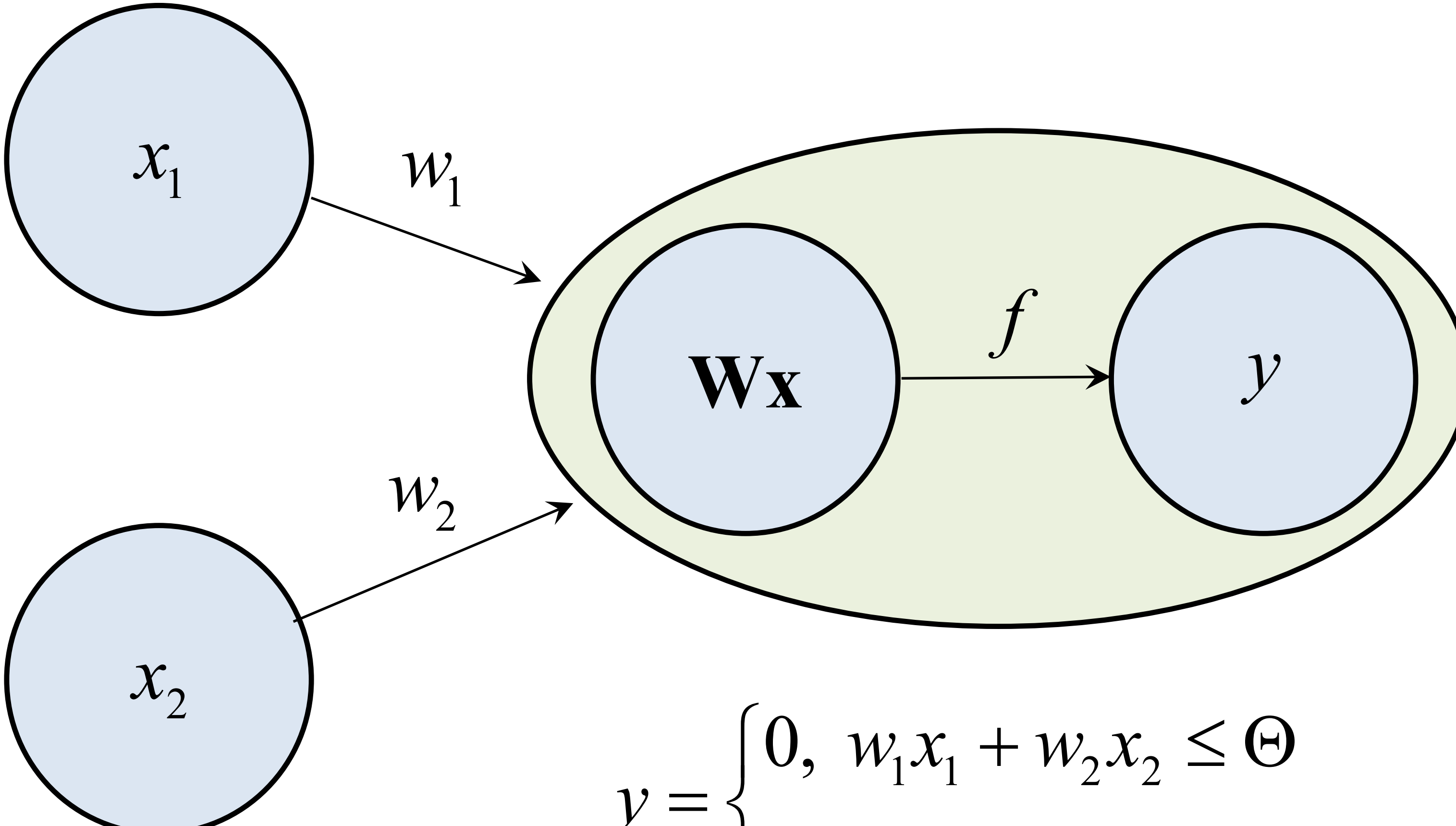
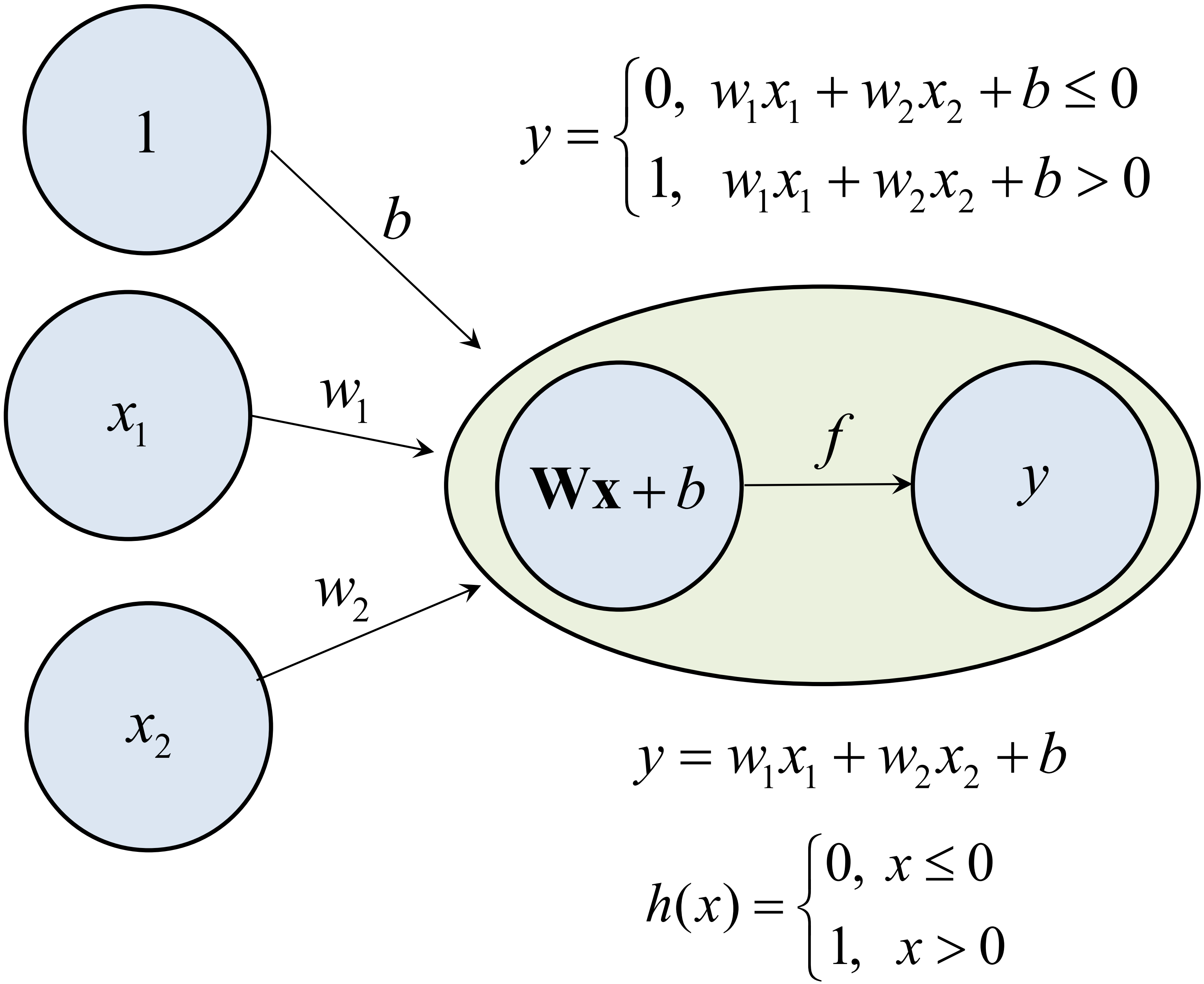
Neural Network

