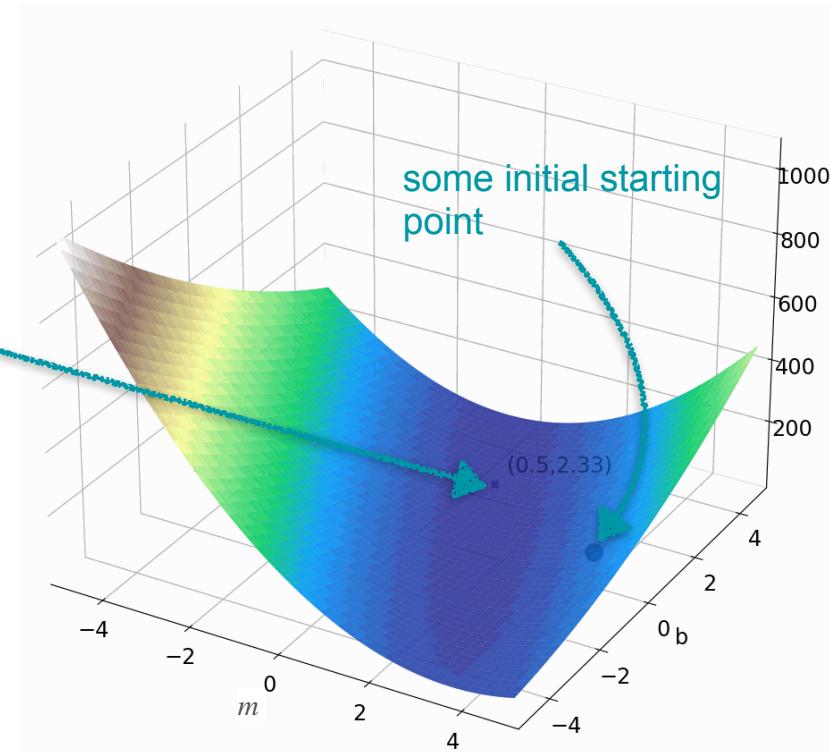
# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

The points  $m, b$  such that  
the cost is minimum



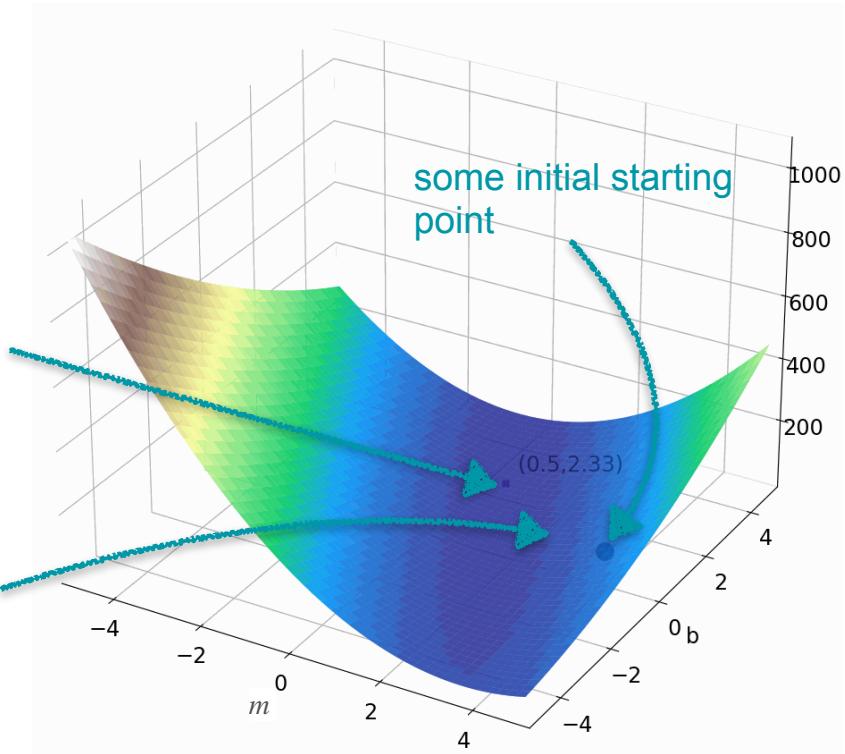
# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

The points  $m, b$  such that  
the cost is minimum



# Linear Regression: Gradient Descent

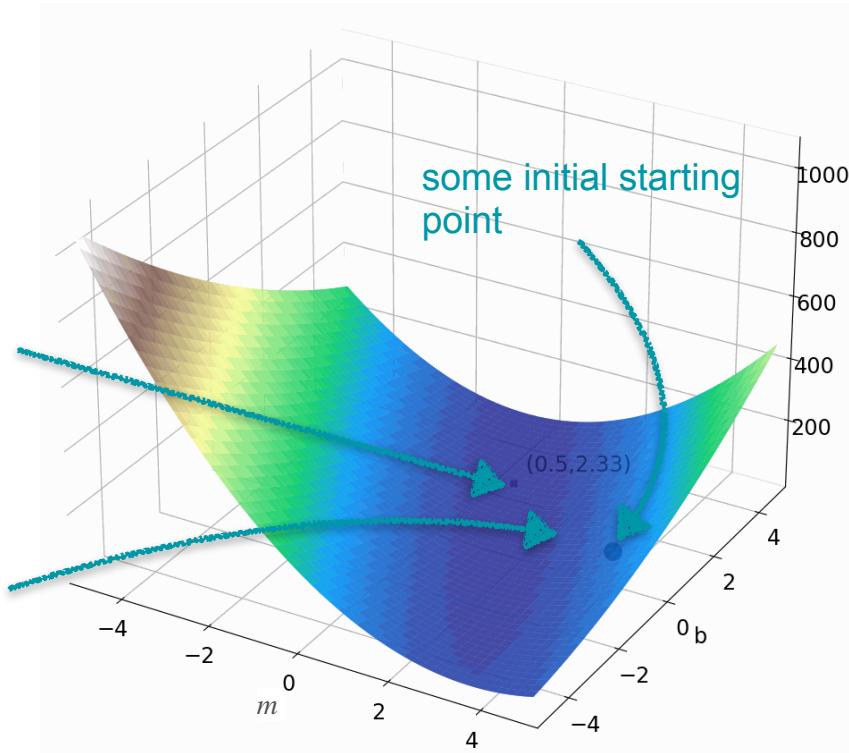
Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

The points  $m, b$  such that  
the cost is minimum

descend until you  
find the minimum



# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

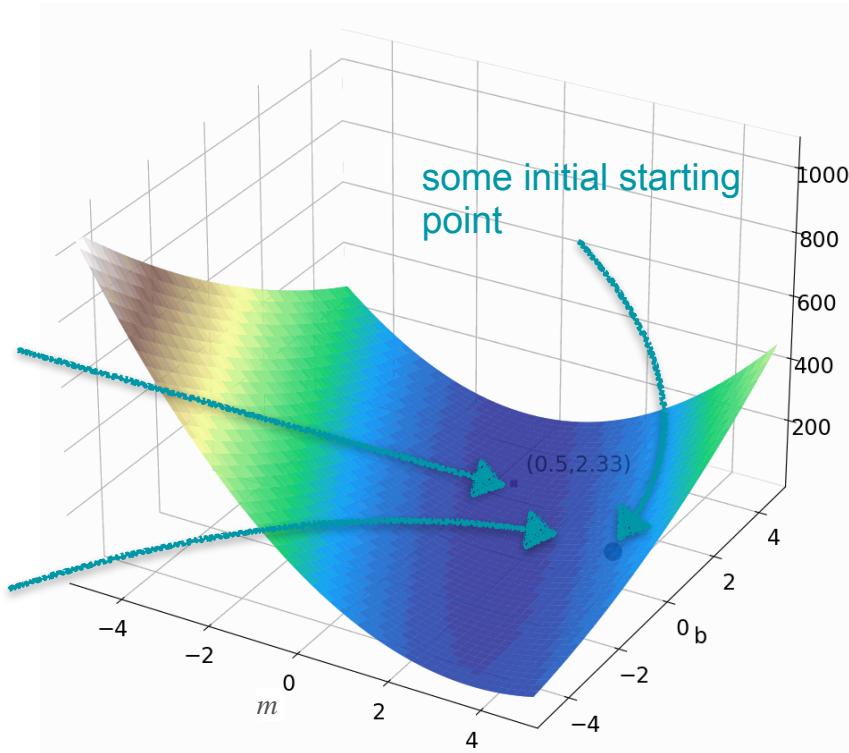
$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

Steps:

descend until you  
find the minimum

The points  $m, b$  such that  
the cost is minimum



# Linear Regression: Gradient Descent

Goal: Minimize sum of squares cost

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

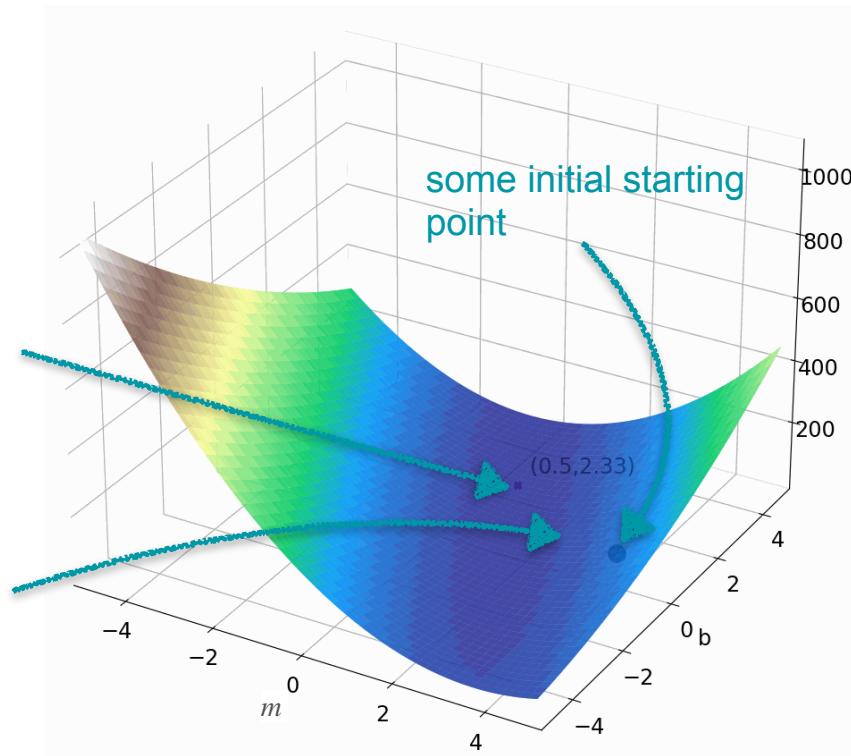
$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

Steps:

Start with  $(m_0, b_0)$

descend until you  
find the minimum

The points  $m, b$  such that  
the cost is minimum



# Linear Regression: Gradient Descent

**Goal: Minimize sum of squares cost**

$$\nabla E = [28m + 12b - 42, 6b + 12m - 20]$$

$$\begin{aligned} m &= ? \\ b &= ? \end{aligned}$$

The points  $m, b$  such that  
the cost is minimum

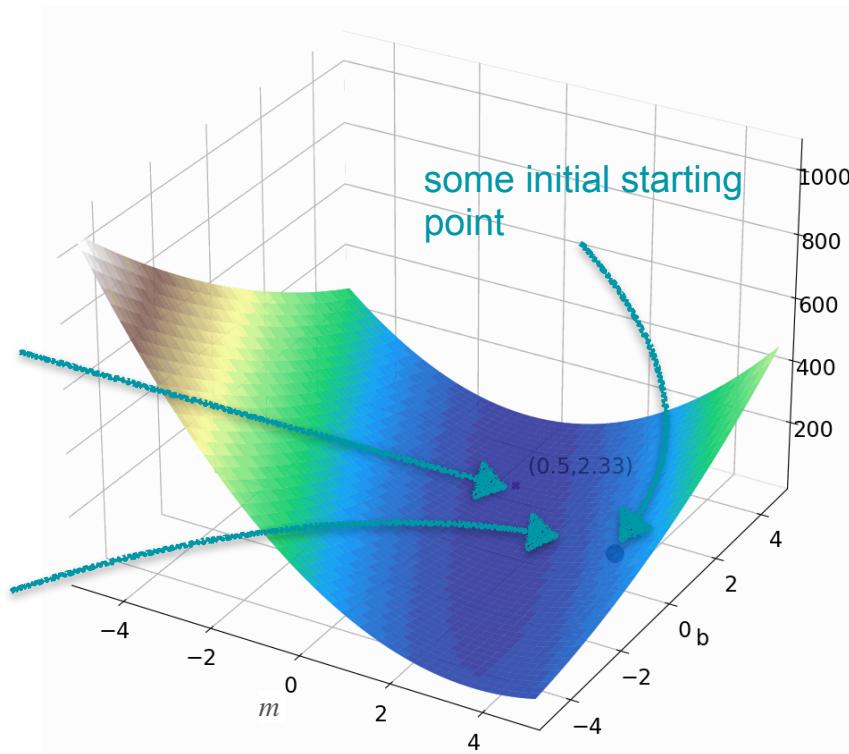
Steps:

Start with  $(m_0, b_0)$

descend until you  
find the minimum

Iterate

$$(m_{k+1}, b_{k+1}) = (m_k, b_k) - \alpha \nabla E(m_k, b_k)$$





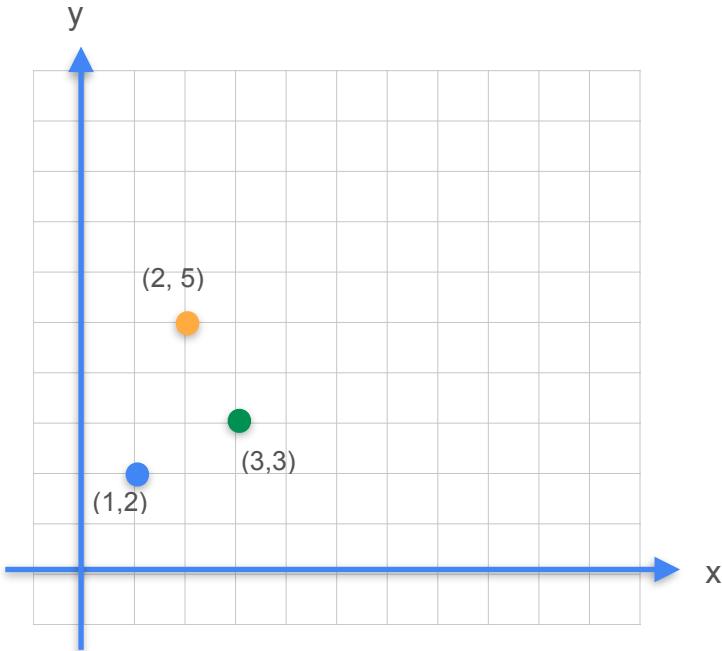
DeepLearning.AI

# Gradients and Gradient Descent

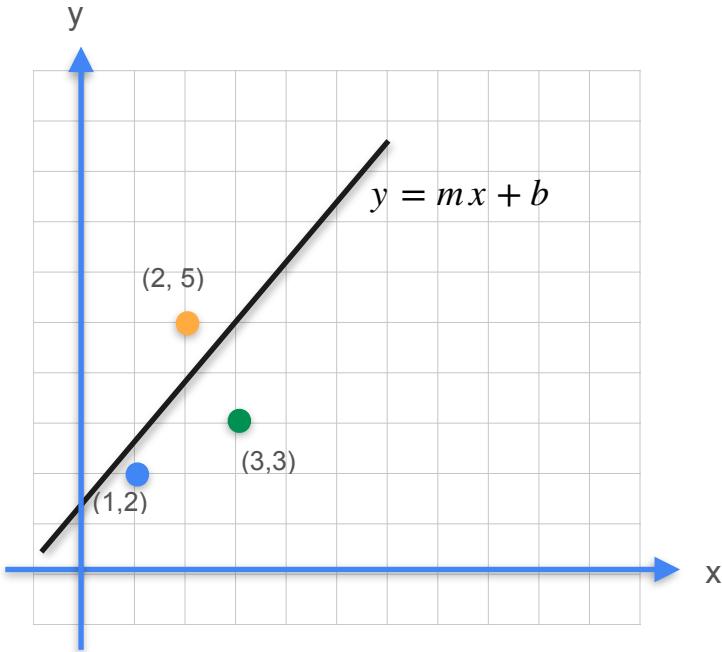
---

**Optimization using Gradient  
Descent - Least squares  
with multiple observations**

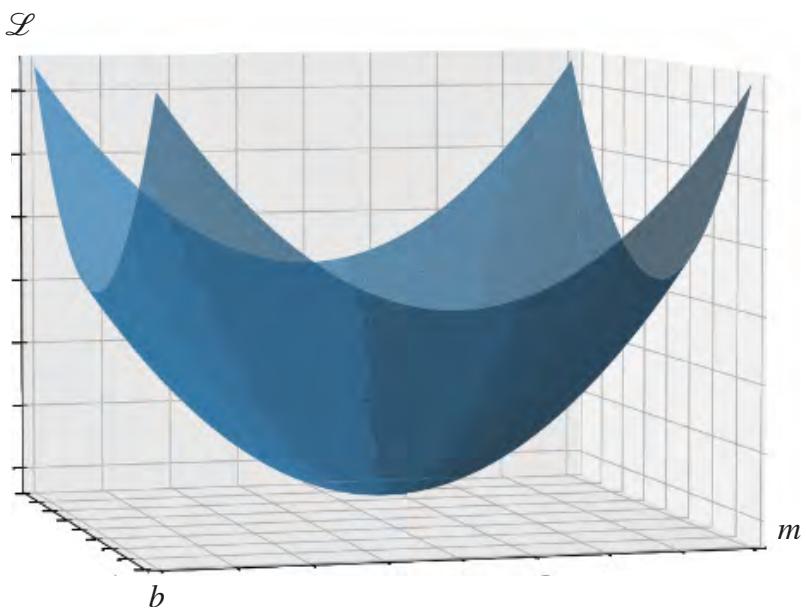
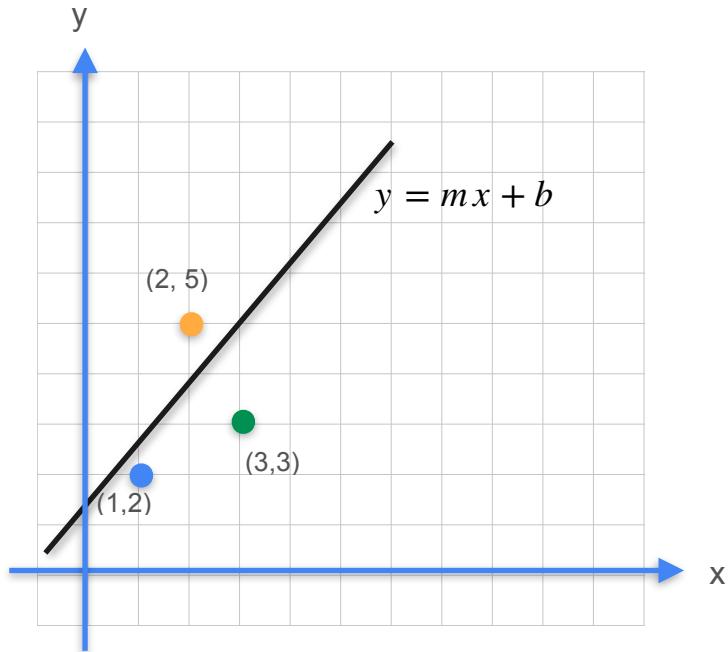
# Gradient Descent



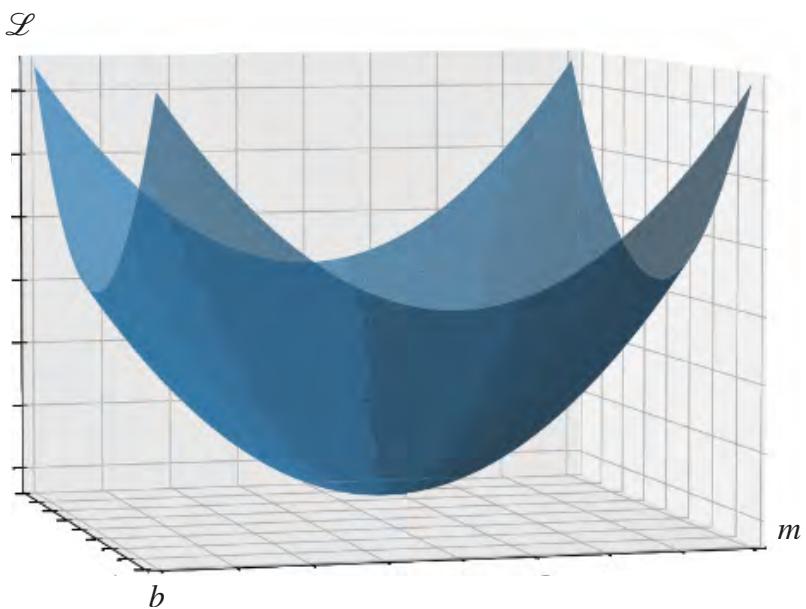
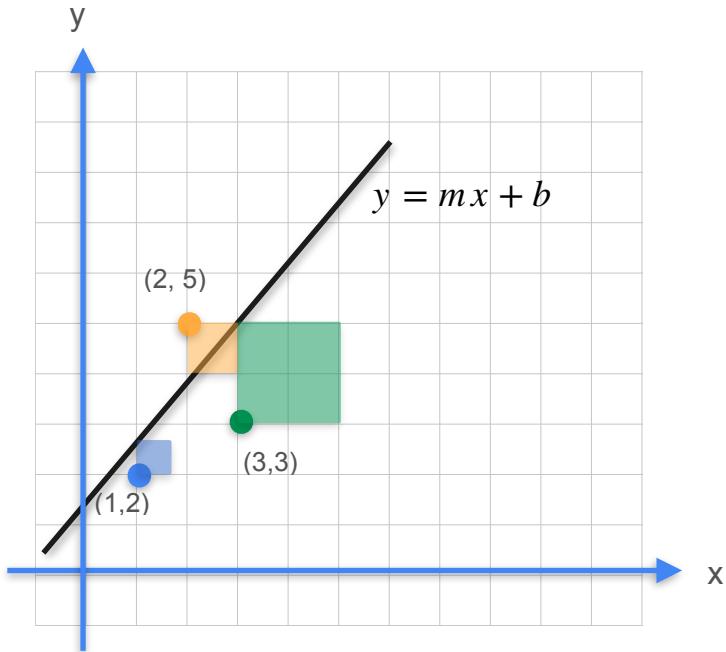
# Gradient Descent



