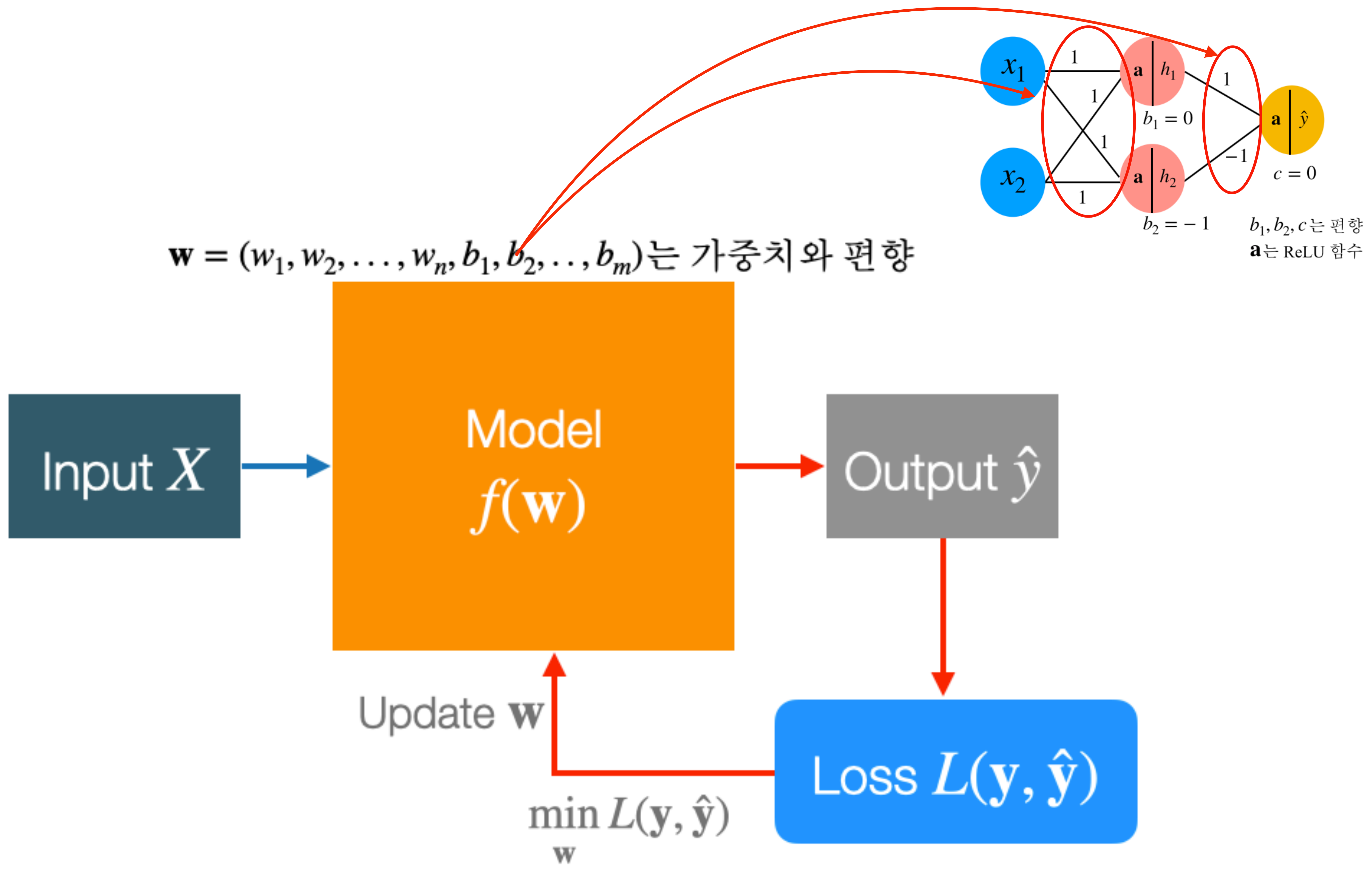
딥러닝 올인원

경사 하강법

9강

딥러닝호형

최적화(Optimization)



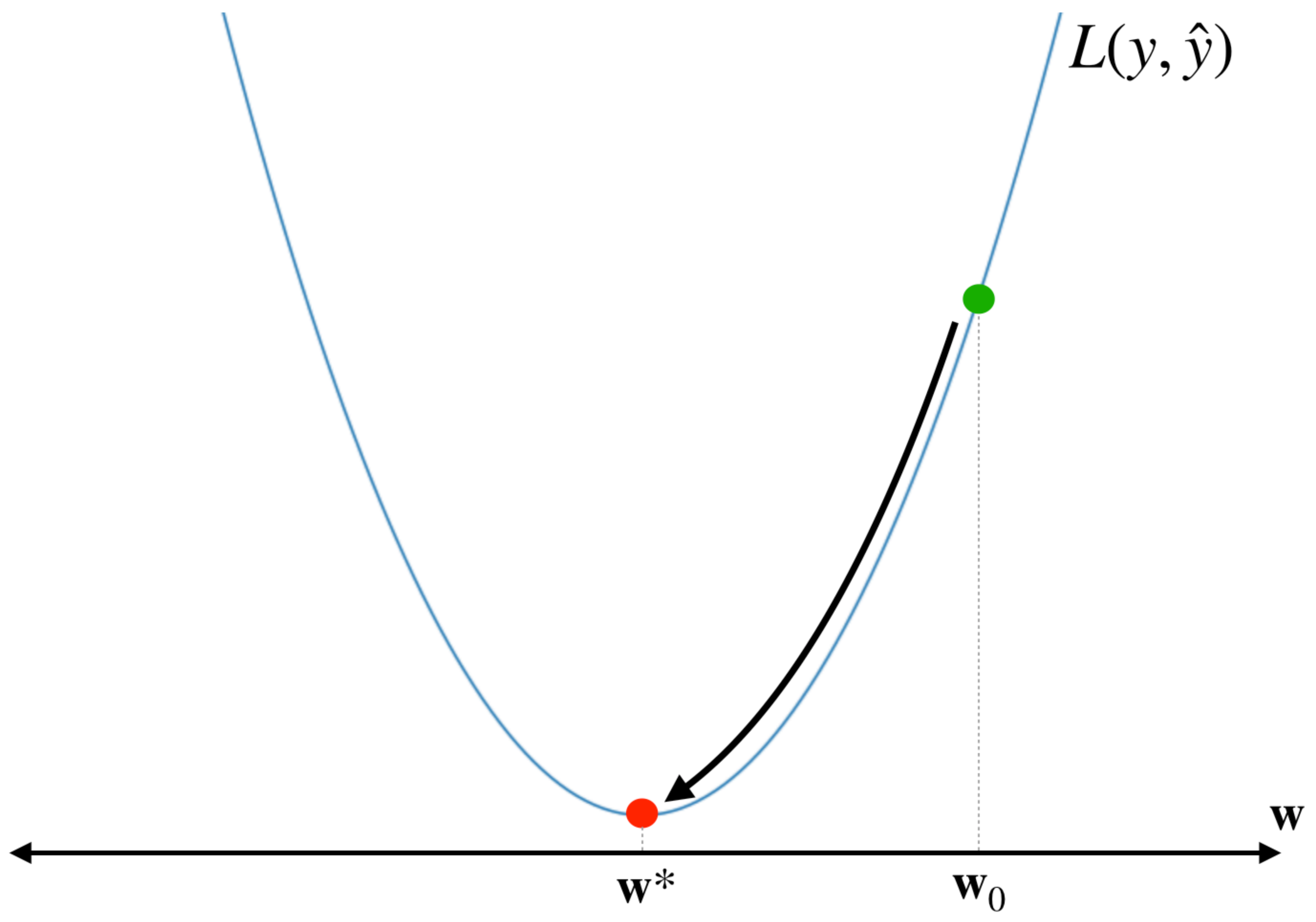
최적화(Optimization)