3. Neural Network

Perceptron



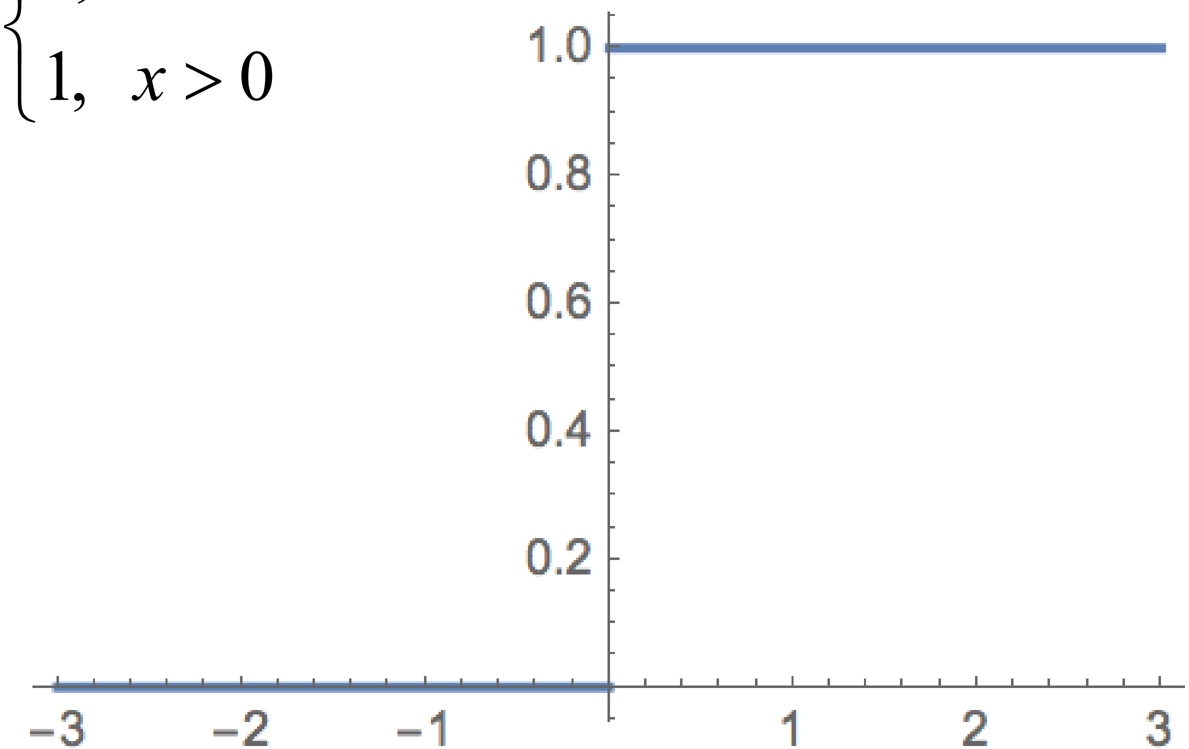
$$y = \begin{cases} 0, & w_1 x_1 + w_2 x_2 \leq \Theta \\ 1, & w_1 x_1 + w_2 x_2 > \Theta \end{cases}$$



1
b
x_1
w_1
x_2
w_2
$\mathbf{Wx}+b$
f
y
$y=\begin{cases}0, & w_1x_1+w_2x_2+b\le 0\\ 1, & w_1x_1+w_2x_2+b>0\end{cases}$
$y=w_1x_1+w_2x_2+b$
$h(x)=\begin{cases}0, & x\le 0\\ 1, & x>0\end{cases}$

