ML/DS 지식 따라가기

경사하강법, Forward / Back Propagation

미분과 도함수

미분 : 어떤 함수의 정의역 속 각 점에서 함숫값의 변화량과 독립 변숫값의 변화량 비의 극한

- $\lim_{\Delta x \to 0} \frac{f(a + \Delta x) - f(a)}{\Delta x}$
- $\lim_{h \to 0} \frac{f(a + h) - f(a)}{h}$
- $\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$

미분과 도함수

도함수 : 어떤 함수의 정의역 속 각 점에서 함숫값의 변화량과 독립 변숫값의 변화량 비의 극한들로 치역이 구성되는 새로운 함수

> 함수 f의 정의역의 원소 x에 다음 극한값: $m_x = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$가 존재하면 m_x 함수를 f의 **도함수**라 한다.

편미분과 편도함수

편미분 : 다변수 함수의 특정 변수를 제외한 나머지 변수를 상수로 간주하여 미분하는 것

$$z=2x^2y \xrightarrow[x\text{를 상수 취급: } z=(2x^2)y]{y\text{에 대하여 편미분}} \frac{\partial z}{\partial y}=2x^2\times 1 =2x^2$$

$$\xrightarrow[y\text{를 상수 취급: } z=(2y)x^2]{x\text{에 대하여 편미분}} \frac{\partial z}{\partial x}=2y\times 2x =4xy$$

