

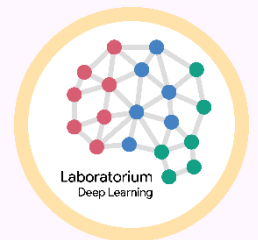
# Gradient dan Gradient Descent

*Dennis A. Christie*



Machine  
Learning  
Course

22/02/2019

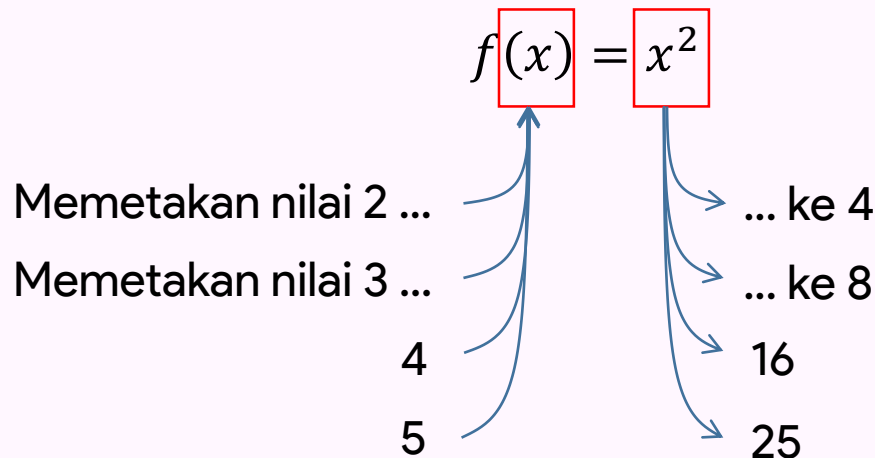




# Gradient Descent

Apa dan Bagaimana

- Cost Function
- **Function**, memetakan nilai dari A ke B



- **Cost Function**, fungsi yang memberitahukan seberapa buruk prediksi yang kita lakukan.



Cost function

$$C(w)$$

Single input

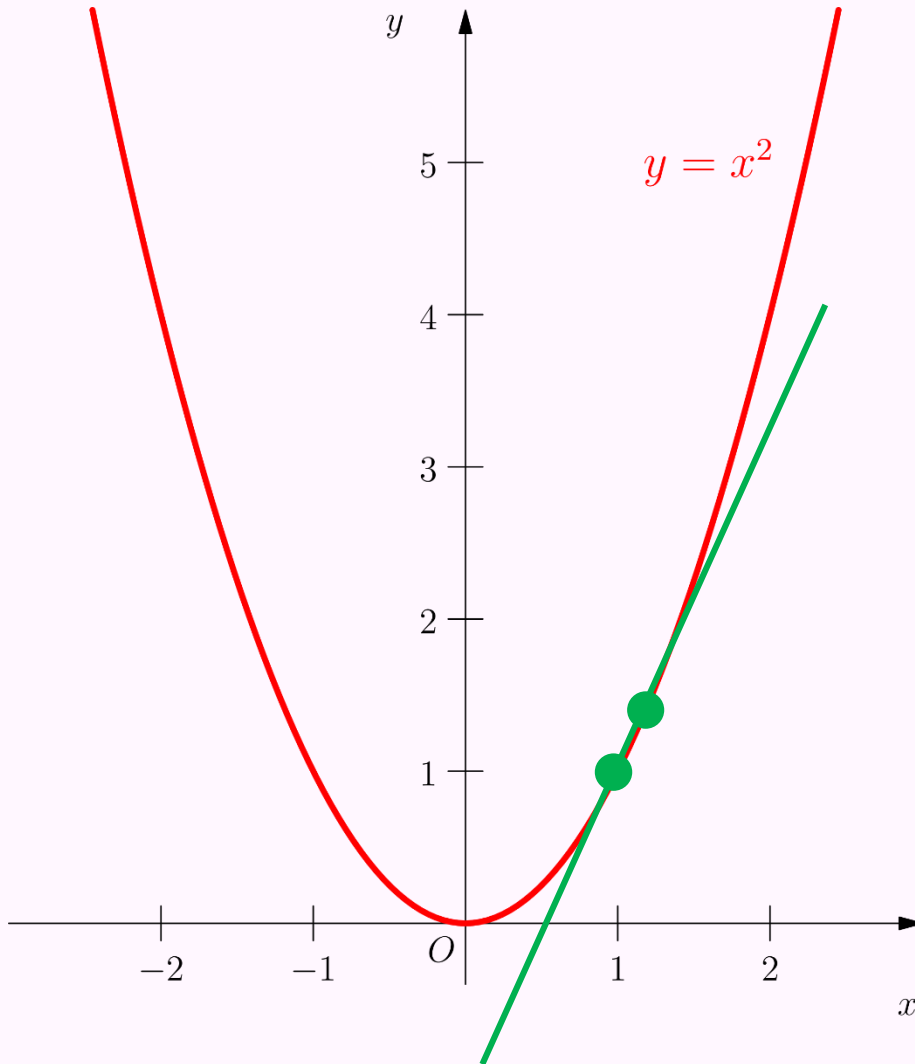


- Gradient merupakan *first order derivative*.
- Dalam bahasa indonesia, kita menyebutnya **turunan pertama**.
- $f(x) = x^2$
- $f'(x) = 2x$
- Informasi apa yang kita dapatkan dari **turunan pertama** ini?

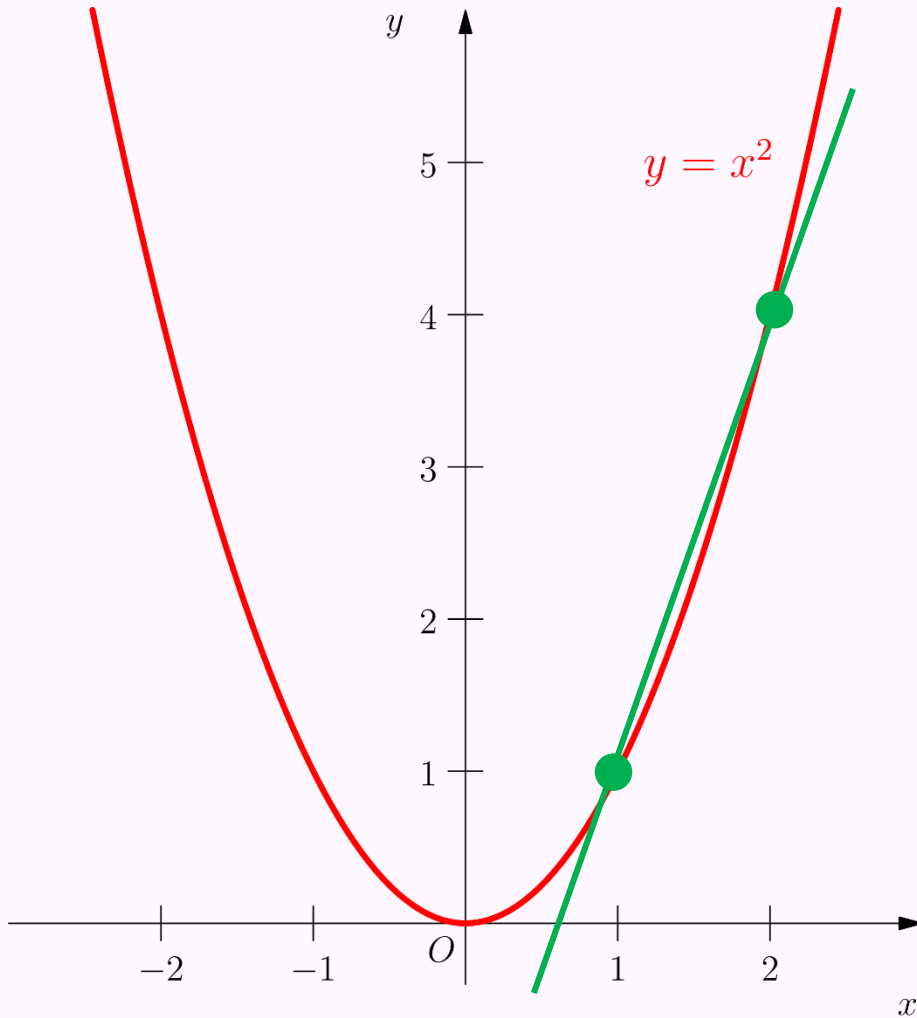


- Formula turunan pertama

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{h}$$



- $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$
- Ambil salah satu poin.
- Ambil poin lainnya.
  - menggeser poin sebelumnya dengan nilai  $h$  yang kecil.
- Selisihkan dan bagi dengan rasio perubahannya.



$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

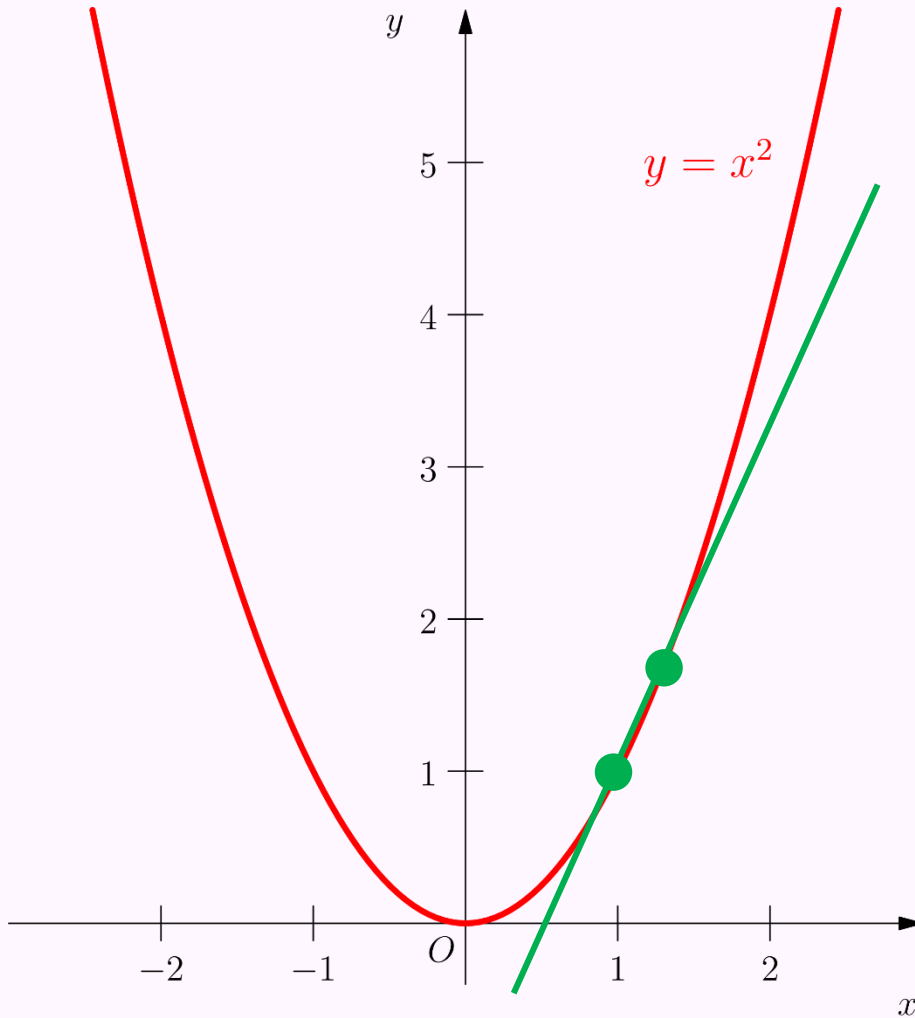
Cara Formulasi

$$f'(x) = 2x$$

Cara Cepat

$$f'(1) = \frac{f(1+1) - f(1)}{1} = \frac{f(2) - f(1)}{1} = \frac{4 - 1}{1} = 3$$