목적 함수(Objective Function)

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

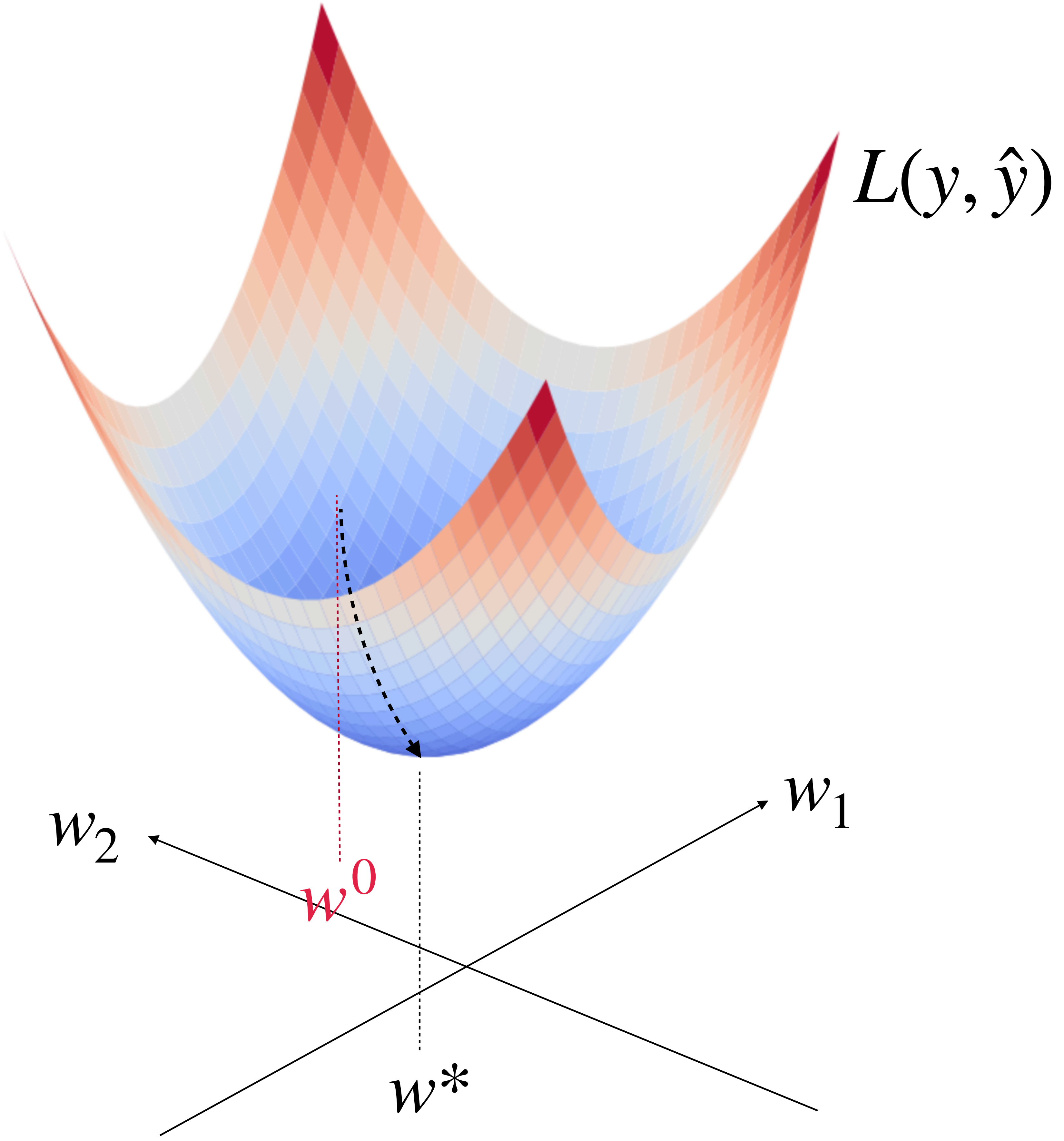
최적화(Optimization)

하강법(Descent Method)