편미분과 편도함수

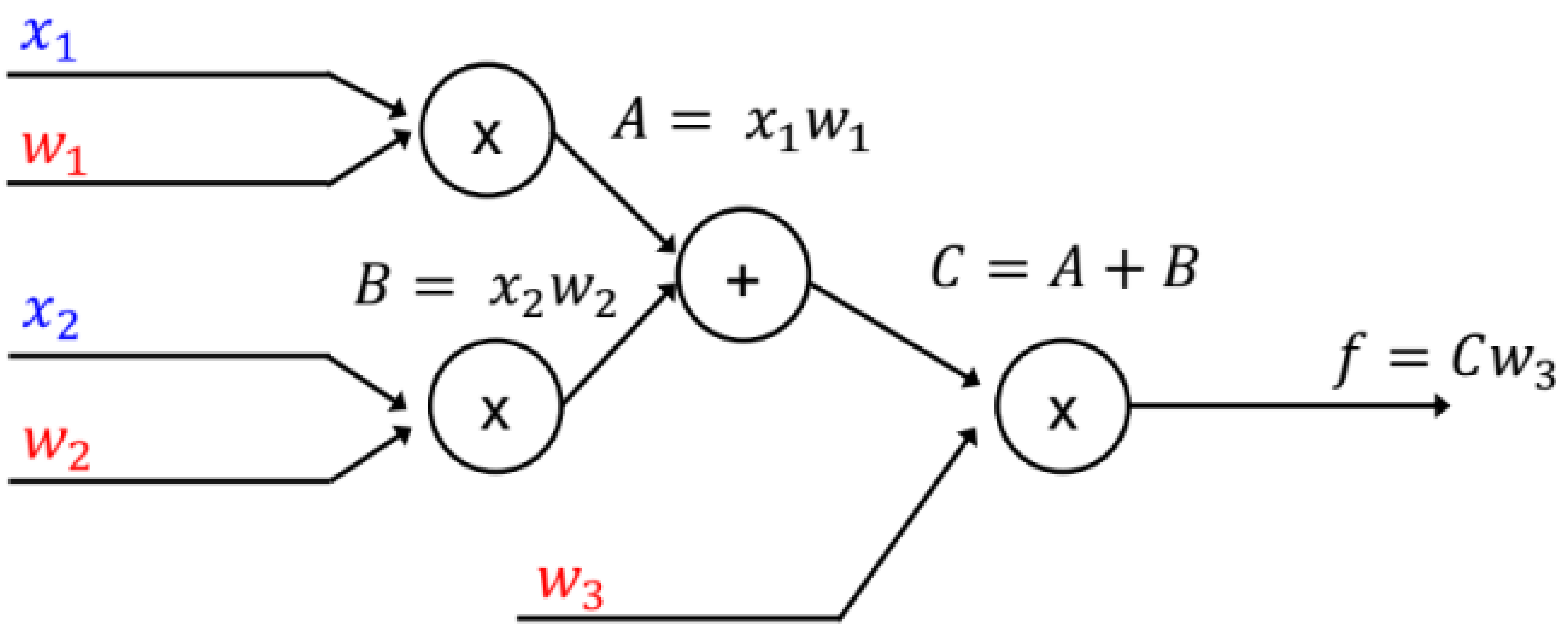
편도함수 : 독립변수가 2개 이상인 다변수함수에서의 도함수

$$z = 2x^2 y \xrightarrow[x\text{를 상수 취급: } z=(2x^2)y]{y\text{에 대하여 편미분}} \frac{\partial z}{\partial y} = 2x^2 \times 1 = 2x^2$$

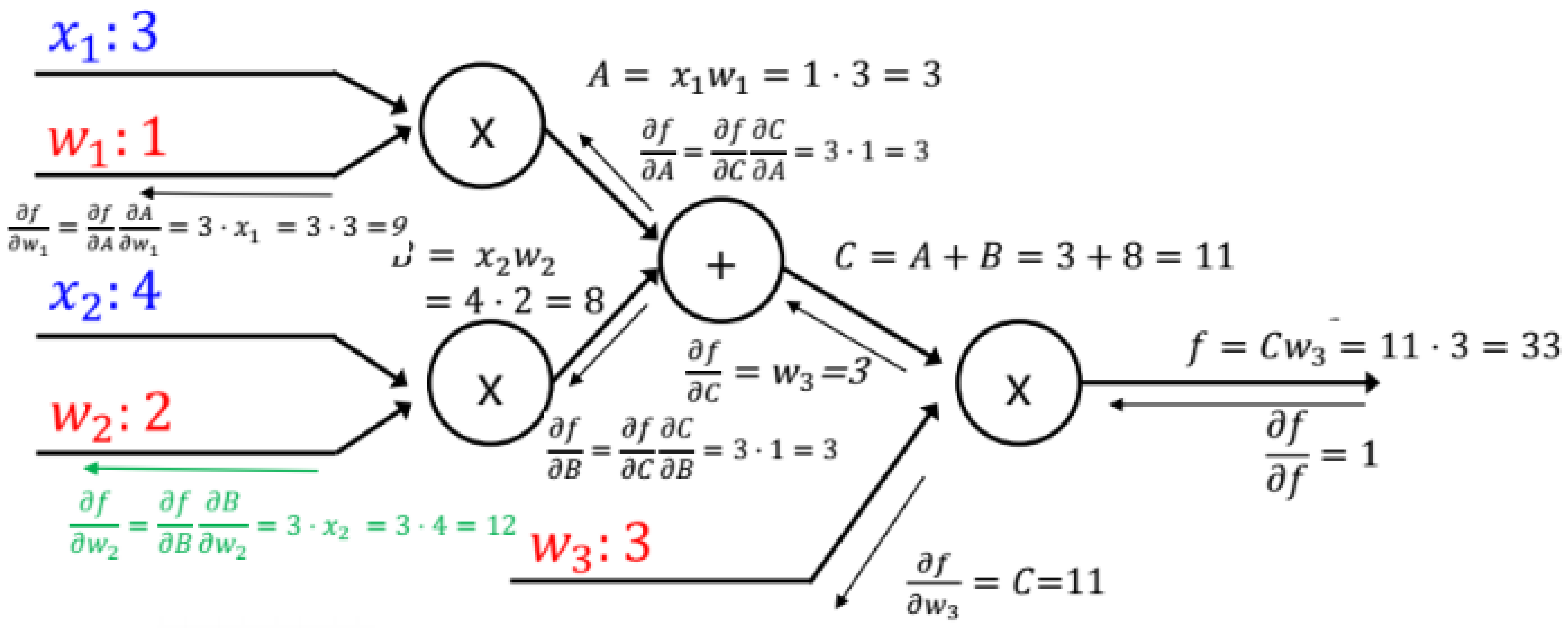
$$\xrightarrow[y\text{를 상수 취급: } z=(2y)x^2]{x\text{에 대하여 편미분}} \frac{\partial z}{\partial x} = 2y \times 2x = 4xy$$

