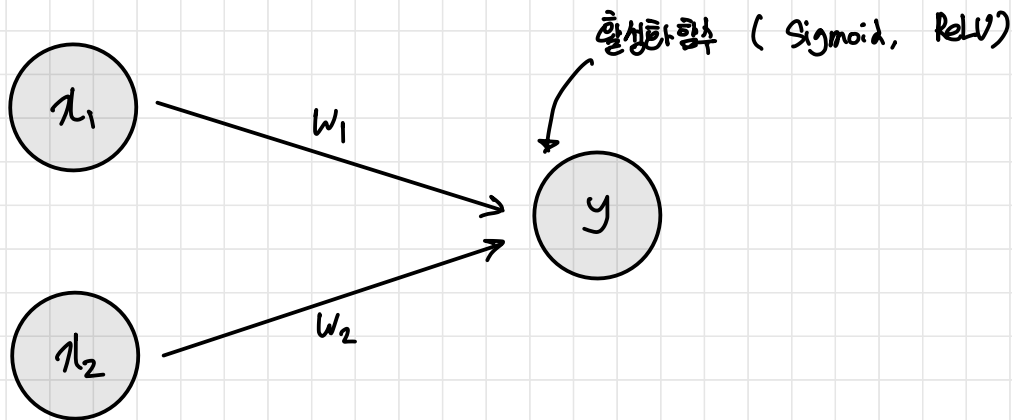


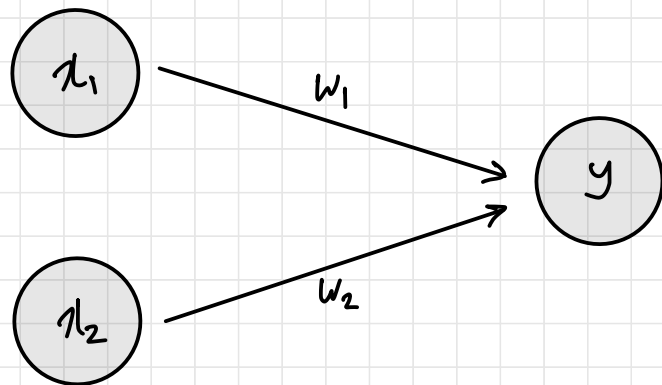
# Ch4. 신경망 학습

→ 데이터로부터 가중치 매개변수의 최적값을 찾는 것.

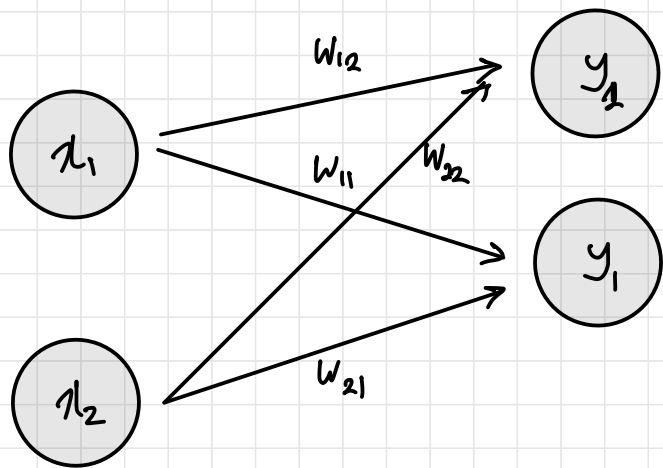
- 학습
- ① 미분배치
  - ② 가중치 산출 → 손실함수 (loss function), 경사하강법 (Gradient Descent)
  - ③ 매개변수 (가중치 값) 갱신



1) 손실함수 (Loss function) → 최적의 매개변수를 결정하는 지표



$$y = [x_1 \ x_2] \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = [x_1 w_1 + x_2 w_2]$$



$$y = [x_1 \ x_2] \cdot \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} = [y_1 \ y_2]$$

$\Downarrow$                        $\Downarrow$                        $\Downarrow$

$$X \cdot W = Y$$

$$y = [a \ b \ c \ d] \rightarrow \text{신경망 예측값}$$

$$t = [0 \ 0 \ 1 \ 0] \rightarrow \text{정답값}$$

\* Mean - Squared Error (평균 제곱 오차)

$$F = \frac{1}{2} \sum_k (y_k - t_k)^2 = \frac{1}{2} (a^2 + b^2 + (c-1)^2 + d^2) \rightarrow \text{결과는 매개변수의 함수로 표현}$$