$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Cara Formulasi

$$f'(x) = 2x$$

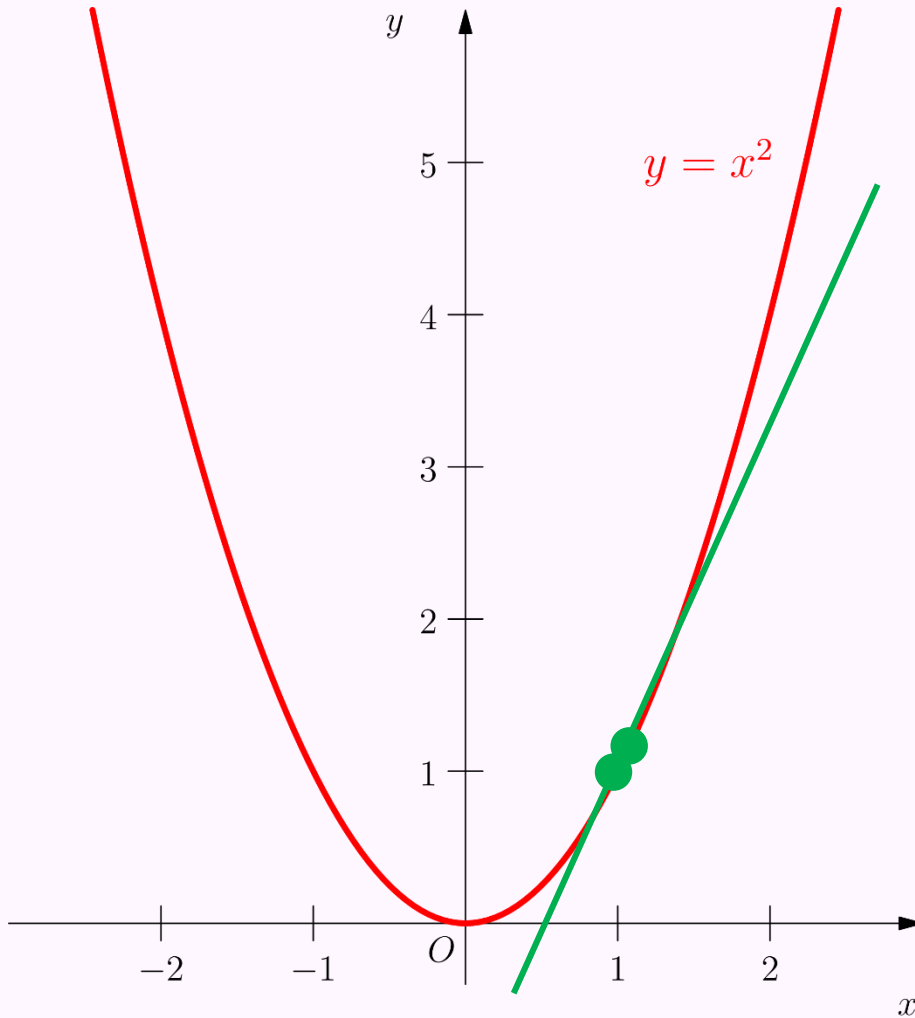
Cara Cepat

$$f'(1) = \frac{f(1+1) - f(1)}{1} = \frac{f(2) - f(1)}{1} = \frac{4 - 1}{1} = 3$$

$$f'(1) = \frac{f(1+0.1) - f(1)}{0.1} = \frac{f(1.1) - f(1)}{0.1} = \frac{1.21 - 1}{0.1} = 2.1$$

Semakin  $h$  mendekati 0

$$f'(1) = 2.00000 \dots 1$$



$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Cara Formulasi

$$f'(x) = 2x$$

Cara Cepat

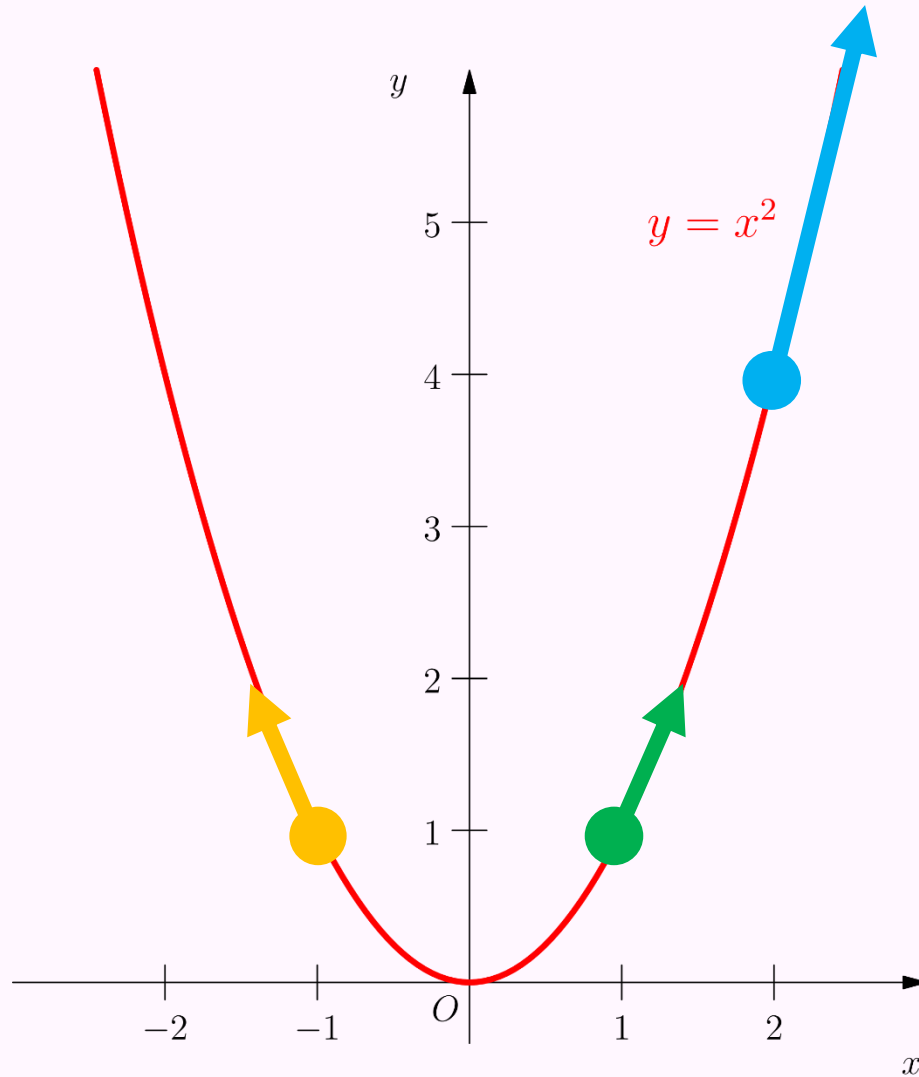
$$f'(1) = \frac{f(1+1) - f(1)}{1} = \frac{f(2) - f(1)}{1} = \frac{4 - 1}{1} = 3$$

$$f'(1) = \frac{f(1+0.1) - f(1)}{0.1} = \frac{f(1.1) - f(1)}{0.1} = \frac{1.21 - 1}{0.1} = 2.1$$

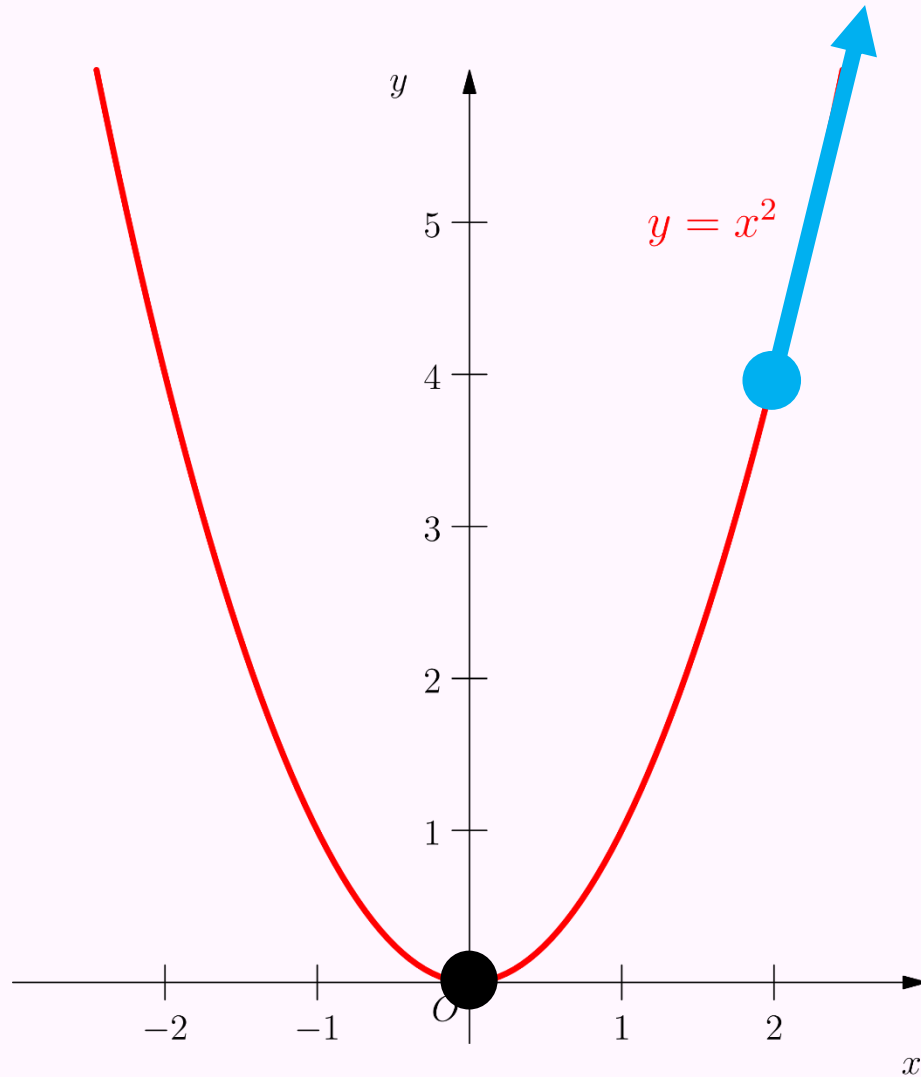
$$f'(1) = \frac{f(1+0.01) - f(1)}{0.01} = \frac{f(1.01) - f(1)}{0.01} = \frac{1.0201 - 1}{0.01} = 2.01$$