FORWARD PROPAGATION

$$f = (x_1 w_1 + x_2 w_2) w_3$$



BACKWARD PROPAGATION

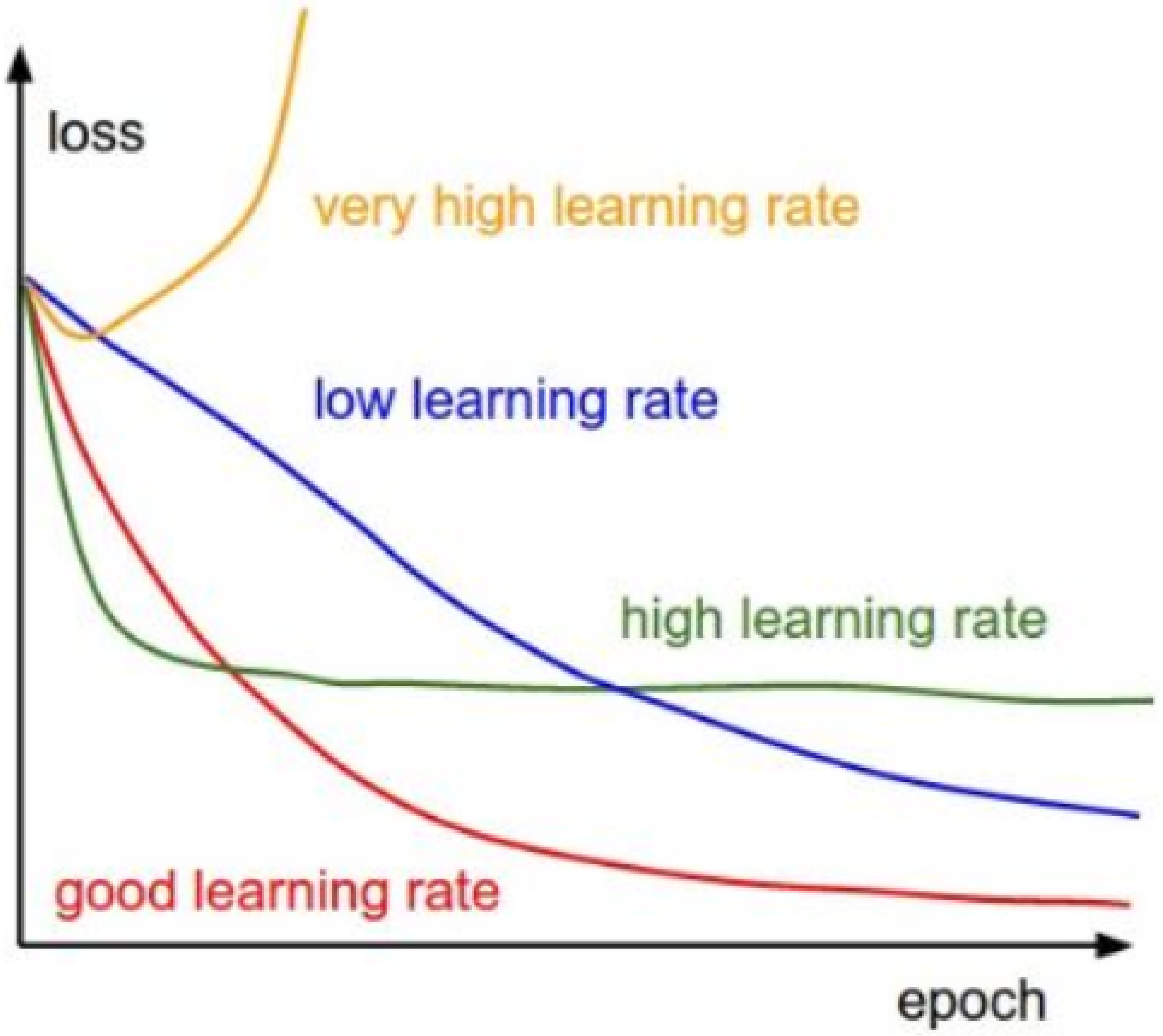


BACKWARD PROPAGATION WITH COST FUNCTION

$$\frac{\partial f}{\partial w_1} = \frac{\partial f}{\partial A}\frac{\partial A}{\partial w1} \longrightarrow \frac{\partial C}{\partial w_1} = \frac{\partial C}{\partial f}\frac{\partial f}{\partial A}\frac{\partial A}{\partial w1}$$