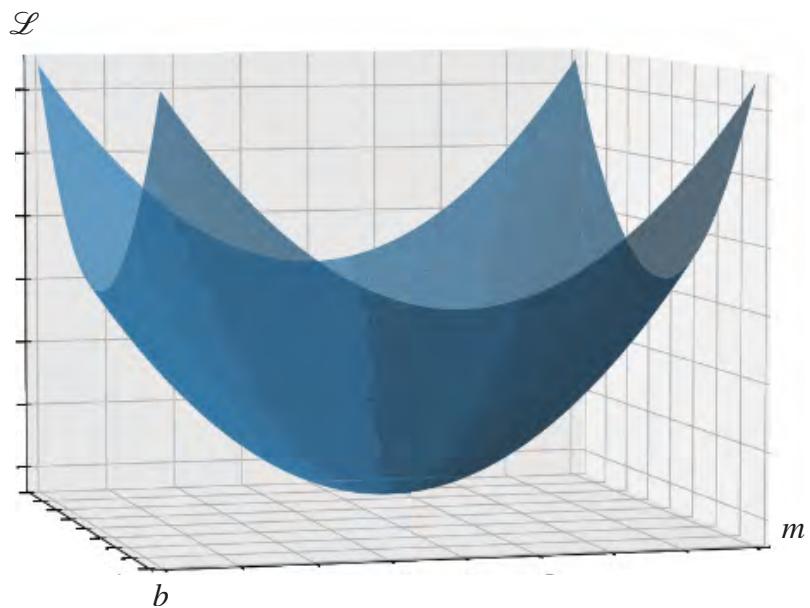
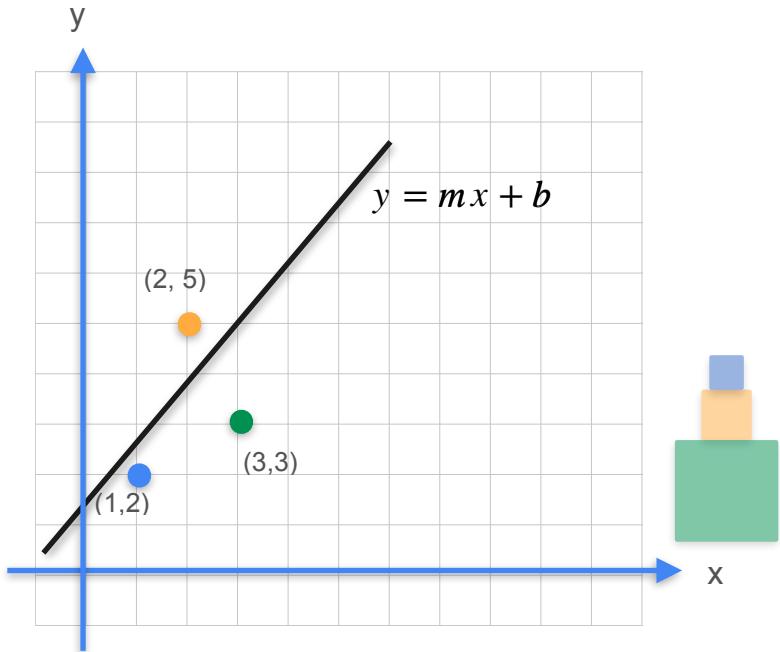
# Gradient Descent



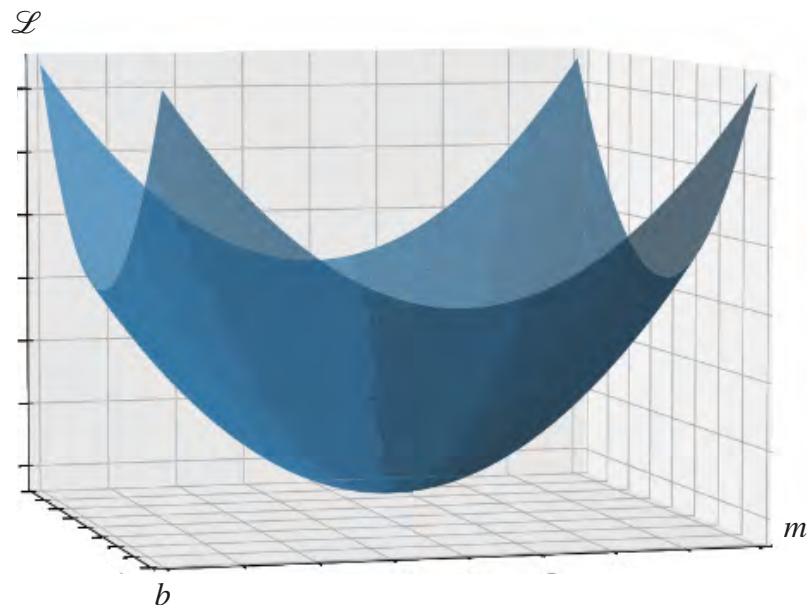
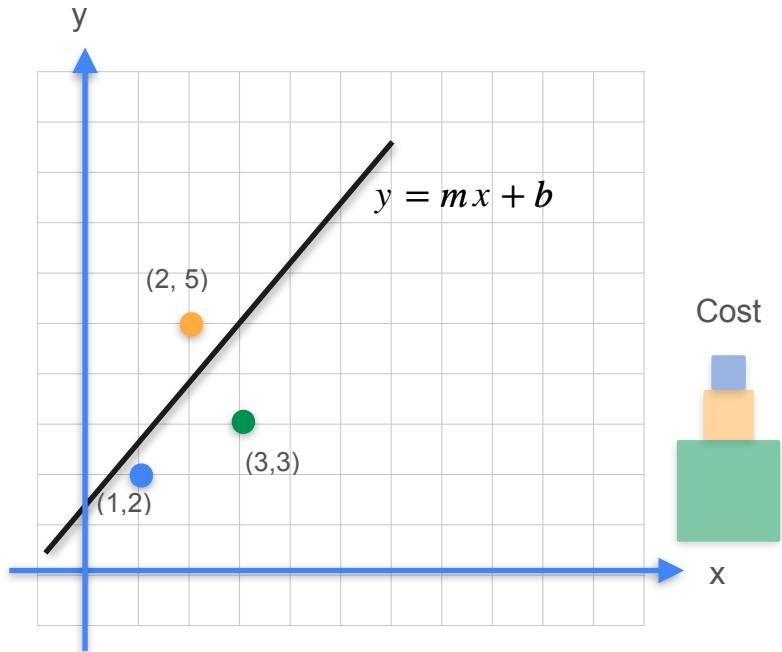
# Gradient Descent



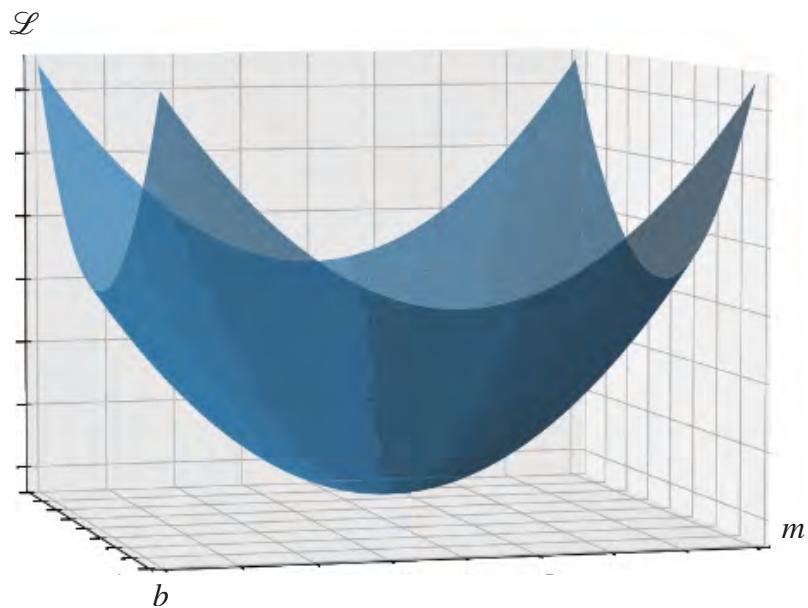
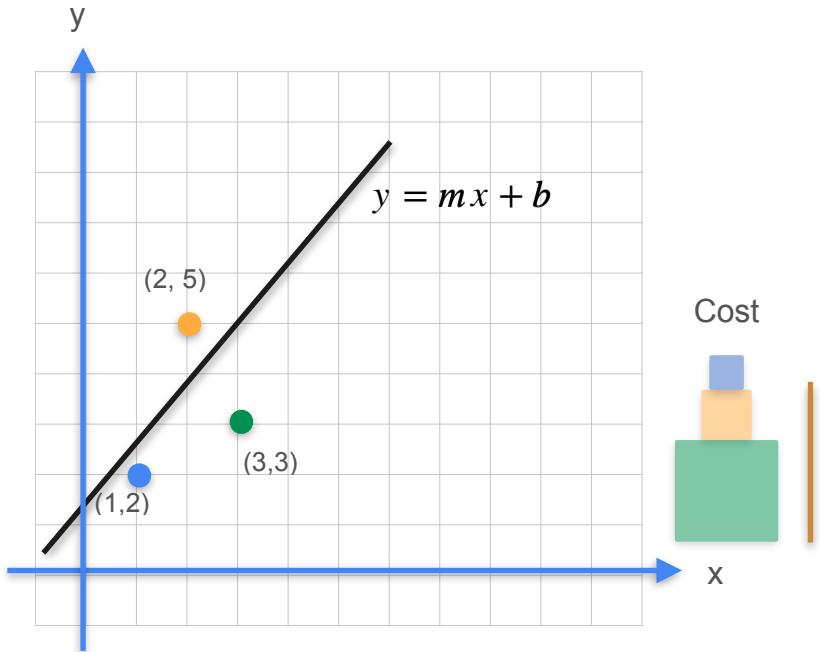
# Gradient Descent