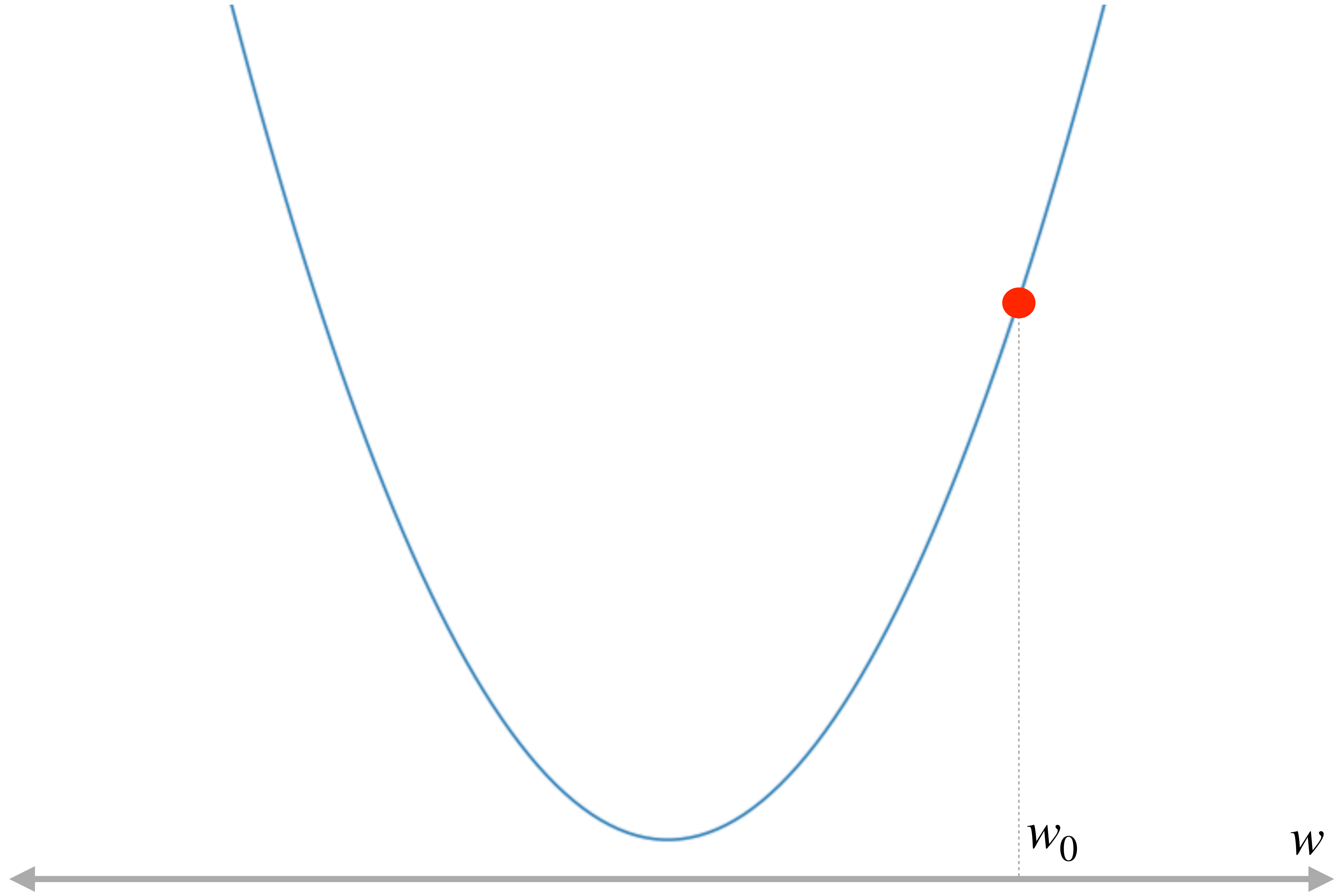
최적화(Optimization)

하강법(Descent Method)



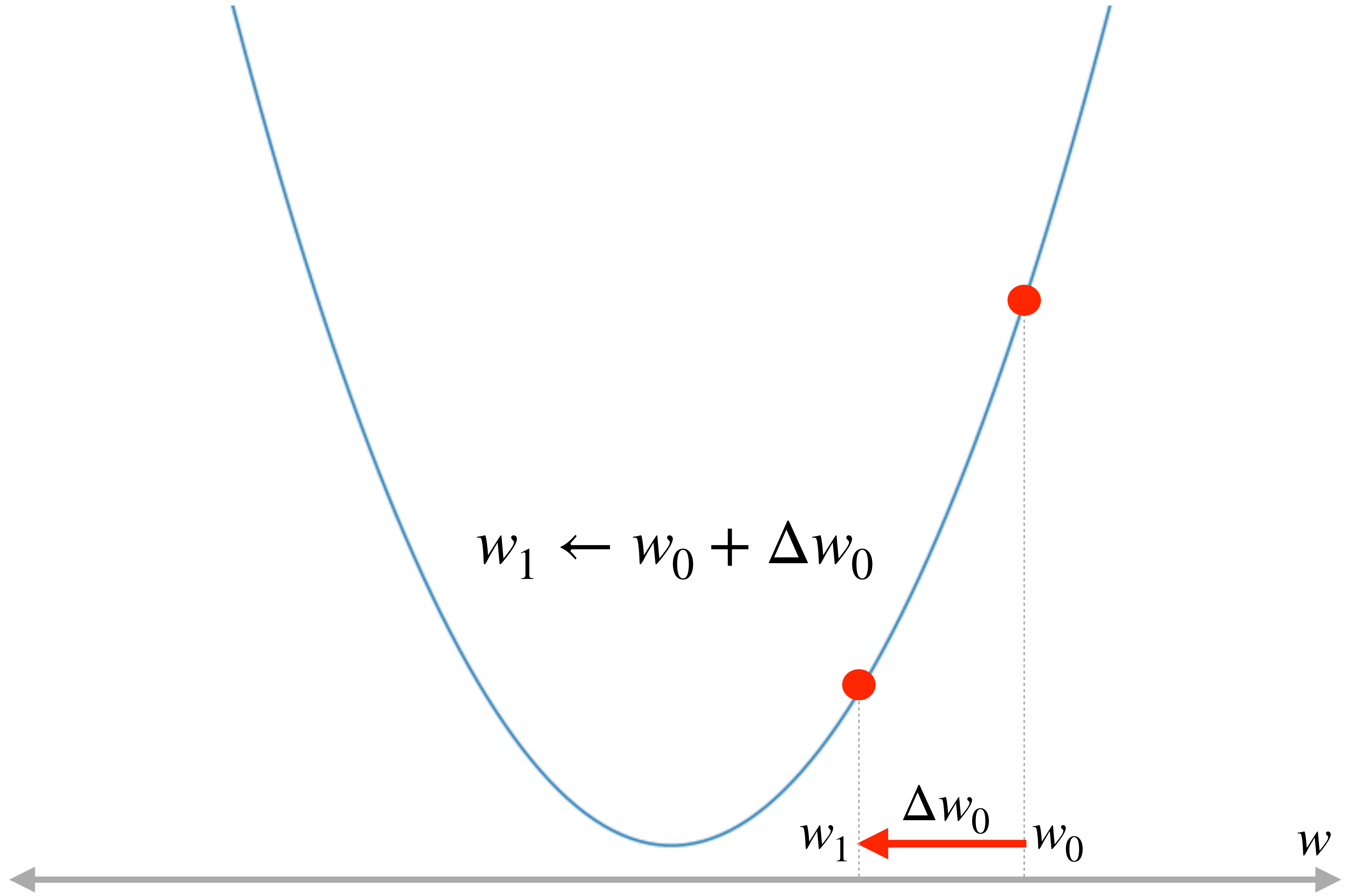
최적화(Optimization)