BEFORE UPDATE - LEARNING RATE

학습률 : loss function의 최소값을 향해 이동할 때 그 보폭을 결정하는 파라미터

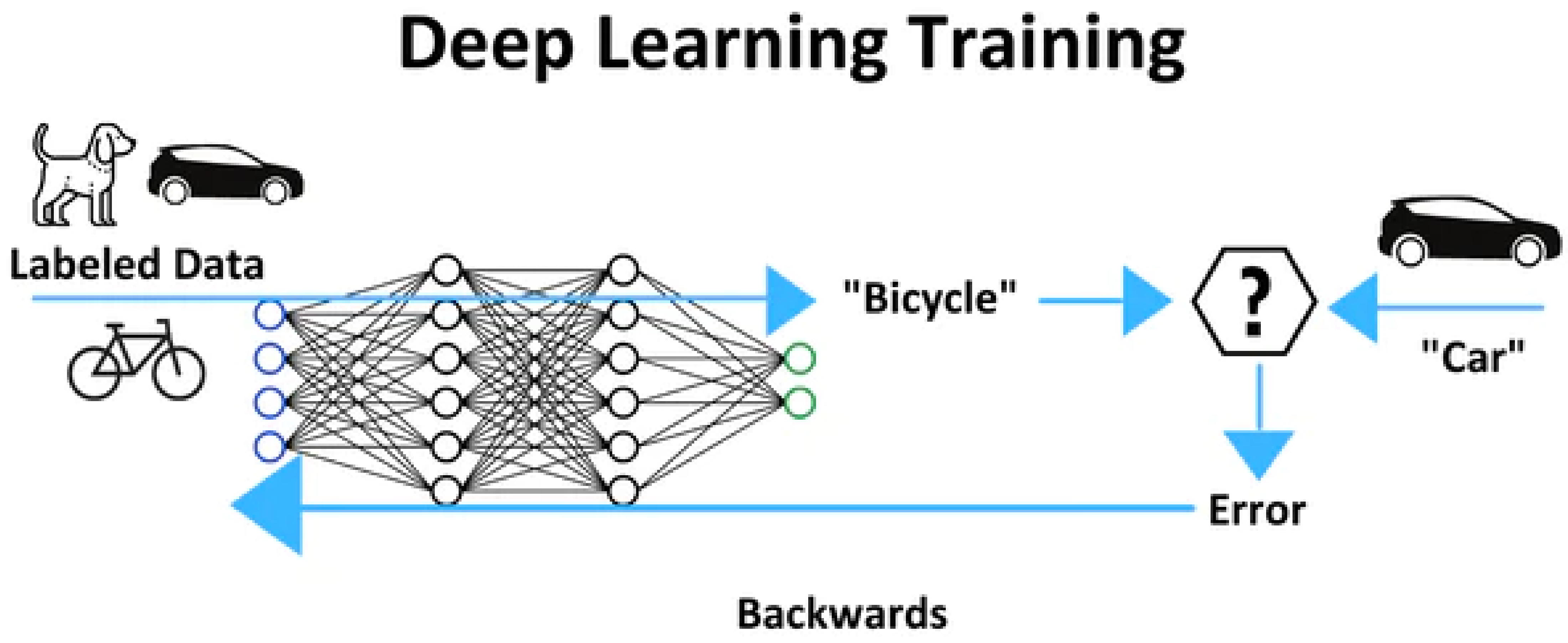


UPDATE THE WEIGHT

$$w := w - \alpha \frac{\partial c}{\partial w}$$

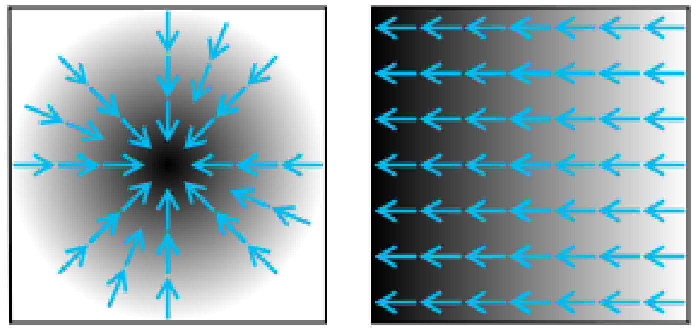
TRAINING의 목표

weight를 바꿔가면서 loss function의 최소값에 도달하는 것



GRADIENT

어렵게 설명하기 : 벡터 미적분학에서 스칼라장의 최대의 증가율을 나타내는 벡터장



쉽게 설명하기 : 그래프에서의 기울기

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