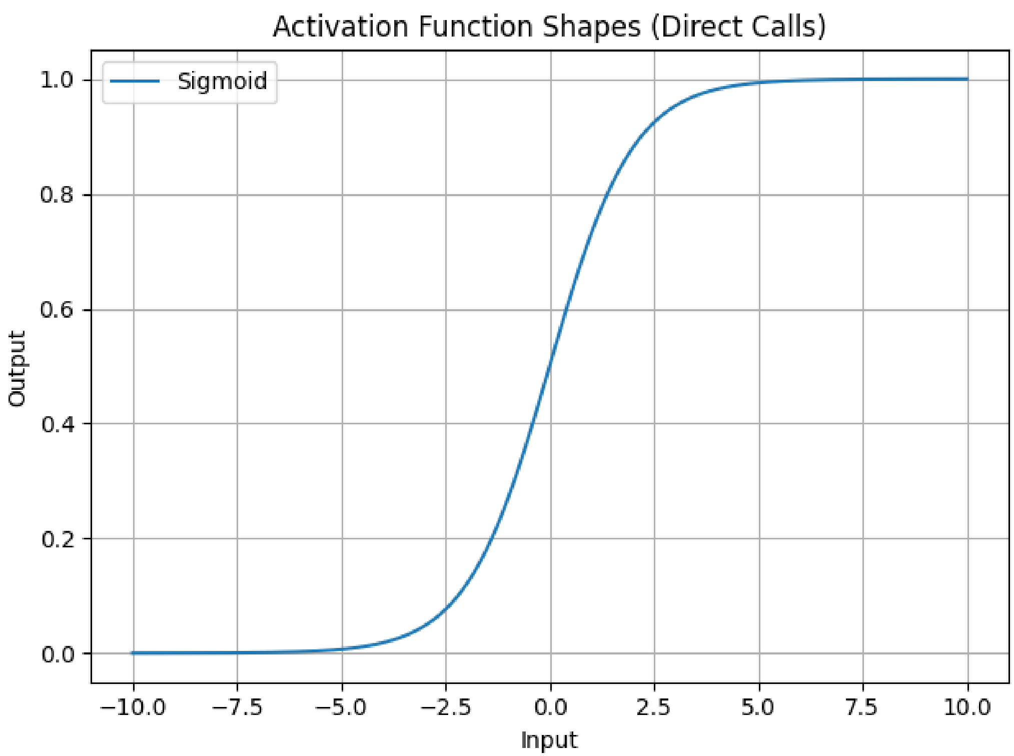
Activation Function

$$h(x) = \begin{cases} 0, & x \le 0 \\ 1, & x > 0 \end{cases}$$



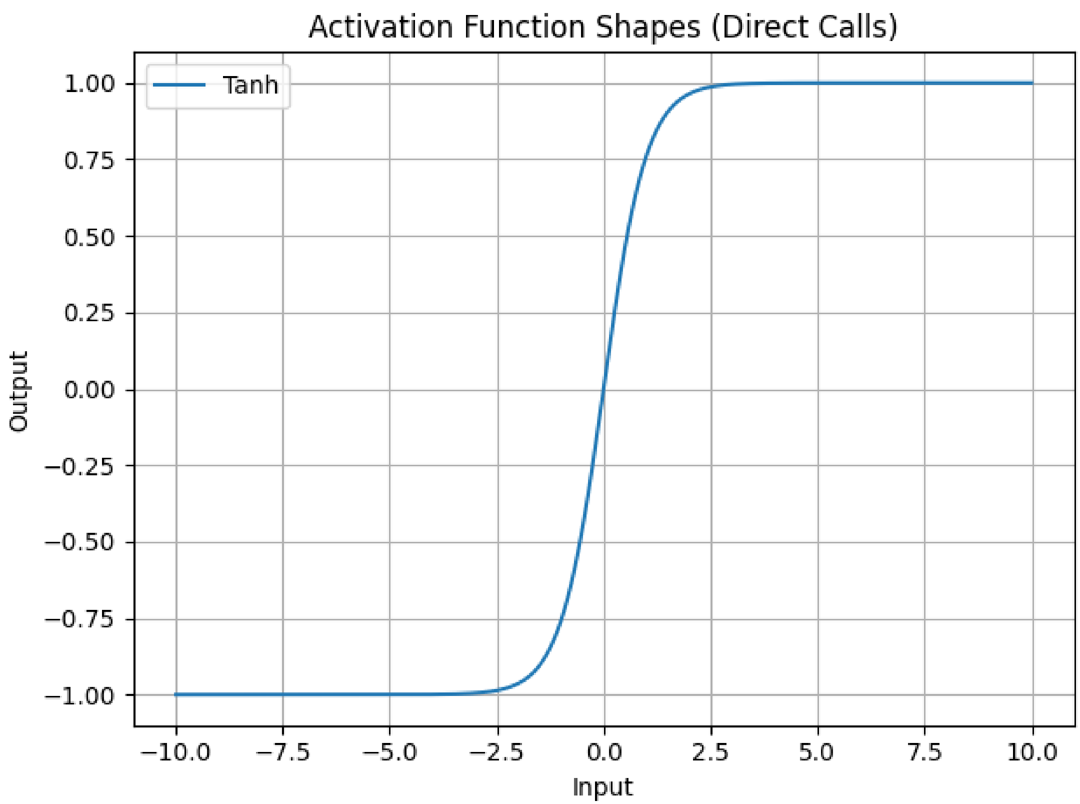
Sigmoid

$$f(x) = \frac{1}{1+e^{-x}}$$



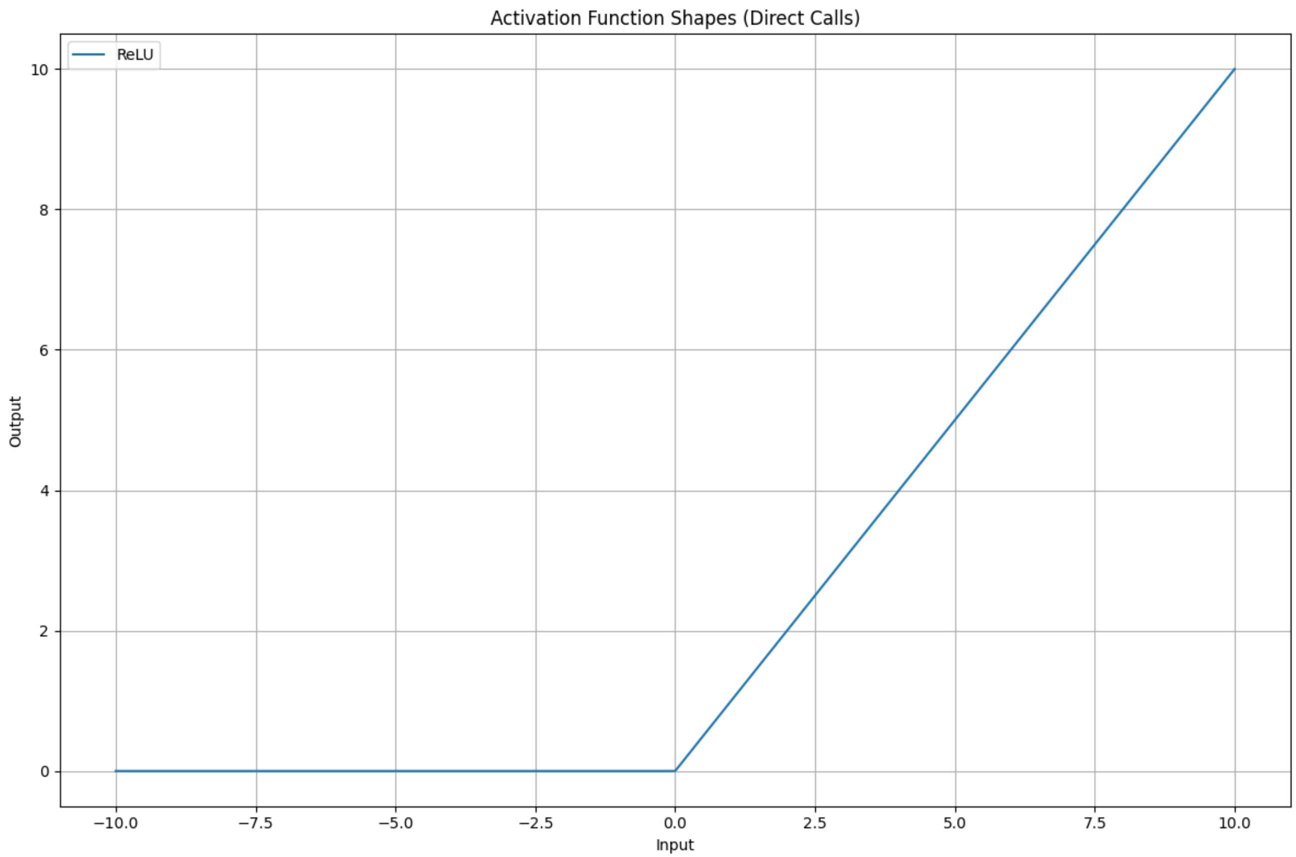
Tanh

$$f(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad sigmoid = \frac{1}{2}\left\{1 + \tanh(\frac{x}{2})\right\}$$



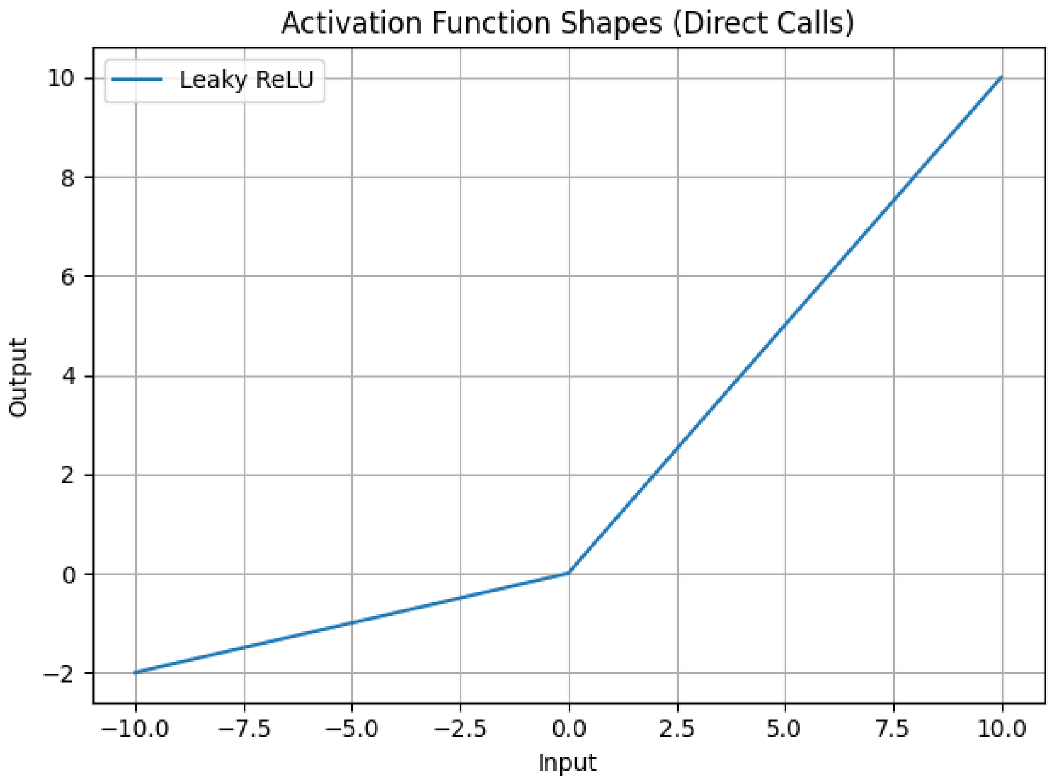
ReLU (Recitified Linear Unit)