하강법(Descent Method)



최적화(Optimization)

하강법(Descent Method)

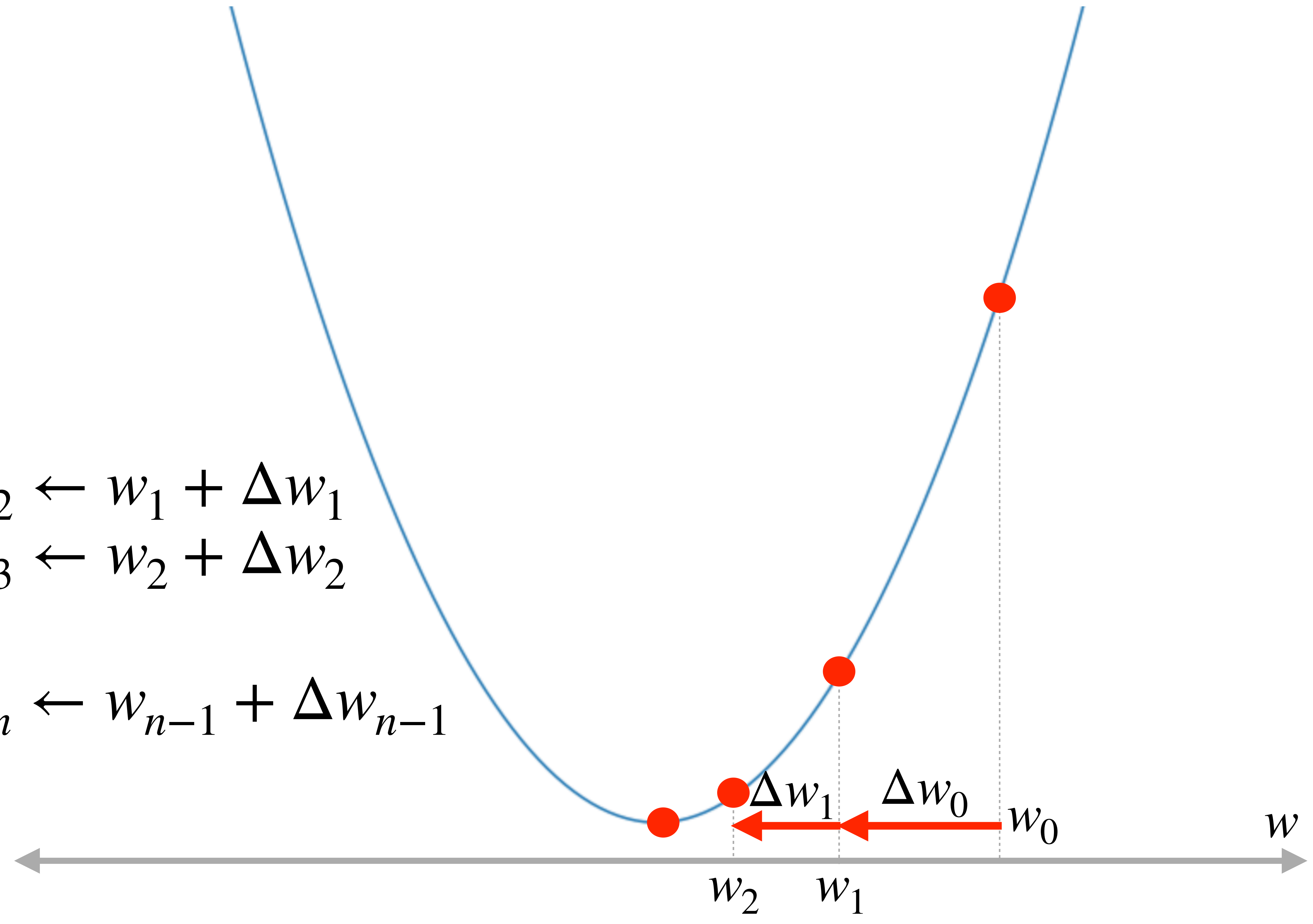


최적화(Optimization)

하강법(Descent Method)