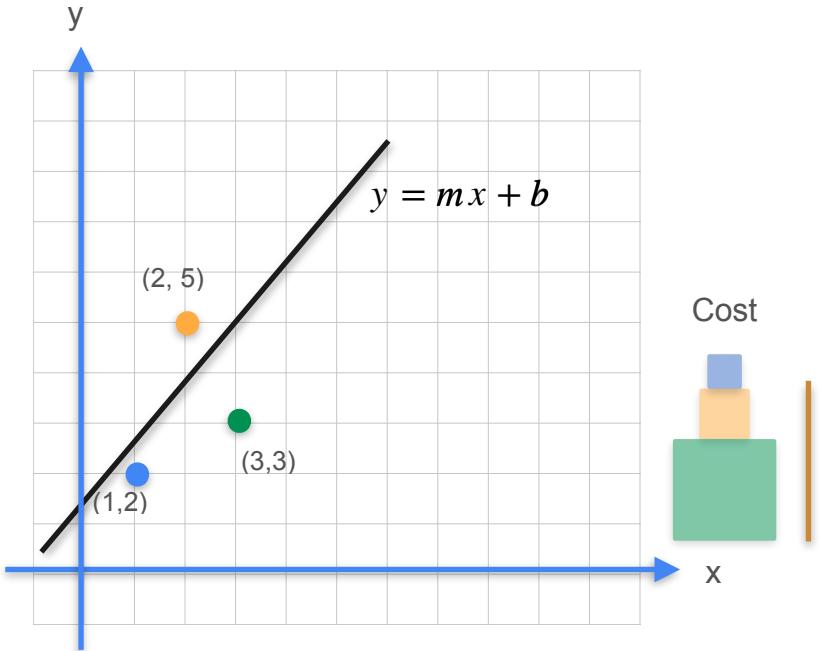
# Gradient Descent



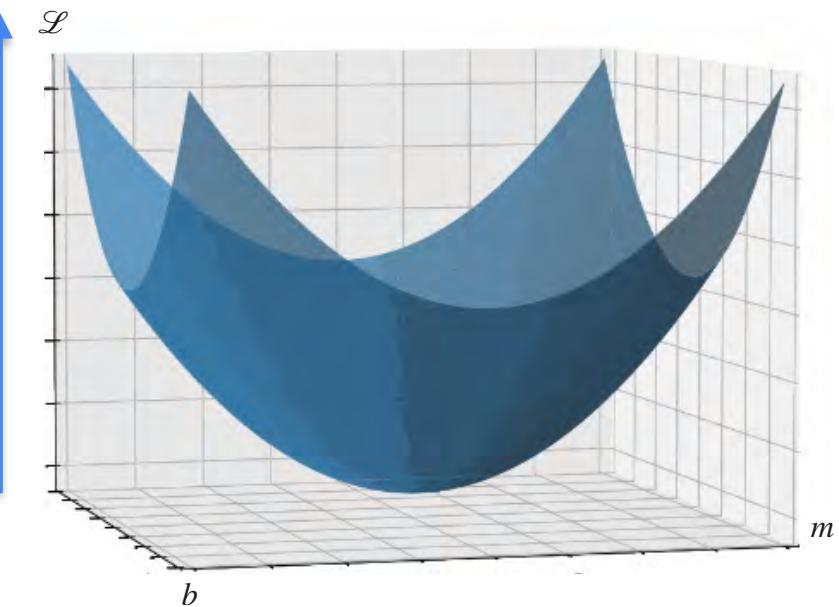
# Gradient Descent



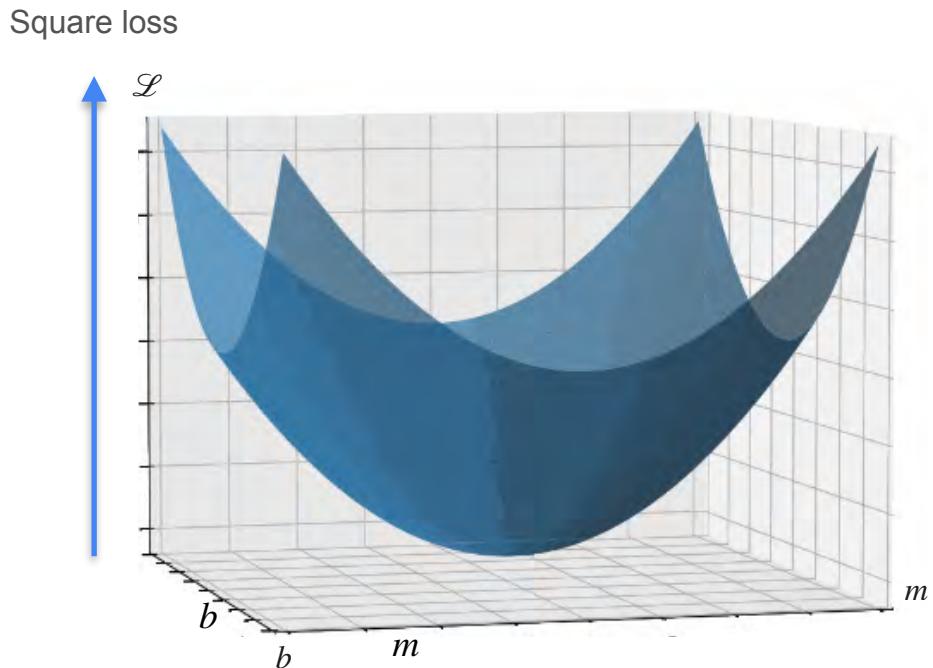
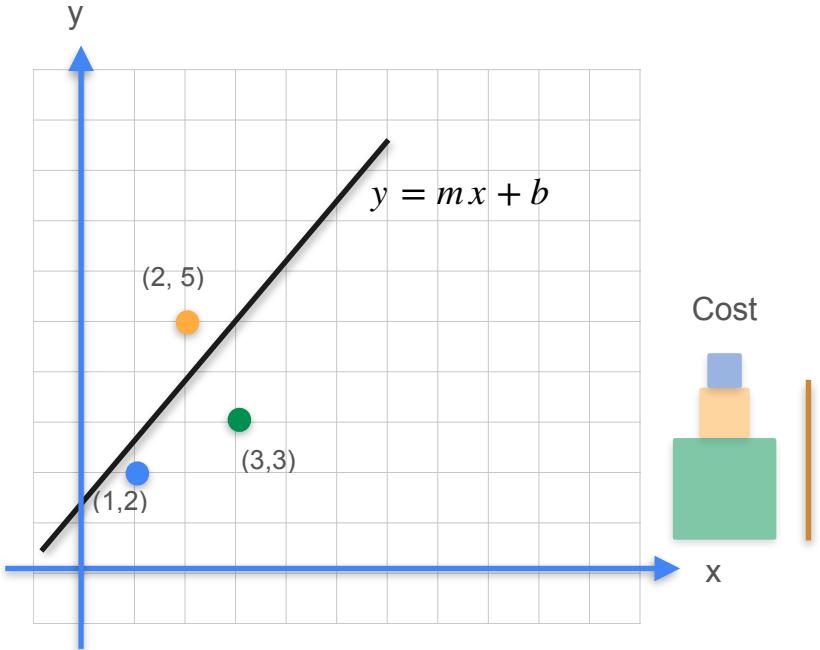
# Gradient Descent



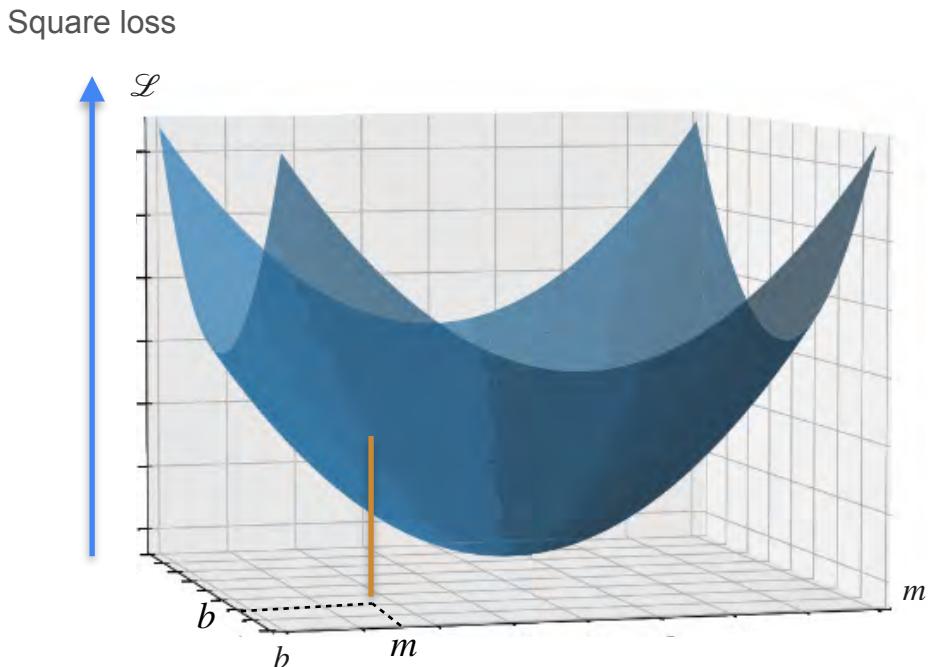
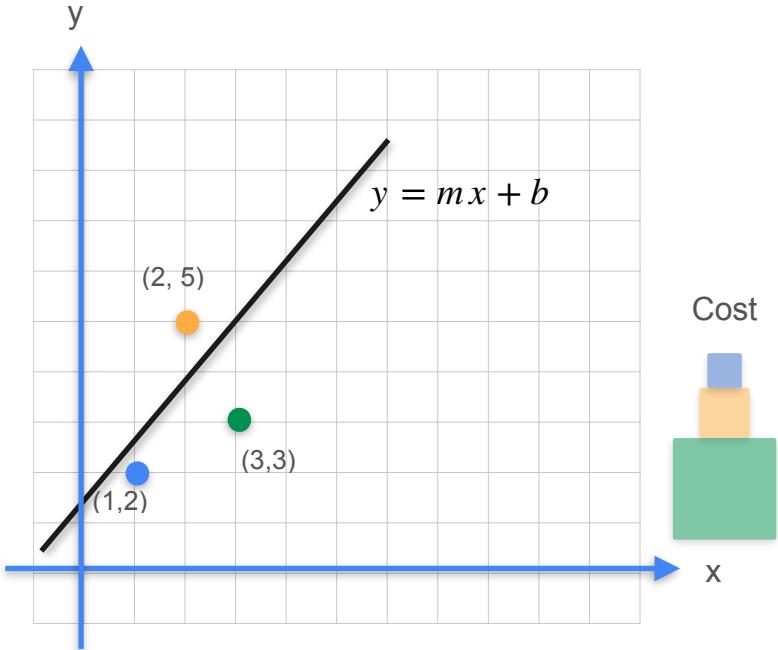
Square loss



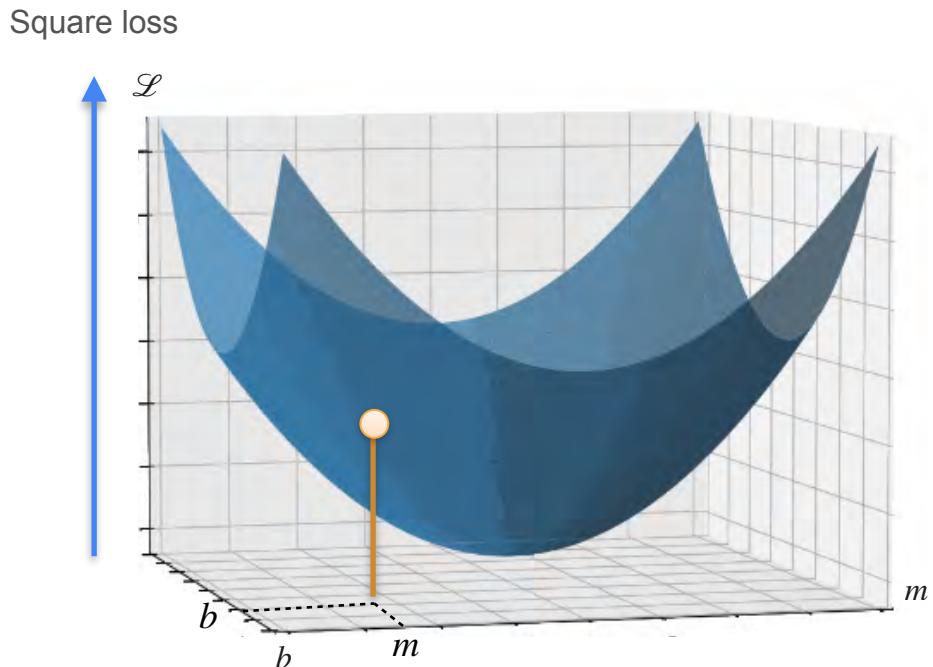
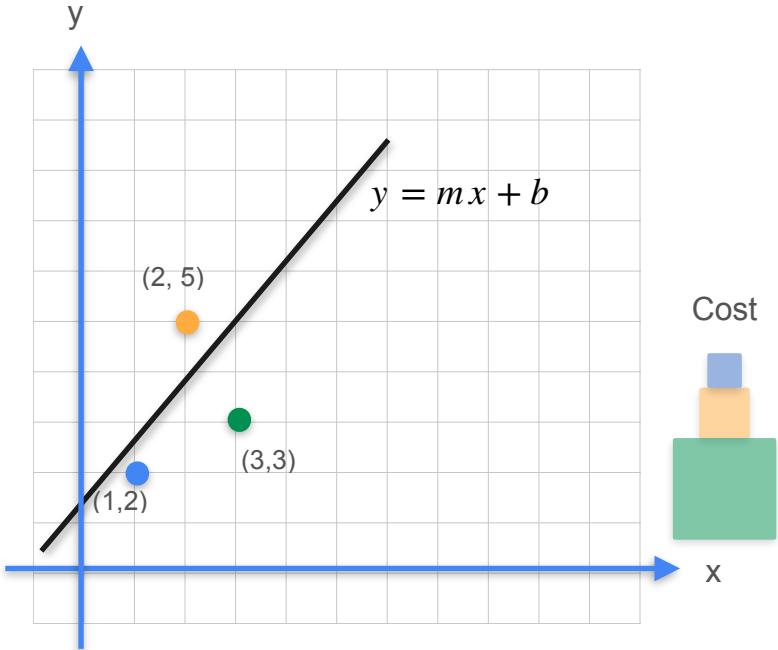
# Gradient Descent



