

Supervised Learning

Yaya Wihardi, S.Kom., M.Kom.

Email: yayawihardi@upi.edu

Department of Computer Science Education

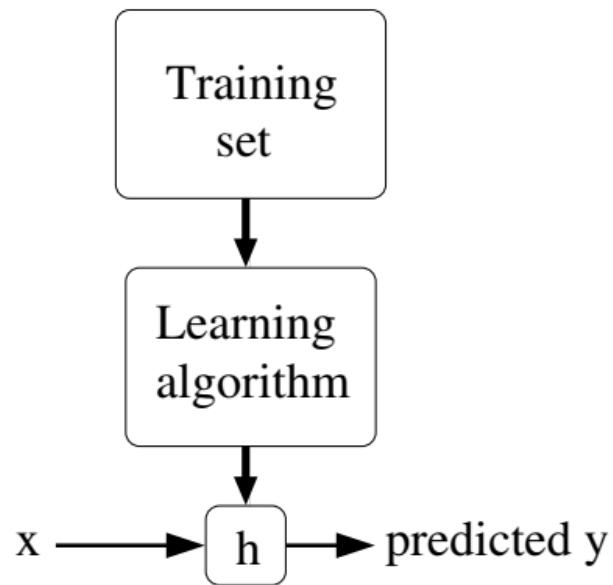
Universitas Pendidikan Indonesia

Outline

- Apa itu Supervised Learning?
- Task dalam Supervised Learning
- Notasi yang sering digunakan
- Regresi Linear
- Least Mean Square
- Gradient Descent

Apa itu Supervised Learning?

- Proses pembelajaran mesin yang mencoba membuat fungsi pemetaan dari input menjadi output berdasarkan data training



Task dalam Supervised Learning

- Regression → memprediksi nilai continue, seperti harga rumah, harga saham, dll



- Classification → Memprediksi nilai diskrit, seperti kategori, kelas, kelompok, dll

Apakah ini foto Susi, Steak, atau Pizza?



Your Turn

- Anda diberikan history data penjualan item-item produk dalam 3 tahun terakhir, prediksi berapa jumlah penjualan setiap item produk dalam 3 bulan kedepan.
- Anda diberikan data akun dan riwayat aktifitasnya dalam sebuah sistem pemerintahan. Prediksi apakah akun-akun tersebut diretas atau tidak.
- Diberikan data email yang telah dilabeli spam atau bukan spam. Buatlah spam filter.

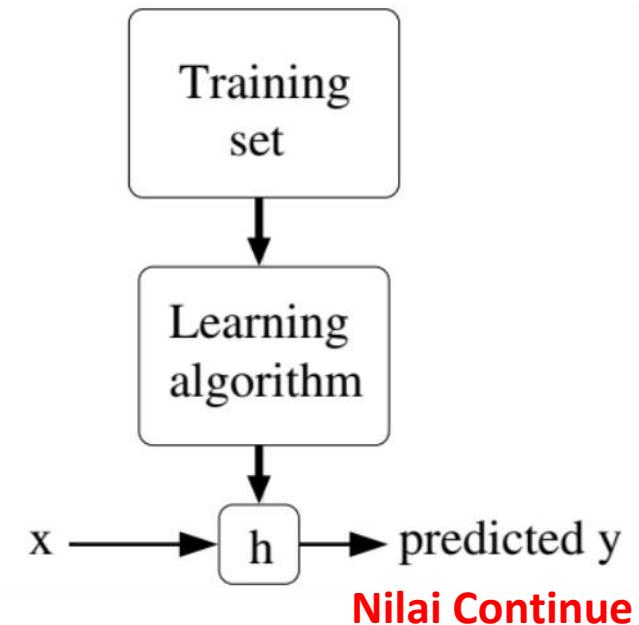
Notasi

Notasi	Keterangan
$x^{(i)}$	Variabel Input/Fitur
$y^{(i)}$	Variabel Output/Target/ Variabel yang akan diprediksi
$(x^{(i)}, y^{(i)})$	Pasangan Training Example/Instance/Data Point
$\{(x^{(i)}, y^{(i)}); i = 1, \dots, n\}$	Dataset/Himpunan Data yg digunakan untuk membangun model
$\dots^{(i)}$	(i) Menyatakan indeks dari data, bukan pangkat
R^n	Ruang vector berdimensi n

Regresi: Studi Kasus Real Data

- Diberikan data luas living area (feet²), jml bedrooms, dan harganya (USD) sbb:

Living area (feet ²)	#bedrooms	Price (1000\$s)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
:	:	:



- Buat model ML yg dapat memprediksi harga rumah berdasarkan variable input luas living area dan jml bedrooms