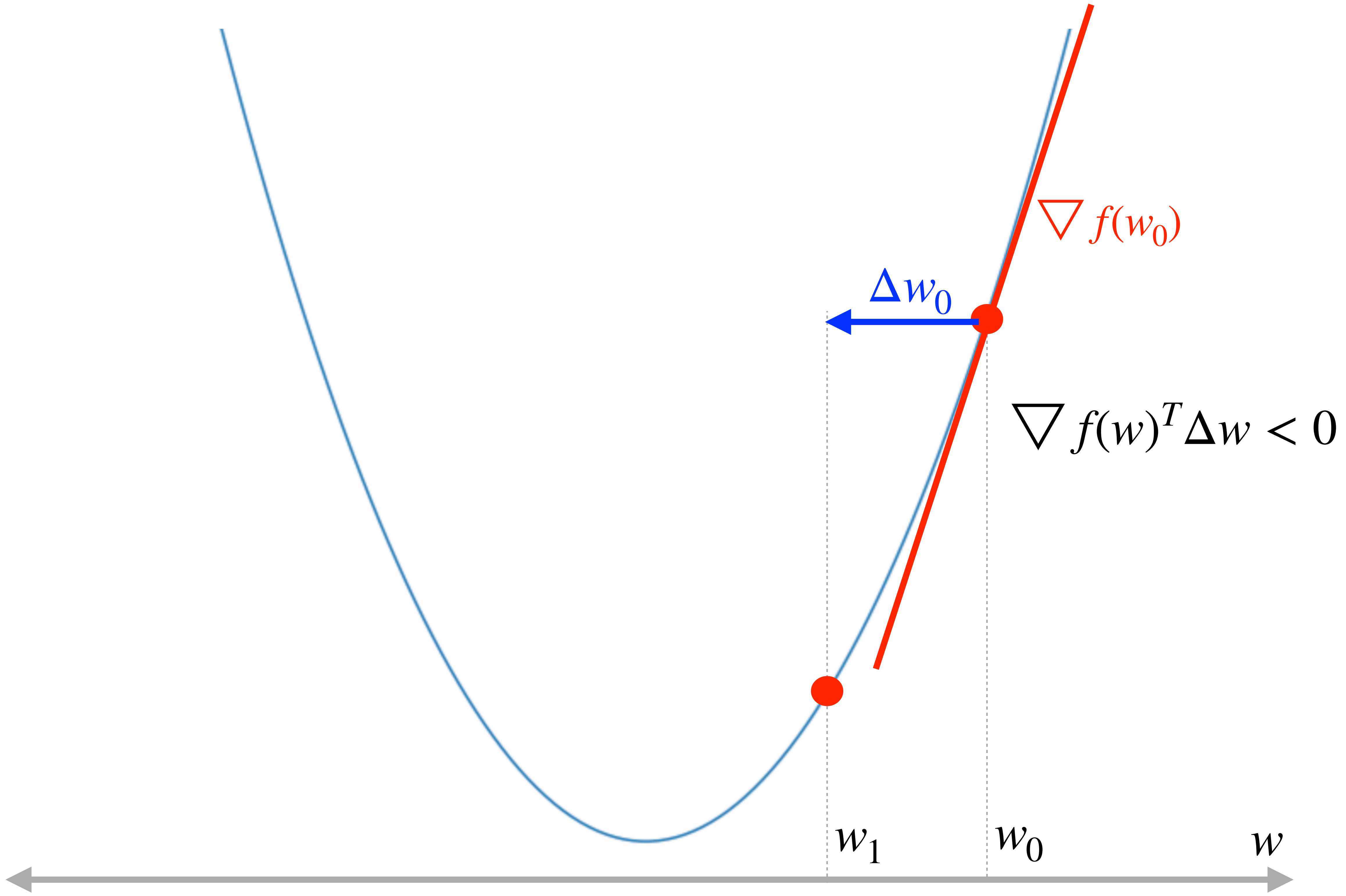
$$w_2 \leftarrow w_1 + \Delta w_1$$
$$w_3 \leftarrow w_2 + \Delta w_2$$

$$w_n \leftarrow w_{n-1} + \Delta w_{n-1}$$



최적화(Optimization)

하강법(Descent Method)



최적화(Optimization)

하강법(Descent Method)

학습률(learning rate, step size)

$$w \leftarrow w + \mu \Delta w, \triangledown f(w)^T \Delta w < 0$$

탐색 방향

- 탐색 방법 Δw에 따라 하강법이 정해진다.
- 경사 하강법, 뉴턴 방법 등

최적화(Optimization)

경사 하강법(Gradient Descent)