Semakin  $h$  mendekati 0

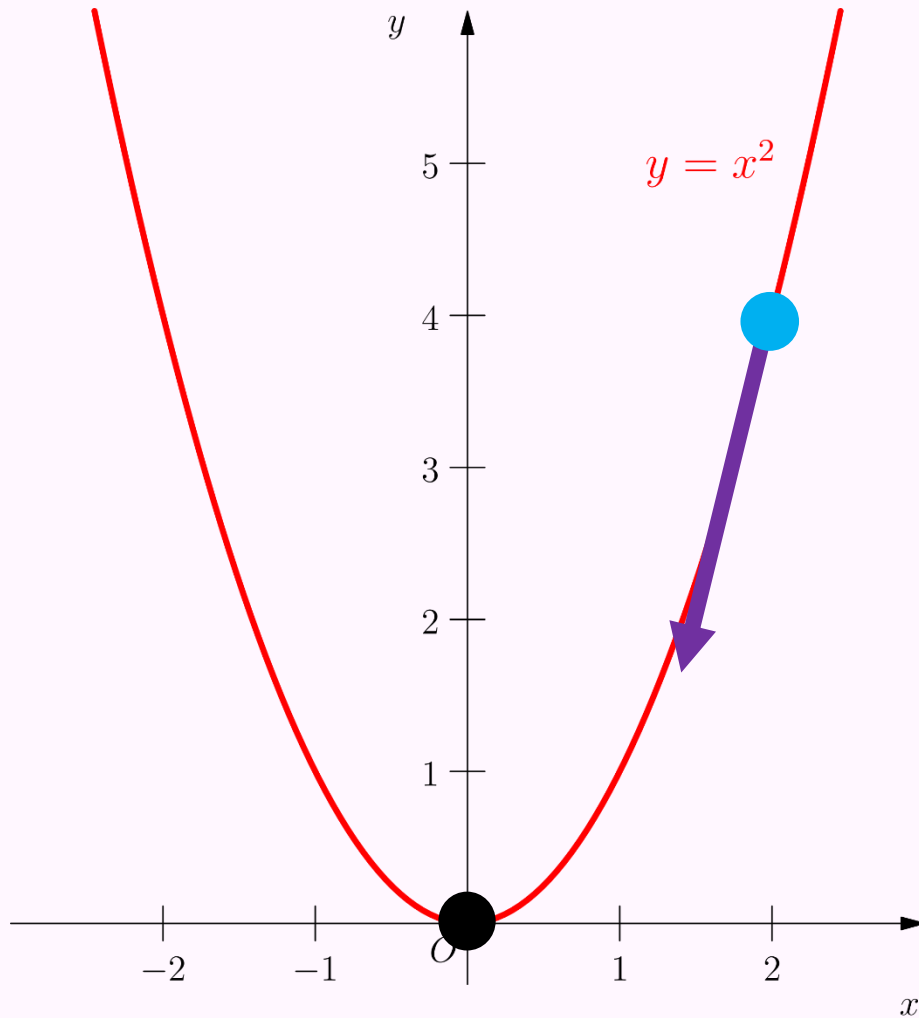
$$f'(1) = 2.00000 \dots 1$$



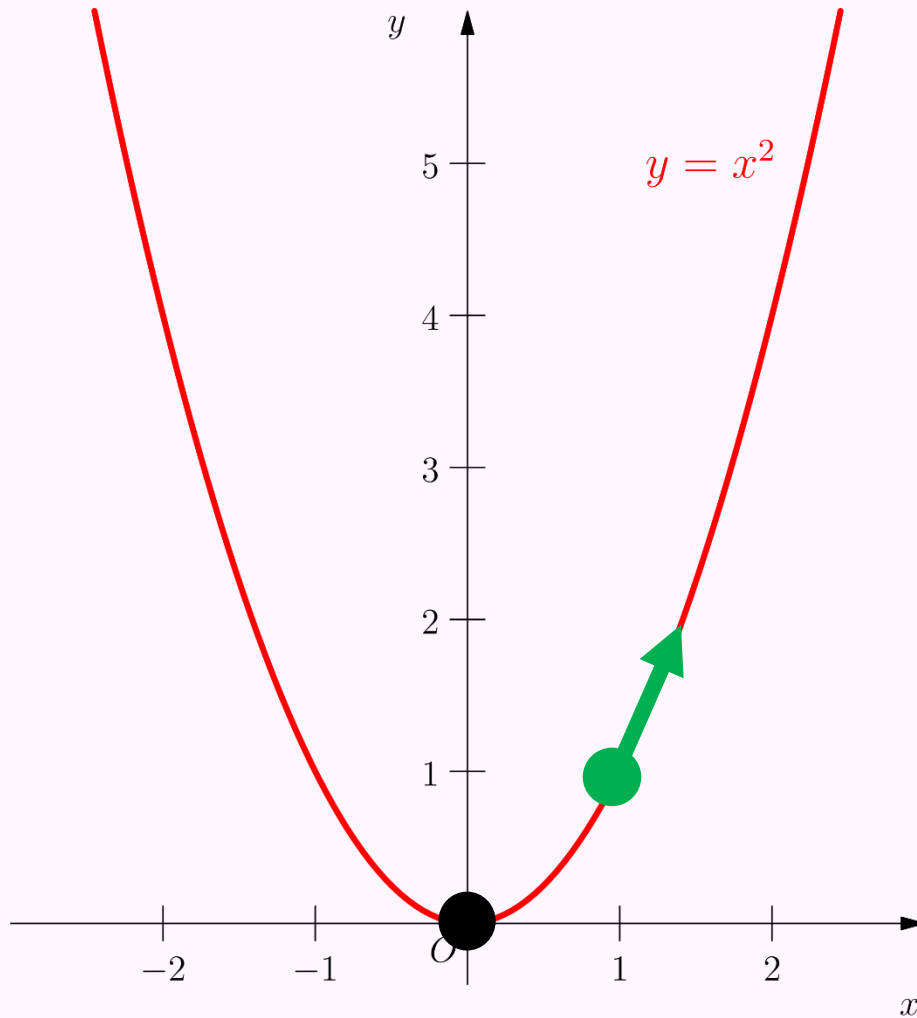
- Tapi, apa maksud  $f'(x) = 2x$  ?
- $f'(1) = 2$
- $f'(2) = 4$
- $f'(-1) = -2$
- **Tanda nilai turunan** (positif atau negatif) menunjukkan arah untuk bergerak ke arah atas.
- **Besar nilai turunan** menunjukkan seberapa cepat perubahannya.
- Nilai turunan = gradient.



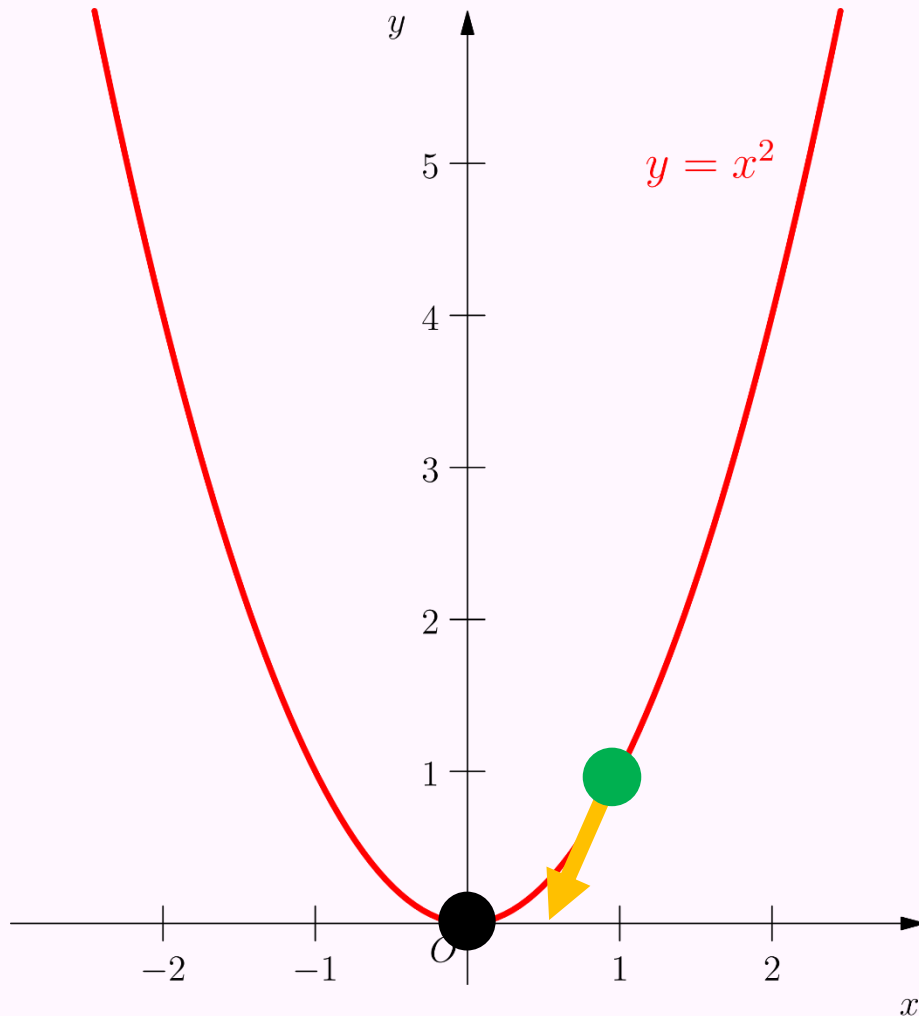
- Nilai  $f'(2)$  masih relatif besar, tanda bahwa masih jauh dari minima.



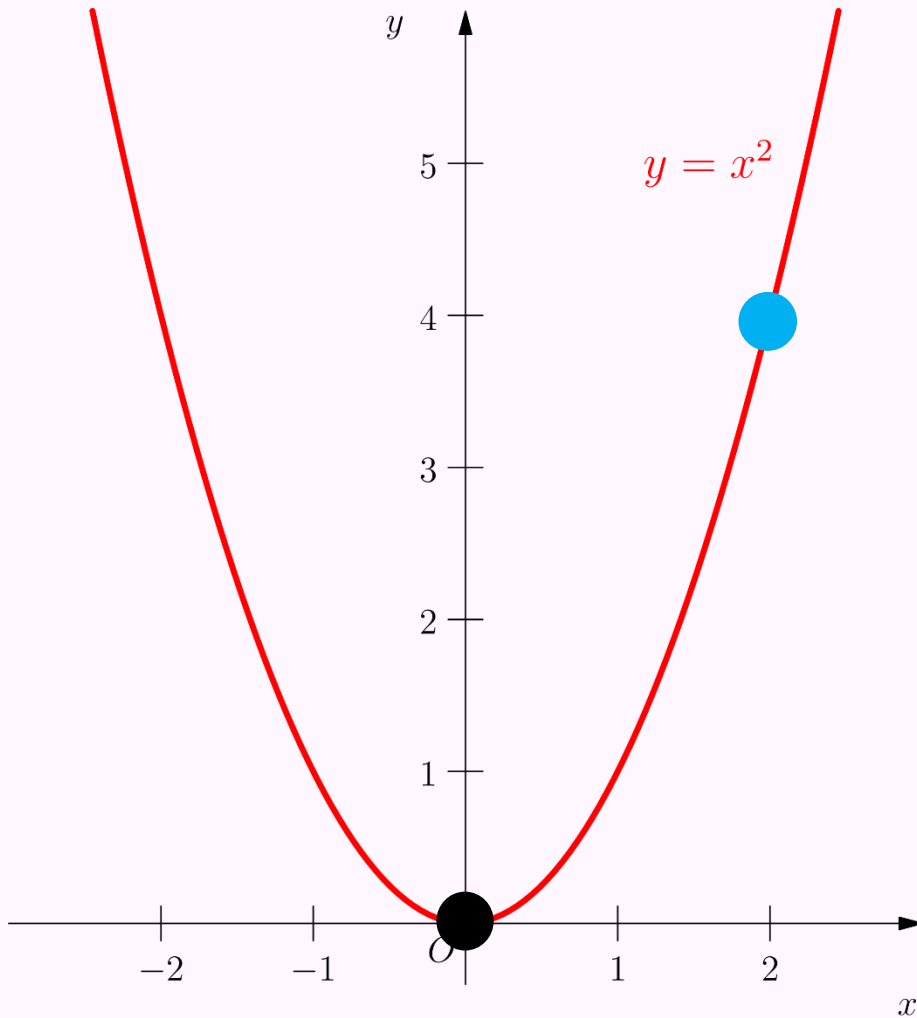
- Nilai  $f'(2)$  masih relatif besar, tanda bahwa masih jauh dari minima.
- Kemana arah menuju titik terendah?
  - Negatif dari gradient!
  - $f'(2) = 4$  (kanan).
  - $-f'(2) = -4$  (kiri).