\* Cross-entropy Error (교차 엔트로피 오차)

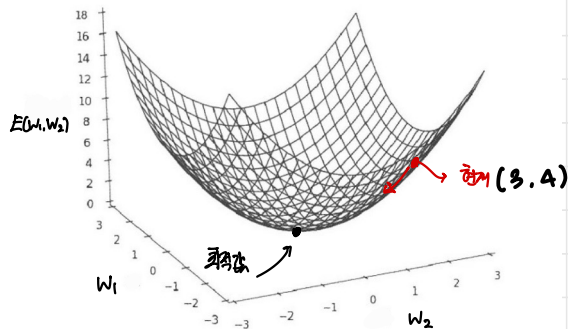
$$F = - \sum_k t_k \cdot \log y_k = - \log C$$

- 최적화 과정에서 이용할 것은 기울기  $\rightarrow$  '정확도'를 이용하면 계단식 함수 이므로 최적화 어려움  
 $\rightarrow$  손실함수를 이용.

## 2) 경사하강법 (Gradient Descent)

○ 평미분  $\Rightarrow$  손실함수 :  $w_1, w_2 \dots w_n$  의 함수로 표현.

$$F(w_1, w_2) = w_1^2 + w_2^2 \text{ 라고 가정}$$



$$F(w_1, 4) = w_1^2 + 16$$

$$F(3, w_2) = 9 + w_2^2$$

$$\frac{\partial F}{\partial w_1} = 2w_1 = 6$$

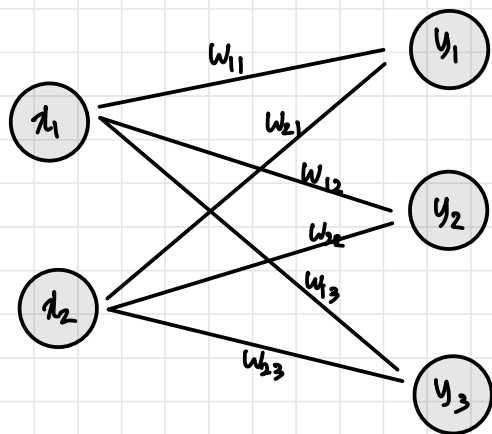
$$\frac{\partial F}{\partial w_2} = 2w_2 = 8$$

$(-6, -8) \rightarrow$  기울기로 접근할 때 optimal point로.

$$\therefore \left. \begin{aligned} w_1 &= w_1 - \eta \cdot \frac{\partial F}{\partial w_1} \\ w_2 &= w_2 - \eta \cdot \frac{\partial F}{\partial w_2} \end{aligned} \right\} \text{경사 하강}$$

\*  $\eta$  (Learning rate)

$\rightarrow$  너무 크면 발산  
너무 작으면 수렴 X



1) 신경망 연산.

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \cdot \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}$$

$$X \cdot W_0 = Y$$

$$t_k = [t_1 \ t_2 \ t_3]$$

2) 손실함수 (MSE)

$$L = \frac{1}{2} \sum_k (t_k - y_k)^2 = \frac{1}{2} ((t_1 - y_1)^2 + (t_2 - y_2)^2 + (t_3 - y_3)^2)$$

$$= \frac{1}{2} ((t_1 - (x_1 \cdot w_{11} + x_2 \cdot w_{21}))^2 + (t_2 - (x_1 \cdot w_{12} + x_2 \cdot w_{22}))^2 + (t_3 - (x_1 \cdot w_{13} + x_2 \cdot w_{23}))^2)$$

3) 경사하강법.

$$W_0 = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}$$

$$\frac{\partial L}{\partial W} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \end{bmatrix}$$

$$W_1 = W_0 - \eta \cdot \frac{\partial L}{\partial W}$$

## 2) 미니배치 학습.

GD

이 전까지는 데이터 1개의 손실함수를 다룬.  $\rightarrow$  전체 데이터 ( $N$ ) 의 손실함수의 평균을 사용.

$\rightarrow$  데이터의 수가 많아지면 계산이 어려움.

$\rightarrow$  일부분만 골라서 학습을 수행 (미니배치 학습)

SGD