$$f(x) = \max(0, x)$$



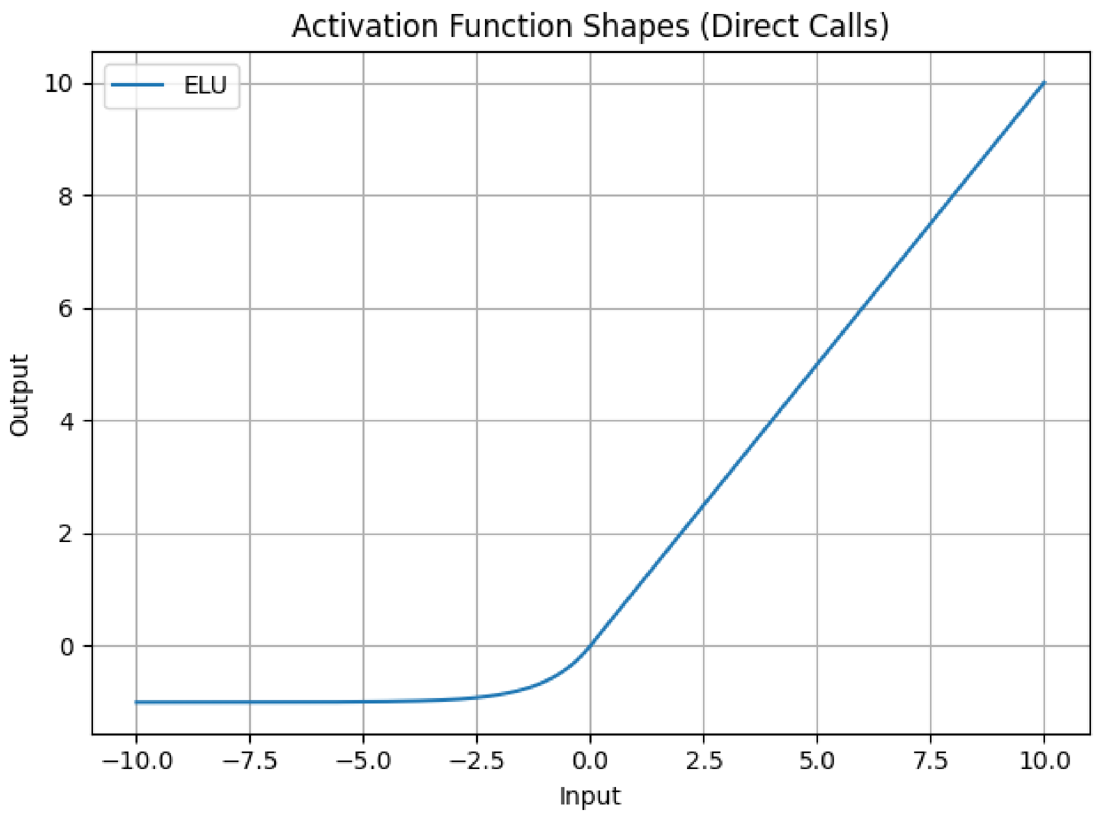
Leaky ReLU

$$f(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$$



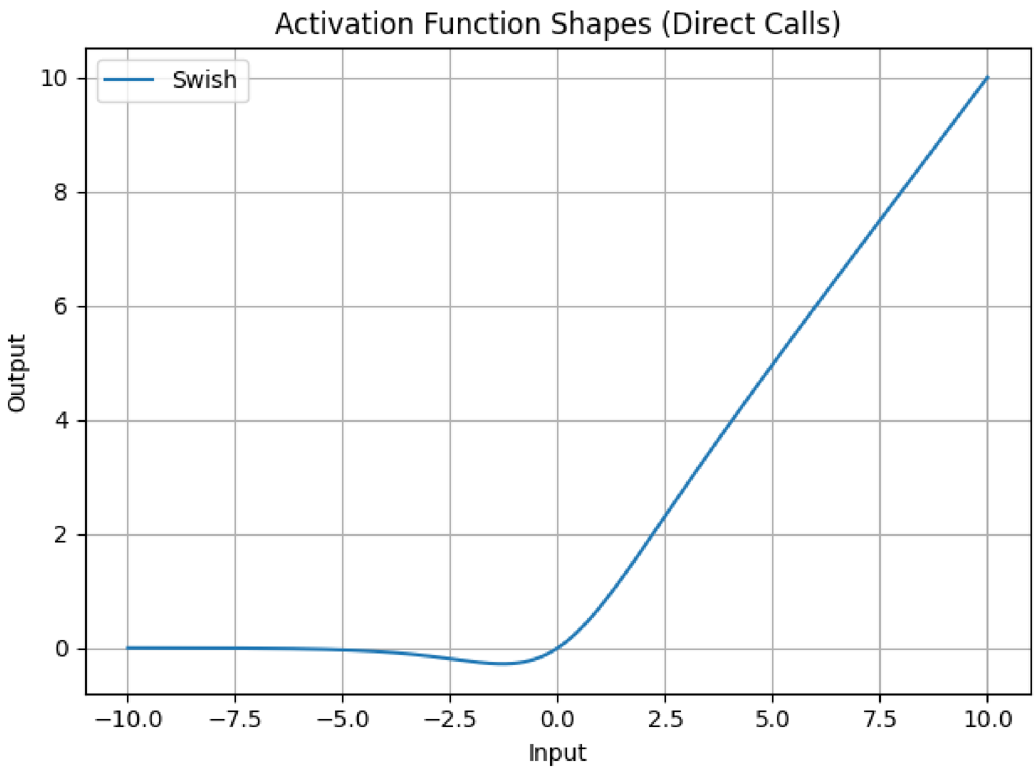
ELU

$$f(x)=\begin{cases} x, \ x \geq 0 \\ \alpha(e^{x}-1), \ \ x<0 \end{cases}$$



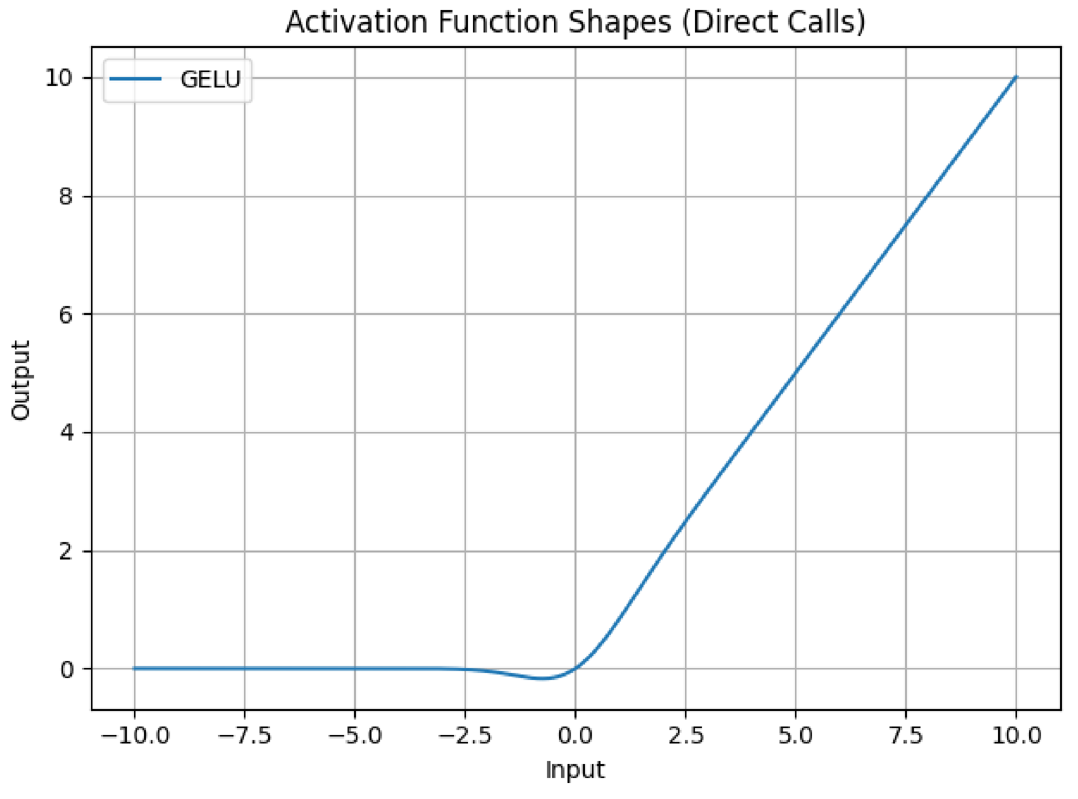
SWISH

$$f(x) = x \cdot sigmoid(x)$$

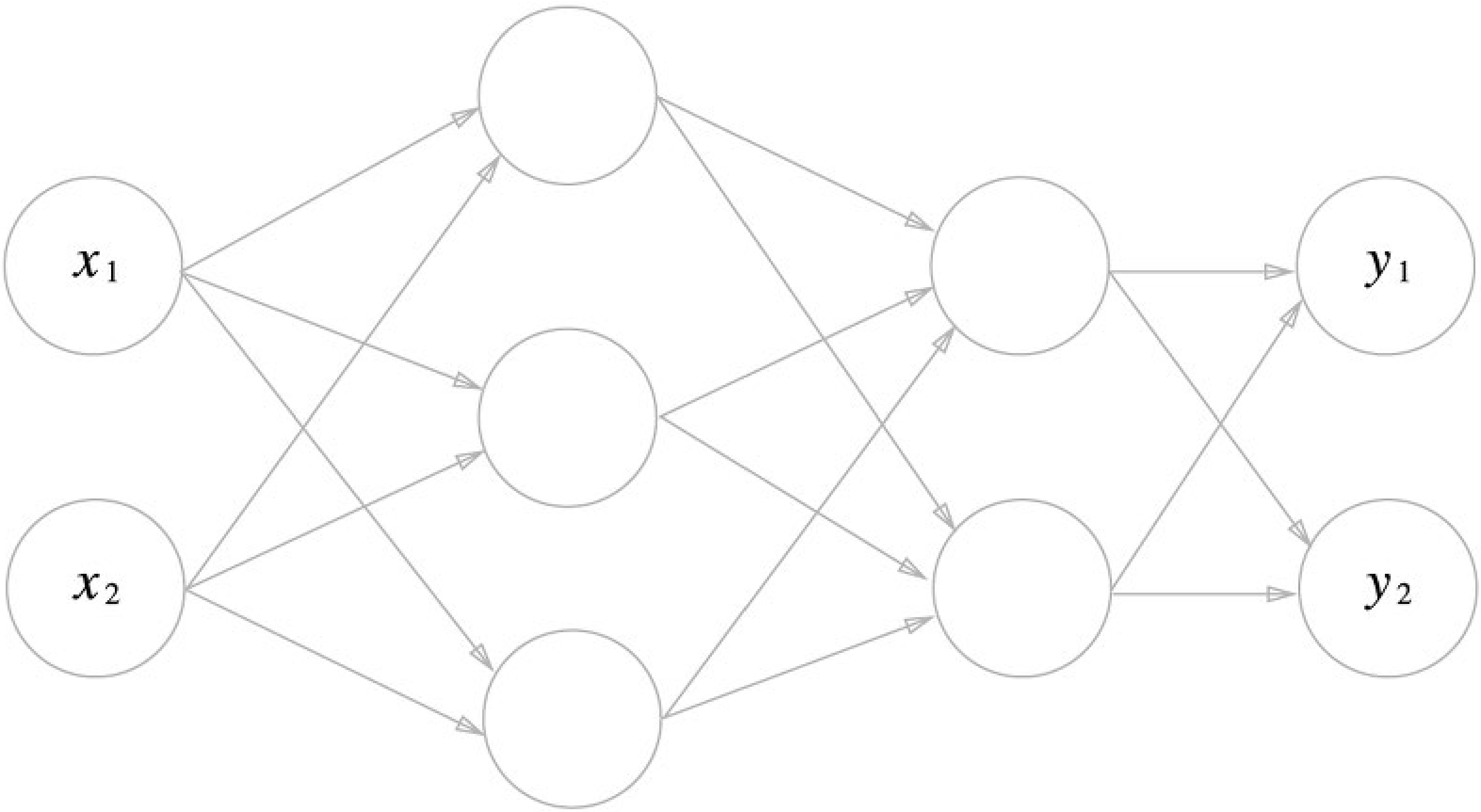


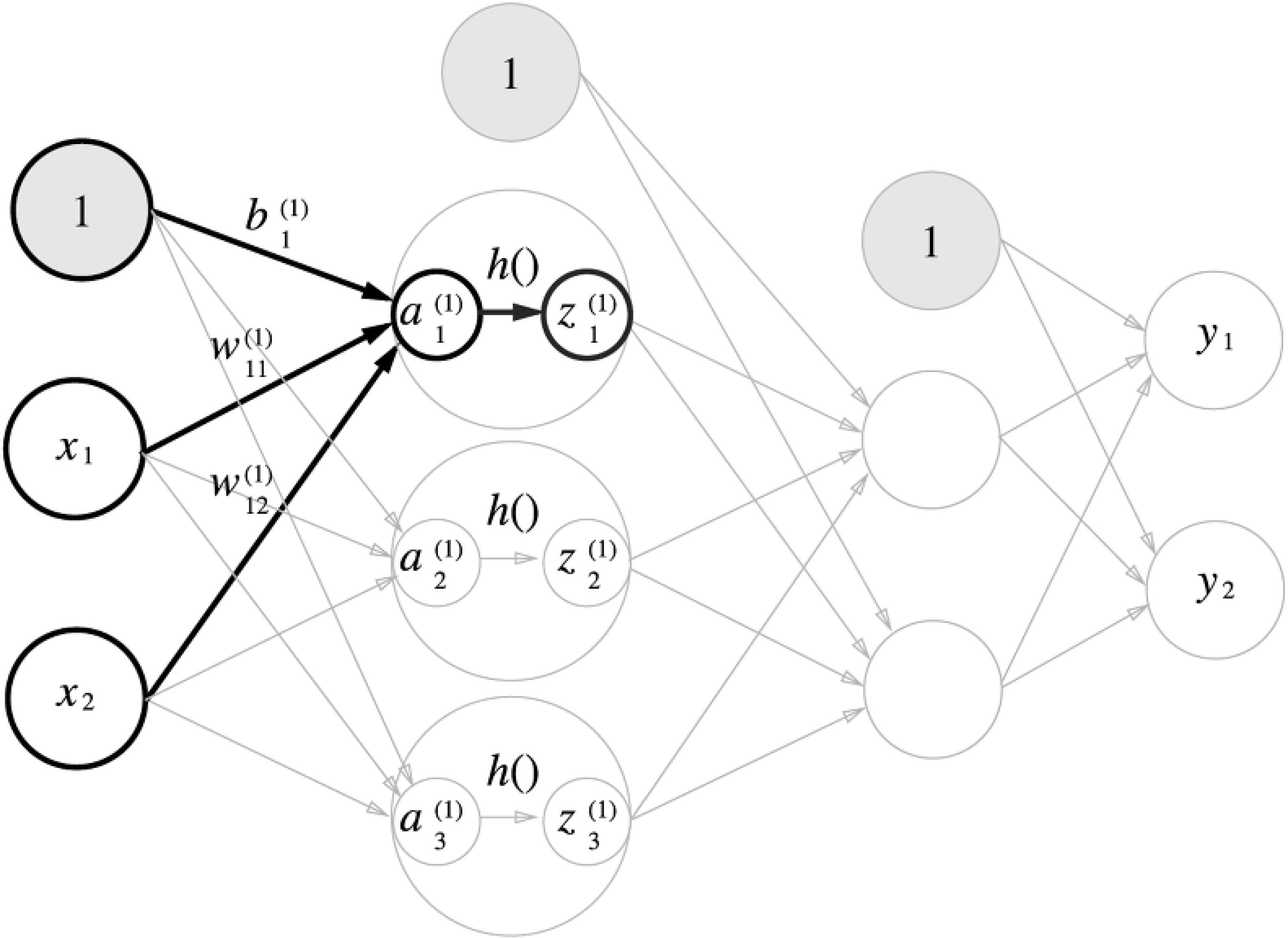
GELU (Gaussian Error LU)