Regresi: Studi Kasus Real Data

- Diberikan data luas living area (feet²), jml bedrooms, dan harganya (USD) sbb:

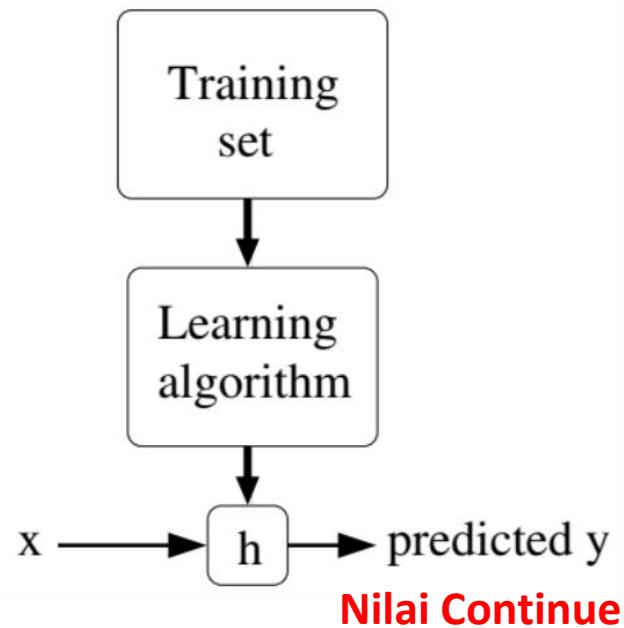
Living area (feet ²)	#bedrooms	Price (1000\$)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
:	:	:

- Utk kasus ini, x merupakan vector berdimensi 2 (R^2)
- $x_1^{(i)}$ menyatakan living area, $x_2^{(i)}$ menyatakan jml bedrooms

Regresi Linear: Studi Kasus Sederhana

- Diberikan data pasangan x dan y sbb:

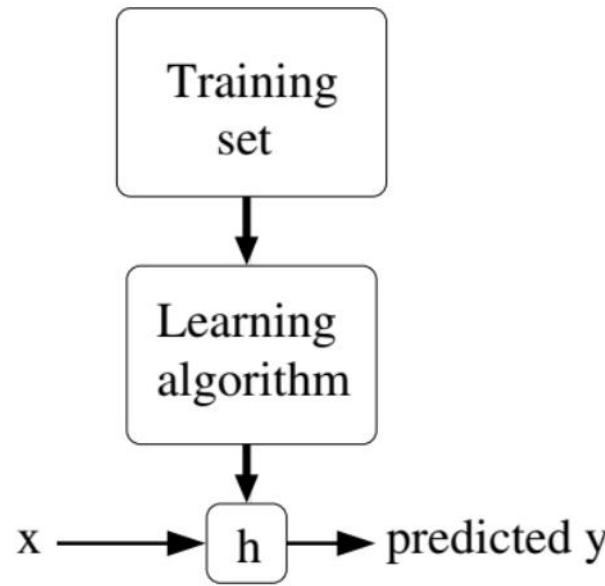
x	y
2	8
5	17
4	13
7	24
8	26
.	.
.	.
.	.



- Buat model ML yg dapat memprediksi nilai y berdasarkan input x

Caranya?

x	y
2	8
5	17
4	13
7	24
8	26
.	.
.	.
.	.



- Cari fungsi hipotesis (h) yang dapat memetakan x ke y
- Ingat prinsip occam's razor: mulailah dari yang sederhana
- Sebagai contoh, bisa digunakan fungsi persaman linear

Model Hipotesis (h)