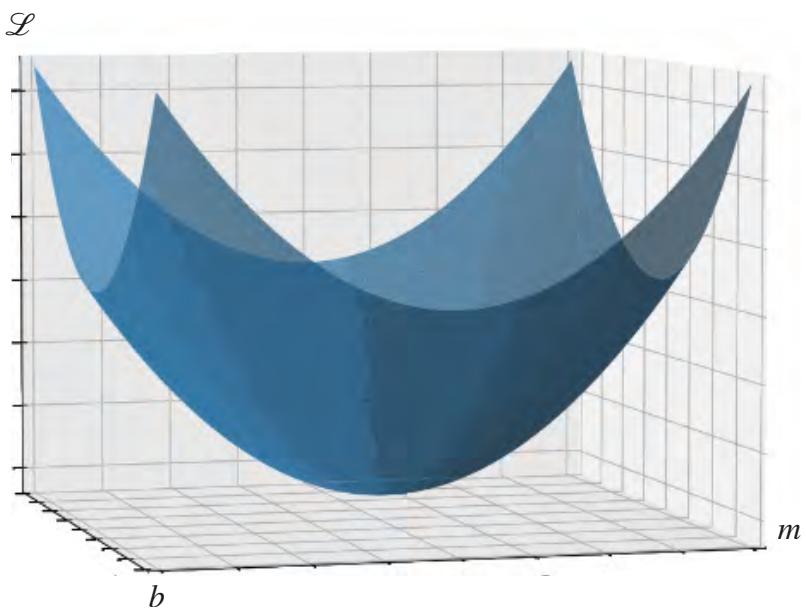
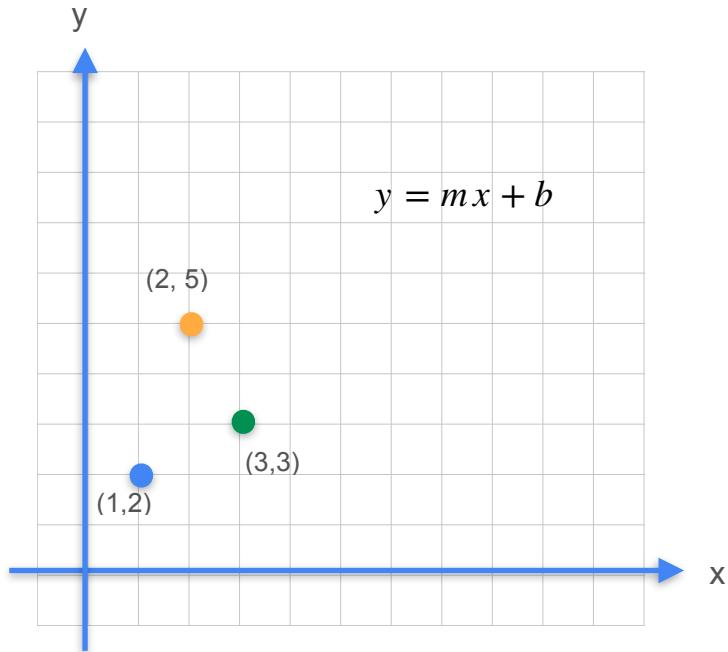
# Gradient Descent



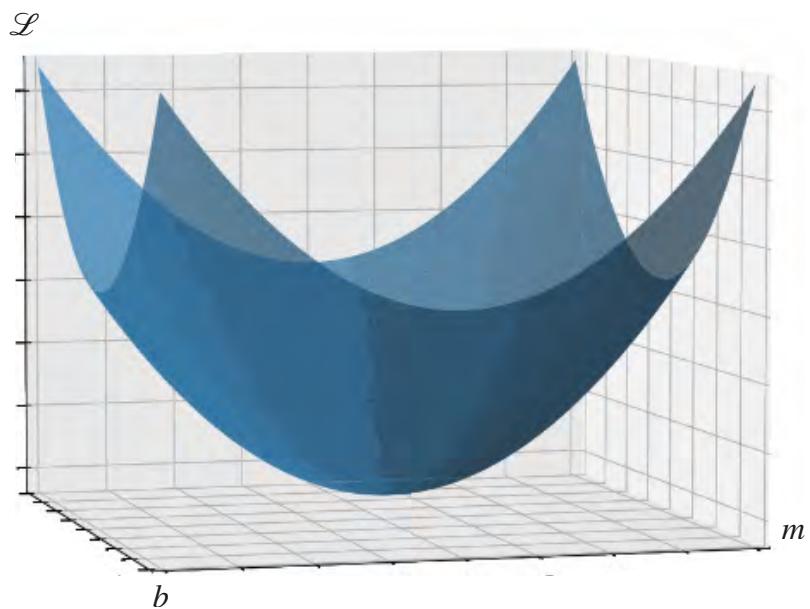
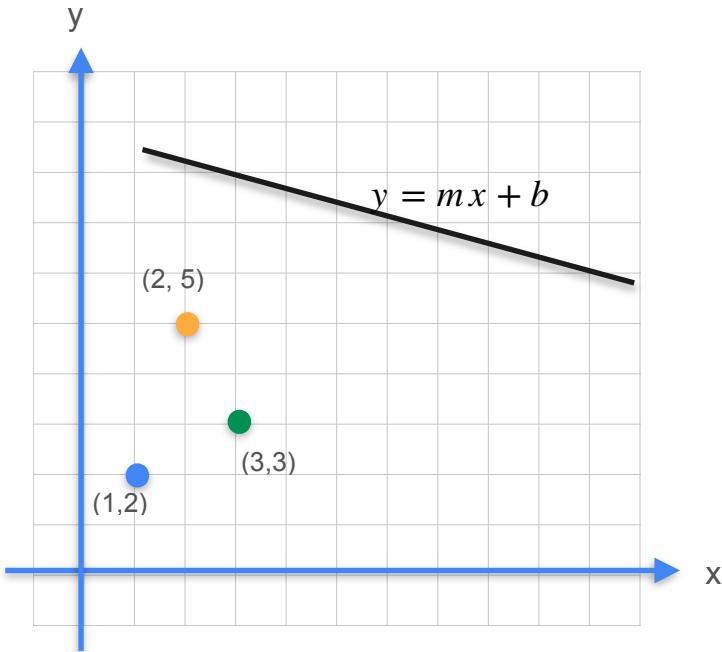
# Gradient Descent



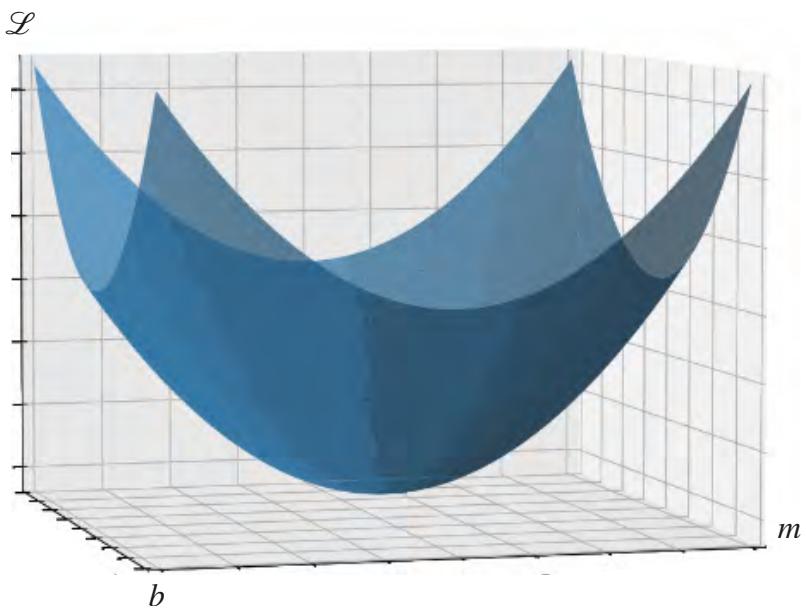
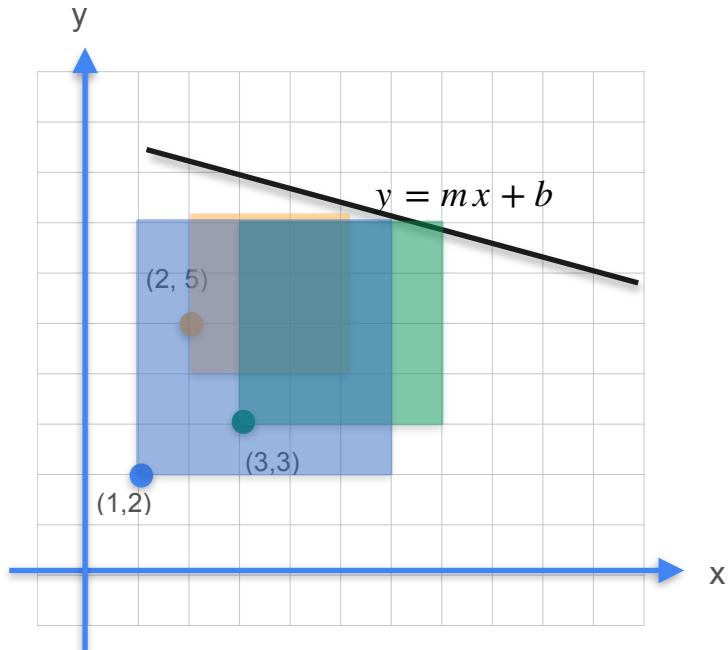
# Gradient Descent



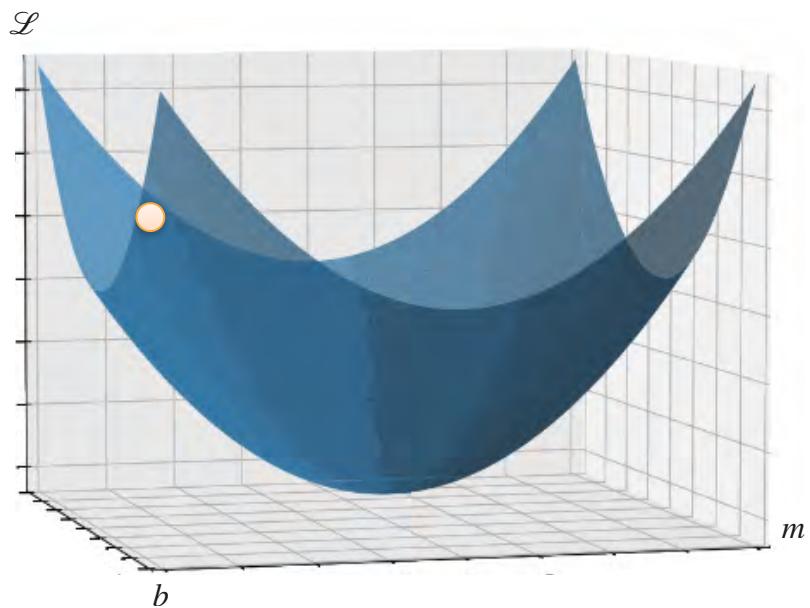
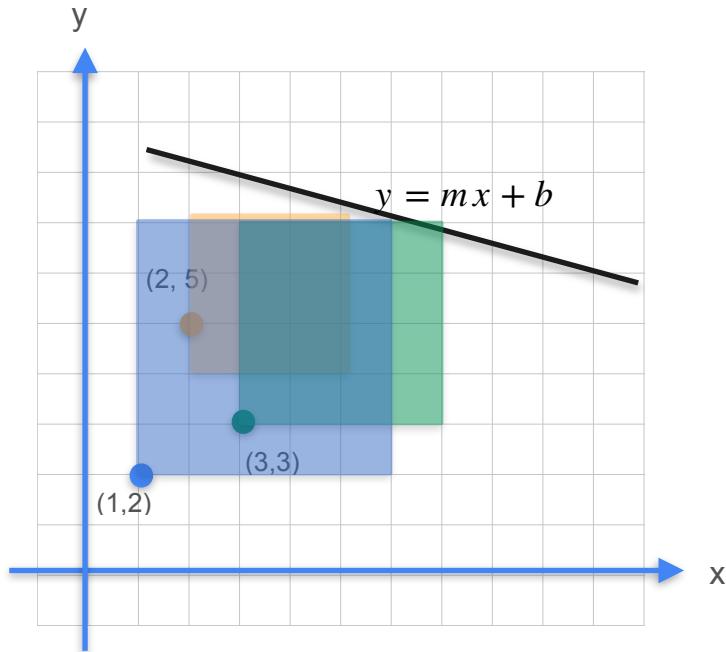
# Gradient Descent