$$w \leftarrow w + \mu \Delta w, \nabla f(w)^T \Delta w < 0$$

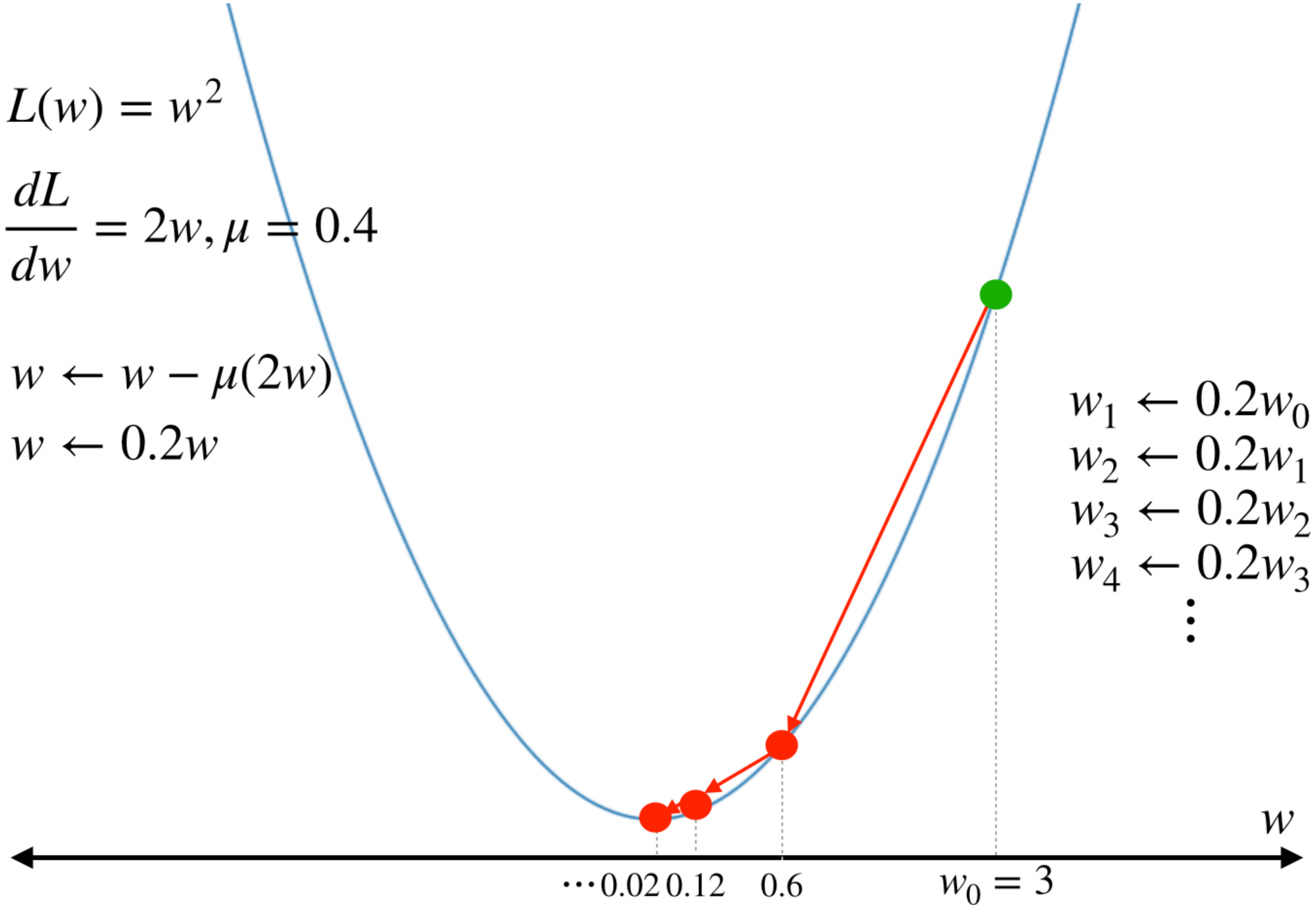
$$\Delta w = -\nabla f(w)$$

$$\nabla f(w)^T(-\nabla f(w)) = -\nabla f(w)^T \nabla f(w) < 0$$

$$w \leftarrow w - \mu \nabla f(w)$$

최적화(Optimization)

경사 하강법(Gradient Descent)



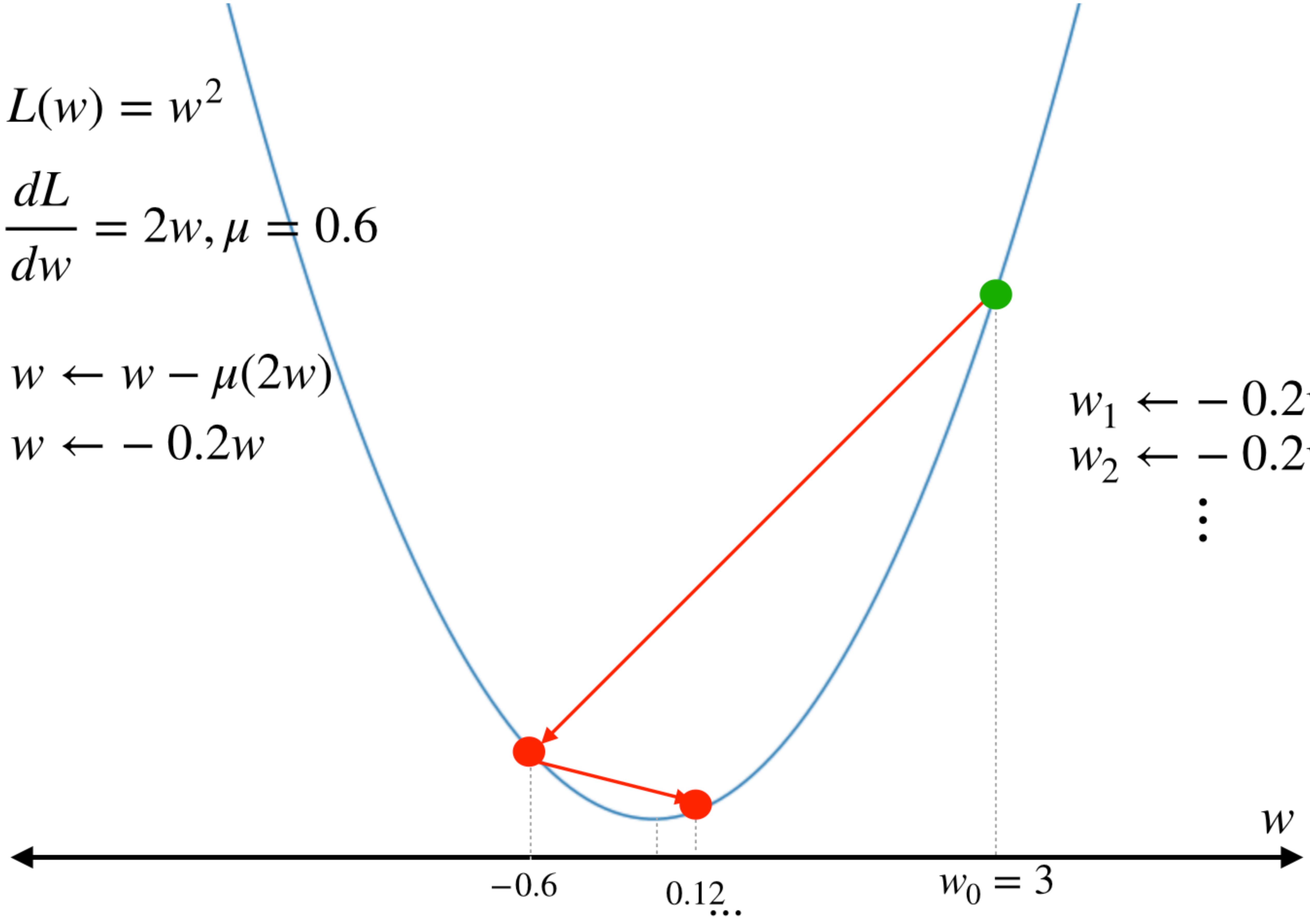
최적화(Optimization)