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

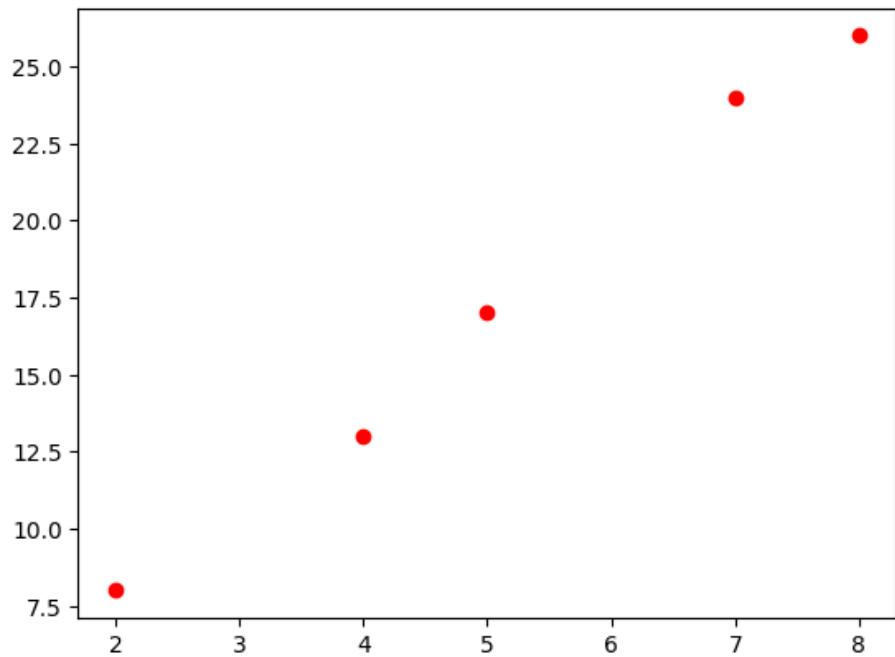
θ_i : Parameter model

$$h_w(x) = w_0 + w_1 x$$

w_i : Parameter model

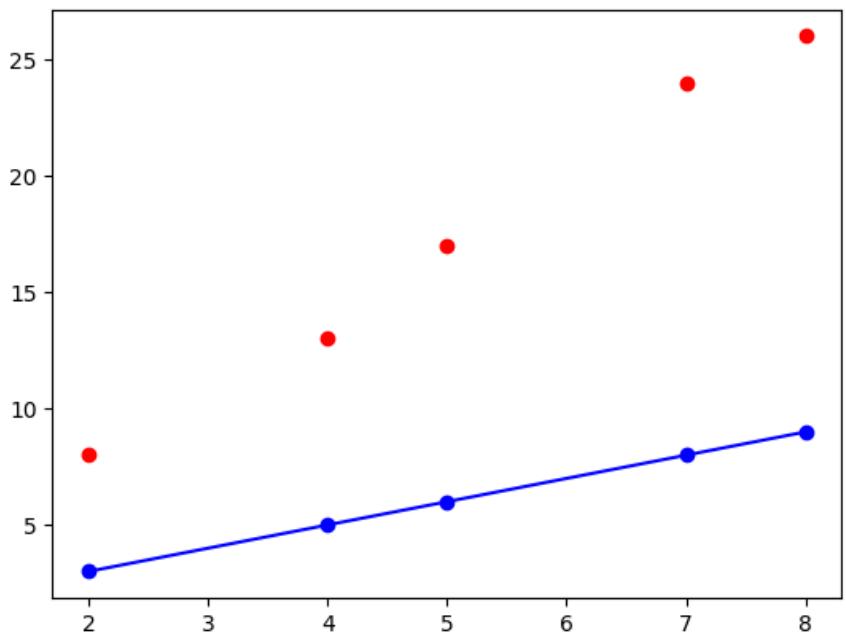
x	y
2	8
5	17
4	13
7	24
8	26
.	.