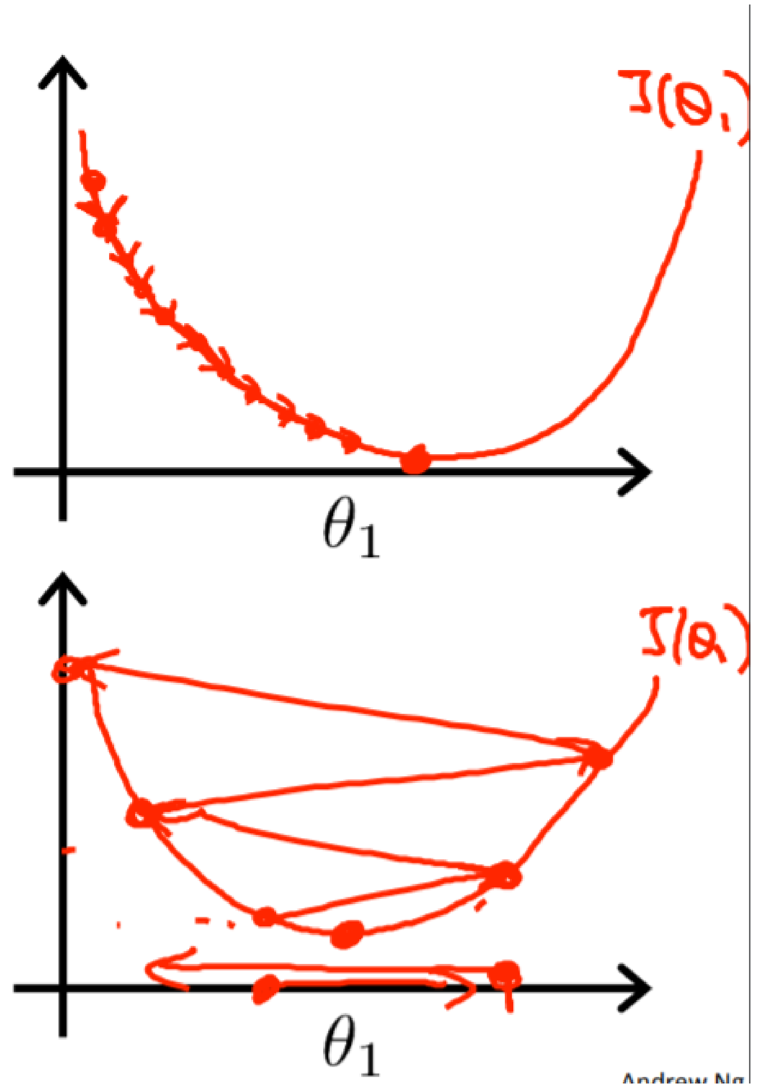
GRADIENT DESCENT

비용함수를 최소화하는 값을 찾기

$$\boxed{\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_1)}$$

If α is too small, gradient descent can be slow.

If α is too large, gradient descent can overshoot the minimum. It may fail to converge, or even diverge.



GRADIENT DESCENT AND ……

비용함수를 최소화하는 값을 찾기

이러한 특성을 가지는 다른 함수들은?

Optimization

Thank you