경사 하강법(Gradient Descent)

$$L(w) = w^2$$

$$\frac{dL}{dw} = 2w, \mu = 0.6$$

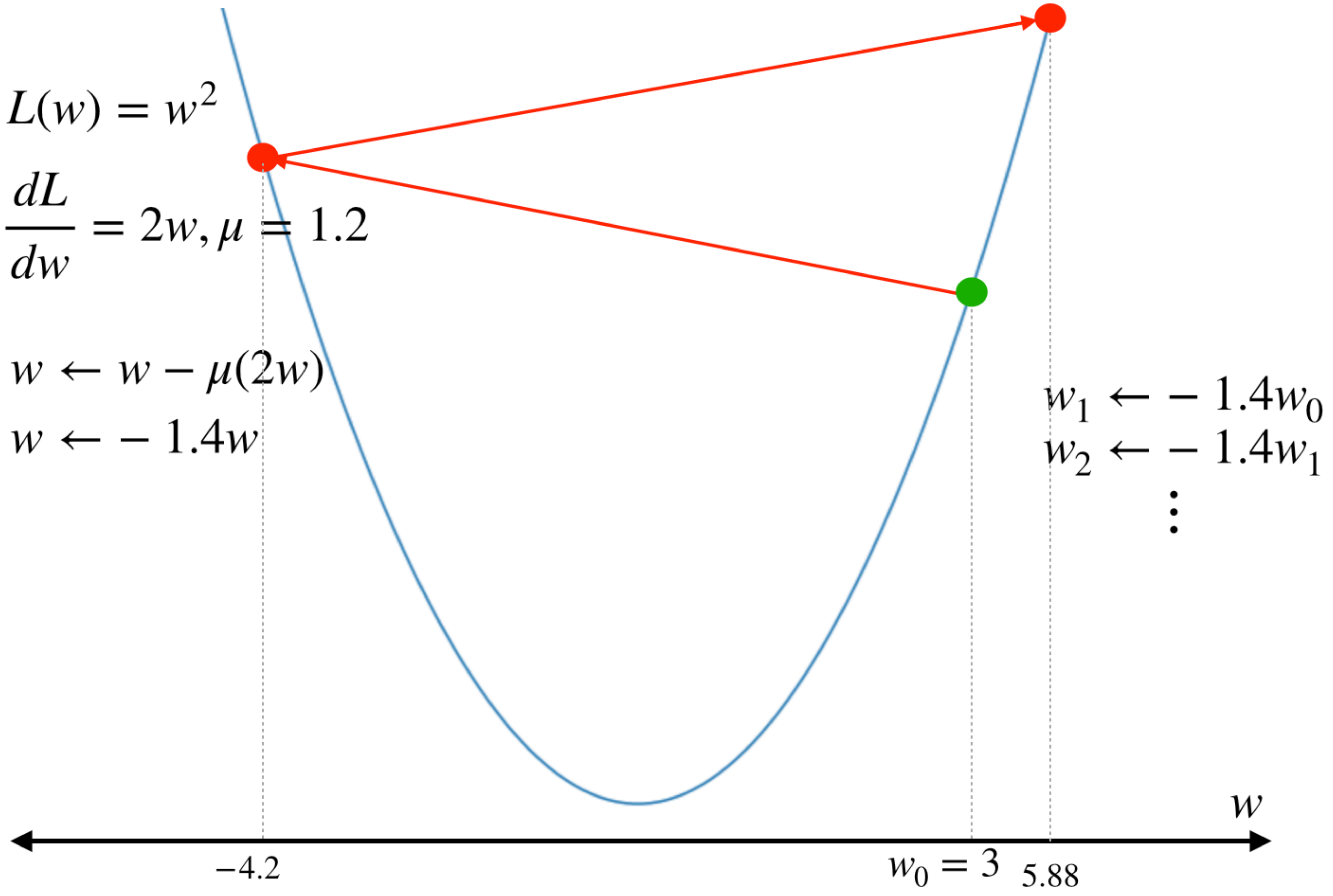
$$w \leftarrow w - \mu(2w)$$
$$w \leftarrow -0.2w$$



$$w_1 \leftarrow -0.2w_0$$
$$w_2 \leftarrow -0.2w_1$$
$$\vdots$$

최적화(Optimization)

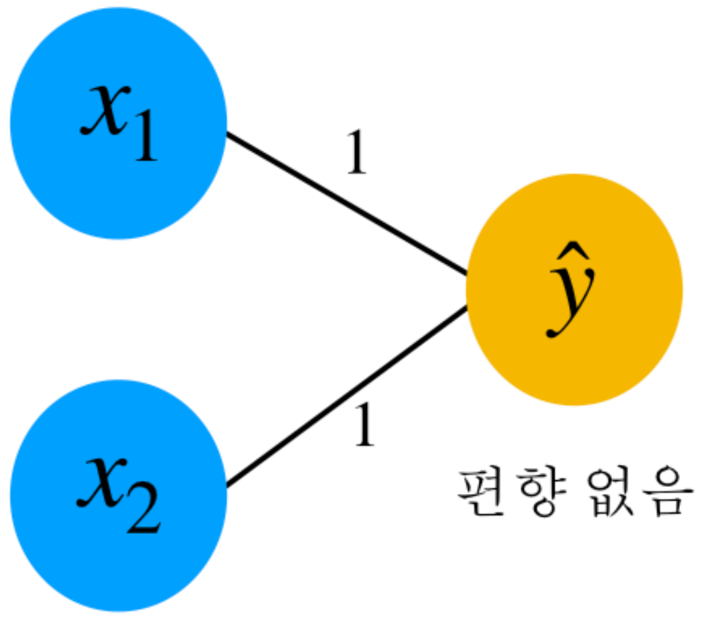
경사 하강법(Gradient Descent)



최적화(Optimization)