**Bagaimana memilih
parameter w_i/θ_i
yang tepat?**

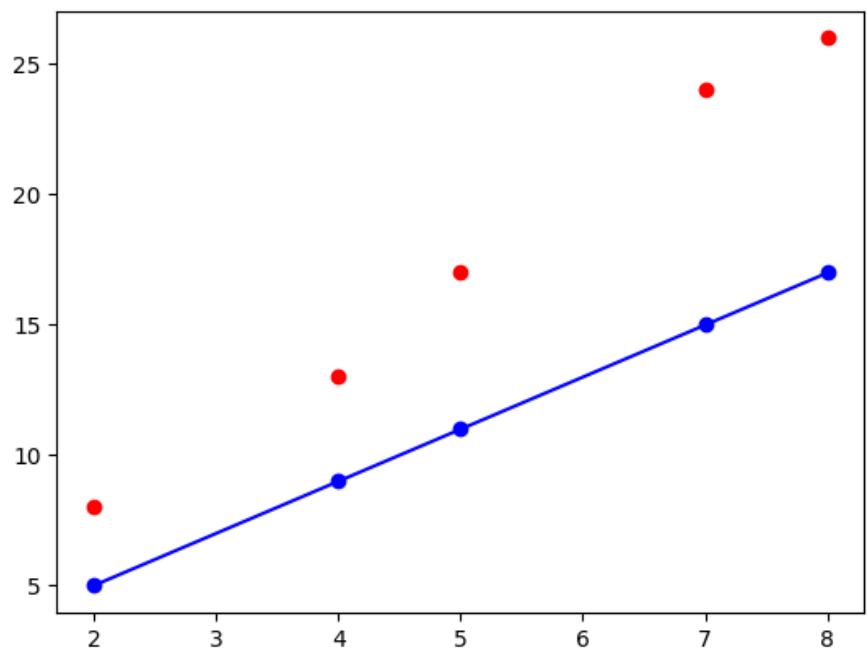


Model Hipotesis

$$h_w(x) = w_0 + w_1 x$$



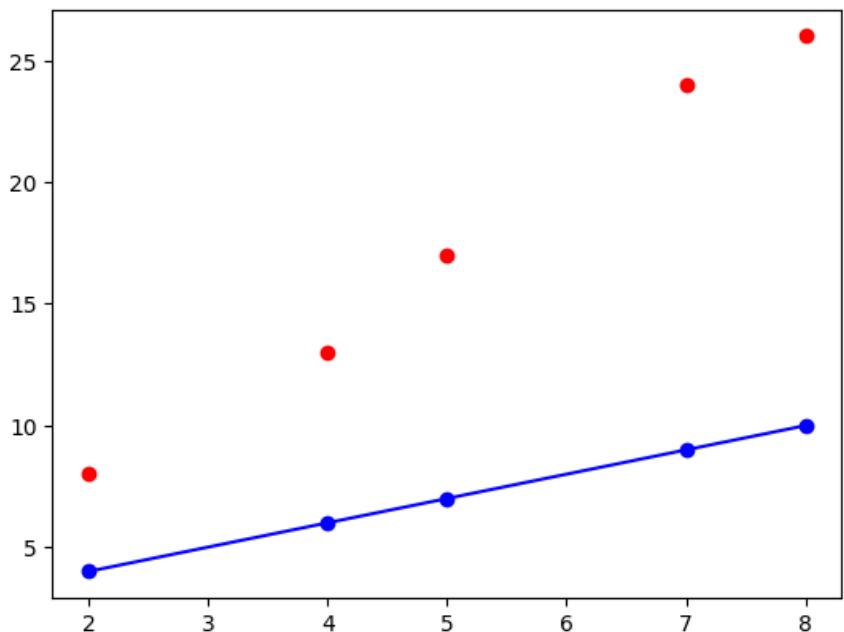
$$\begin{aligned}w_0 &= 1 \\w_1 &= 1\end{aligned}$$



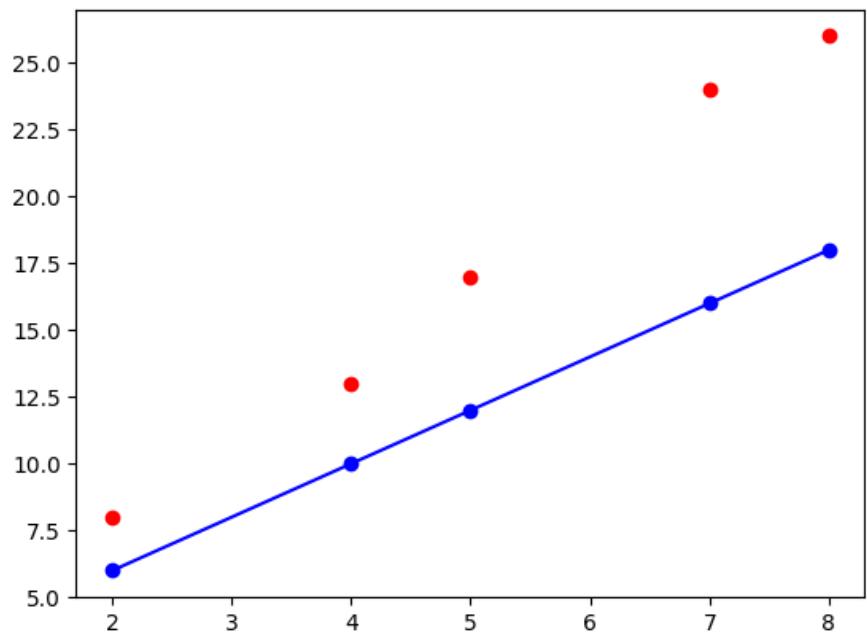
$$\begin{aligned}w_0 &= 1 \\w_1 &= 2\end{aligned}$$

Model Hipotesis

$$h_w(x) = w_0 + w_1 x$$



$$\begin{aligned}w_0 &= 2 \\w_1 &= 1\end{aligned}$$



$$\begin{aligned}w_0 &= 2 \\w_1 &= 2\end{aligned}$$

Memilih Model Hipotesis (h)