- Nilai  $f'(2)$  masih relatif besar, tanda bahwa masih jauh dari minima.
- Kemana arah menuju titik terendah?
  - Negatif dari gradient!
  - $f'(2) = 4$  (kanan).
  - $-f'(2) = -4$  (kiri).
- Nilai  $f'(1)$  sudah cukup kecil, tanda bahwa sudah dekat dengan minima.



- Nilai  $f'(2)$  masih relatif besar, tanda bahwa masih jauh dari minima.
- Kemana arah menuju titik terendah?
  - Negatif dari gradient!
  - $f'(2) = 4$  (kanan).
  - $-f'(2) = -4$  (kiri).
- Nilai  $f'(1)$  sudah cukup kecil, tanda bahwa sudah dekat dengan minima.
- Arah menuju titik terendah?



- Formula gradient descent.

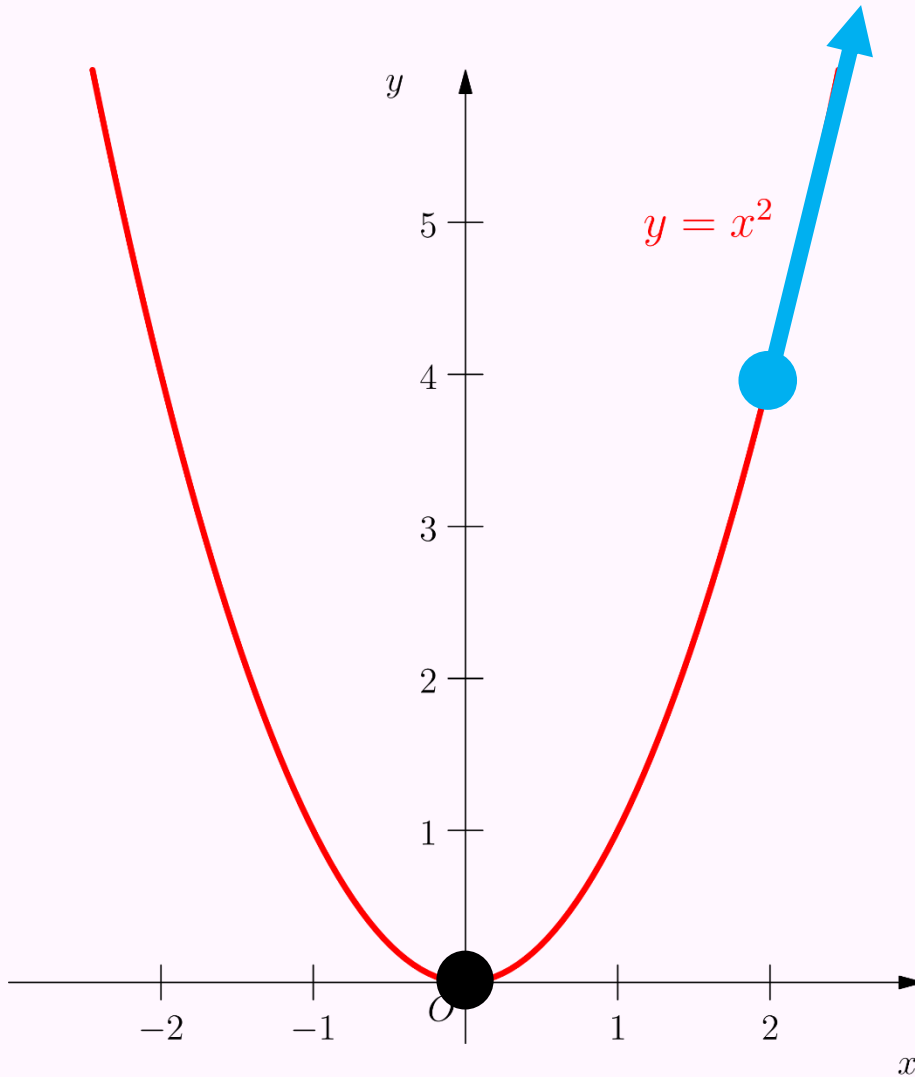
$$x_t = x_{t-1} - \alpha f'(\boxed{x_{t-1}})$$

Lokasi poin  
sebelumnya

- Formula dengan notasi yang baik.

$$x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$$





- Formula gradient descent.

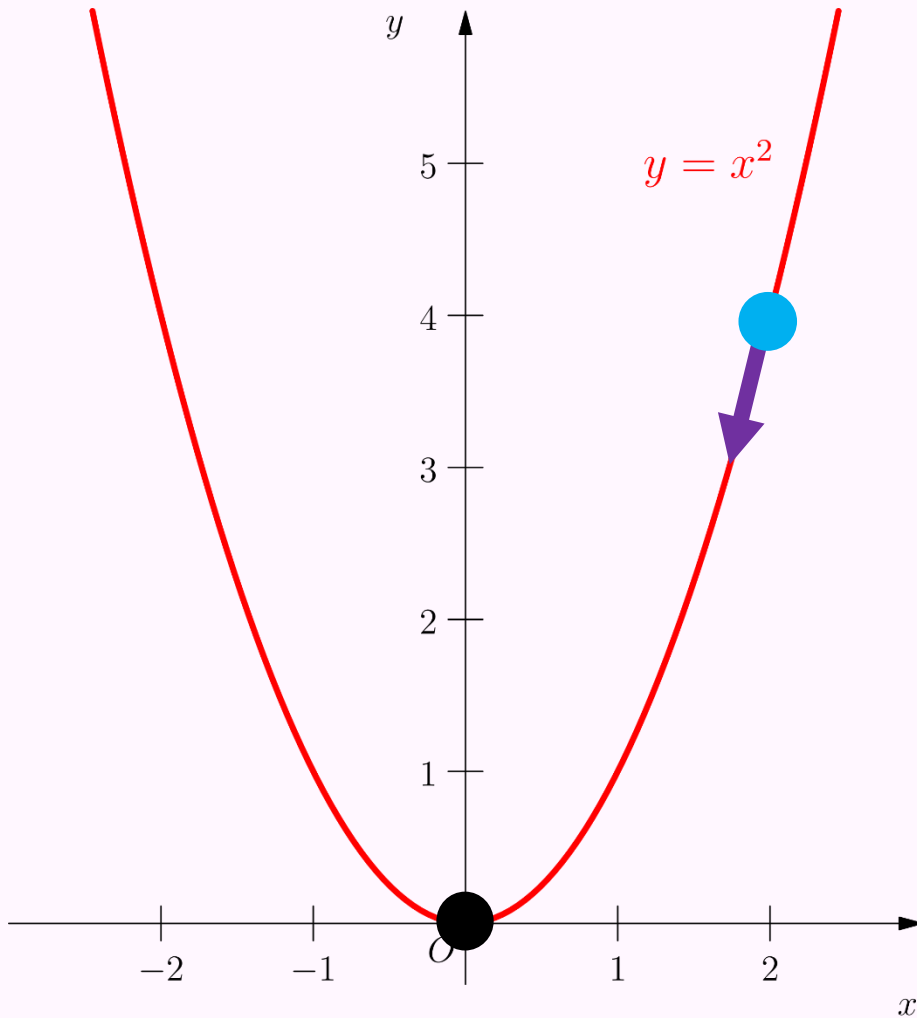
$$x_t = x_{t-1} - \alpha f'(x_{t-1})$$

Lokasi poin sebelumnya

Gradient fungsi  
(naik ke atas)

- Formula dengan notasi yang baik.

$$x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$$



- Formula gradient descent.

$$x_t = x_{t-1} - \alpha f'(x_{t-1})$$

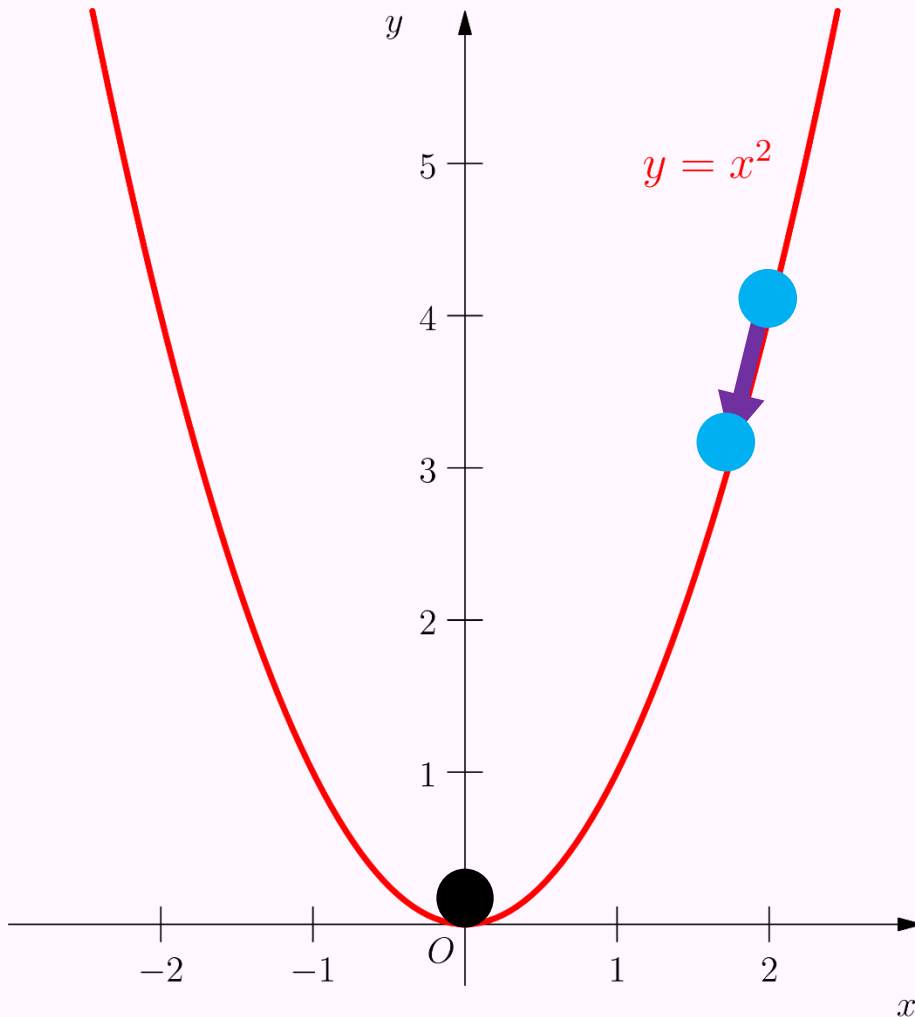
Skala langkah (points to  $\alpha$ )

Lokasi poin sebelumnya (points to  $x_{t-1}$ )

Gradient fungsi (naik ke atas) (points to  $f'(x_{t-1})$ )

- Formula dengan notasi yang baik.

$$x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$$



- Formula gradient descent.