$$f(x) = x \cdot \Phi(x)$$

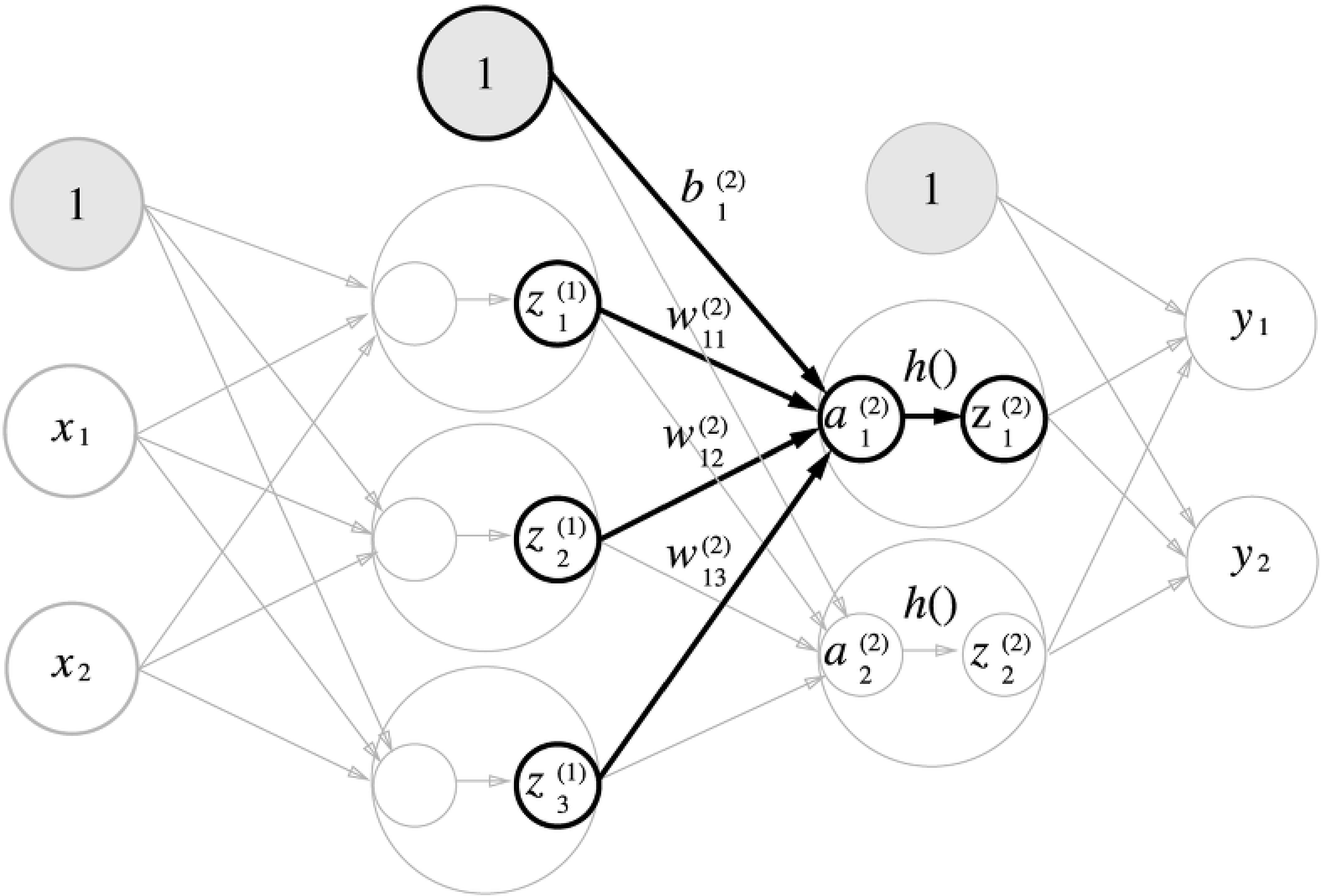


3층 신경망 구조

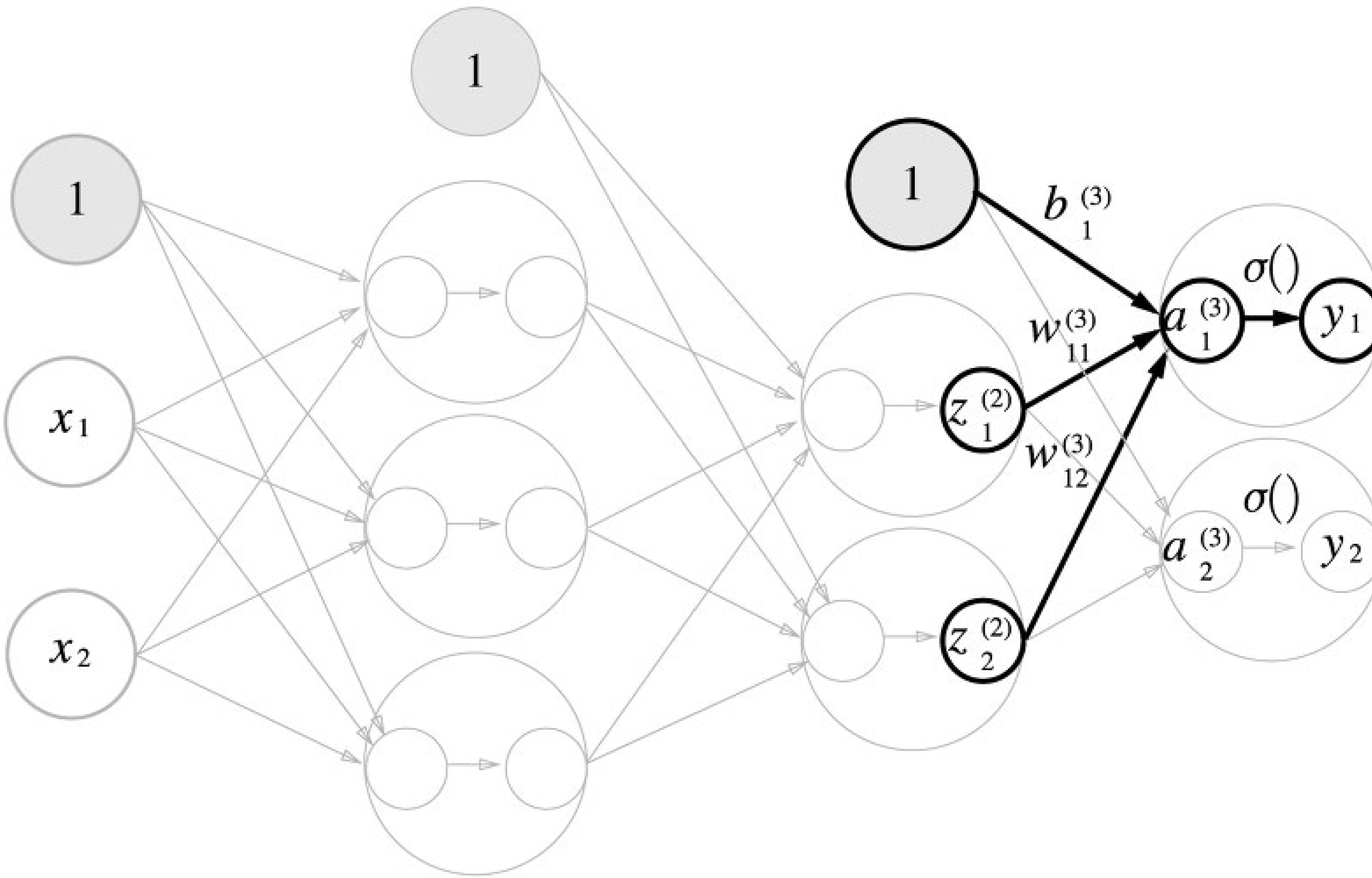




1
1
1
$b_1^{(1)}$
$w_{11}^{(1)}$
$w_{12}^{(1)}$
x_1
x_2
$h()$
$h()$
$h()$
$a_1^{(1)}$
$z_1^{(1)}$
$a_2^{(1)}$
$z_2^{(1)}$
$a_3^{(1)}$
$z_3^{(1)}$
y_1
y_2



1
1
1
x_1
x_2
$z_1^{(1)}$
$z_2^{(1)}$
$z_3^{(1)}$
$b_1^{(2)}$
$w_{11}^{(2)}$
$w_{12}^{(2)}$
$w_{13}^{(2)}$
$h()$
$a_1^{(2)}$
$z_1^{(2)}$
$h()$
$a_2^{(2)}$
$z_2^{(2)}$
y_1
y_2


1
1
1
x_1
x_2
$b_1^{(3)}$
$w_{11}^{(3)}$
$w_{12}^{(3)}$
$z_1^{(2)}$
$z_2^{(2)}$
$a_1^{(3)}$
$a_2^{(3)}$
σ()
σ()
y_1
y_2

Matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = (a_{ij})$$

$m \times n$ matrix

Operation

If $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ then

$$\mathbf{A} + \mathbf{B} = (a_{ij} + b_{i\,j}) \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}$$

$$\mathbf{A} - \mathbf{B} = (a_{ij} - b_{i\,j}) \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} - \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} -4 & -4 \\ -4 & -4 \end{pmatrix}$$

$$c\mathbf{A} = (ca_{ij}) \qquad 2\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$$

Product

$\mathbf{A} = (a_{ij})$ m x p matrix

$\mathbf{B} = (b_{ij})$ p x n matrix

$\mathbf{AB} = (c_{ij})$ m x n matrix

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$$

Product

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} =$$

$$\mathbf{AB} \neq \mathbf{BA}$$

Transpose

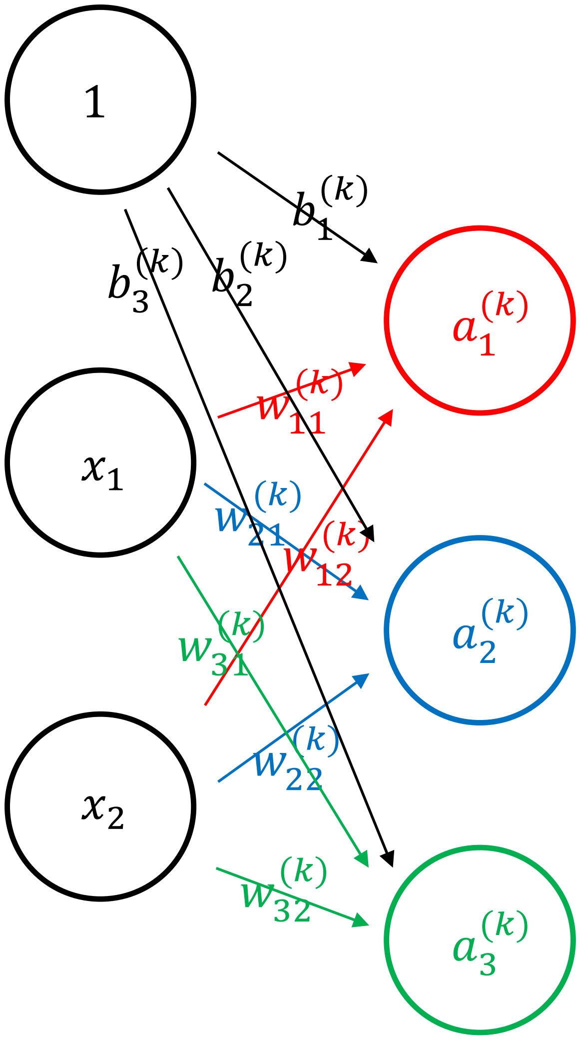
$$\mathbf{A}^T = (a_{ji})$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

$$(\mathbf{A}^T)^T = \mathbf{A}$$

$$(\mathbf{AB})^{\mathbf{T}} = \mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$$