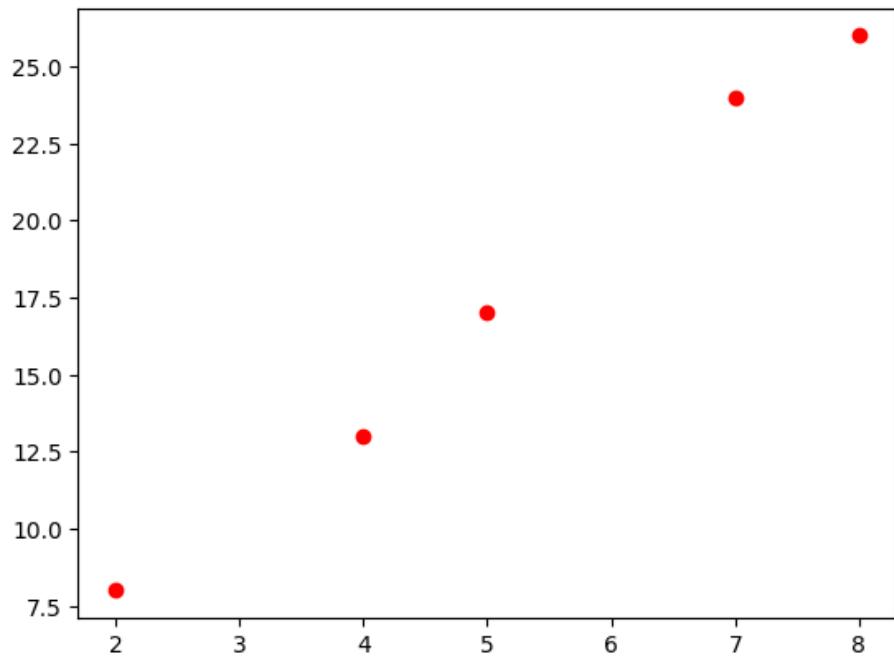
$$h_w(x) = w_0 + w_1 x$$

w_i : Parameter model

Bagaimana memilih
parameter w_i
yang tepat?

IDE: Pilihlah w_0, w_1 yang
membuat $h_w(x)$ paling
mendekati y untuk semua
pasangan (x, y)

x	y
2	8
5	17
4	13
7	24
8	26
.	.



Cost Function: Least Mean Square (LMS)

Hipotesis:

$$h_w(x) = w_0 + w_1 x$$

Parameter:

$$w_0, w_1$$

Cost Function:

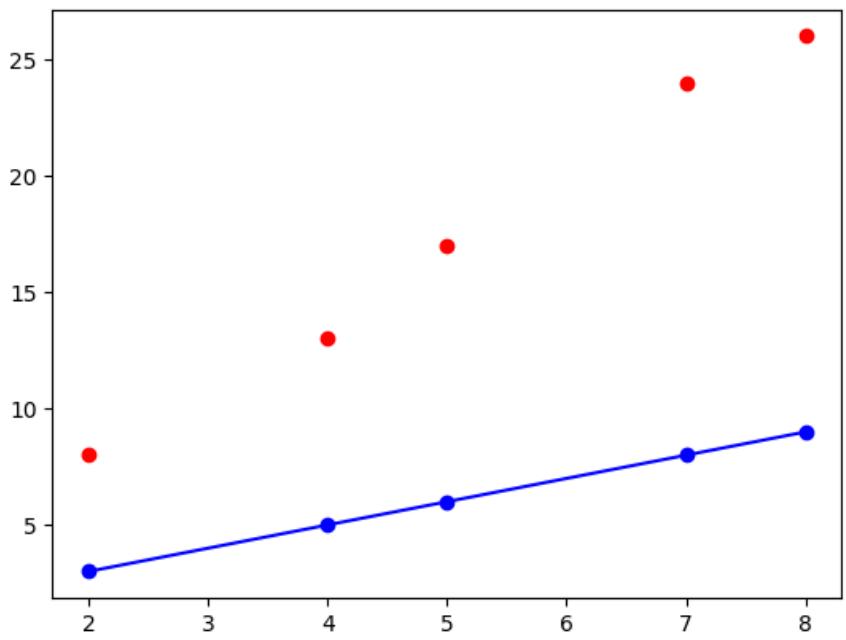
$$J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)})^2$$

Goal:

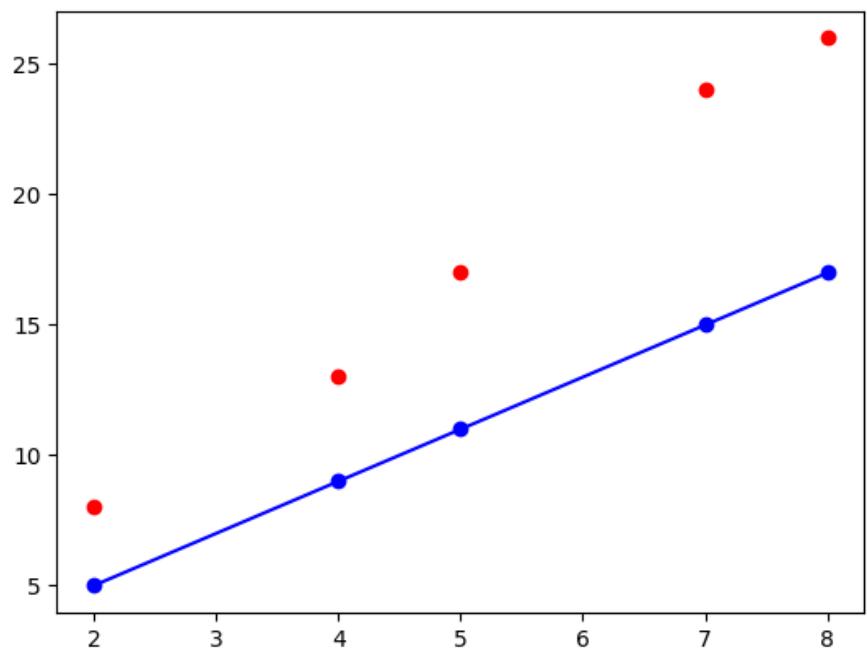
$$\underset{w_1, w_1}{\text{minimize}} \ J(w_0, w_1)$$

Model Hipotesis

$$h_w(x) = w_0 + w_1 x$$



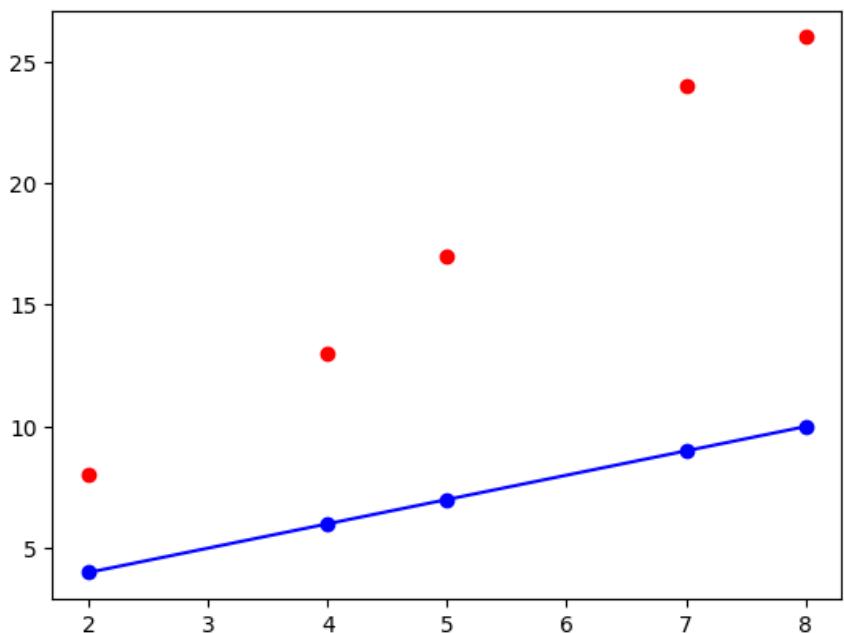
$$\begin{aligned}w_0 &= 1 & J &= 75.5 \\w_1 &= 1\end{aligned}$$



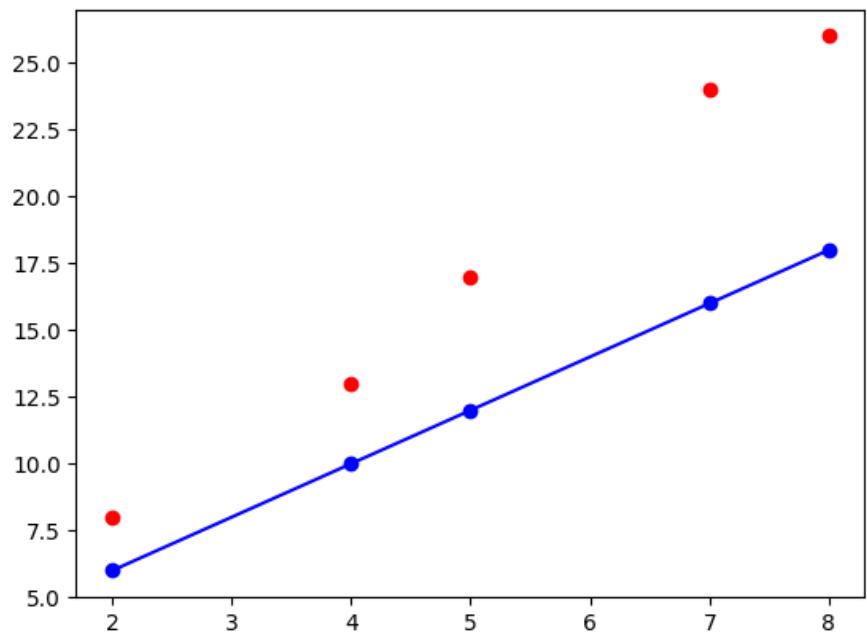
$$\begin{aligned}w_0 &= 1 & J &= 22.3 \\w_1 &= 2\end{aligned}$$