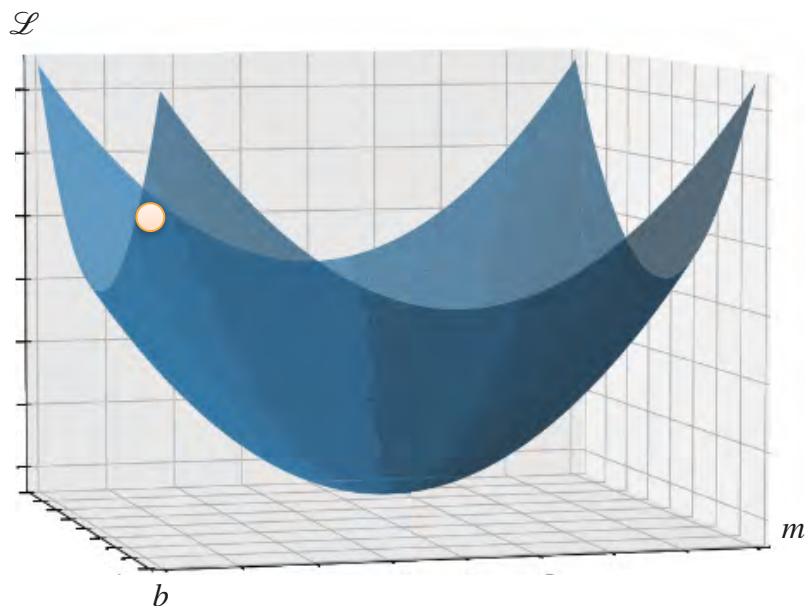
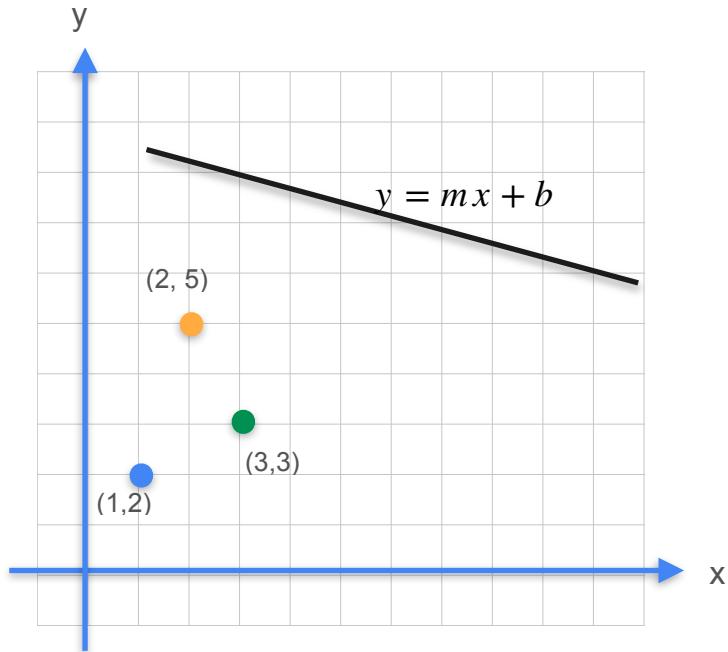
# Gradient Descent



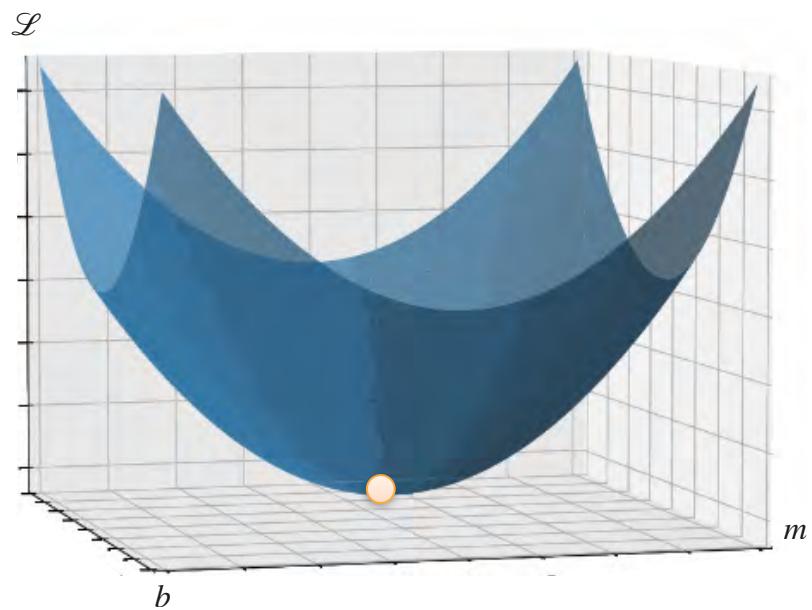
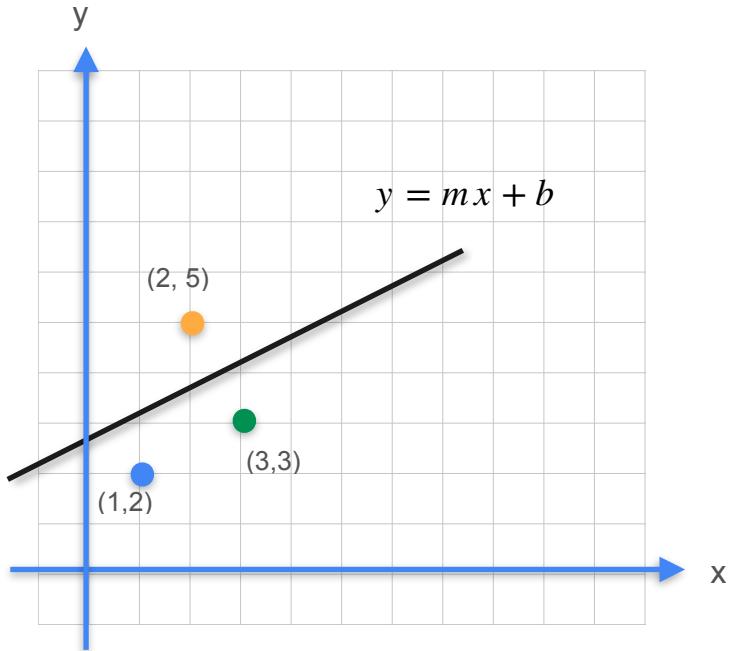
# Gradient Descent