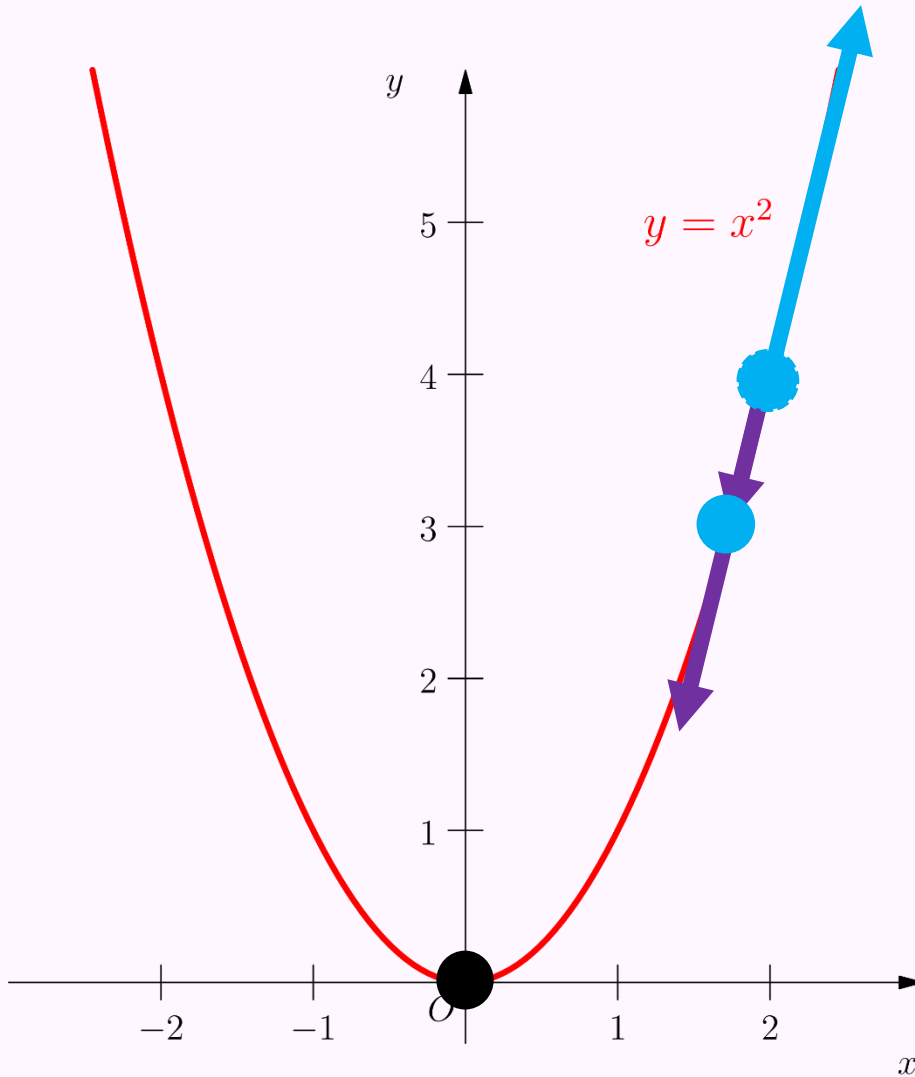
Lokasi poin sebelumnya      Skala langkah      Lokasi poin sebelumnya

$$x_t = x_{t-1} - \alpha f'(x_{t-1})$$

Gradient fungsi (naik ke atas)

- Formula dengan notasi yang baik.

$$x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$$



- Formula gradient descent.

Lokasi poin sebelumnya      Skala langkah      Lokasi poin sebelumnya

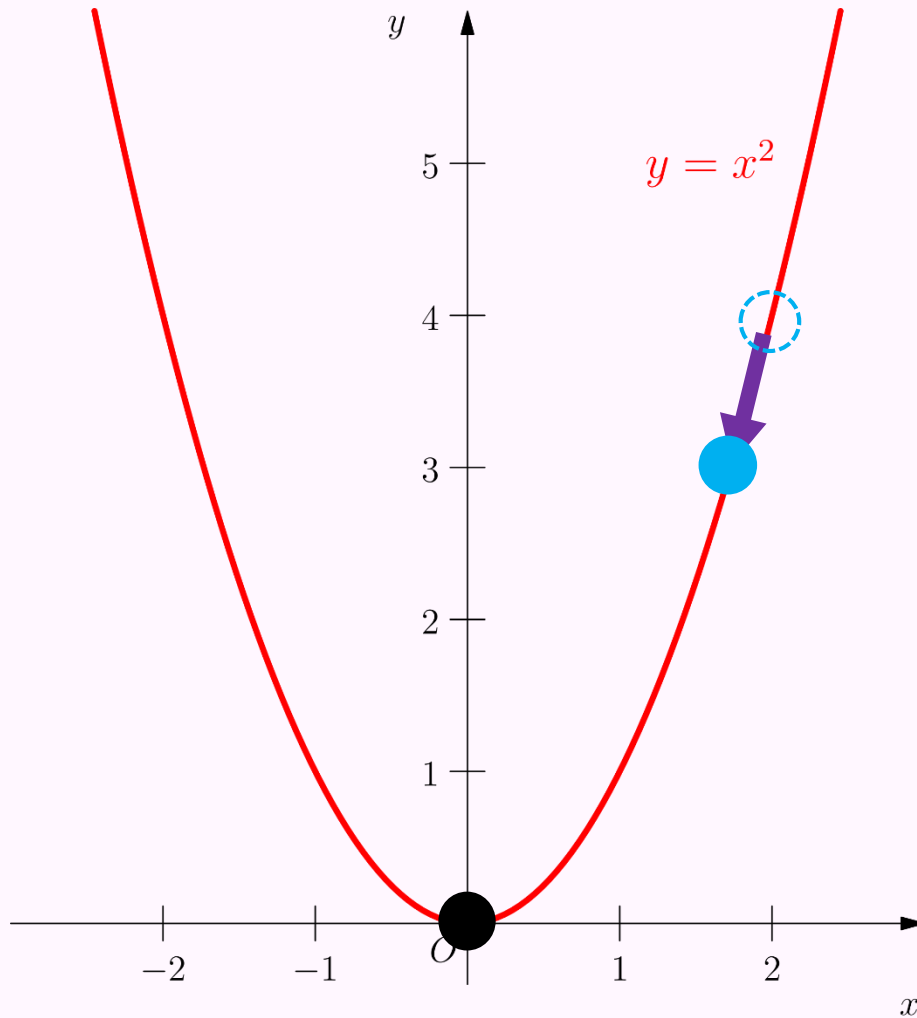
$$x_t = x_{t-1} - \alpha f'(x_{t-1})$$

Gradient fungsi (naik ke atas)

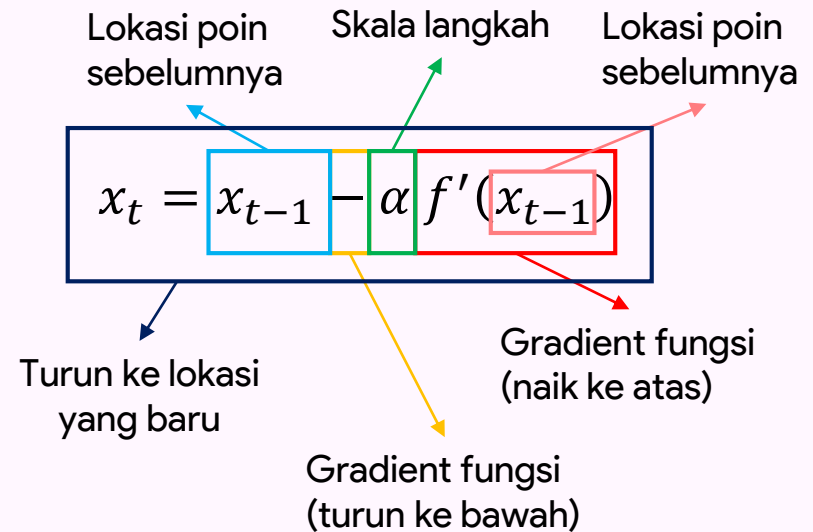
Gradient fungsi (turun ke bawah)

- Formula dengan notasi yang baik.