Model Hipotesis

$$h_w(x) = w_0 + w_1 x$$



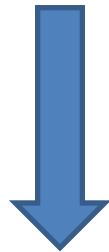
$$\begin{aligned}w_0 &= 2 & J &= 64.6 \\w_1 &= 1\end{aligned}$$



$$\begin{aligned}w_0 &= 2 & J &= 16.6 \\w_1 &= 2\end{aligned}$$

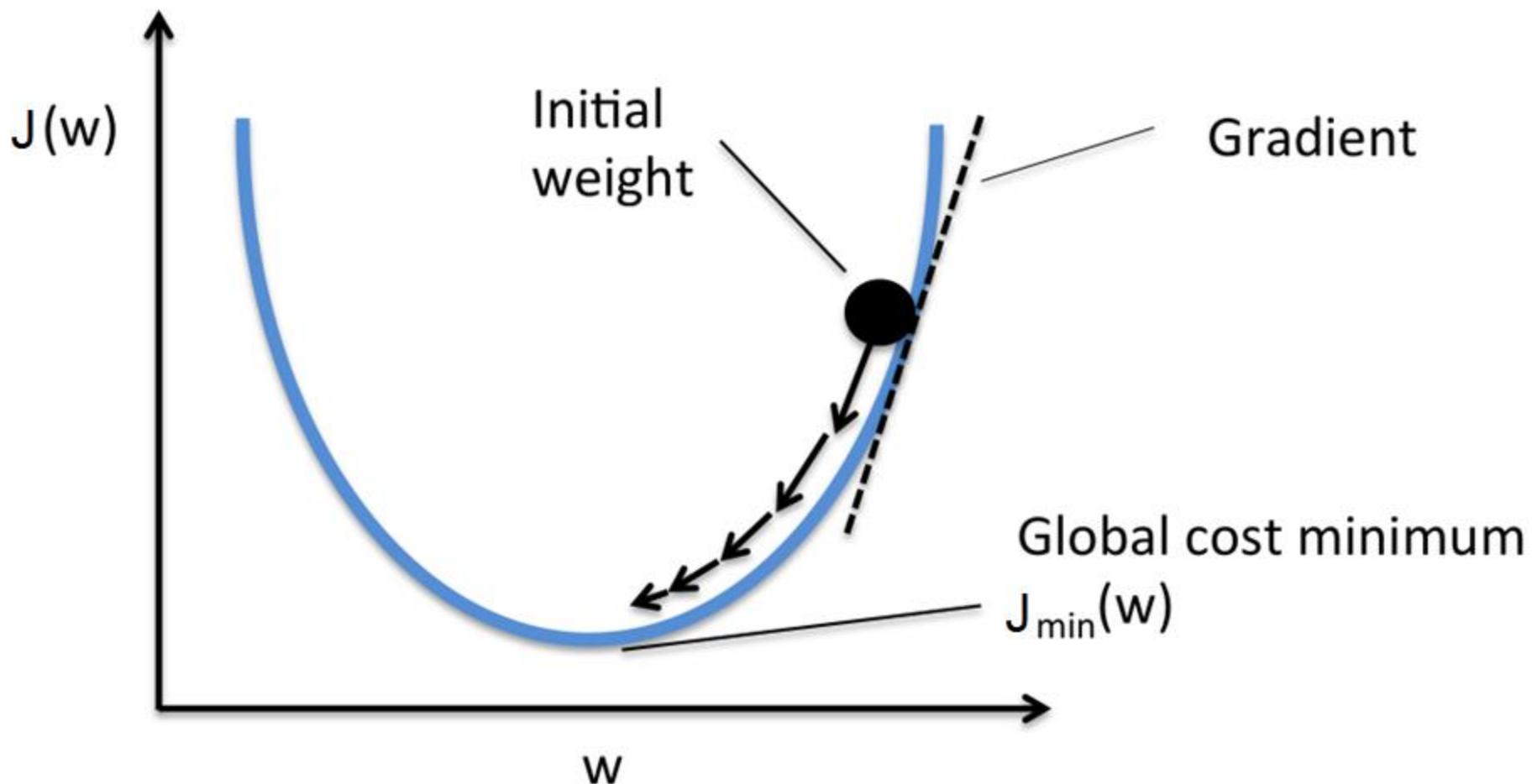
Tujuan Training

Meminimalkan Cost Function

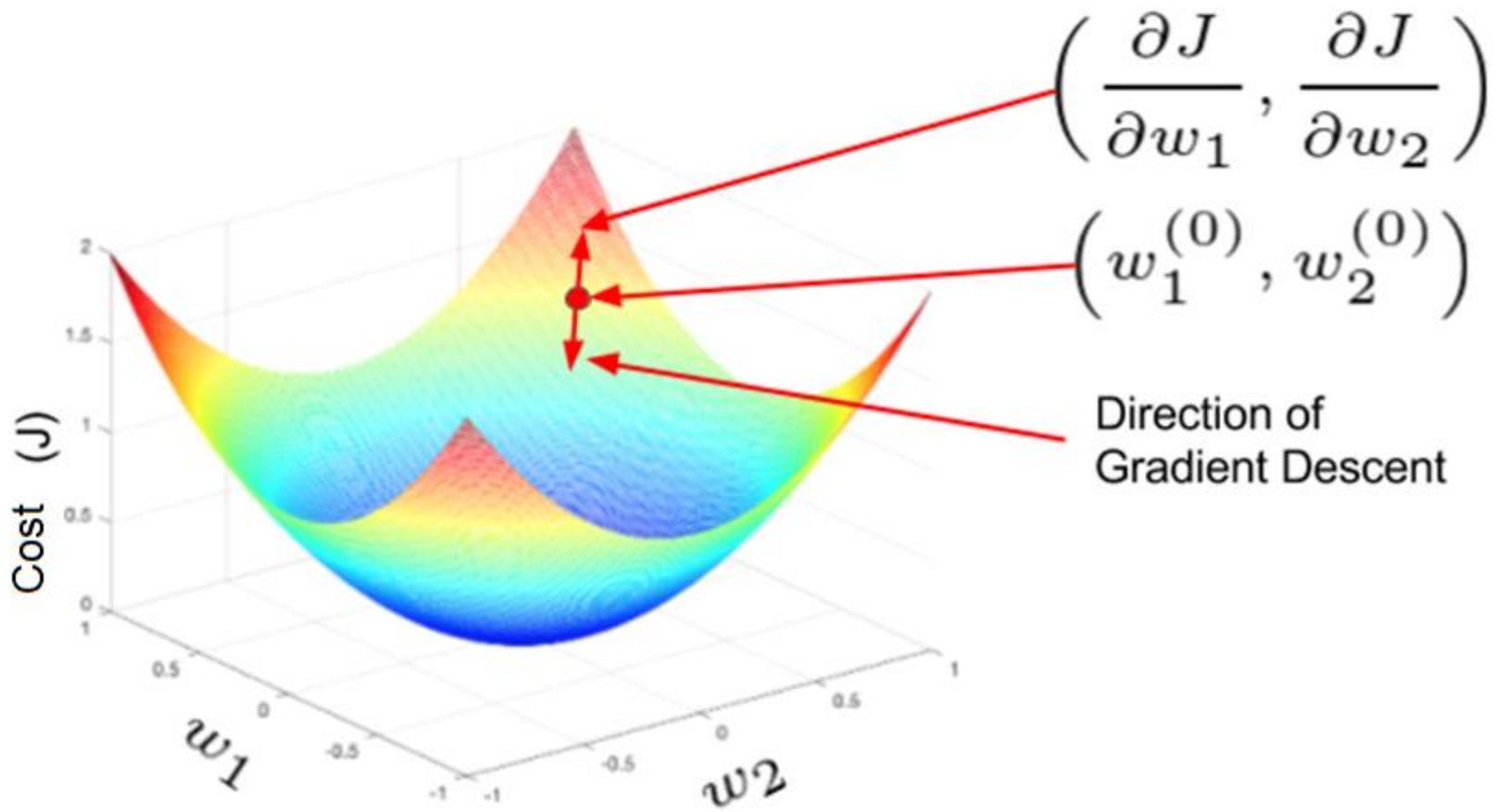


Dg Merubah Nilai Parameter Model

Gradient Descent



Gradient Descent



Thank You