# Gradient Descent



# Gradient Descent



# Another Example

# Another Example



# Another Example



TV advertisement  
budget

# Another Example



TV advertisement  
budget



# Another Example



TV advertisement  
budget



Number of sales

# Another Example

# Another Example

TV budget	Sales
-----------	-------

# Another Example

TV budget	Sales
230.1	22.1

# Another Example

TV budget	Sales
230.1	22.1
44.5	10.4

# Another Example

TV budget	Sales
230.1	22.1
44.5	10.4
17.2	9.3

# Another Example

TV budget	Sales
230.1	22.1
44.5	10.4
17.2	9.3

**Goal:** Predict sales in terms of TV budget

# Another Example

TV budget	Sales
230.1	22.1
44.5	10.4
17.2	9.3

**Goal:** Predict sales in terms of TV budget

**Tool:** Linear regression

# Another Example

TV budget	Sales
230.1	22.1
44.5	10.4
17.2	9.3

**Goal:** Predict sales in terms of TV budget

**Tool:** Linear regression

$$y = mx + b$$

# Another Example

TV budget	Sales
230.1	22.1
44.5	10.4
17.2	9.3

**Goal:** Predict sales in terms of TV budget

**Tool:** Linear regression

$$y = mx + b$$

# Another Example

TV budget	Sales
230.1	22.1
44.5	10.4
17.2	9.3



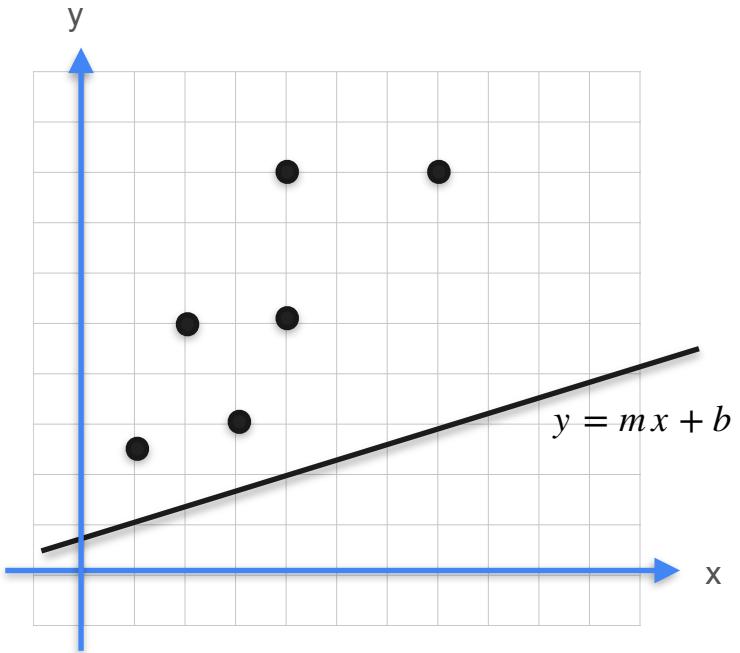
Multiple observations

**Goal:** Predict sales in terms of TV budget

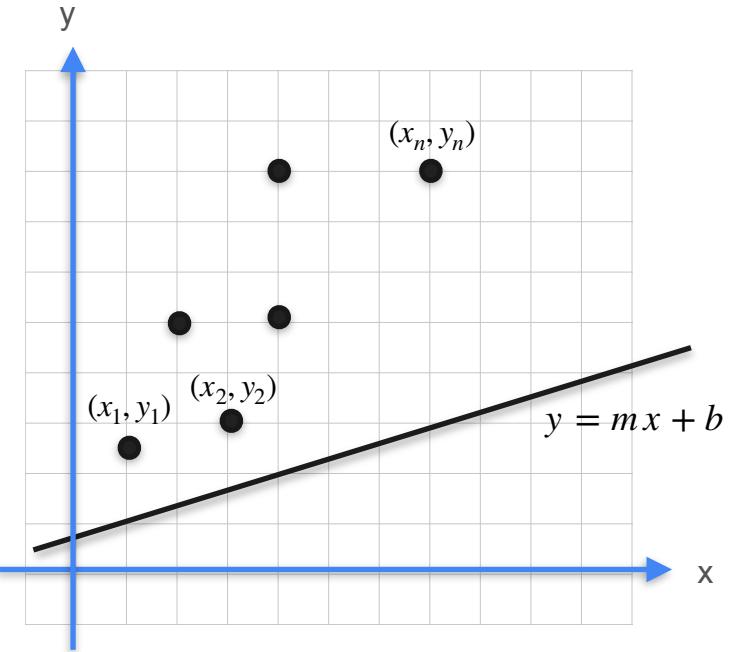
**Tool:** Linear regression

$$y = mx + b$$

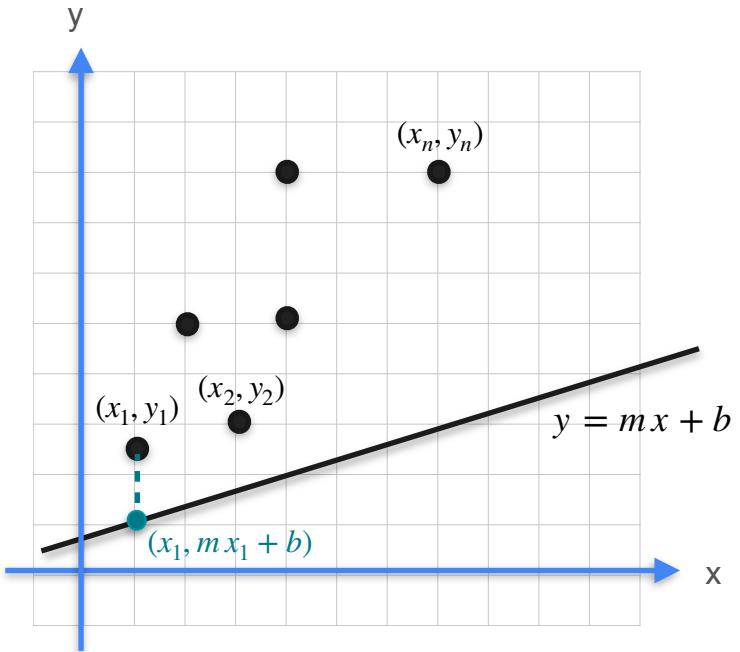
# Gradient Descent



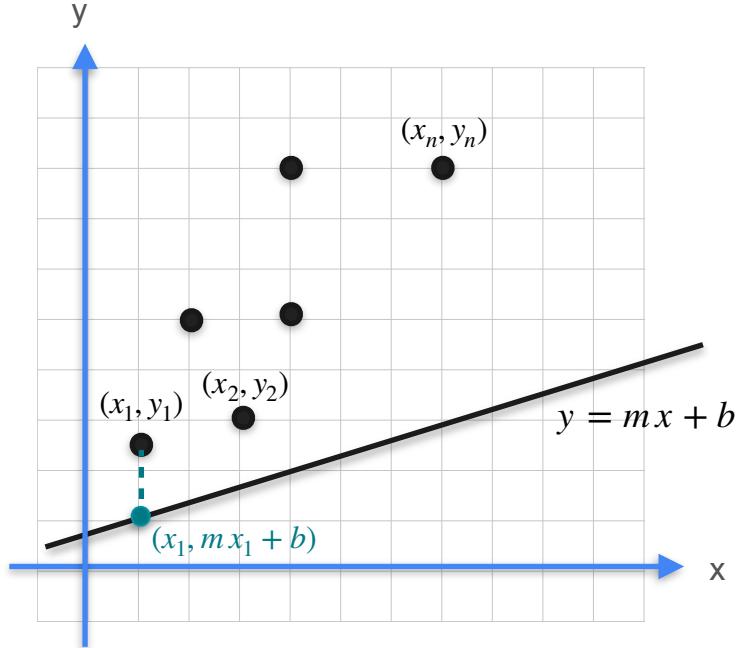
# Gradient Descent



# Gradient Descent

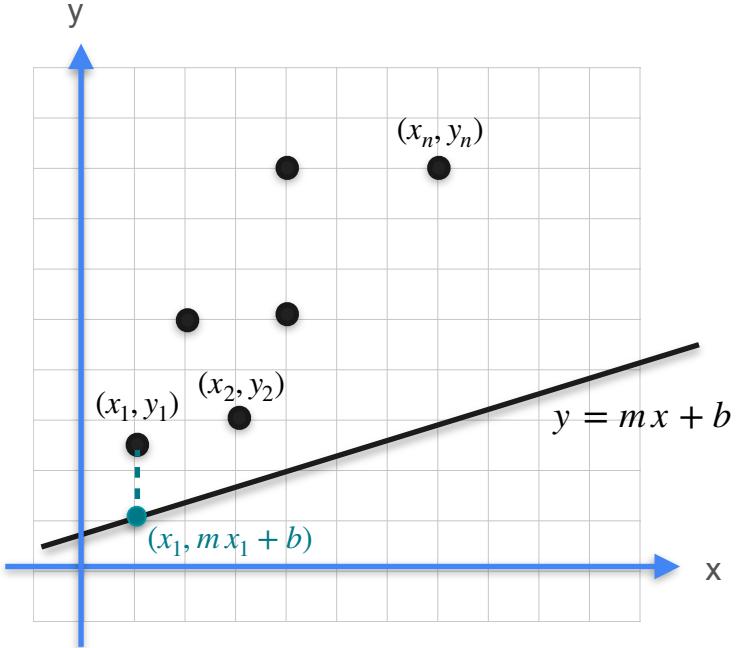


# Gradient Descent



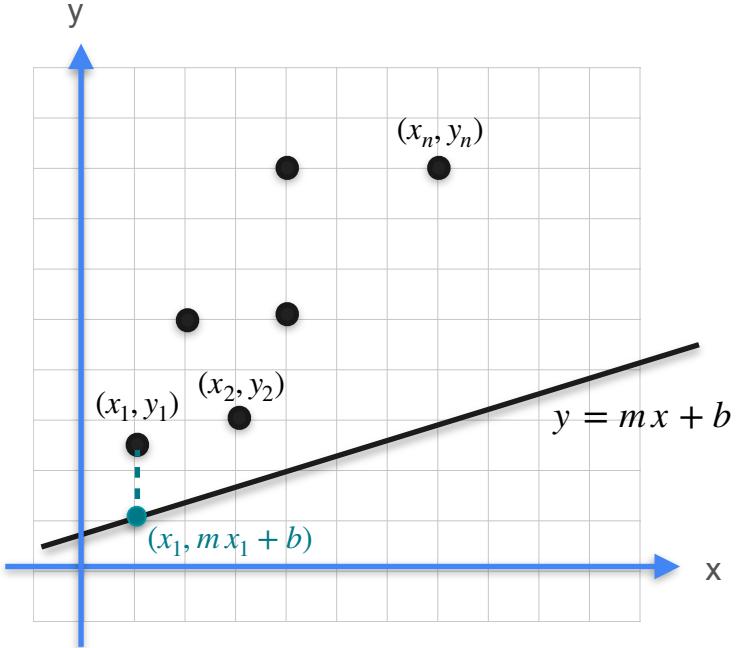
Loss  
↓  
 $mx_1 + b - y_1$

# Gradient Descent



Loss  
↓  
 $(mx_1 + b - y_1)^2$

# Gradient Descent



Cost

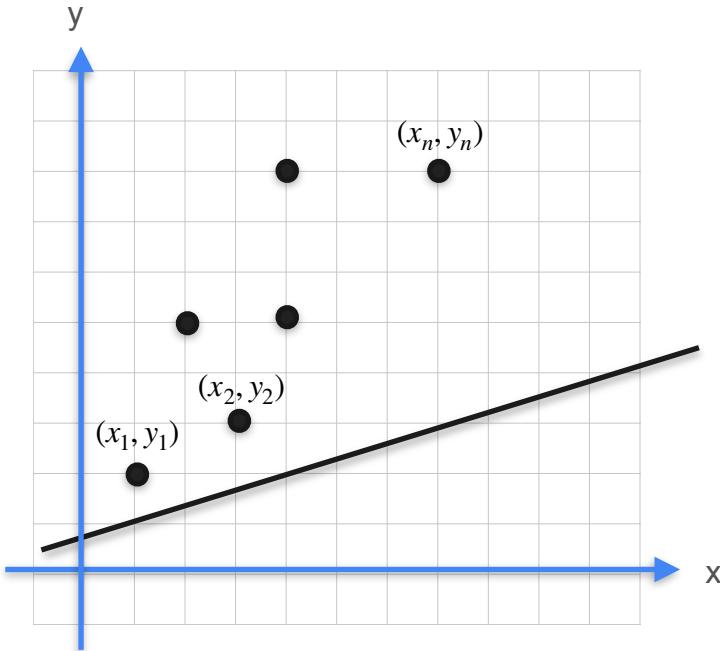


$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

Loss

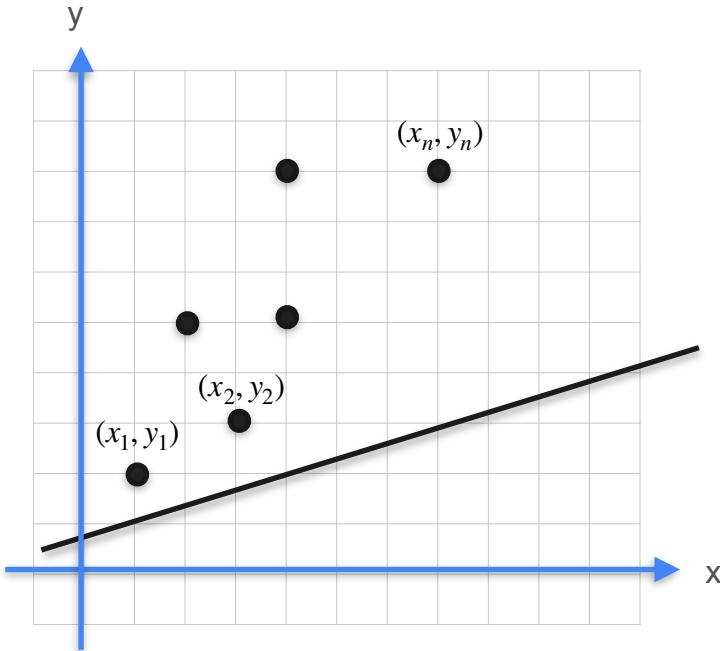


# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

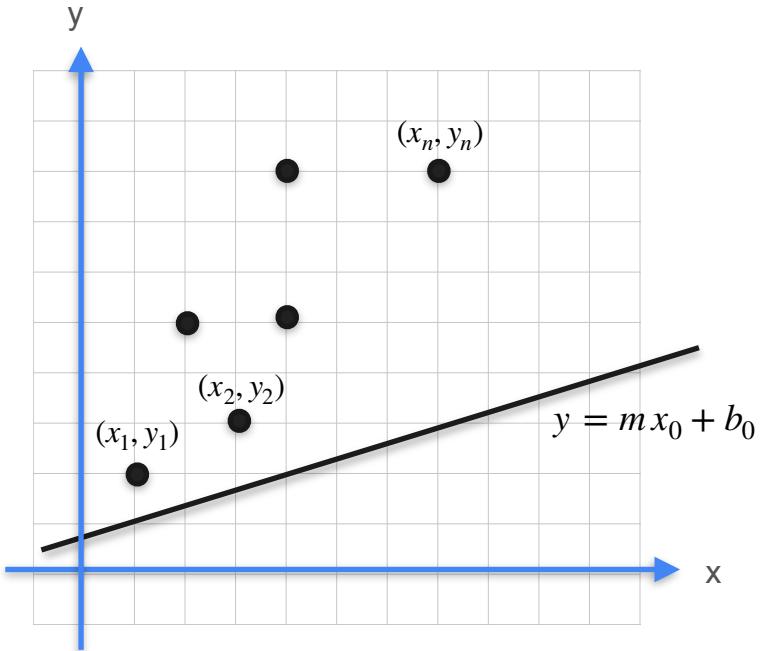
# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_0 \\ b_0 \end{bmatrix}$$

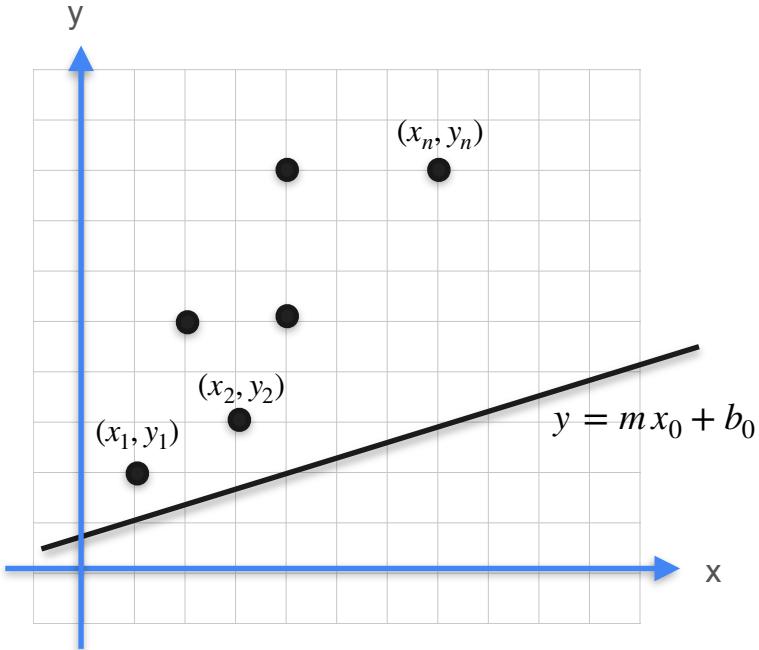
# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_0 \\ b_0 \end{bmatrix}$$