$$x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$$

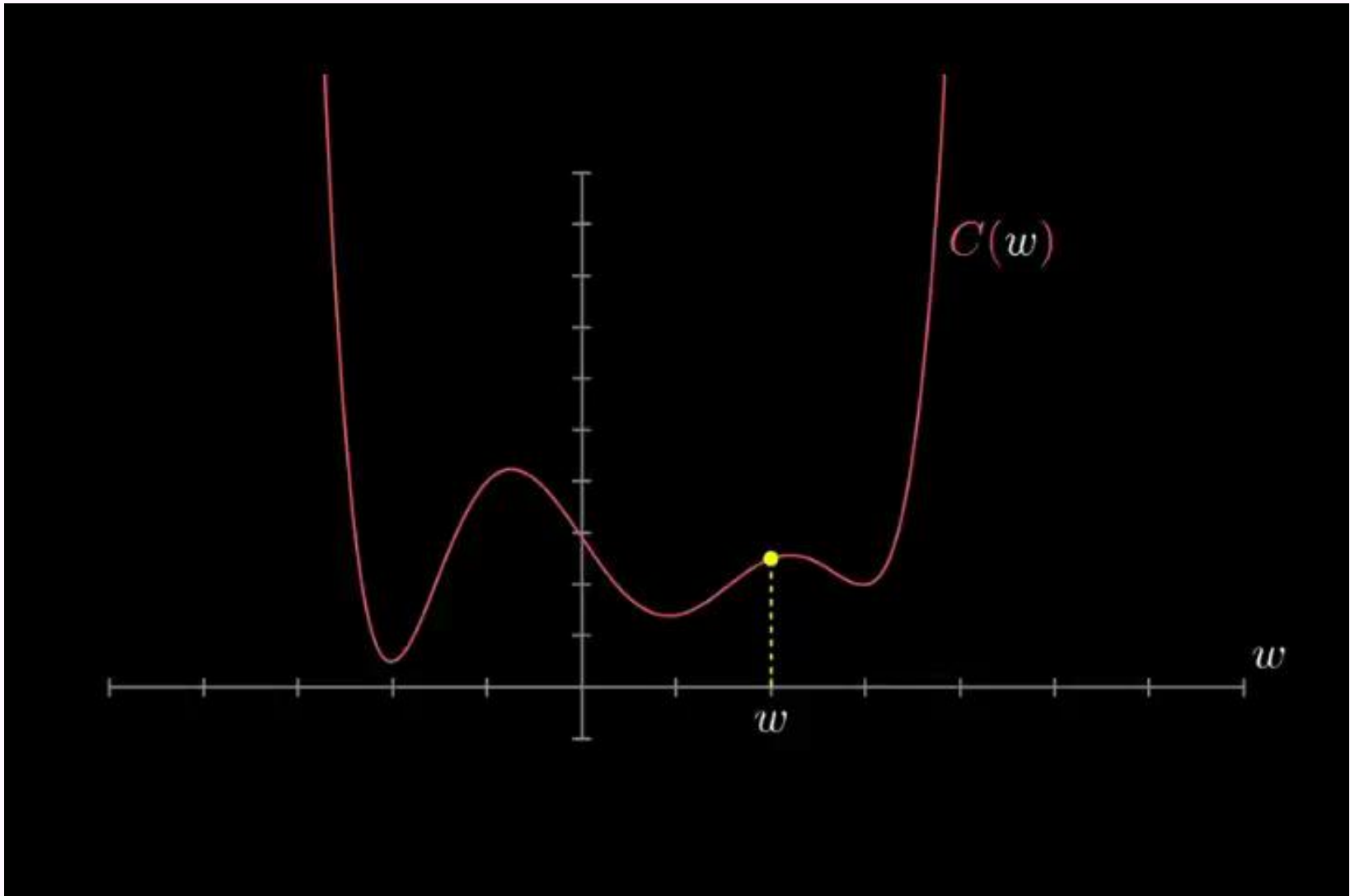


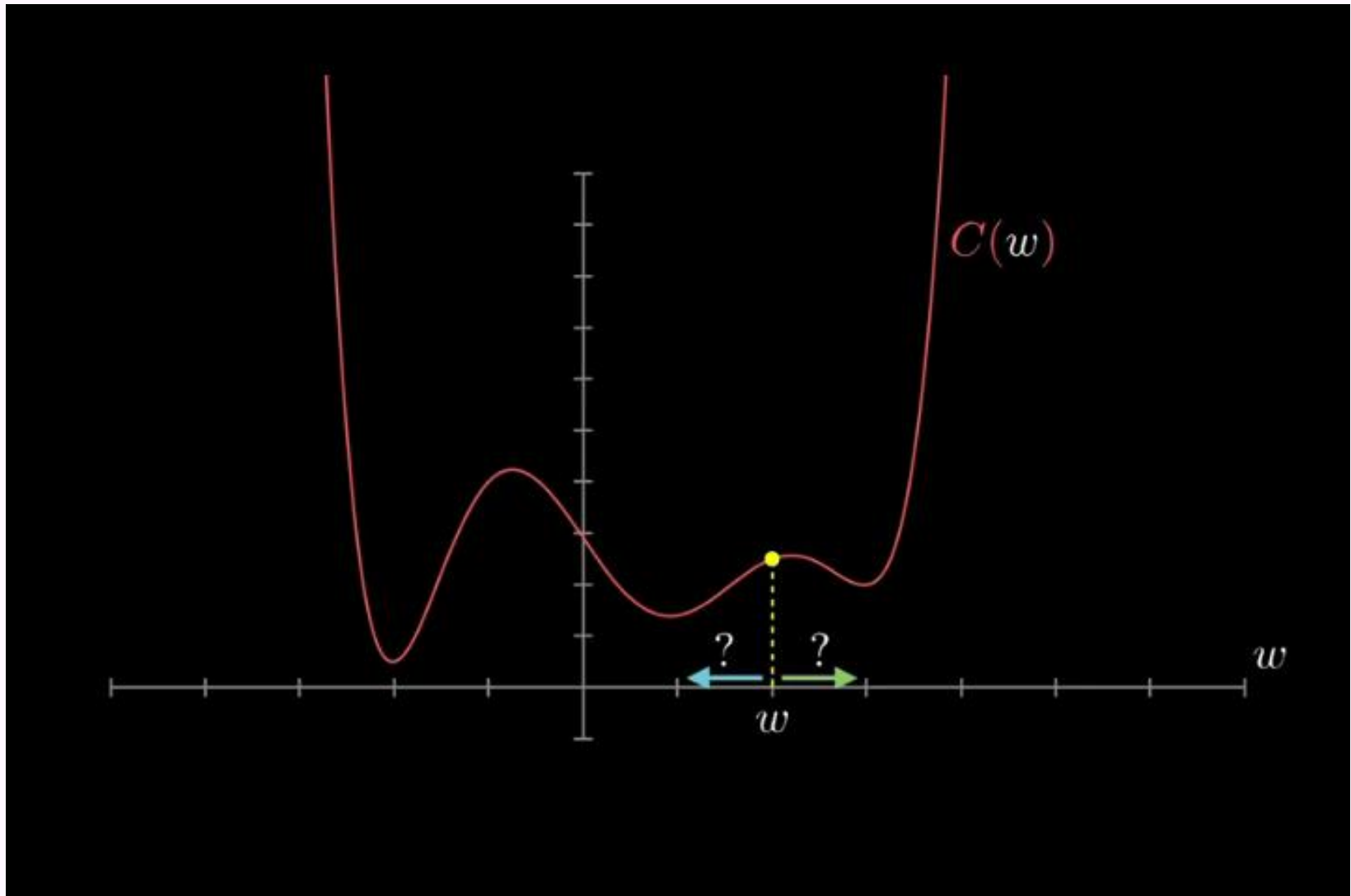
- Formula gradient descent.

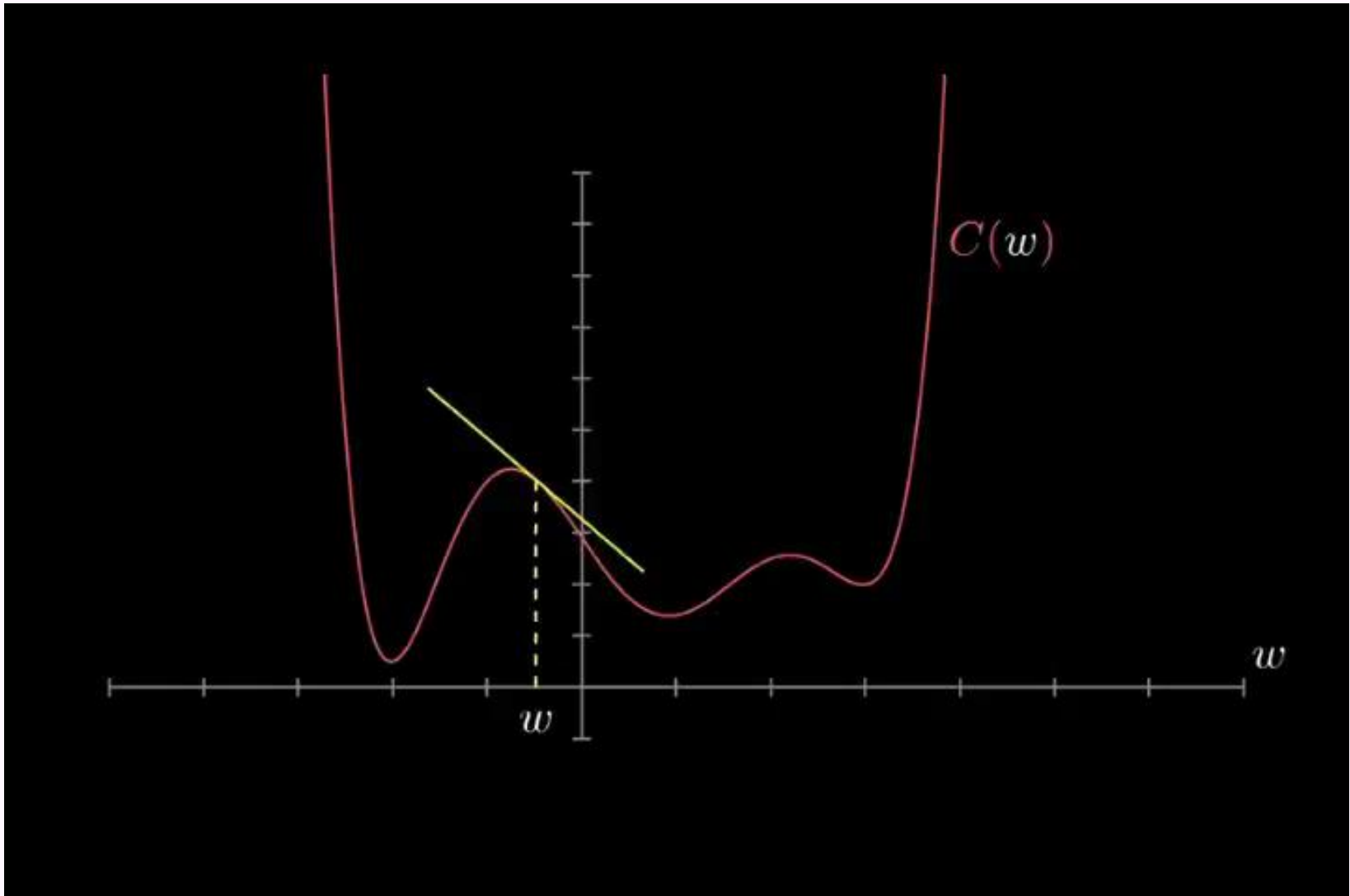


- Formula dengan notasi yang baik.

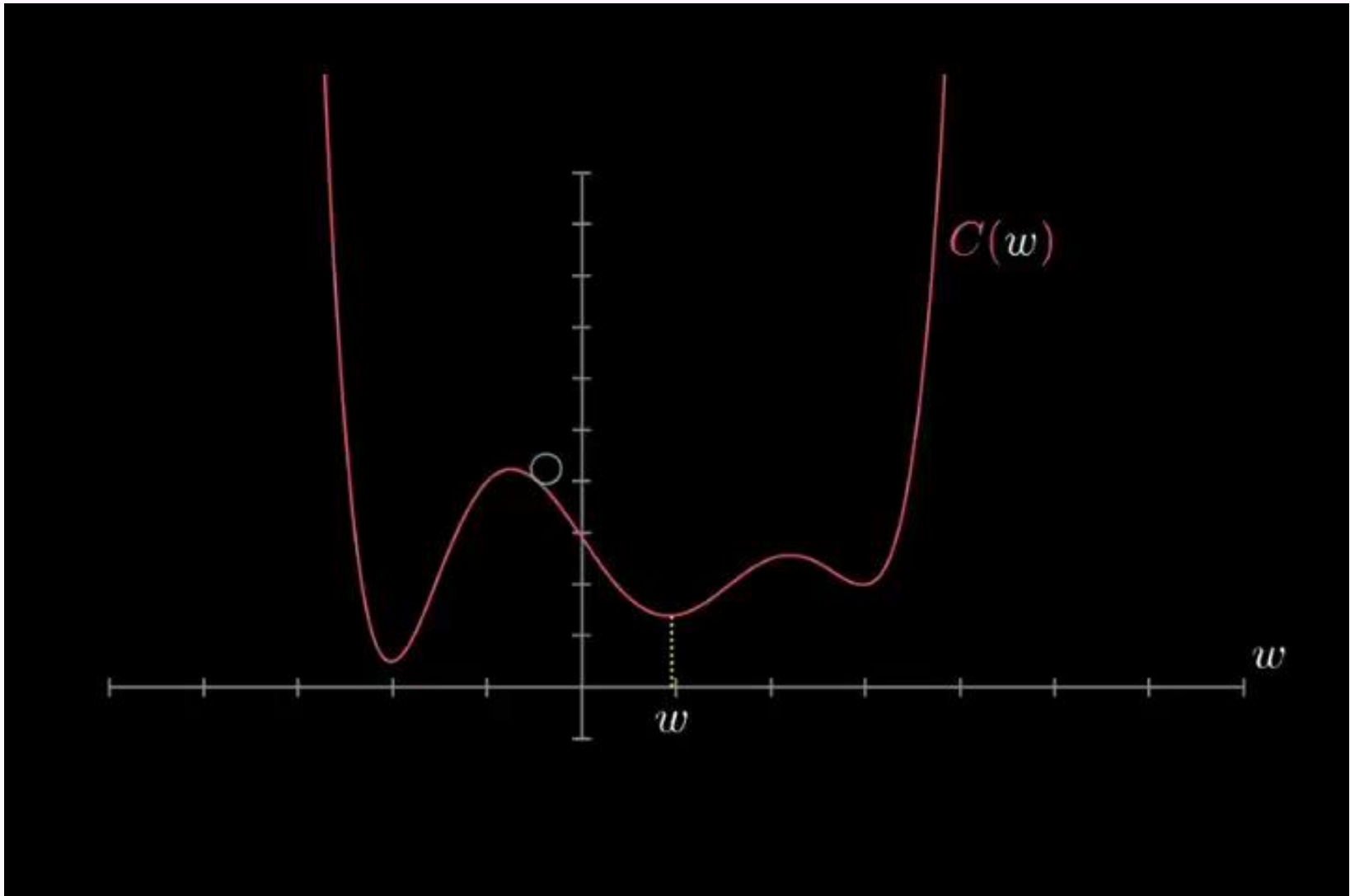
$$x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$$











- Bagaimana mencari gradient dalam 2D?

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

- Kita gunakan **Partial Derivative**!