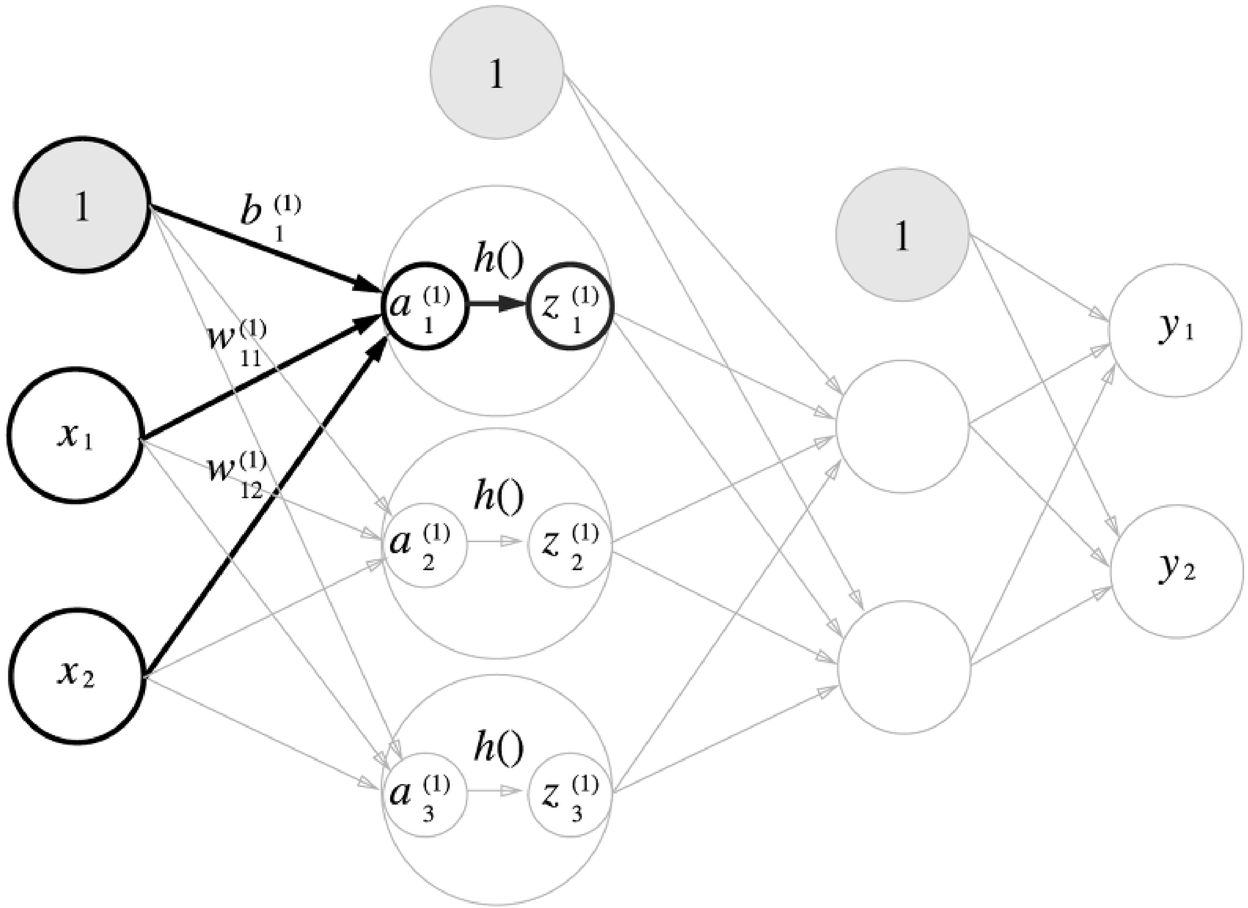
$$w_{out,in}^{(layer)}$$

$$a_1^{(k)} = w_{11}^{(k)} x_1 + w_{12}^{(k)} x_2 + b_1^{(k)}$$

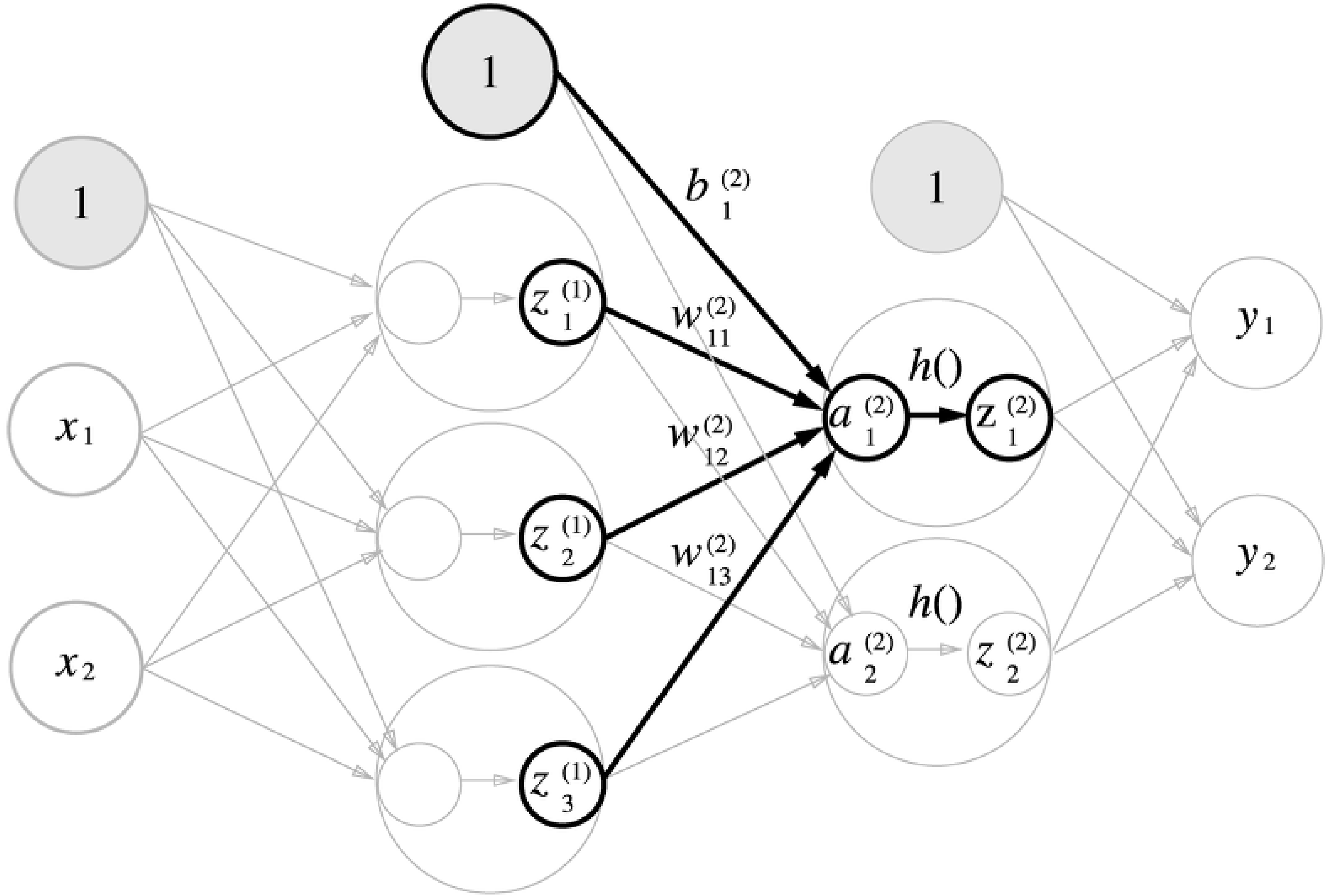
$$a_2^{(k)} = w_{21}^{(k)} x_1 + w_{22}^{(k)} x_2 + b_2^{(k)}$$

$$a_3^{(k)} = w_{31}^{(k)} x_1 + w_{32}^{(k)} x_2 + b_3^{(k)}$$

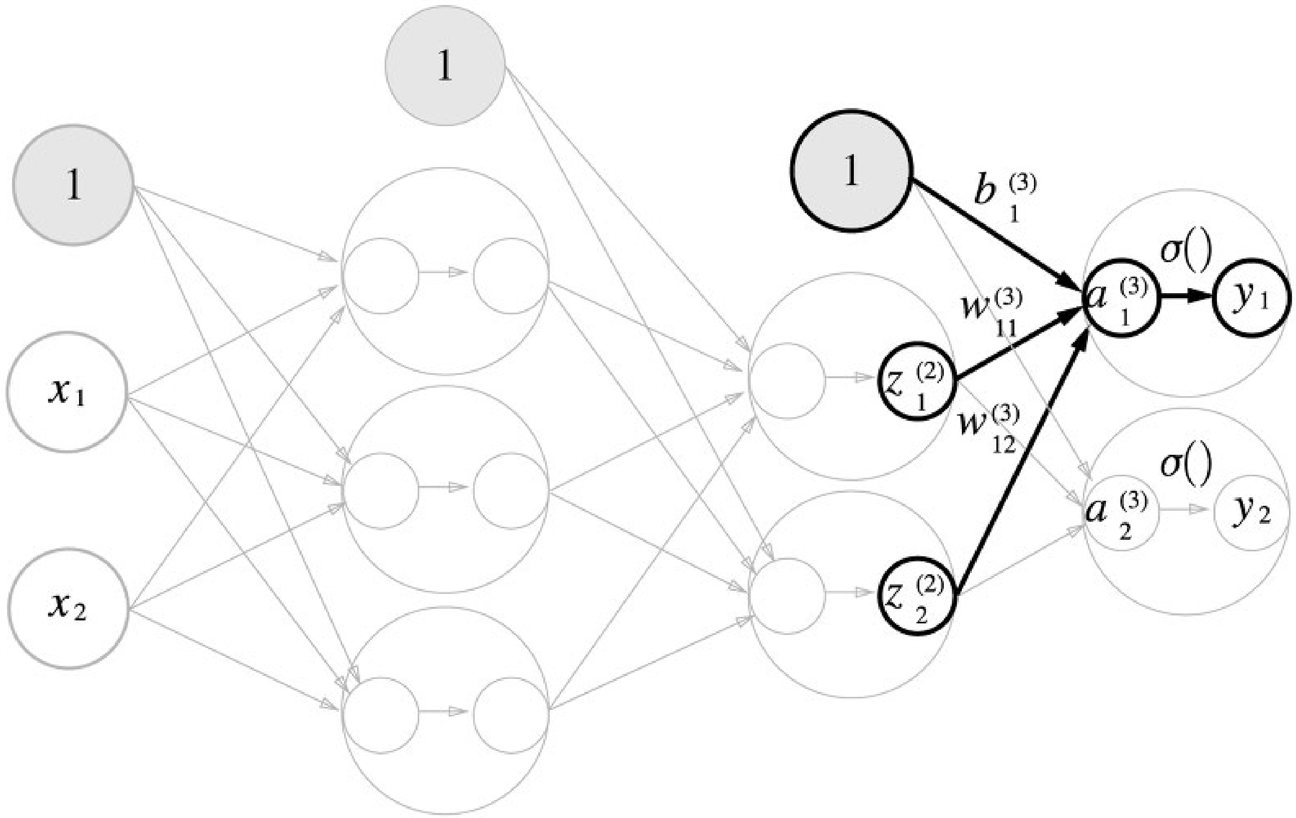
$$\begin{pmatrix} a_1^{(k)} \\ a_2^{(k)} \\ a_3^{(k)} \end{pmatrix} = \begin{pmatrix} w_{11}^{(k)} & w_{12}^{(k)} \\ w_{21}^{(k)} & w_{22}^{(k)} \\ w_{31}^{(k)} & w_{32}^{(k)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1^{(k)} \\ b_2^{(k)} \\ b_3^{(k)} \end{pmatrix}$$



$$\begin{pmatrix} a_1^{(1)} \\ a_2^{(1)} \\ a_3^{(1)} \end{pmatrix} = \begin{pmatrix} w_{11}^{(1)} & w_{12}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \end{pmatrix}$$



$$\begin{pmatrix} a_1^{(2)} \\ a_2^{(2)} \end{pmatrix} = \begin{pmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{pmatrix} b_1^{(2)} \\ b_2^{(2)} \end{pmatrix}$$



$$\begin{pmatrix} a_1^{(3)} \\ a_2^{(3)} \end{pmatrix} = \begin{pmatrix} w_{11}^{(3)} & w_{12}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} \end{pmatrix} \begin{pmatrix} z_1^{(2)} \\ z_2^{(2)} \end{pmatrix} + \begin{pmatrix} b_1^{(3)} \\ b_2^{(3)} \end{pmatrix}$$