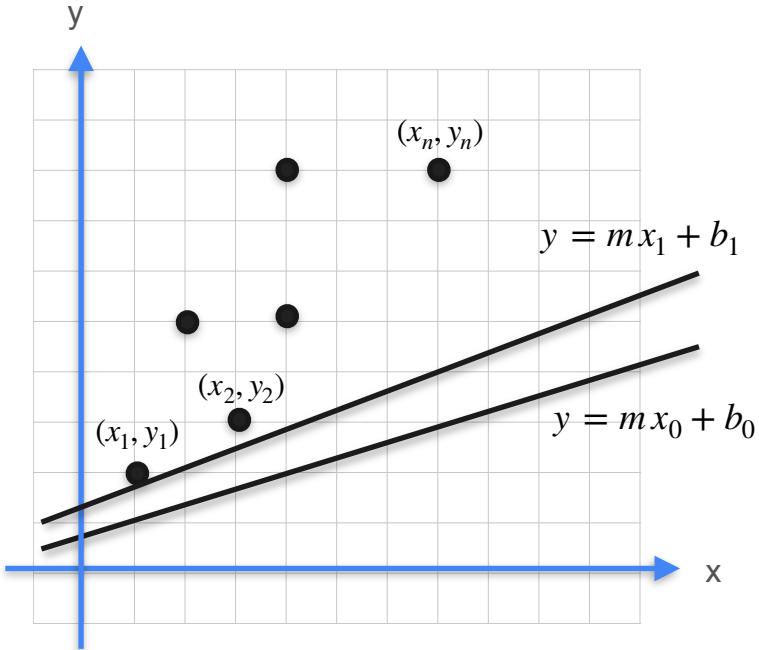
# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_0 \\ b_0 \end{bmatrix} \xrightarrow{\text{ }} \begin{bmatrix} m_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} m_0 \\ b_0 \end{bmatrix} - \alpha \nabla \mathcal{L}_1(m_0, b_0)$$

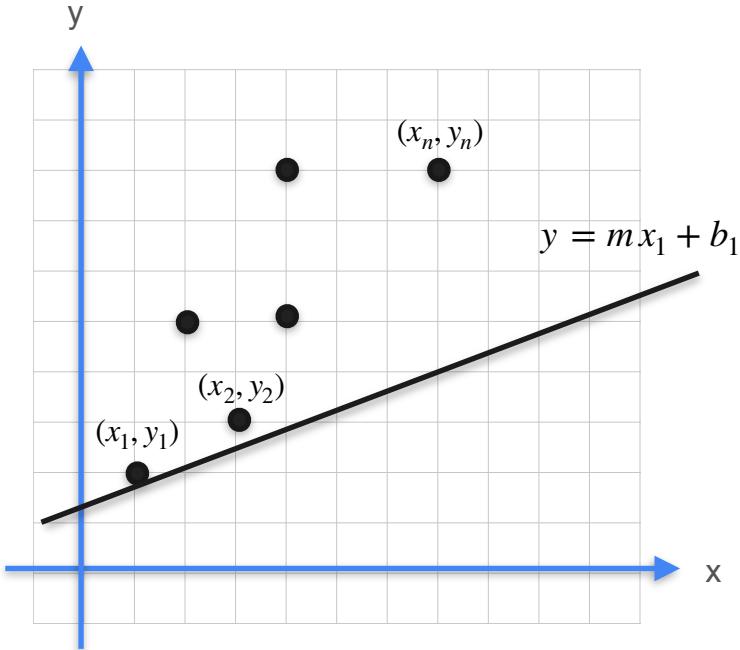
# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_0 \\ b_0 \end{bmatrix} \xrightarrow{\text{ }} \begin{bmatrix} m_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} m_0 \\ b_0 \end{bmatrix} - \alpha \nabla \mathcal{L}_1(m_0, b_0)$$

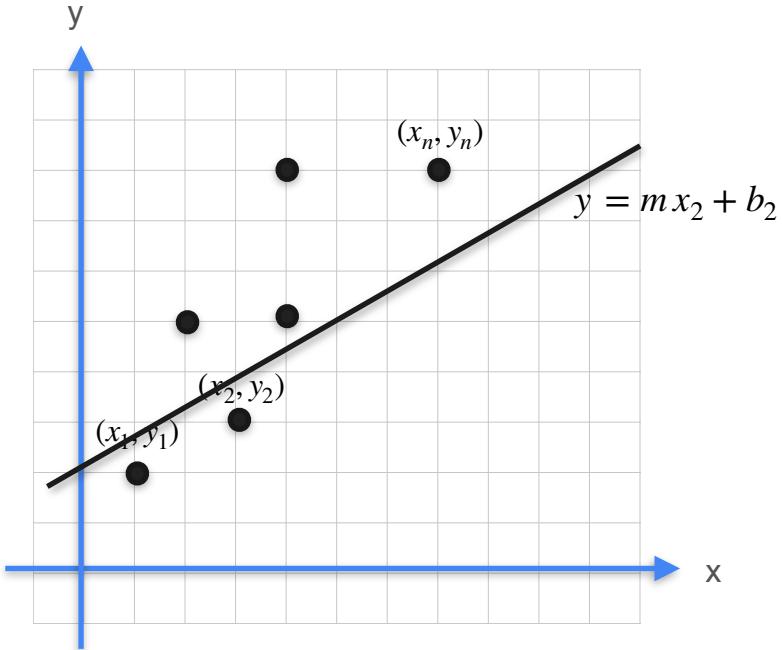
# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_0 \\ b_0 \end{bmatrix} \xrightarrow{\text{ }} \begin{bmatrix} m_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} m_0 \\ b_0 \end{bmatrix} - \alpha \nabla \mathcal{L}_1(m_0, b_0)$$

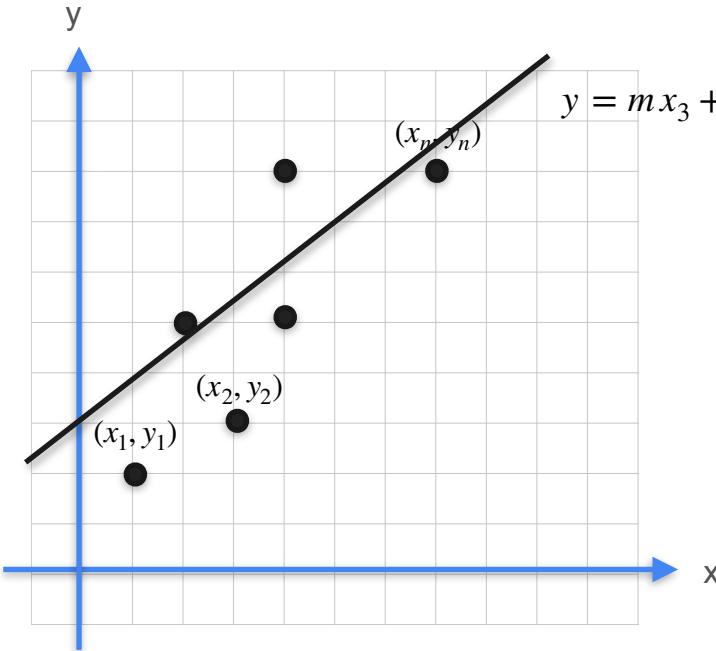
# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_1 \\ b_1 \end{bmatrix} \xrightarrow{\text{ }} \begin{bmatrix} m_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} m_1 \\ b_1 \end{bmatrix} - \alpha \nabla \mathcal{L}_1(m_1, b_1)$$

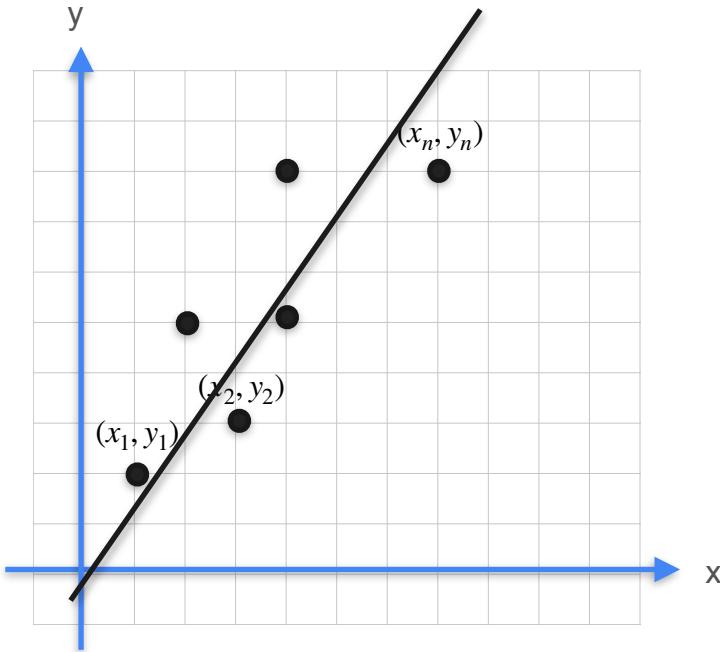
# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_2 \\ b_2 \end{bmatrix} \rightarrow \begin{bmatrix} m_3 \\ b_3 \end{bmatrix} = \begin{bmatrix} m_2 \\ b_2 \end{bmatrix} - \alpha \nabla \mathcal{L}_1(m_2, b_2)$$

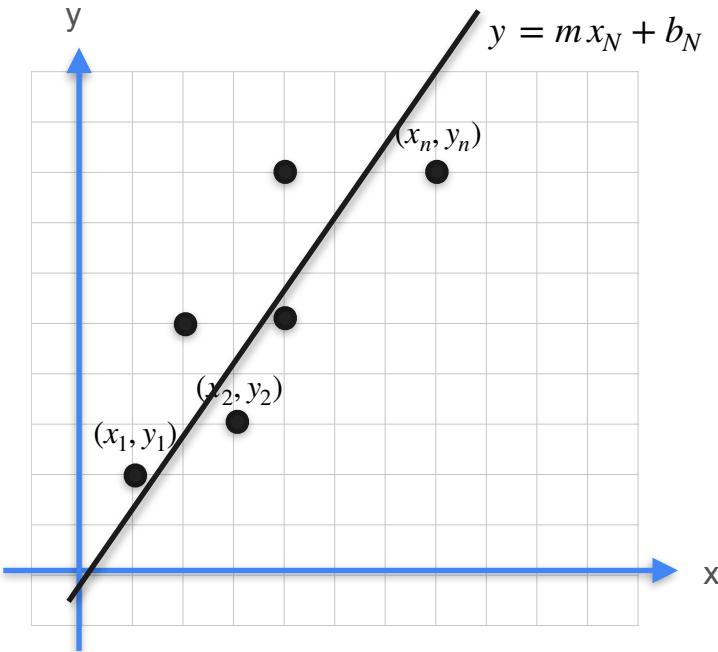
# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_N \\ b_N \end{bmatrix} \xrightarrow{\text{ }} \begin{bmatrix} m_N \\ b_N \end{bmatrix} = \begin{bmatrix} m_{N-1} \\ b_{N-1} \end{bmatrix} - \alpha \nabla \mathcal{L}_1(m_{N-1}, b_{N-1})$$

# Gradient Descent



$$\mathcal{L}(m, b) = \frac{1}{2m} [(mx_1 + b - y_1)^2 + \dots + (mx_n - y_n)^2]$$

$$\begin{bmatrix} m_N \\ b_N \end{bmatrix} \rightarrow \begin{bmatrix} m_N \\ b_N \end{bmatrix} = \begin{bmatrix} m_{N-1} \\ b_{N-1} \end{bmatrix} - \alpha \nabla \mathcal{L}_1(m_{N-1}, b_{N-1})$$



DeepLearning.AI

# Gradients and Gradient Descent

---

## Conclusion