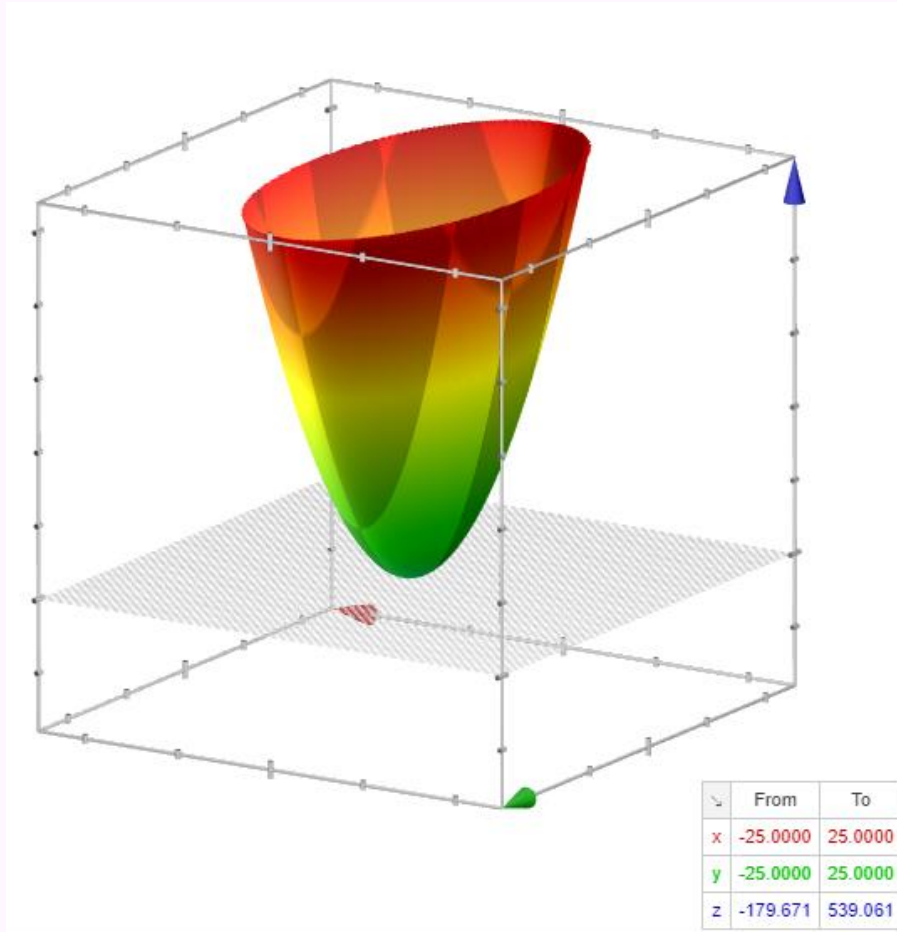
$$\nabla f(x, y) = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

$$\nabla f(x, y) = \left( \frac{4x^2 + y^2}{\partial x}, \frac{4x^2 + y^2}{\partial y} \right) = (8x, 2y)$$

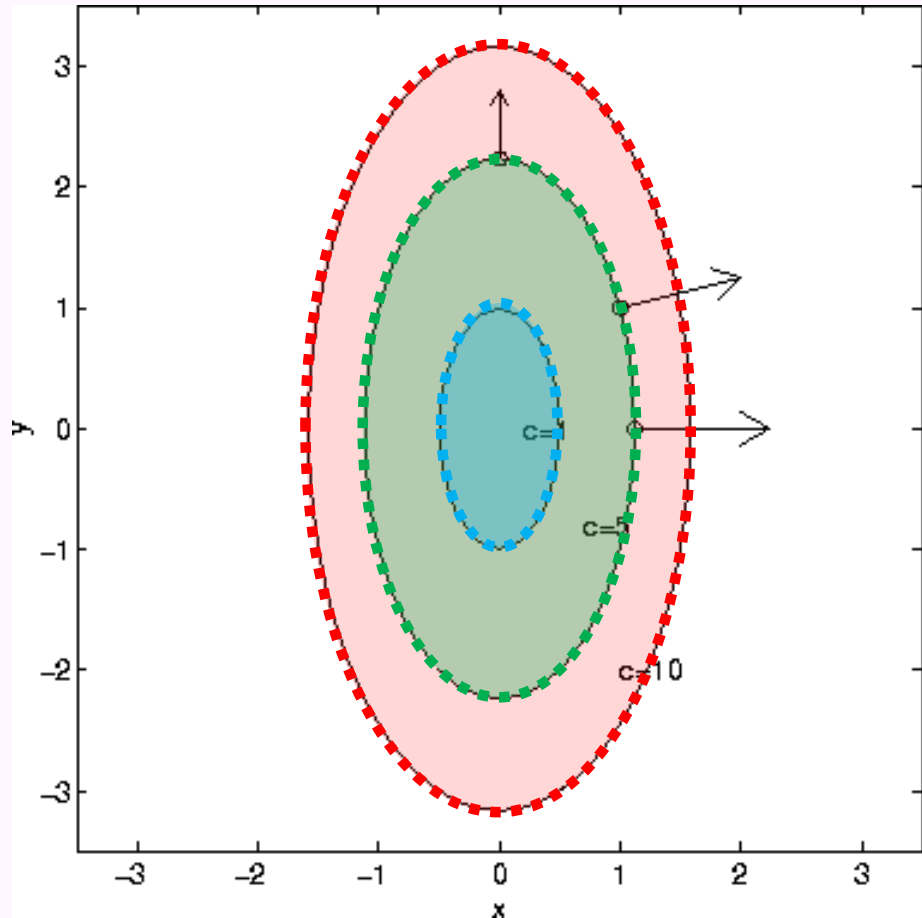
Turunkan berdasarkan  $x$

Turunkan berdasarkan  $y$

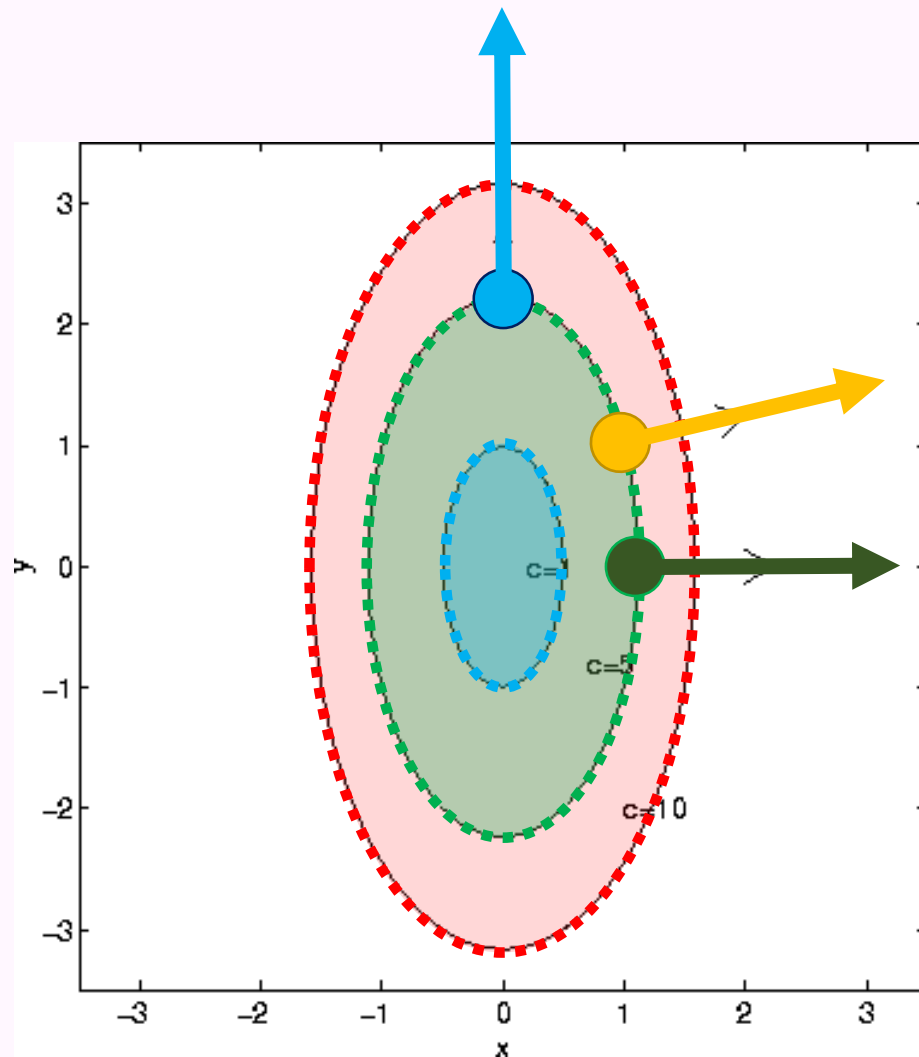
## Tampak 3D



## Tampak Atas



$$\nabla f(x, y) = (8x, 2y)$$



- $\nabla f(1,0) = (8,0)$
- $\nabla f(0,2) = (0,4)$
- $\nabla f(1,1) = (8,2)$
- **Tanda nilai turunan** menunjukkan arah untuk bergerak ke arah atas.
  - Sekarang arahnya 2D.
  - Bukan hanya kiri-kanan.
  - Tapi juga atas-bawah.
- **Besar nilai turunan** menunjukkan seberapa cepat perubahannya.

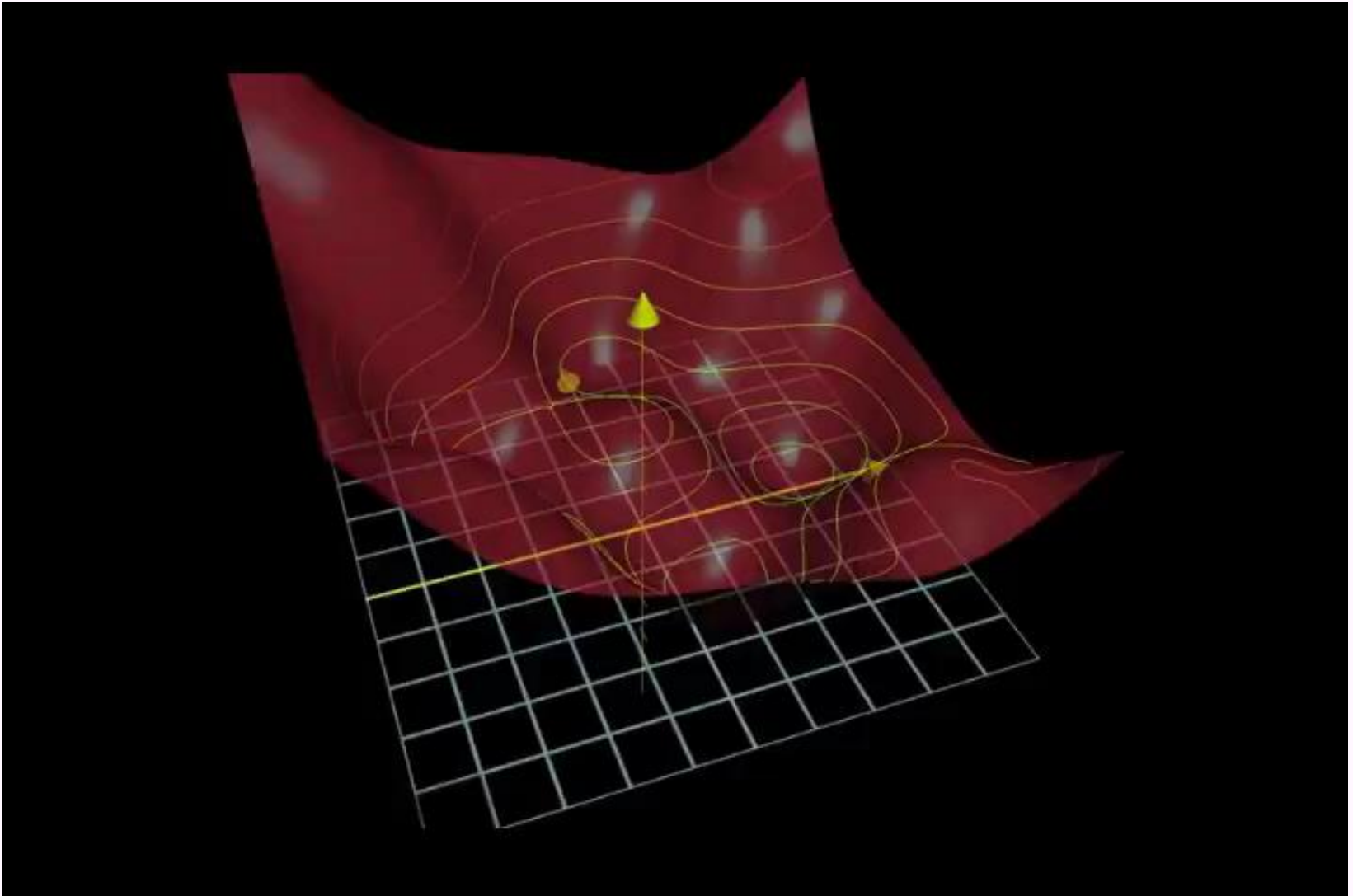


- Formula gradient descent 2D.