경사 하강법(Gradient Descent)

x_1	x_2	y
1	0	2
-1	-1	1
2	0	5



$$\hat{\mathbf{y}} = X\mathbf{v}^T = \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$$

$$L = \frac{1}{2}\sum_{i=1}^{3}(\hat{y}_i - y_i)^2 = \frac{1}{2}((1-2)^2 + (-2-1)^2 + (2-5)^2) = 9.5$$

$$\frac{\partial L}{\partial w} = \begin{pmatrix} \sum_{i=1}^{3}(\hat{y}_i - y_i)x_{i1} \\ \sum_{i=1}^{3}(\hat{y}_i - y_i)x_{i2} \end{pmatrix} = \begin{pmatrix} (1-2)\cdot 1 + (-2-1)\cdot(-1) + (2-5)\cdot 2 \\ (1-2)\cdot 0 + (-2-1)\cdot(-1) + (2-5)\cdot 0 \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \end{pmatrix}$$

$$w \leftarrow w - \mu\frac{\partial L}{\partial w} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 0.01\begin{pmatrix} -4 \\ 3 \end{pmatrix} = \begin{pmatrix} 1.04 \\ 0.97 \end{pmatrix}$$

최적화(Optimization)

경사 하강법(Gradient Descent)

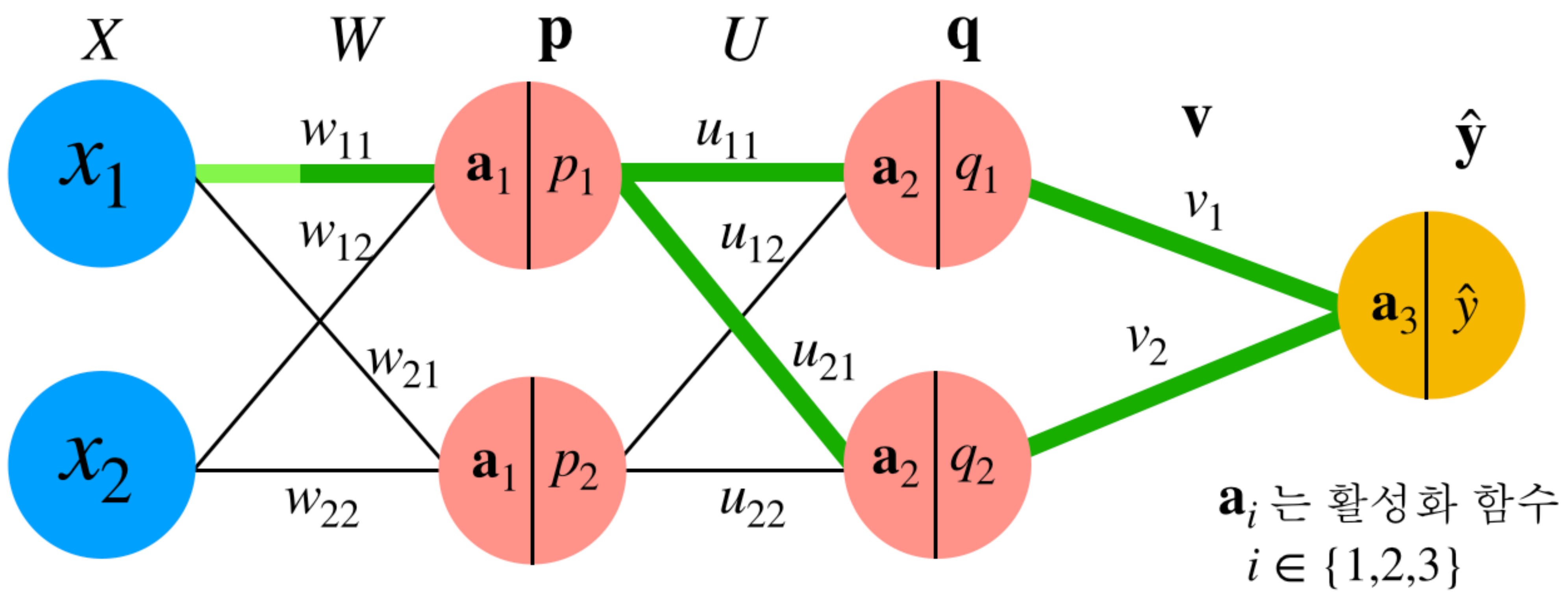
$$\hat{\mathbf{y}} = X\mathbf{v}^T = \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix} \xrightarrow{\text{업데이트}} \hat{\mathbf{y}} = X\mathbf{v}^T = \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1.04 \\ 0.97 \end{pmatrix} = \begin{pmatrix} 1.04 \\ -2.01 \\ 2.08 \end{pmatrix}$$

$$\hat{\mathbf{y}} = X\mathbf{v}^T = \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1.04 \\ 0.97 \end{pmatrix} = \begin{pmatrix} 1.04 \\ -2.01 \\ 2.08 \end{pmatrix}$$

$$L = \frac{1}{2}\sum_{i=1}^{3}(\hat{y}_i - y_i)^2 = \frac{1}{2}((1.04-2)^2 + (-2.01-1)^2 + (2.08-5)^2) = 9.25$$

최적화(Optimization)

경사 하강법(Gradient Descent)



$$\frac{\partial \hat{y}}{\partial w_{11}} = \frac{\partial \hat{y}}{\partial q_1}\frac{\partial q_1}{\partial w_{11}} + \frac{\partial \hat{y}}{\partial q_2}\frac{\partial q_2}{\partial w_{11}} = \frac{\partial \hat{y}}{\partial q_1}\frac{\partial q_1}{\partial p_1}\frac{\partial p_1}{\partial w_{11}} + \frac{\partial \hat{y}}{\partial q_2}\frac{\partial q_2}{\partial p_1}\frac{\partial p_1}{\partial w_{11}}$$

최적화(Optimization)

경사 하강법(Gradient Descent)

$$\begin{pmatrix} w_{11} \\ w_{12} \\ \vdots \\ v_1 \\ v_2 \\ b_1 \\ b_2 \\ c \end{pmatrix} \leftarrow \begin{pmatrix} w_{11} \\ w_{12} \\ \vdots \\ v_1 \\ v_2 \\ b_1 \\ b_2 \\ c \end{pmatrix} - \mu \begin{pmatrix} \frac{\partial L}{\partial w_{11}} \\ \frac{\partial L}{\partial w_{12}} \\ \vdots \\ \frac{\partial L}{\partial v_1} \\ \frac{\partial L}{\partial v_2} \\ \frac{\partial L}{\partial b_1} \\ \frac{\partial L}{\partial b_2} \\ \frac{\partial L}{\partial c} \end{pmatrix}$$