Softmax

$(p_1, p_2, p_3, \cdots, p_n)$: Probability vector

$$p_i \geq 0 \qquad \sum p_i = 1$$

$$(a_1, a_2, a_3, \cdots, a_n) \qquad (0, -x)$$

$$(e^{a_1}, e^{a_2}, e^{a_3}, \cdots, e^{a_n}) \qquad (1, e^{-x})$$

$$\left(\frac{e^{a_1}}{\sum e^{a_i}}, \frac{e^{a_2}}{\sum e^{a_i}}, \frac{e^{a_3}}{\sum e^{a_i}}, \cdots, \frac{e^{a_n}}{\sum e^{a_i}}\right) \qquad \left(\frac{1}{1+e^{-x}}, \frac{e^{-x}}{1+e^{-x}}\right)$$

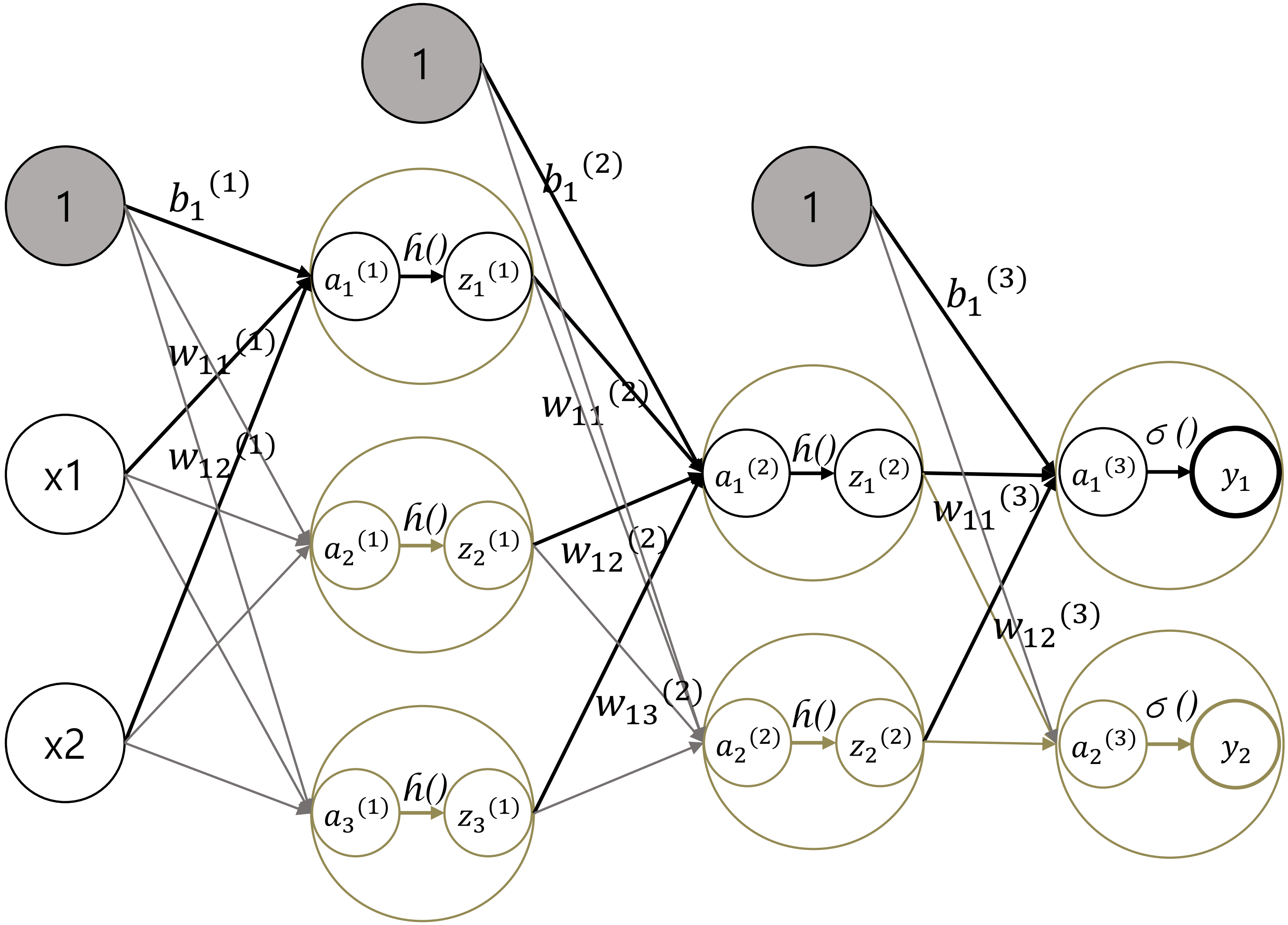
Softmax

$$(0, -x) \qquad\qquad (0, x)$$

$$(1, e^{-x}) \qquad\qquad (1, e^{x})$$

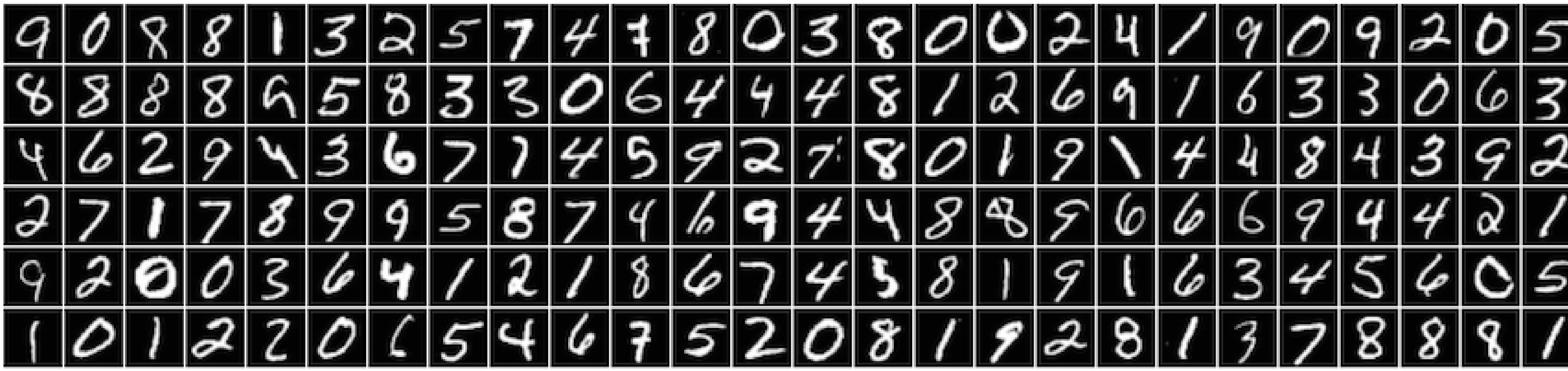
$$\left(\frac{1}{1+e^{-x}}, \frac{e^{-x}}{1+e^{-x}}\right) \qquad\qquad \left(\frac{1}{1+e^{x}}, \frac{e^{x}}{1+e^{x}}\right)$$

$$\begin{aligned} &\mathrm{softmax}(a_1, a_2, a_3, \cdots, a_n) \\ &= \mathrm{softmax}(a_1 + c, a_2 + c, a_3 + c, \cdots, a_n + c) \end{aligned}$$



Affine → Sigmoid → Affine → Sigmoid → Affine → Softmax

MNIST



0부터 9까지의 손 글씨 이미지로 구성

훈련 데이터가 6만장, 테스트 데이터가 1만장

각 데이터는 이미지와 라벨로 이루어짐

각 이미지는 28×28 해상도의 흑백 사진

각 픽셀은 0에서 255로 밝기 표현

Question?

자료 출처

Deep learning from scratch, 한빛미디어, 사이토고키

https://github.com/youbeebee/deeplearning_from_scratch