$$x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$$

$[x \ y]^T \longleftarrow \quad \quad \quad \longrightarrow \left[ \frac{\partial f}{\partial x} \ \frac{df}{dx} \right]^T$

- Serupa, yang membedakan hanya dimensi  $x$  dari  $\nabla f$



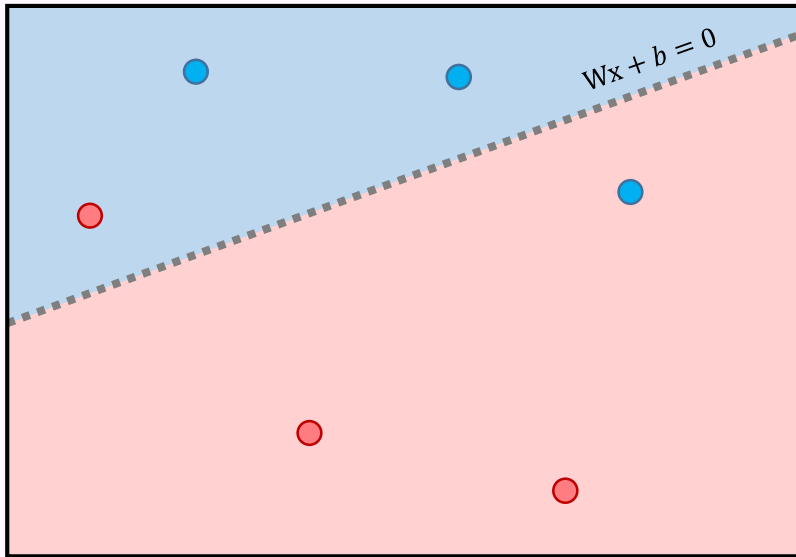


# ANN Gradient Descent

Gradient Descent untuk Neural Network

1. Awali dengan random weight:

$$w_1, \dots, w_n, b$$



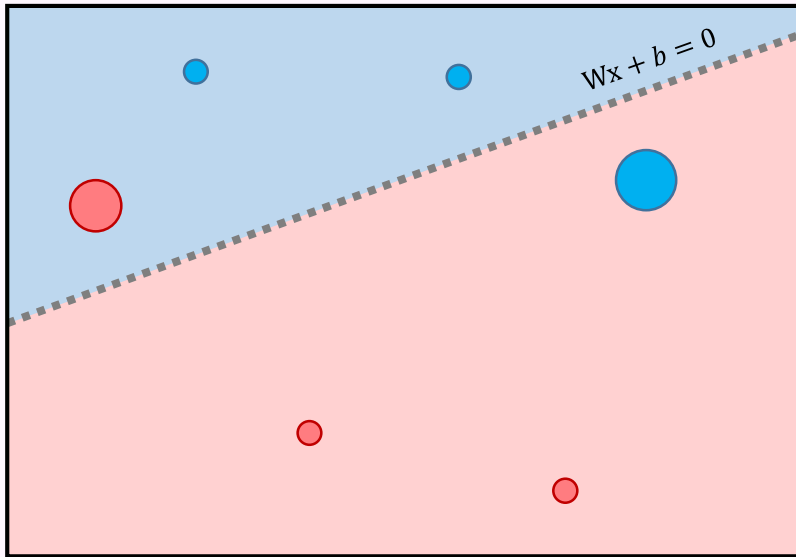


1. Awali dengan random weight:

$$w_1, \dots, w_n, b$$

2. Untuk setiap poin  $(x^{(1)}, \dots, x^{(n)})$ :

2.1. Hitung Error



1. Awali dengan random weight:

$$w_1, \dots, w_n, b$$

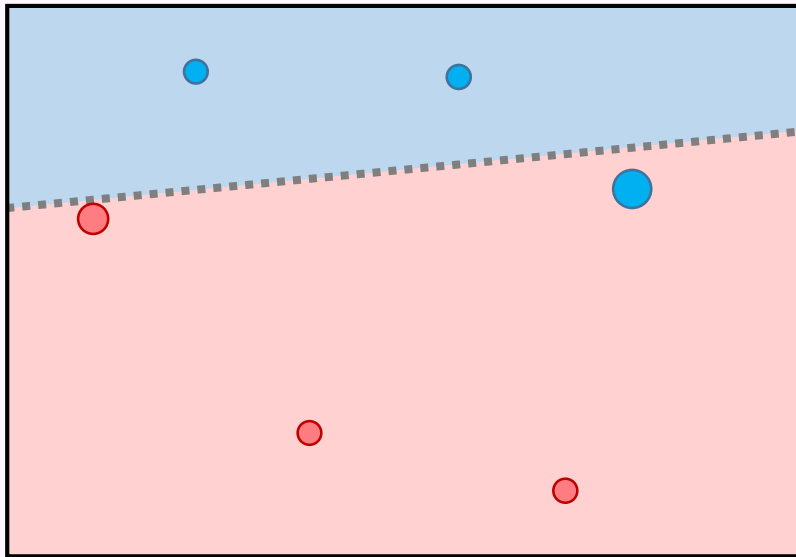
2. Untuk setiap poin  $(x^{(1)}, \dots, x^{(n)})$ :

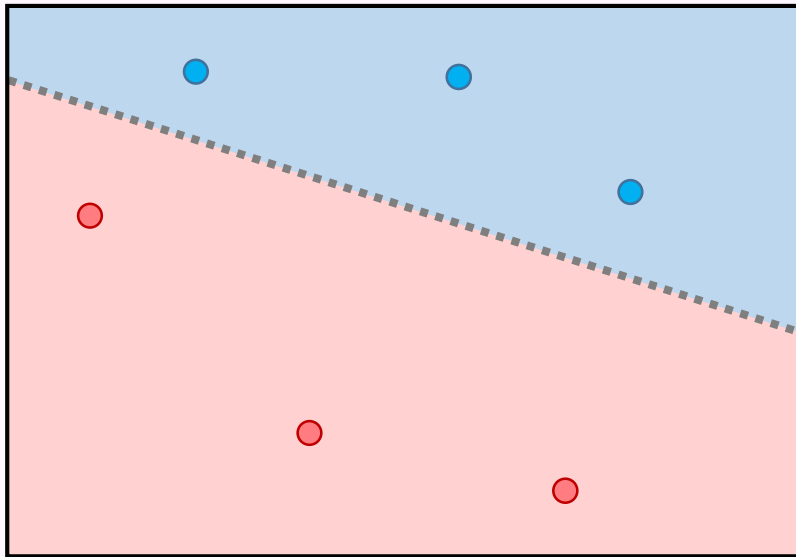
2.1. Hitung Error

2.2. Untuk setiap dimensi  $i = 1 \dots n$

2.2.1 Update  $w'_i = w_i - \alpha \frac{\partial E}{\partial w_i}$

2.2.2 Update  $b' = b - \alpha \frac{\partial E}{\partial b}$





1. Awali dengan random weight:

$$w_1, \dots, w_n, b$$

2. Untuk setiap poin  $(x^{(1)}, \dots, x^{(n)})$ :

2.1. Hitung Error

2.2. Untuk setiap dimensi  $i = 1 \dots n$

2.2.1 Update  $w'_i = w_i - \alpha \frac{\partial E}{\partial w_i}$

2.2.2 Update  $b' = b - \alpha \frac{\partial E}{\partial b}$

3. Ulangi sampai convergence

- Error Formula  $E = -\frac{1}{m} \sum_{i=1}^m (y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i))$
- Prediction  $\hat{y}_i = \sigma(Wx^{(i)} + b)$
- Input  $x^{(i)} = [x_1^{(i)} \quad x_2^{(i)} \quad \dots \quad x_n^{(i)}]^T$
- Weights  $W = [w_1 \quad w_2 \quad \dots \quad w_n]$

- Error Formula  $E = -y \ln(\hat{y}) - (1 - y) \ln(1 - \hat{y})$
- Prediction  $\hat{y} = \sigma(Wx + b)$
- Input  $x = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$
- Weights  $W = [w_1 \quad w_2 \quad \cdots \quad w_n]$

- Target kita adalah mencari gradient dari Cost Function kita.

$$\nabla E = \left( \frac{\partial}{\partial w_1} E, \dots, \frac{\partial}{\partial w_n} E, \frac{\partial}{\partial b} E \right)$$

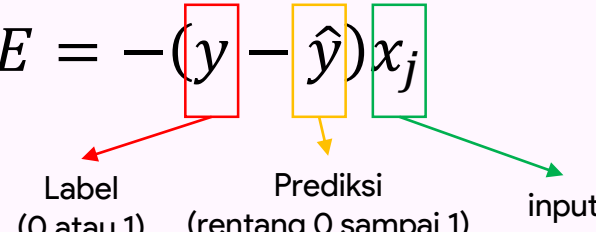
⋮

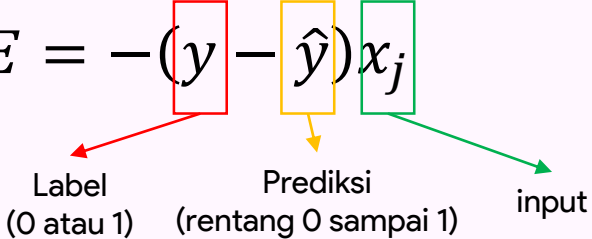
*Melalui sekian baris penurunan persamaan*

*First order derivative dari fungsi  $E$  terhadap  $w$  dan  $b$  adalah*

⋮

$$\frac{\partial}{\partial w_j} E = -(y - \hat{y})x_j \qquad \frac{\partial}{\partial b} E = -(y - \hat{y})$$



$$\frac{\partial}{\partial w_j} E = -(y - \hat{y})x_j$$


Mana diantara berikut ini yang benar?

- a. Semakin dekat label ke prediksi, semakin besar nilai gradien.
- b. Semakin dekat label ke prediksi, semakin kecil nilai gradien.
- c. Semakin jauh label ke prediksi, semakin besar gradien.
- d. Semakin jauh label ke prediksi, semakin kecil gradien.

- Formula generic gradient descent

$$x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$$

- Gradient descent untuk melakukan update *weight* pada ANN

$$w'_i = w_i - \alpha \nabla E$$

$$w'_i = w_i - \alpha (-(y - \hat{y})x_i)$$

$$w'_i = w_i + \alpha (y - \hat{y})x_i$$

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